SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(144th Edition)

DECEMBER 2024

BY: CENTRAL LIBRARY

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

PREFACE

This is the **One Hundred Forty four** 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 **December, 2024**. The arrangement of the contents is alphabetical. The full text of the article which is either subscribed by the university or available in the web is provided in this bulletin.

Central Library

CONTENTS

1. An energy storage system with SOA-based FONPID controller for AGC Study in restructured power systems, ***6. Ram Naresh Mishra, 3. Narendra Kumar,*** Electrical, DTU

2. A low-power single ended half-select free 7T SRAM cell with improved write margin at 32nm technology node, ***8.Arunima Tripathi, 6. Ayush Dahiya*** and ***3.Poornima Mittal***, ECE, DTU

3. A novel deep learning framework for PVmodule thermal condition monitoring, ***6.Rahma Aman, 3.Mohammad Rizwan,*** and Astitva Kumar, Electrical, DTU

4. An Optimized Multi-kernel Based Extreme Learning Machine for Authentication Threat Detection with Feature Reduction Scheme in IoT***, 6.Prachi Dahiya, 3.Vinod Kumar***, Computer Science, DTU

5. Artificial intelligence based broken instrument from root canal and pulp retrieval of tooth using 33kHz ultrasonics signal, ***6. Rahul Kumar, 3. Rajiv Kapoor***, ECE, DTU

6. Artificial Neural Network based Modelling for Variational Effect on Double Metal Double Gate Negative Capacitance FET, ***6.Yash Pathak***, Laxman Prasad Goswami***, 3. Rishu Chaujar*** and ***3.Bansi Dhar Malhotra,*** Applied Physics, Biotech, DTU

7. Annual energy, energy, and efficiency analyses for conical solar Still combined with N number of evacuated collectors, ***6. Abhishek Kumar, 3.Rajesh Kumar***, Mechanical, DTU

8. *Accelerating cosmological ƒ (R,Lm,T) gravitymodel along with bulk viscous fluid,* **3. Shivani Sharma***, Pratik V. Lepse and Mehul Raj Sharma, Mathematics, DTU*

9. Black Phosphorus-Based Slotted D-Shaped PCF SPR Sensor for Cancer Detection ***7. Arhee Bhuyan, 6.Akash Khamaru*** and ***3. Ajeet Kumar***, Applied Physics, DTU

10. Bivariate generalization for operators involving Appostol-Genocchi polynomial, ***6. Km. Lipi, 3. Naokant Deo***, Applied Mathematics, DTU

11. Circular RFID Antenna for White Space TV Band, ***3. Richa Sharma, 3. N. S. Raghava*** and ***3.Asok De,*** ECE, DTU

12. Comparison of Tin and Indium Thin Film_s Structural, Morphological, Adhesion, and Electrical Properties for Ohmic Contact in HgCdTe-Based Infrared Detectors***, 7.1 Shailendra Kumar Gaur***, ***3. Qasim Murtaza,*** and ***3. R.S. Mishra***, Mechanical, DTU

13. Comprehensive study on thermal, structural,and luminescent properties of BiYWO6:Eu3+phosphors synthesized by various methods, ***6. Indrajeet Maurya and 3. M. Jayasimhadri,*** Applied Physics, DTU

14. Comprehensive study of Design attributes forsustainable packaging design: Analytical techniquesand expert validation***, 3. Taruna Singh, 3. Ranganath M. Singari*** and Sumer Singh, Design, DTU

15. Computation studies of phytocompounds of *Papaver somniferum* and *Boswellia serrata* for diabetes management, Madhura Roy, Garvit Satija, Ritika Luthra***, 6. Shreeja Datta***, Arpita Roy, Mohamed A.M. Iesa, Sarvesh Rustagi, Devvret Verma, Vaseem Raja, Cheng Wan Hee, Humaira Ja Swetha Raj, Biotechnology, DTU

16. Design of a Low-Power Dynamic Latched Comparator for Biomedical Applications**, 6. Mohit Tyagi,, 3. Poornima Mittala**, and Parvin Kumar, ECE, DTU

17. Digital twin: securing IoT networks using integrated ECC with blockchain for healthcare ecosystem**, 6.Vikas Sharma, 3. Kapil Sharma**, and Akshi Kumar , IT, DTU

18. Digital Transformation: Using Technology to Reshape Industries, ***3. Dr. Arjun Reddy,*** CSE, DTU

19. Dual hop Malaga underwater wireless optical communication system with pointing errors for IoUT, ***6. Dushyant Singh Chauhan, 3. Gurjit Kaur*** and ***3.Dinesh Kumar***, ECE, DTU

20. EmoMBTI-Net: Introducing and Leveraging a NovelEmoji Dataset for Personality Profiling with LargeLanguage Models, ***6. Dipika Jain***, CSE, DTU

21. Effect of Number of Probes on Spherical Near-field Measurement System DONGMING WEI, JUNWEI DONG, ***8.1 RAVI KUMAR ARYA***, YIDA FAN, SHIYUAN KONG AND LIJUAN LI, ECE, DTU

22. Electron Paramagnetic Resonance and Optical Absorption Studies of Cu2+ Ion-doped ZnAlBiB Glasses, K. Chandrika, G. Snegitha · M. Venkateswarlu, M. V. V. K. Srinivas Prasad, G. G. Dhar, S. K. Mahamuda, K. Swapna and ***3. A. S. Rao,*** Applied Physics, DTU

23. Elucidating the Effects of Aromatic Mutations on the RNA Binding Efficacy of CELF2 Protein***, 6.N. Kumari***, N. Sarovar Bhavesh, and ***3. Y. Hasijaa****, Biotechnology, DTU*

24. Energy Transfer Studies on Red Emitting $CaGdSbWO_8$: Sm3+/Eu3+ Phosphor for w-LEDs and Indoor Plant Growth Lighting System, ***7.1Jatin Parashar, 7.1 Gaurank Yadav***, Seema, Anu, Aman Prasad and ***3.A. S. Rao,*** Applied Physics, DTU

25. GAN-Based Novel Approach for Generating Synthetic Medical Tabular Data, Rashid Nasimov, Nigorakhon Nasimova, Sanjar Mirzakhalilov, Gul Tokdemir, ***3. Mohammad Rizwan*** ,Akmalbek Abdusalomov and Young-Im Cho,

26. High isolation dual band circularly polarized CPW based MIMO antenna with improved diversity performance for WLAN and C band wireless application, ***6. Rakesh Kumar, 3. Asok De*** & ***3. Priyanka Jain,*** ECE, DTU

27. Hyper-personalized employment in urban hubs: multimodal fusion architectures for personality-based job matching***, 6. Dipika Jain***, Saurabh Raj Sangwan and Akshi Kumar, CSE, DTU

28. Interconnection of SDN node with conventional network using an ONOS controller, ***6. Samiullah Mehraban, 3.Rajesh K Yadav***, CSE, DTU

29. Investigations on structural, optical and mechanical parameters of Pr3+ -doped oxyfluoro antimony borat glasses, Ch. Pravallika, K. Swapna , Sk. Mahamuda, M. Venkateswarlu and ***3. A. S. Rao***, Applied Physics, DTU

30. Investigation of MoS2-hydrogen interaction using in-situ X-ray diffraction studies, 6***. Ramesh Kumar***,· Nitesh K. Chourasia, Pawan K. Kulriya, Mahesh Kumar and ***3.Vinod Singh,*** Applied Physics, DTU

31. KRAFS-ANet: A novel framework for EEG-based stress classification using channel selection and optimized ensemble stacking, ***6. Shikha Shikha, 3. Divyashikha Sethia*** and ***3. S. Indu***, CSE, Software, ECE, DTU

32. Multimodal Integration of Mel Spectrograms and Text Transcripts for Enhanced Automatic Speech Recognition: Leveraging Extractive Transformer-Based Approaches and Late Fusion Strategies, ***6.1 Sunakshi Mehra, 3.Virender Ranga*** and ***3. Ritu Agarwal,*** IT, DTU

33. Microplastics in Agricultural Soil and Their Impact: A Review, P. Solank, S. Jain, R. Mehrotra, P. Mago and ***6.S. Dagar***, Environment, DTU

34. Mitigation of Residential EV Charging Effects on Power Distribution Networks with Optimal Allocation of DGs Using Coati Optimization***, 6. Shefali Painuli, 3. Suman Bhowmick*** and ***6.Radheshyam Saha***, EE, DTU

35. Mobile secured IoT sensors-driven network using efficient QoS management, ***7.1 MOHAMMAD SIRAJ, 7.1 MAJID ALTAMIMI*** and ZEESHAN ABBASI, Computer Technology, ECE, DTU

36. Modified six-state cryptographic protocol with entangled ancilla states, ***6. Rashi Jain*** and ***3. Satyabrata Adhikari,*** Applied Mathematics, DTU

37. Major vulnerabilities in Ethereum smart contracts: Investigation and statistical analysis, Qin Xin, ***3.Rajeev Kumar***, Mahdi Bahaghighat and Mohammad Pishdar, CSE, DTU

38. Molecularly Imprinted polypyrrole decorated Ti3C2Tx electrochemical sensor for highly selective and sensitive detection of levofloxacin***, 6.Divya Hudda and 3. Devendra Kumar***, Applied Chemistry, DTU

39. Numerical optimization of cold gas dynamic spray process parameters through response surface methodology Chirag Singhal & ***3.Qasim Murtaza***, Mechanical, DTU

40. Optimizing passive micromixers: Enhancing mixing efficiency through computational fluid dynamics and metaheuristic algorithms, 6. Md Gulam Mustafa, ***3. Mohammad Zunaid*** and ***3.Samsher Gautam***, Mechanical, DTU

41. ON COEFFICIENT PROBLEMS FOR S∗, _ ***3. S. SIVAPRASAD KUMAR, 7.1 ARYA TRIPATHI***, and ***6. SNEHAL PANNU***, Mathematical, DTU

42. On novel Hexa-compound combination synchronization over nineteen n-dimensional category-B chaotic systems and electronic circuit schematic, Dinesh Khattar, ***3. Naokant Deo*** and ***6. Mukul Sirohi***, Applied Mathematics, DTU

43. Optimizing Secure Data Transmission in 6G-Enabled IoMT Using Blockchain Integration, Aruna Malik, Vikas Tyagi, Samayveer Singh, ***3.Rajeev Kumar***, Huaming Wu, Senior and Sukhpal Singh Gill, Comuter Science, DTU

44. Predator Prey Scavenger Model using Holling's Functional Response of Type III and Physics-Informed Deep Neural Networks, Aneesh Panchal, ***6.Kirti Beniwal***, and ***3.Vivek Kumar,*** Mathematics, DTU

45. Performance Evaluation of Gate Engineered Ferroelectric MIMOS for Analog/Electrical IC Applications, ***6. Yash Pathak, 6. Kajal Verma***, ***3. Bansi Dhar Malhotra*** and ***3. Rishu Chaujar***, Physics, Biotechnology, DTU

46. REHAS: Robust and Efficient Hyperelliptic Curve-Based Authentication Scheme for Internet of Drones, ***6. Bhanu Pratap***, Ayush Singh and ***3. Pawan Singh Mehra***, CSE, DTU

47. Starlike Functions Associated with a Non-Convex Domain, ***3. S. Sivaprasad Kumar*** and ***6. Surya Giri,*** Applied Mathematics, DTU

48. Secular trends in leprosy burden: Comparative insights from India and Brazil in the last three decades using the Global Burden of Disease study Data**,** Shalini, Ravi P. Jha, Deepak Dhamnetiya, Krittika Bhattacharyya & Abhishek Mehra, Mathematics and Computing, DTU

49. Stochastic Neural Network Control for Stochastic Nonlinear Systems With Quadratic Local Asymmetric Prescribed Performance, Yu Xia, Ke Xiao , Jinde Cao, Radu-Emil Precup,***6.1 Yogendra Arya*** , Hak-Keung Lam and Leszek Rutkowski, Electrical, DTU

50. Synergistic Antimicrobial Effects of Waste Cotton Cloth Extracted Cellulose with Phytofabricated TiO2 for Potential Application in Agriculture, Chhavi Sharma,, Archana Rana, Amit Kumar Kesharwani, Dinesh Singh, Ritu Srivastava and ***3. Shailesh Narain Sharma***

51. Tb3+/Eu3+ co-doped NaLa(MoO4)2 phosphors: photoluminescence and optical thermometry studies for solid-state lighting and sensor applications Sonali Tomar1 , Vaibhav Chauhan, ***3. M. Jayasimhadri*** and C. Shivakumara, Applied Physics, DTU

52. Traffic-induced nanoparticle emissions and associated respiratory risk analysis using measurements conducted in a roadside environment, ***6. Kanagaraj Rajagopal*** · S. Ramachandran and ·***3. Rajeev Kumar Mishra,*** Environmental, DTU

53. Ultra-sensitive NO2 gas sensor at room temperature based on glycerolcross- linked PEDOT: PSS-MoS2 nanocomposite, Priyanka Dutta, Anuj Sharma, ***6. Videsh Kumar*** and Govind Gupta, Applied Physics, DTU

54. UnMA-CapSumT: Unified and Multi-Head Attention-driven Caption Summarization Transformer, ***6.Dhruv Sharma, 3. Chhavi Dhiman*** and ***3. Dinesh Kumar***, ECE, DTU

55. Validation of Nonviolent Work Behavior (NVWB) Scale and Its Relation With Workplace Conflict, Suhans Bansal, Sarika Kumari and ***3. Naval Garg***, School of Management, DTU

56. Visible emission studies of samarium doped Ca2Ga2SiO7 phosphor for luminescent device applications, Anand Parasar, ***6.Vikas Sangwan***, Sunil Kumar,· Sanjay Kumar, Kaushal Jha and Kusum Rawat, Applied Physics,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*

Physica Scripta

INCLUSIVE PUBLISHING
TRUSTED SCIENCE

PURPOSE-LED
PUBLISHING™

PAPER

A low-power single ended half-select free 7 T SRAM cell with improved write margin at 32 nm technology node

To cite this article: Arunima Tripathi *et al* 2025 *Phys. Scr.* **100** 015015

View the article online for updates and enhancements.

You may also like

- Large time dynamics of 2D semi-dissipative Boussinesq equations
Bing Li, Fang Wang and Kun Zhao
- The Neon Gap: Probing Ionization with Dwarf Galaxies at z 1
John Pharo, Yicheng Guo, David C. Koo et al.
- Ag/Fe_3O_4-N-Doped Ketjenblack Carbon Composite as Highly Efficient Oxygen Reduction Catalyst in Al-Air Batteries
Fuzhi Li, Liang Fu, Jingsha Li et al.

This content was downloaded from IP address 14.139.251.98 on 09/12/2024 at 06:40

Physica Scripta

PAPER

A low-power single ended half-select free 7 T SRAM cell with improved write margin at 32 nm technology node

RECEIVED
14 September 2024

REVISED
13 November 2024

ACCEPTED FOR PUBLICATION
25 November 2024

PUBLISHED
6 December 2024

Arunima Tripathi, Ayush Dahiya and Poornima Mittal

Department of Electronics and Communication Engineering, Delhi Technological University, New Delhi, India

E-mail: poornimamittal@dtu.ac.in

Keywords: SRAM cell, static noise margin, process corner, monte carlo, write margin

Abstract

A single-ended single port 7T SRAM bit-cell with a process-variation tolerant architecture and half-select resilience at 32 nm CMOS technology node is presented. The proposed 7 T (7TP) SRAM bit-cell integrates a low threshold voltage (LV_{TH}) feedback cutting transistor along with nominal V_{TH} devices. The reliability of the 7TP cell is presented for process, voltage and temperature (PVT) variations to account for local/global variations incurred during CMOS fabrication. 7TP SRAM bit-cell consistently exhibits 6-σ deviation for static noise margin (SNM), Write Margin (WM) and delay using 1000-point Monte Carlo simulations. The 7TP bit-cell shows improved WM, SNM, power consumption and delay compared to other SRAM cell architectures. The 7TP has the lowest normalized WM-1(WM- 0) of 8.5025%(8.5025%) compared to 17.0150%(46.0875%), 36.5800% (36.5800%), 17.0150%(53.5725%), 53.5450%(14.3650%), 11.8925%(11.8925%) for 7Tn, 7Ti, 7Tj, 8T and 11T SRAM cells respectively. The read-1(read-0) and write-1(write-0) power of the 7TP cell is 887.2 nW(10.09 μW) and 1087 nW(26.6 μW) respectively. The write-1 power of 8 T, 9 T and 10 T cells is 112.2%, 113.70%, and 136.11% of 7TP cell respectively. The proposed cell has the second-best HSNM/RSNM value of 0.269 V at $V_{DD} = 0.8$ V. The area of the proposed cell is 0.104 sq. μm, the least among other SRAM bit-cells.

1. Introduction

There has been a great deal of growth in the use of SRAM in the ever-increasing web of neural network. Digital devices are surrounding us all. The need for faster and less power-draining devices has pushed digital designers to explore deeper in the realm of SRAM [1, 2]. Technology has been continuously worked upon to increase the performance and efficiency of devices in general and system-on-chip (SoC) design in particular. This has resulted in a reduction in the size of technology nodes, eventually moving into the sub-45nm, with sizes as low as a few nm. This rapid scaling down of classical CMOS (such as Bulk-Si) [3] has certainly increased the cell density. Still, it has also caused various reliability issues, corner effects, and short channel effects [3, 4] and also increased the complexity of the fabrication process. Further reduction in channel length faces various issues leaving the only alternatives of non-classical devices [3] or improved cell architectures [5–8]. FinFET [9], TFET [10], Ge-DLTFET [11] and various other devices are also under scrutiny for utilization in low-power VLSI technologies. Even though advancing, the work in these areas is still in its infancy stage and more exploration of the whole novel design flows and their acceptance by well- established foundries remains a question. Also, moving to newer, radical technologies will require a huge investment in research, fabrication and manufacturing [12], further pushing researchers, at least for the near future, to continue to uncover the potential of CMOS devices.

The conventional SRAM architecture includes the differential-ended 6 T SRAM cell [13]. It has two cross-coupled inverter pairs and two access transistors to assist the read/write operation. Bitline (BL) and its complement BLB (bitline bar) assist in the differential read and write operation. The 6 T cell has a poor read margin and easily suffers from read disturbance [14]. To increase the performance of SRAM, researchers focused on developing new SRAM architectures having higher speed and stability. The relatively newer architectures

© 2024 IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

such as 7 T [15], 8 T [16], 9 T [17], 10 T [18] and 11 T [19] enable path optimization for each operation of read, write, and data retention. These newer bit-cells with increased transistor count offer higher read/write stability and have lower read and write access times but they also result in increased area. This decreases the cell density. Also, a higher number of transistors requires a greater number of control signals and complex peripheral circuitry for its operation. Thus, compelling researchers to dive into SRAM cell architectures with lower cell counts like 7 T which offer several benefits such as better stability and lower read/write delay than the 6 T cell and also have a lower area when compared to bit-cells of higher transistor count such as 10 T or 11 T.

An SRAM cell's performance can also be enhanced by increasing the efficiency of the peripheral devices connected to it [20]. Devices such as address decoders, column multiplexers [21] and sense amplifiers (SA) [22] have been proven to enhance the performance of a cell. The SA is a necessary part of the SRAM peripheral circuitry as it provides an amplified output analogous to the change in BL voltages [23]. The SA topology employed to perform the read operation has an immense impact on the read delay of the SRAM. Therefore, a sense amplifier together with the SRAM cell is a much better indicator of the overall latency exhibited by the SRAM rather than an isolated cell read performance.

This paper is organized into 5 sections as follows: section 2 describes the proposed single-ended, single port 7 TP SRAM bit-cell and also showcases various results of static noise margin (SNM), write margin (WM), delay, half-select error resilience and power at various supply voltages as well as the impact of process corners on the cell's performance. A comparison of the proposed cell with existing 7T SRAM bit-cells is presented in section 3. The comparative analysis of the proposed seven transistor (7TP) cell is done with cells of increasing transistor count from 8 T to 11 T. Section 5 concludes this work.

2. Proposed single-ended, single-port half-select free 7 T SRAM cell

A single-ended, single port seven transistor SRAM bit-cell (7TP) architecture is presented in this work. In the 7TP SRAM bit-cell, two inverters INV1 (M1 and M2) and INV2 (M3 and M4) are linked via the feedback transistor M5 and through the low threshold voltage (LV_{TH}) MOS M6. The Wordline (WL) signal is applied to the gate terminal of access transistor M7, while the Read (R) signal is applied to the gate terminal of the M5 transistor. The gate terminal of the LV_{TH} MOS M6 (PMOS) is grounded, and thus it remains on during all operations. The architecture of the proposed 7TP SRAM bit-cell along with transistor sizing is depicted in figure 1(a) and its layout is shown in figure 1(b). During the hold operation, the bit-cell stands disconnected from the bit line (BL), rendering the cell dysfunctional to read or write operations on the bit-cell data. Hold operation is carried out when the WL signal goes low and the R signal is high thus turning off the access transistor M7 while turning on the FB transistor M5. The bit line is disconnected from the remainder of the SRAM cell and the data stored at the storage nodes Q and QB of the cell stays intact. The use of multi-threshold transistors has various benefits such as LV_{TH} MOS devices has the benefit of allowing faster switching speeds albeit the cost of increased static power dissipation whereas the HV_{TH} device forgo delay requirements while consuming reduced leakage power [24, 25]. Therefore, there is a trade-off between speed and power when using different V_{TH} devices in a single circuit.

Read operation is brought about when the R and WL signals are kept high, consequently turning on both M5 and M7 respectively, thus allowing the data stored at the storage node Q to be sent to the BL via the access transistor M7. A single-ended SA is employed to read the data at the storage node Q of the SRAM cell. The architecture of the single-ended PMOS sense amplifier (SPSA) [26] is illustrated in figure 1(e). It uses two complementary control signals (Ctr and CtrB) and takes the input to be sensed as BL and gives the output at node Z. During the precharge phase, the signal Ctr of the SPSA is increased to $|V_{DD}|$, turning on transistor M1 and turning off M3 which causes the BL to precharge but only to $V_{DD} - |V_{TH,M1}|$ where $|V_{TH,M1}|$ is the threshold voltage of M1. This happens due to the PMOS being diode connected where its drain and gate are tied together while its source is at V_{DD}. The PMOS M1 remains turned on until the gate-source voltage (V_{GS}) equates to the threshold voltage. Therefore, the gate of M1 cannot be precharged beyond $V_{DD} - |V_{TH,M1}|$ because the PMOS turns off if its V_{GS} becomes greater than V_{TH}. In the evaluation phase, the Ctr signal is pulled down to GND therefore M1 is turned off and M3 is turned on. Now, the sense amplifier works as an inverter such that the voltage of the output Z goes from V_{DD} to GND when the voltage of BL switches from GND to V_{DD}.

The transient response of output Z of the sense amplifier for read-1 and read-0 operations are as illustrated in figure 1(c). Initially, the data in the Q storage of the 7TP cell is logic 0 and that in the QB storage node is logic 1. To read this data stored, the Ctr signal is pulled down to GND and CtrB is pulled to V_{DD} as required during the evaluation phase. Consequently, the output Z of the SPSA rises to V_{DD}. Since the SPSA works as an inverter, the output Z rises to V_{DD} in a read-0 operation, signifying its successful completion. Now, to show the read-1 operation of the 7TP cell, we flip the data in the storage nodes Q and QB of the bit-cell. As Q now holds logic 1,

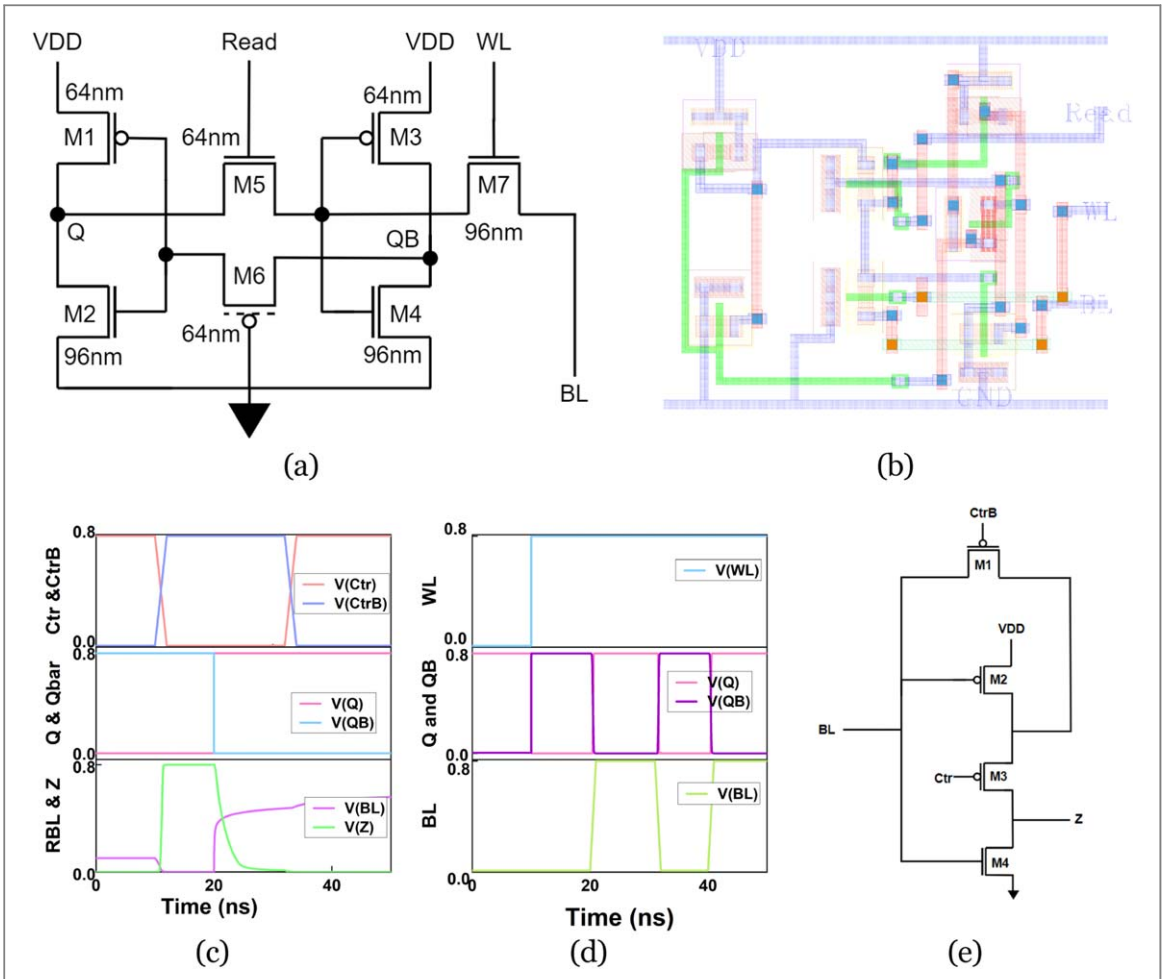


Figure 1. Proposed single-ended, single port 7T SRAM bitcell (a) 7TP circuit schematic, (b) layout of the 7TP bit-cell, (c) transient simulation with control signals for SPSA during read-0 operation of 7TP, (d) transient simulation during subsequent write-1 and write-0 operations, and (e) schematic of single-ended SA used for performing read operation.

the output Z of the SPSA gives an inverted value of logic 0, thus performing both read-0 and read-1 operations effectively.

To write data into the bit cell, it is required to drive the cell into the write mode to accomplish the write operation. To drive the cell into write mode, the WL signal is high and the Read input signal is turned low, which turns on the access transistor M7, and turns off the feedback transistor M5 respectively. With the FB off, the connection between the inverter pairs INV1 and INV2 is disrupted, reducing the circuit to its effective write mode equivalent circuit. The data present on the BL now appears at the storage node Q, and its complement appears at the storage node QB. The write-0 and write-1 transient response of the 7TP cell along with the control signal WL and output BL is shown in figure 1(d). As the WL signal increases to V_{DD}, the write operation on the cell begins. The data on the BL is logic 0 initially and thus storage node Q flips from its initial state of logic 1 to logic 0. It traces the data stored in BL as long as WL is pulled to V_{DD}.

2.1. Static noise margin analysis of 7TP SRAM bit-cell

Keeping a bit cell stable is vital to ensure that the data in a cell remains intact and doesn't flip as a result of noise or other factors. Static noise margin (SNM) is a measure of the stability of the SRAM cell. The maximum DC noise voltage after which, the data present in the cell can be inverted is called the SNM of an SRAM cell [27] which is calculated through the value of the side of the largest square that fits into the smaller lobe of the butterfly curve obtained by overlapping the voltage transfer characteristics of the inverter pair (INV1 and INV2). The observed HSNM and RSNM values of the 7TP cell are identical owing to the activation of similar transistors for both operations. Thus, the cell is said to be read SNM-free [28]. The hold SNM (HSNM) values are derived from the butterfly curve which is plotted by keeping the bit cell in the hold mode and sweeping the value at the storage node Q to then observe the value at storage node QB. This is then repeated for the storage node QB being swept and Q observed thereby resulting in both the traces for the butterfly curve. The side of the largest square that can

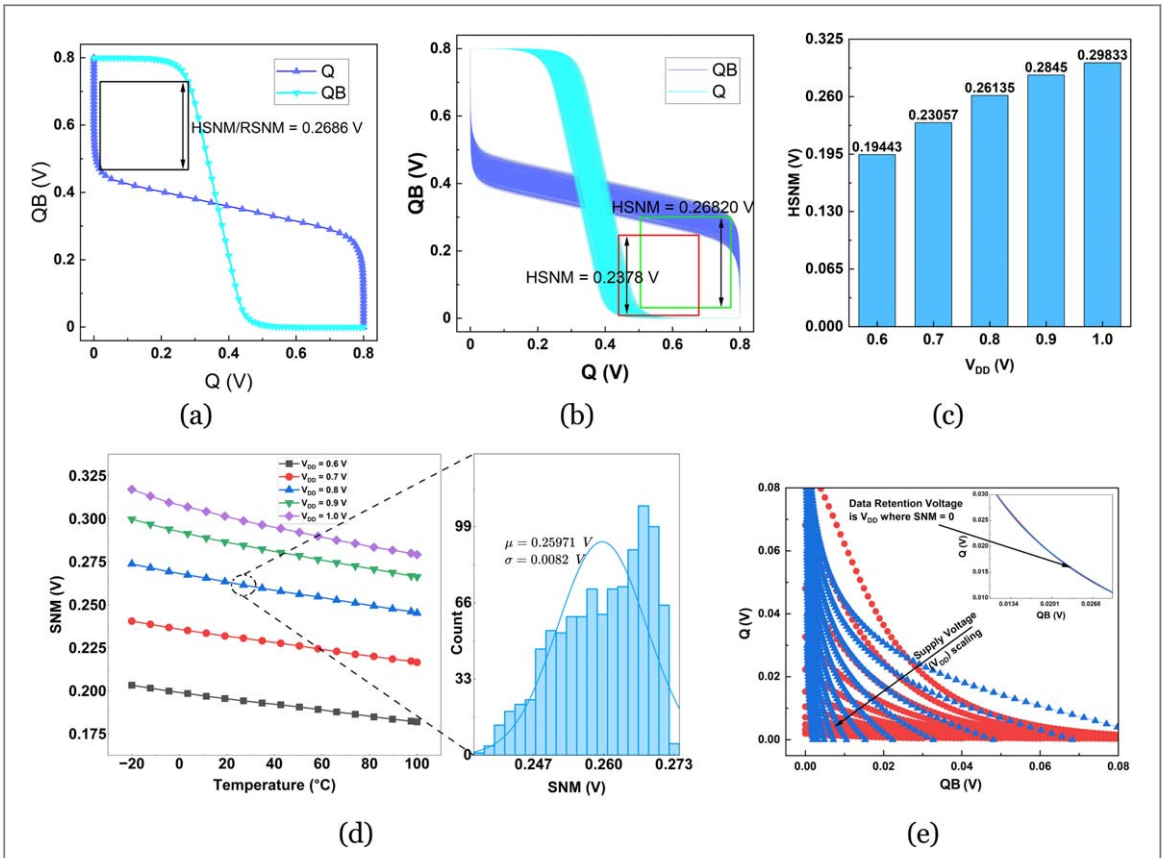


Figure 2. Static noise margin butterfly curve of the 7TP SRAM cell for (a) HSNM/RSNM calculation, (b) butterfly curve for Monte Carlo (1000-point) analysis at $V_{DD} = 0.8$ V, (c) HSNM/RSNM as a function of supply voltage, (d) SNM as a function of temperature (Inset: distribution of SNM at T = 27 °C for $V_{DD} = 0.8$ V) and (e) impact of supply voltage scaling on the SNM of the SRAM cell and data retention voltage for SNM = 0.

fit in the smaller of the two lobes of the butterfly curve gives the value of the HSNM. The HSNM/RSNM value of the 7TP cell is 0.27 V at = 0.8 V and its SNM curve is shown in figure 2(a).

A 1000-point Monte Carlo analysis of HSNM/RSNM at 0.8 V was also conducted to ensure that the proposed 7TP bit-cell would function properly under circumstances of noise or voltage fluctuations which is illustrated in figure 2(b). The SRAM cell was subjected to a variation of threshold voltage with a tolerance of 10%. The best and the worst possible HSNM/RSNM values recorded are 0.268 V and 0.237 V respectively. To verify the stability of the proposed SRAM cell at multiple V_{DD}, we have calculated the HSNM/RSNM at multiple V_{DD} ranging from 0.6 V to 1 V. As the V_{DD} increases, the SNM value also increases. The magnitude of the HSNM/RSNM values obtained in the above range of V_{DD} are 0.194 V, 0.230 V, 0.261 V, 0.284 V and 0.298 V respectively which is shown in figure 2(c). To depict the temperature dependence of SNM, the SNM value for the proposed bit-cell was calculated while varying temperature from −20 °C to 100 °C at multiple V_{DD} varying from 0.6 V to 1 V. As evident from figure 2(d), SNM follows an inverse relationship with temperature. As temperature increases from −20 °C to 100 °C, the SNM value falls. The SNM of a bit-cell depends only on V_{DD}, V_{TH} and β ratio and as temperature increases, the V_{TH} decreases [29]. This makes the bit-cell more susceptible to voltage fluctuations and thus decreases its stability and SNM. The 7TP SRAM bit-cell exhibits 6-σ deviation for SNM values when analysed using 1000-point Monte Carlo simulations which is depicted in figure inset in figure 2(d). Additionaly, the continuous scaling of the supply voltage results in a decrease in the SNM of the 7TP cell as is evident from the butterfly curve shown for different V_{TH} values. The data retention voltage (DRV) is another measure of the stability of a SRAM cell which is equal to the lowest supply voltage where SNM can be calculted and comes out to be nearly zero [30, 31]. The DRV is calculated for the 7TP SRAM cell and equals to 112.7 mV at T = 25 °C for TT process corner.

2.2. Write margin analysis of 7TP SRAM bit-cell

The write margin (WM) or write ability of a cell is its measure of stability during a write operation. For the proper functioning of the cell, it is expected to retain the value stored at its storage nodes in the face of any DC noise or voltage fluctuations and flip only when required. The write margin is the difference between V_{DD} and the write trip point (WTP) [32], as illustrated in figures 3(a) and (b), where the WTP is the voltage at which Q and QB flip.

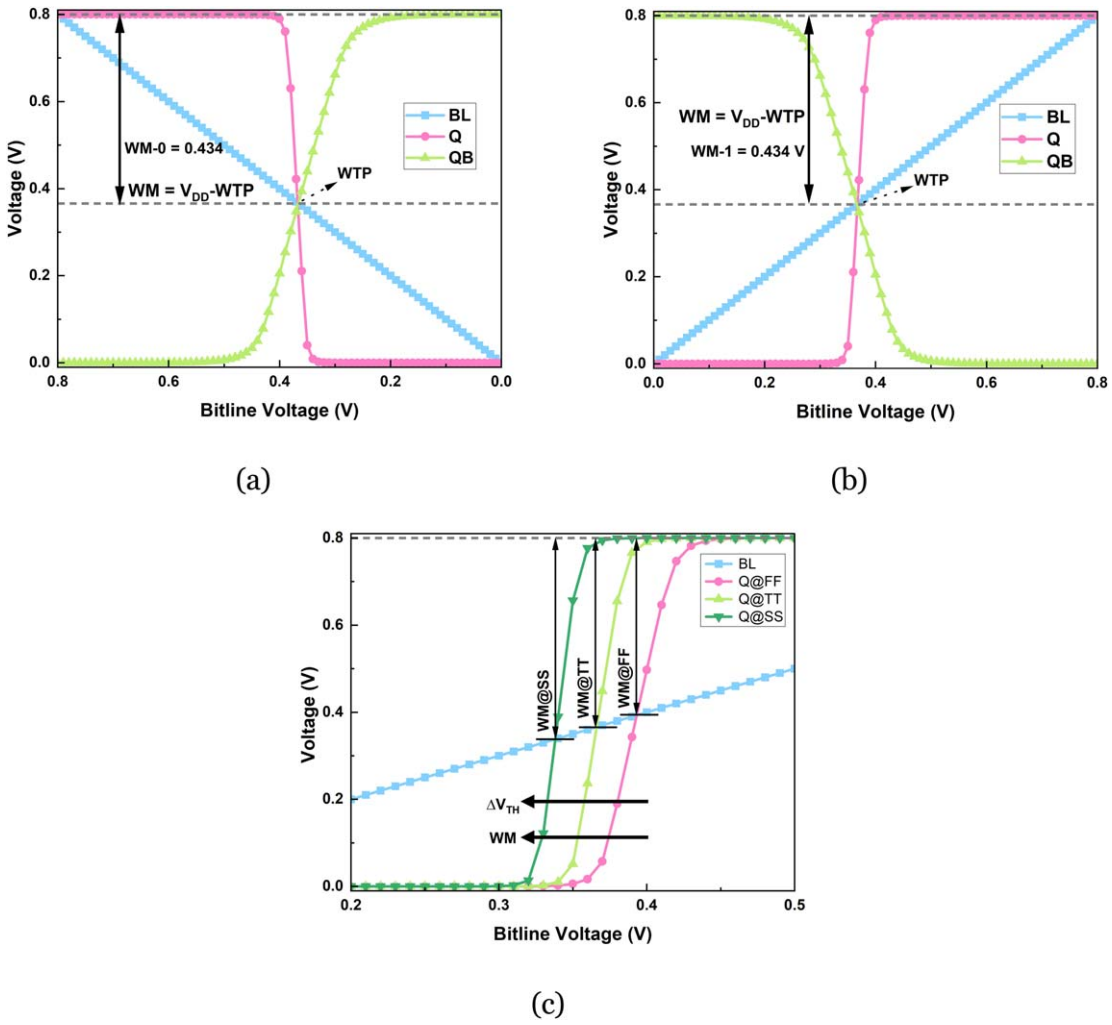


Figure 3. Write Margin and it's calculation for the proposed 7TP SRAM bit-cell (a) WM-0, (b) WM-1, and (c) the impact of process corners on WM for $V_{DD} = 0.8$ V.

For the 7TP SRAM bit-cell, during the write operation, the transistor M5 remains off and the data is sent from INV2 (M3 and M4) to INV1 (M1 and M2) through the LV_{TH} MOS (M6). The WM for the 7TP cell is 0.434 V. The ideal value of WM should be $0.5^{*}V_{DD}$ i.e. 0.4 V, signifying that the proposed cell has a good write ability. A low WM would have caused the cell to flip the data stored at the storage nodes Q and QB at low BL voltages or due to voltage variations or noise; while a high WM signifies that it will be very tough to write data into the storage nodes of the cell. The decoupled feedback loop of the proposed 7TP cell allows for improved WM by reducing the inverter's hold strength during the write operation. This enables the pass transistors to flip the stored data values more easily. In the 7TP cell, the identical write margins (WM-0 and WM-1 = 0.434 V) are a result of the deliberate symmetry included in the cell design. This allows robust performance under write operation which leads to predictable performance under process variations.

While manufacturing integrated circuits, many fabrication errors might creep in which may lead to process variation in the fabricated circuit. To establish the 7TP cell's proper working as a part of an SRAM cell array, it now becomes necessary to check the 7TP proposed SRAM cell for PVT variations. The proposed cell is analysed for FF (fast NMOS, fast PMOS), SS (slow NMOS and slow PMOS) and the typical (TT) process corner assuming a 30 mV variation in threshold voltage. Figure 4 shows the Monte Carlo variations of WM of the proposed 7 T SRAM cell for process variations (SS, TT, and FF; left to right respectively) at three different supply voltages of V_{DD}, $V_{DD}+\Delta V_{DD}$ and $V_{DD}-\Delta V_{DD}$ with ΔV_{DD} being 25% of nominal V_{DD} at 0.8 V.

The mean WM values for each process corner at a particular V_{DD} are also shown in the figure 4. For the TT process corner, the mean WM value is close to the obtained WM at 0.8 V (0.434 V). The mean value of WM distribution, closest to its ideal value at 0.8 V occurs in the FF process corner (0.40665 V). As devised theoretically, this difference in the WM values is caused due changes in the value of parameters in each process corner. As shown in the figure 3(c), as we move from left to right i.e. from the SS process corner to the FF process corner, we see a decrease in the mean value of the WM for the same V_{DD} across different process corners and as

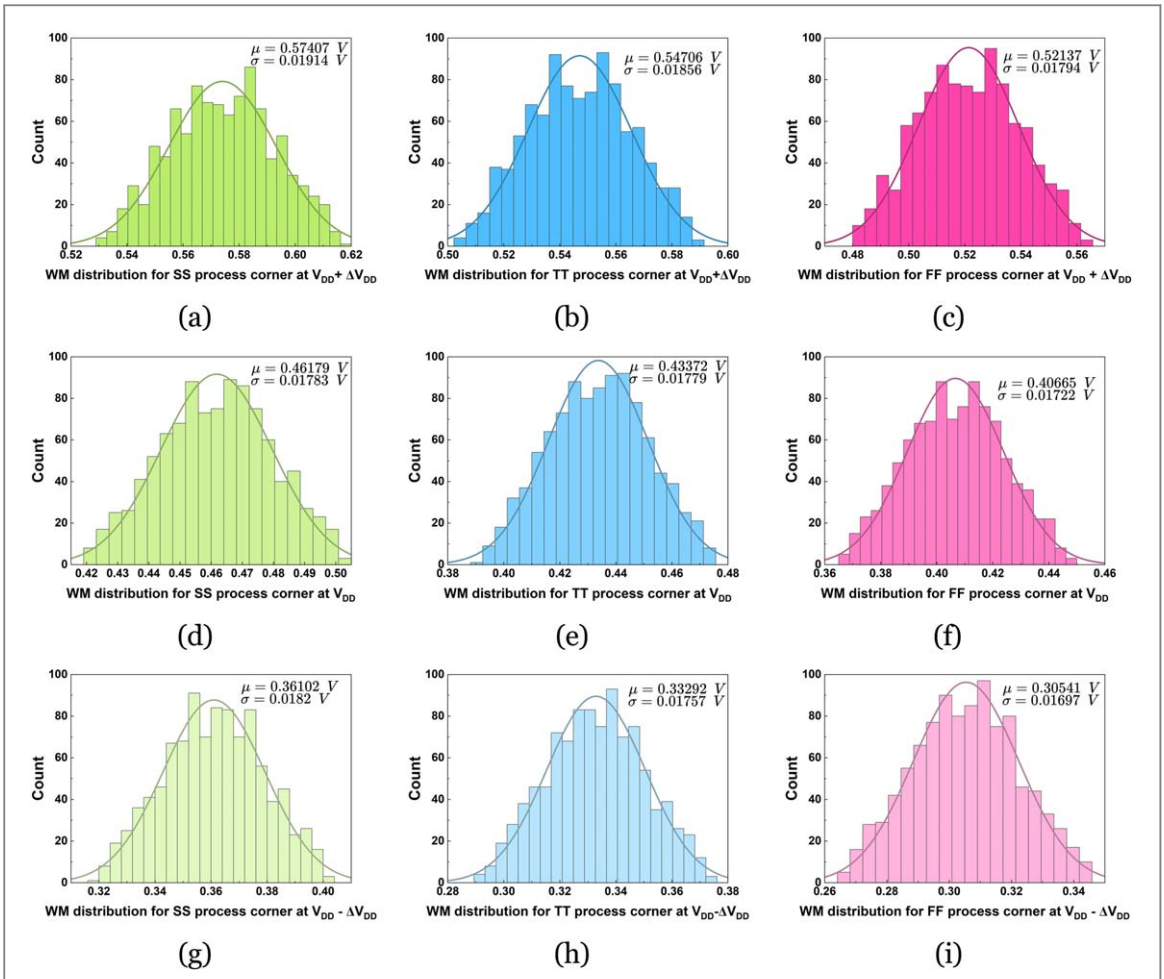


Figure 4. Monte Carlo simulations across process corners under supply voltage and threshold voltage variations for the proposed 7TP. The WM is plotted under V_{TH} deviation of 30 mV with $\Delta V_{DD} = 25\%$ of nominal V_{DD} at 0.8 V.

we move upwards to higher V_{DD} from 0.6 V to 1 V, we further see an increase in the mean of WM values in a particular process corner. As we move from the TT process corner to the SS process corner, the threshold value of the MOS increases and while moving from the TT process corner to the FF corner, the threshold value of the MOS decreases. The impact of process corners on WM is shown in figure 3(c). The write trip point at each process corner is also indicated. The difference between V_{DD} and the write trip point is called the write margin. Therefore, the lesser the magnitude of the trip point, the higher the write margin and vice versa. Therefore, from the illustration of WM at the three process corners, we can reiterate the observed WM trend and conclude that the proposed 7TP SRAM cell is robust to process and voltage variations.

2.3. Power consumption and leakage analysis of 7TP SRAM cell

A major denomination of power of an SoC is contributed by SRAM bit-cells. Decreasing the power consumption of the respective SRAM bit-cell will bring down the overall power consumption of the device. The SRAM bit-cell has three modes of operation, thus giving rise to three distinct classes of power based on the mode of operation. Power consumed during the write operation is referred to as write power similarly, read power is the power consumed during the read operation. The power consumed by the cell when it is in hold mode is known as leakage power.

The read-and-write power consumption of the proposed 7TP SRAM cell is illustrated in figure 5(a). Figure 5(a) also shows the read and write power consumption of the 7TP bit-cell at varying V_{DD} from 0.6 V to 1 V. As the V_{DD} increases, the write-1 and write-0 power consumed by the cell also increases whereas read-1 power and read-0 power of the cell decreases. The power consumed by the 7TP bit-cell during the read- 1 operation at 0.8 V V_{DD} is 0.89 μW while, during the read-0 operation, it is 10.10 μW. During the read operation, some power is also consumed in the operation of the sense amplifier. The overall power consumed is the summation of the power consumed by the 7TP cell and the power consumed by the sense amplifier and other peripheral devices. The leakage power consumed during the hold operation at 0.8 V V_{DD} is 12.73 μW. Leakage

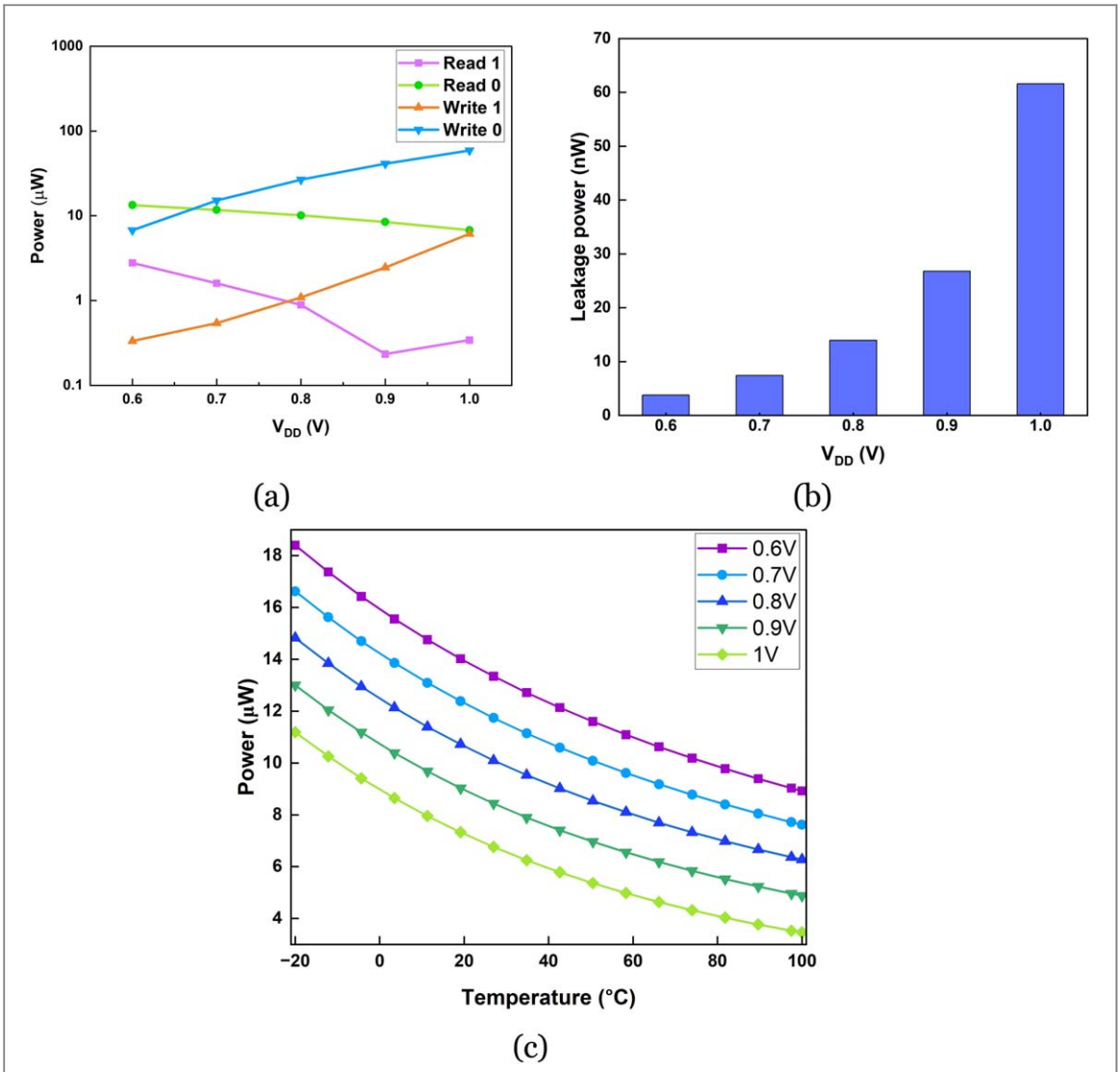


Figure 5. Power consumption of the 7TP SRAM cell as a function of supply voltage (a) read and write power variation, (b) leakage power, and (c) power as a function of temperature at multiple V_{DD}.

power also follows the same trend as of write-0 and write-1 powers when increasing V_{DD} as shown in figure 5(b). Power consumption is proportional to the square of the supply voltage, so even a small increase in the supply voltage results in a significant increase in power consumption. The leakage current is comprised of the sub-threshold, gate oxide and junction leakage components. For a given increment in V_{DD}, the threshold voltage also increases slightly due to body effect, reducing sub-threshold leakage while causing a rise in the gate oxide leakage as a result of the increased electric fields across the gate dielectric.

1.08 μW power is consumed during the write-1 operation whereas, write-0 power is 26.61 μW. The write 1 power is far lower than the write 0 power owing to the fact that large capacitive bit lines are discharged to ground during the write-0 operation while the write-1 only requires charging the bit lines to V_{DD}. During a write-0 operation, there is a longer time when both transistors in an inverter core are partially on. The NMOS transistor has to pull the node low while the PMOS transistor slowly turns off, causing them to simultaneously conduct. This results in an increase in short circuit current which yields a larger power consumption during write-0 operation. In contrast, during a write-1 operation, the PMOS transistor quickly turns on and the NMOS transistor quickly turns off, resulting in a shorter overlap time and less power wasted as short-circuit power. The power consumption of an SRAM bit-cell also varies inversely with temperature i.e. as temperature increases, the power consumption of the bit-cell decreases. This is illustrated in figure 5(c) which shows the variation of power as a function of temperature while varying V_{DD} from 0.6 V to 1 V at increments of 0.1 V.

2.4. Read and write delay analysis of 7TP SRAM bit-cell

Delay can be defined as the amount of time that it takes to perform a given action. We require a delay to have a minimum value to increase the speed of operation. There are various classifications of delay depending on the

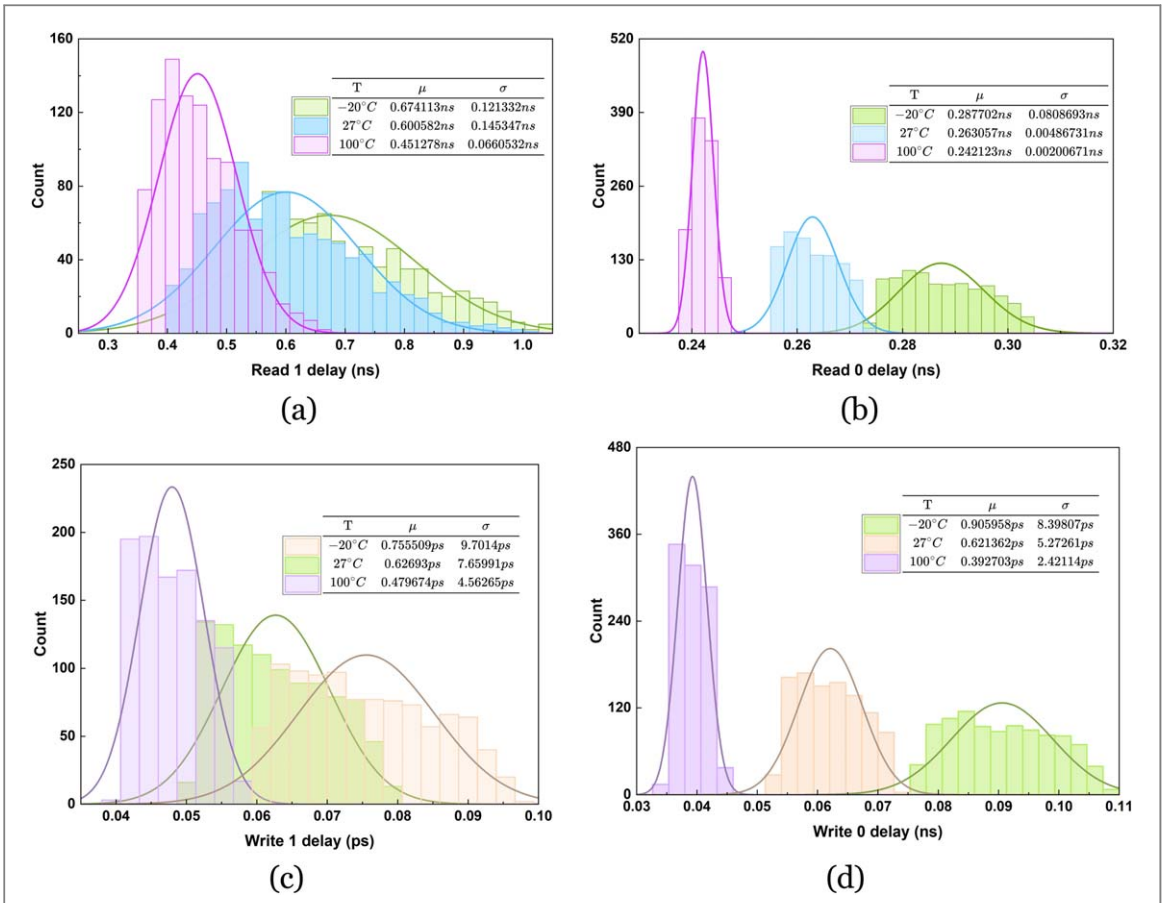


Figure 6. Delay of the proposed 7TP SRAM bit-cell for read and write operations at different temperatures. (a) Read-1 delay, (b) read-0 delay, (c) write-1 delay, and (d) write-0 delay histograms at $V_{DD} = 0.8$ V.

distinct operations performed. The read delay is the interval between the time when the read signal turns high (for read operation) and the time when the output of the sense amplifier (Z) reaches half of V_{DD}. The read-0 delay of the proposed 7TP SRAM cell is 0.2190 ns while the read-1 delay is 0.1826 ns. The time difference between the instant when data appears on the bit line and the instant when the voltage difference between Q and QB is half of V_{DD} known as the write delay. The write-0 delay of the proposed cell is 32.37 ps while the write-1 delay is 19.66 ps.

An SoC may suffer from many defects and fabrication errors when it is constructed, which weren't anticipated during the design phase. These fabrication errors are random and have a fairly low possibility of imposing any detrimental effect on the working of the cell but temperature variations and other external factors occur in everyday life and thus it is important to see the resistance of the cells working against temperature variations. We have performed the Monte Carlo simulation on the read delay (read-1 delay and read-0 delay) and write delay (write-1 and write-0 delay) of the proposed 7TP SRAM cell at three different temperatures of −20 °C, 27 °C and 100 °C with a tolerance of 10% keeping into account the temperature variation (from 0 °C to 70 °C) faced by commercial devices in everyday life.

Figure 6 shows this deviation of delay of the proposed 7TP cell. The mean value of delay at 27 °C for read-1, read-0, write-1 and write-0 operation are 0.60 ns, 0.263 ns, 62.693 ps and 62.136 ps respectively. As evident from figure 6, for any given operation, the delay value increases when the temperature decreases and the delay decreases when the temperature increases. There is an increase in delay value by 1.122, 1.093, 1.205 and 1.457 times when the temperature is decreased from 27 °C to −20 °C for read-1, read-0, write-1 and write-0 operations respectively. Similarly, for an increase in temperature value from 27 °C to 100 °C, the delay values decrease by 0.751, 0.920, 0.765 and 0.632 times respectively for the above-mentioned operations. Thus, from observing the above temperature variation trends, we can conclude that the proposed 7TP SRAM cell performs fairly well in the face of temperature variations.

2.5. Half-select error resilience of 7 TP SRAM bit-cell

In an SRAM cell array, various control signals are common between SRAM bit-cells. In the 2 × 2 array shown in figure 7(a), the Read0 signal is common to the bit-cell 00 and bit-cell 10 while, the Read1 signal is common to

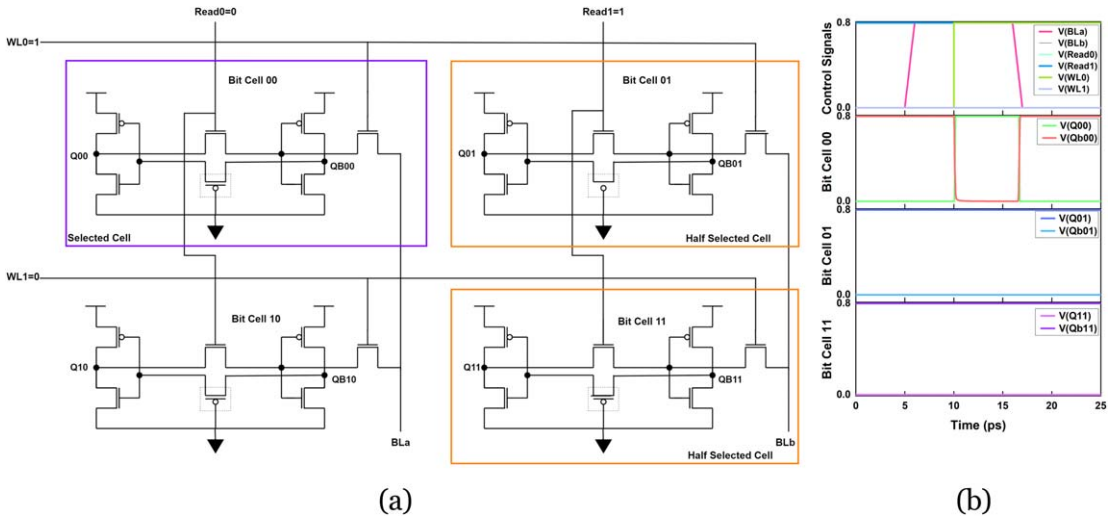


Figure 7. Depiction of half-select resilience of the proposed 7 TP SRAM bit-cell. (a) A 2 × 2 array of 7TP SRAM cells, and (b) transient response of control signals and data at storage nodes of selected and half selected cells.

Table 1. Parameters of the 32 nm PTM models used in simulation.

Parameters	Low V_{TH}	Nominal V_{TH}
^{L}eff	13 nm	13 nm
^{t}ox	0.75 nm	1.15 nm
^{V}DD	0.8 V	0.9 V
^{N}dep,n	8.7×10^{18} cm^{-3}	4.12×10^{18} cm^{-3}
^{N}dep,p	3.5×10^{18} cm^{-3}	3.07×10^{18} cm^{-3}
^{X}J,n	1×10^{-8} m	1×10^{-8} m
^{X}J,p	1.008×10^{-8} m	1×10^{-8} m
^{V}T H,n	0.3558 V	0.4936 V
^{V}to,p	− 0.24123 V	− 0.49155 V

cells 01 and 11. The WL0 signal is common to bit-cells 00 and 01 whereas, WL1 is shared in bit-cells 10 and 11. The control signals, when turned on for the bit-cell 00, could also turn on the surrounding cells (bit-cell 01 and bit-cell 11) known as half-selected cells. If the data at the storage nodes of the half-selected cells gets flipped/disturbed, due to operations carried out on the selected cell (bit-cell 00), then it is called half-select disturbance.

A superior SRAM cell array is mandated to be half-select error-free. To validate that the proposed cell is half-select error-free, a 2 × 2 cell array was constructed as demonstrated in figure 7(a). The half-selected cells, bit-cell 01 and bit-cell 11, are also highlighted. Figure 7(b) shows the variation in control signals and their effect on the storage nodes of the half-selected cells. All four modes of operation i.e., read- 1, read-0, write-1, and write-0 were performed on the bit-cell 00 and the data in the storage nodes of three neighboring bit-cells (bit-cell 01, bit-cell 10, and bit-cell 11) remained intact. From the figure 7(b), it is evident that the proposed 7 TP cell is half error-free as the values of storage nodes of the half-selected cells remains intact and are not disturbed when any operation is performed on any bit-cell thus it is resilient to half-select disturbance.

3. Comparison of 7TP cell with existing 7T SRAM cells

The proposed 7T SRAM cell is compared with multiple single-ended 7T SRAM bit- cells at the same technology node of 32 nm using Predictive Technology Model [33]. The parameters used for simulation at 32 nm PTM node are highlighted in table 1. There are a number of single-ended single port and dual port 7 T cells architecture presented in literature. The single port cells have only one port for both the read/write operations while the dual port cells utilize isolated read and write paths for enabling simultaneous read and write access. The lower area footprint and reduced complexity of the single port cells make them a good choice for high density general purpose SRAM whereas the dual port cells facilitate high performance applications requiring simultaneous access. Figure 8 shows various 7 T cells under comparison. The single-ended dual port 7Tn SRAM

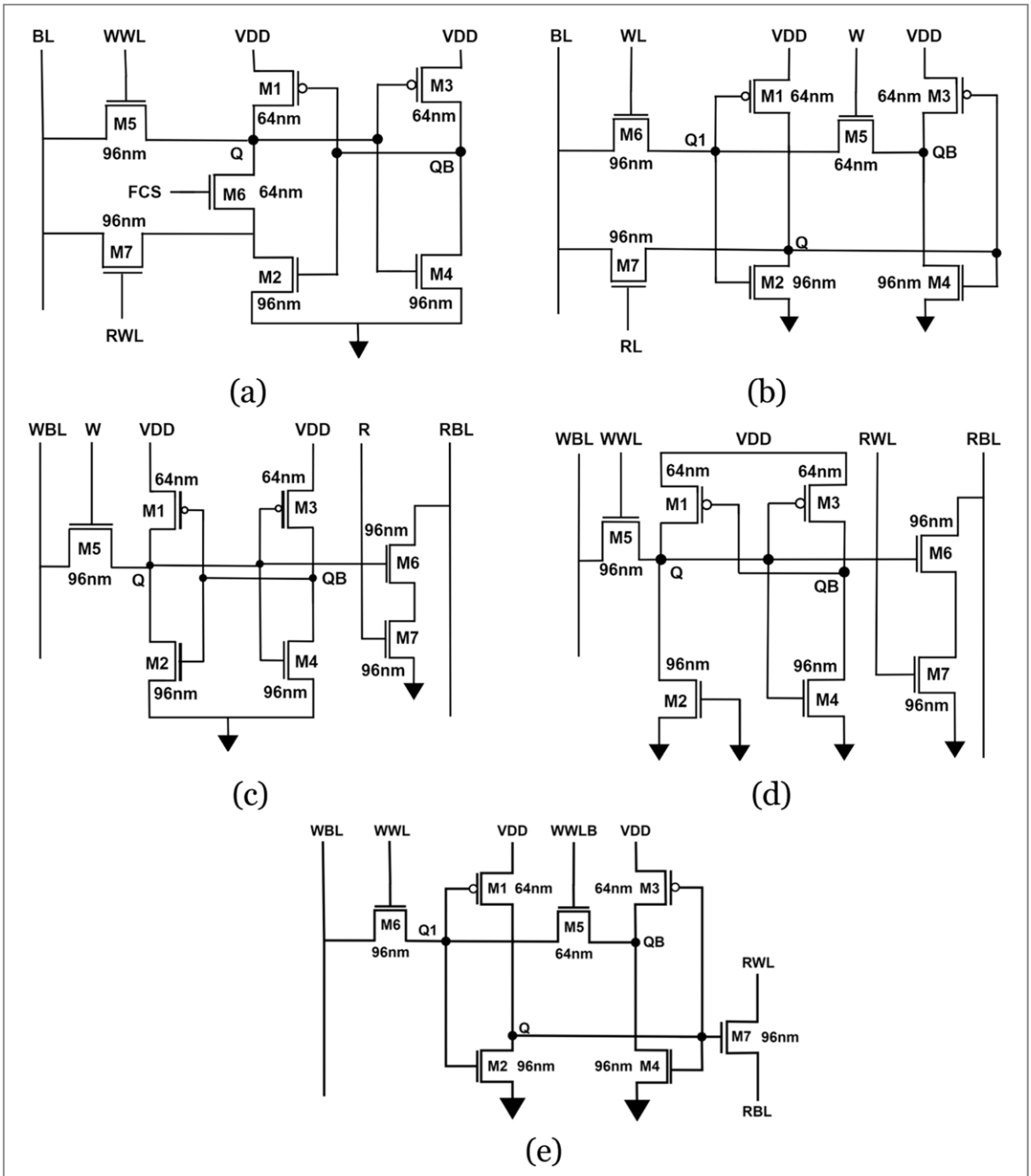


Figure 8. Circuit schematics of existing 7 T SRAM bit-cells (a) 7Tn, (b) 7To, (c) 7Ti, (d) 7Tj and (e) 7Tr SRAM bit-cells.

cell, illustrated in figure 8(a) was proposed by Kushwaha *et al* [7]. It uses three control signals, WWL, RWL and FCS. During the read operation, RWL is on while WWL and FCS are off, thus the cell uses dynamic read decoupling to reduce the read disturbance. In the write-0 operation, all the three control signals are high. Whereas, in the write-1 operation, only WWL is high while RWL and FCS are low. The 7Tn cell uses dynamic feedback cutting and also shows a reduction in power consumption of the cell when compared to the conventional 6 T SRAM cell.

The single-ended dual port 7To SRAM bit cell depicted in figure 8(b) was put forth by Rawat *et al* [8]. The cell uses a feedback transistor (M5) between the two cross-coupled inverters. The feedback inverter is controlled by the signal W which is low during the write operation and high otherwise. The cell uses 2 additional control signals, WL and RL, which are used in write and read operations respectively. The cell has shown higher SNM values when compared to the 6 T SRAM cell.

The 7Ti cell depicted in figure 8(c) was proposed by Tawfik *et al* [5] in 2008. It uses three high V_{TH} MOS in its cross-coupled inverter memory core and has isolated storage node from the bit lines during the read operation thus enhancing data stability of the cell. It has a separate read and write port, operated by the control signal R and W respectively. During the read-1 operation, the RBL voltage decreases down to 0 while during the read-0

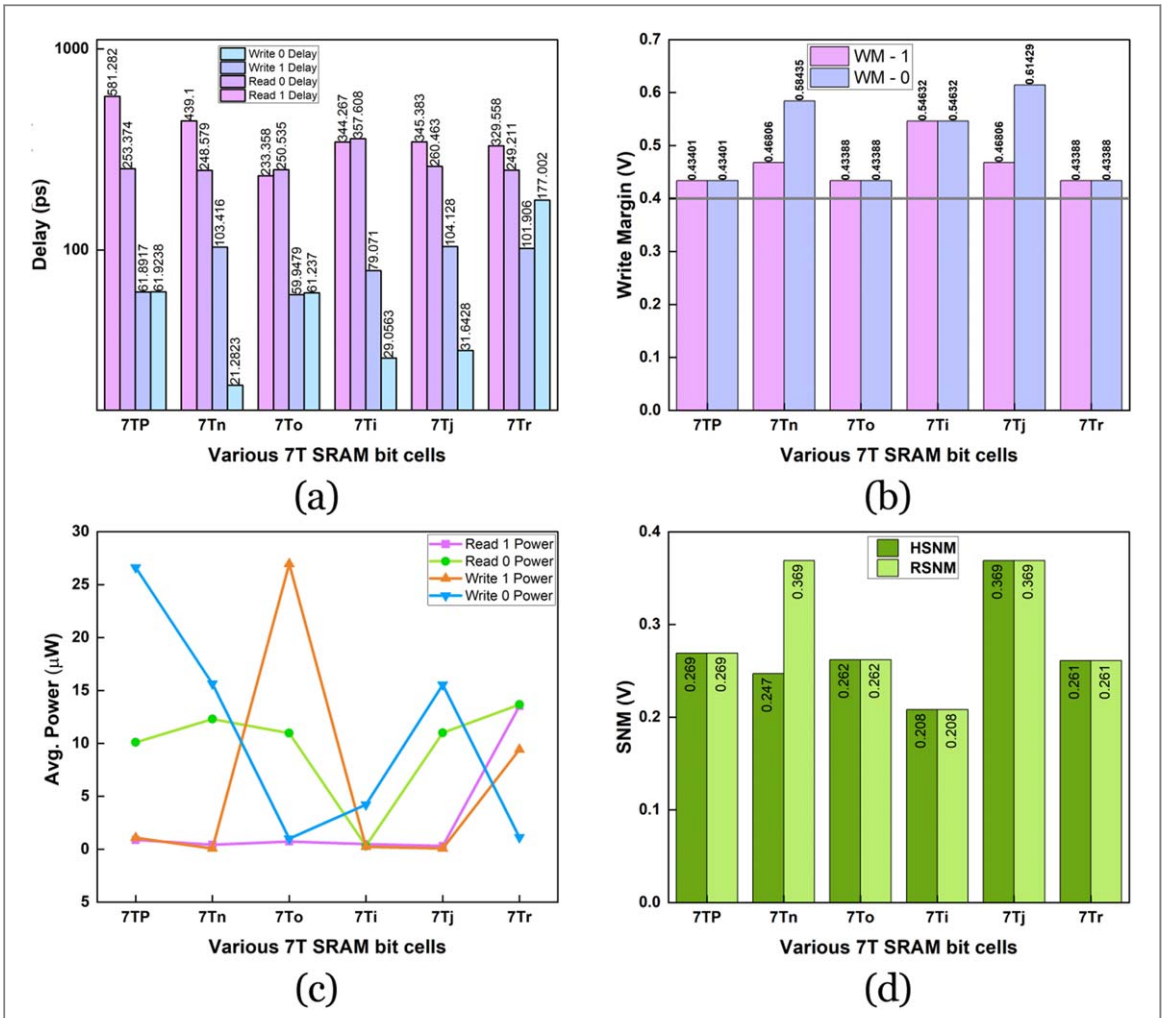


Figure 9. Performance analysis of 7 TP with existing 7T cells (a) read/write delay, (b) WM, (c) power consumption, and (d) SNM.

operation, the RBL voltage remains floating. The cell has shown increased SNM and read speed compared to the conventional 6 T SRAM cell [13].

Proposed by Gupta *et al* [6], the 7Tj cell illustrated in figure 8(d) operates in the sub-threshold region till 0.4 V. It has separate read and write ports operated by the control signals RWL and WWL respectively. The MOS M2 is driven into the cutoff region by connecting the gate terminal of the MOS to the ground. The dual port single- ended cell has shown a reduction in the leakage current through the read port and reduced static power consumption.

Another modified architecture for a single-ended seven transistor SRAM cell (7Tr) is shown in figure 8(e) [34]. The cell uses decoupled read and write ports with a feedback-cutting tranistor M5 controlled by WWLB control signal to optimize the cross- coupled inverter pair in the different operational states. The read assist scheme used here wherein the gate of the tranistor M7 is connected to the data node Q provides substantial reduction in read discharge current.

The delay values for read and write operations of various 7 T SRAM bit-cells are shown in figure 9(a). The 7TP cell has one of the lowest write-1 delays of 61.89 ps, a meagre 1.99 ps slower than 7To cell having the lowest write-1 delay. The read-0 and write-0 delay of the cells under comparison are similar. The read-1 delay of the proposed cell is 0.23 ns more than the median read-1 delay of the 7 T SRAM cells under comparison but can be fairly compensated for by the other parameters.

Figure 9(b) depicts the write margin values (WM-1 and WM-0) of various 7 T SRAM bit-cells. The WM of the proposed cell is 0.434 V, one of the lowest WM amongst other cells. The WM-1 and WM-0 of the 7 Tn cell differ due to the different magnitude of the control signal FCS, which is connected to the gate terminal of transistor M6, during the write-1 and write-0 operation. The WM of the 7 Ti cell is degraded by using LV_{TH} MOS. The difference in WM of the 7Tj bit-cell arises due to the grounded gate terminal of its transistor M2. During the write-1 operation, transistor M1 pulls up Q to V_{DD} but during the write-0 operation, since MOS M1 is driven into cutoff, it leaves the node Q floating, thus degrading the WM-1.

The power (μW) of various 7T SRAM bit-cells is shown in figure 9(c). The cells under comparison have similar read-1 power. The proposed 7T cell has a low write- 1 power of magnitude 1.09 μW. The read-0 power

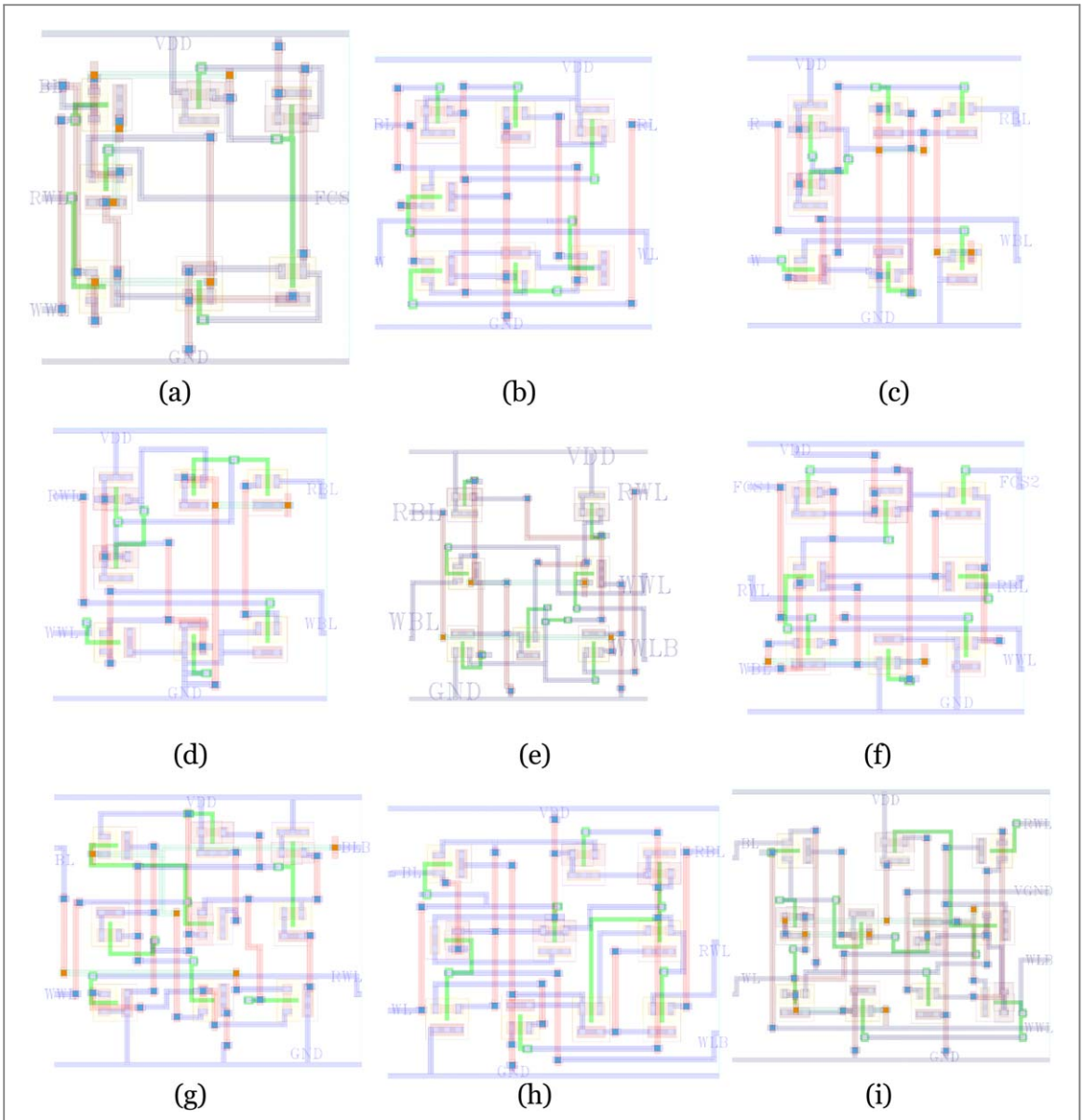


Figure 10. Layout of various SRAM bit-cells under comparison (a) 7Tn, (b) 7To, (c) 7Ti, (d) 7Tj, (e) 7Tr (f) 8 T, (g) 9 T, (h) 10 T, and (i) 11 T.

(10.1 μW) of the proposed cell is comparable to the read-0 power of the other cells except for the 7Ti SRAM bit-cell. In the 7Ti cell, during the read-0 operation, the RBL is set floating or M6 is turned off, as the input to the gate terminal of transistor M6 is 0, thus limiting the change in voltage of RBL leading to reduced read-0 power. The write-0 power of the 7To cell is the lowest (1.01 μW) while its write-1 power is the maximum amongst the other cells as the cell takes QB on the bit line, the opposite of the data that is to be written as the input i.e. for write-1(0), a zero (1) is sent over the BL, thus causing the increased value of write-1 power.

The hold SNM (HSNM) and read SNM (RSNM) values of the proposed 7 TP SRAM cell are compared with the HSNM and RSNM values of other 7T SRAM cells. Figure 9(d) depicts the obtained SNM values. The RSNM of the 7TP cell is 0.2686 V. The 7TP cell has similar HSNM and RSNM values due to the similar hold and read path of the cell. The 7TP cell has one of the largest HSNM value, the maximum HSNM is obtained for 7Tj cell at 0.3687 V at V_{DD} = 0.8 V. The HSNM of 7Tn, 7To and 7Ti cells are 0.246 V, 0.2616 V and 0.208 V respectively. The magnitude of RSNM of 7Tn, 7To, 7Ti and 7Tj cells are 0.3688 V, 0.2616 V, 0.208 V and 0.3687 V respectively.

Area footprint of SRAM cells in an important factor of its practical feasibility for mass production and usage. Cells with lower area have higher cell density and are the preferred choice for production. The layout and area analysis of all the cells under comparison has been carried out and the layout of the bit-cells are shown in figure 10. The area occupancy of the 7 Tn, 7To, 7Ti, and 7Tj bit-cells are 0.127 sq. μm, 0.128 sq. μm, 0.126 sq. μm, and 0.127 sq. μm respectively. The normalised value of area occupied by various 7T SRAM cells is shown in table 1. 7TP bit-cell has the lowest cell area, thus being the preferred choice.

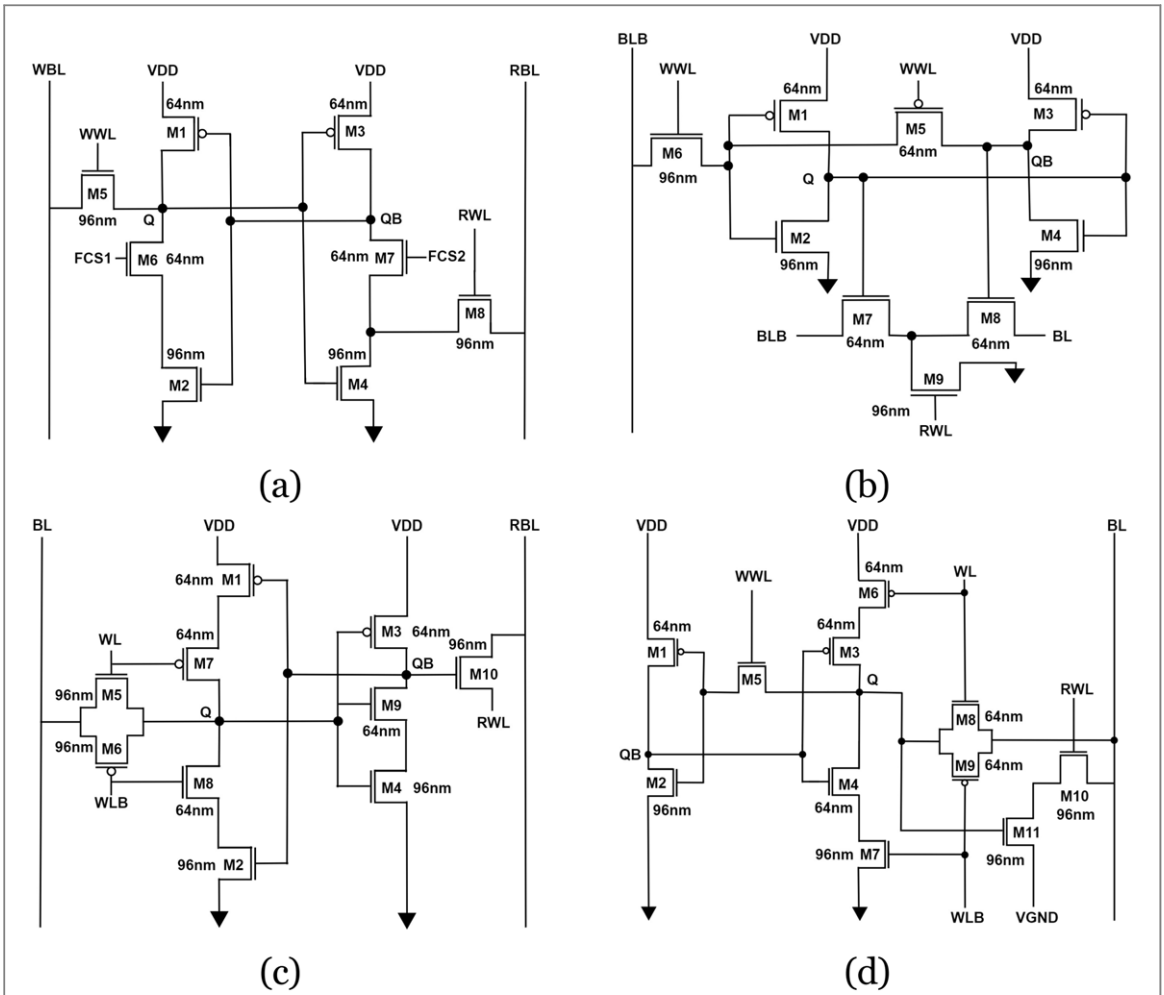


Figure 11. Circuit schematics of existing SRAM bit-cells with varied transistor count (a) 8 T, (b) 9 T, (c) 10 T, and (d) 11 T.

4. Comparison of 7TP cell with other SRAM cells of varied transistor count

Various cells with different transistor counts have been put forth by researchers in the past few decades. All the SRAM cells of varied transistor count compared to the proposed cell are simulated at identical values of cell sizing for similarly located transistors in the cell architecture as is depicted in figure 11 along with consistent switching times for the timing control signals in order to provide a common platform for analysis. The 8 T single-ended with dynamic feedback control SRAM cell as illustrated in figure 11(a) was proposed by Kushwah *et al* [16]. It performed the read and write operations from separate bit lines RBL and WBL respectively with the help of feedback control signals FCS1 and FCS2. The cell has a separate read and write path and exhibits read decoupling. It proved advantageous because of improved data stability and SNM in sub-threshold operation.

The 9 T SRAM cell as shown in figure 11(b) was put forth in [17]. The memory core of the cell is similar to that of a standard 6 T SRAM cell [13] but with an added feedback transistor M5 which is turned on during the write operation. The single-ended 9 T SRAM cell exhibits a lower write power due to the reduction of activity factor and breakup of the feedback path between two inverters during the write operation.

The 10 T SRAM cell as shown in figure 11(c) was proposed by Eslami [18] in 2020. The cell has a dual port nature. Transistor M10 with a separate RBL and RWL constitutes the read port. The write port is created by the pmos-nmos transmission gate, with M6 and M5 respectively, which is connected to the BL. The cell has low leakage power and improved read, write and hold stability in the sub-threshold region. Lorenzo *et al* proposed the 11 T SRAM cell [19] which is shown in figure 11(d).

The feedback transistor M5 is turned on during the write operation while during the read and hold operation, the inverters 1 (M1 and M2 pair) and inverter 2 (M3 and M4) pair stand disconnected by the transistor M5. The 11 T cell exhibits higher write and read SNM and delay values as compared to standard 6 T SRAM cell.

The delay values for the read and write operations for the various cells are shown in figure 12(a). The 8 T SRAM cell has the highest read-1 delay of 5.114 ns. The read-1 delay of the 7TP cell is 0.581 ns. The 7TP cell also

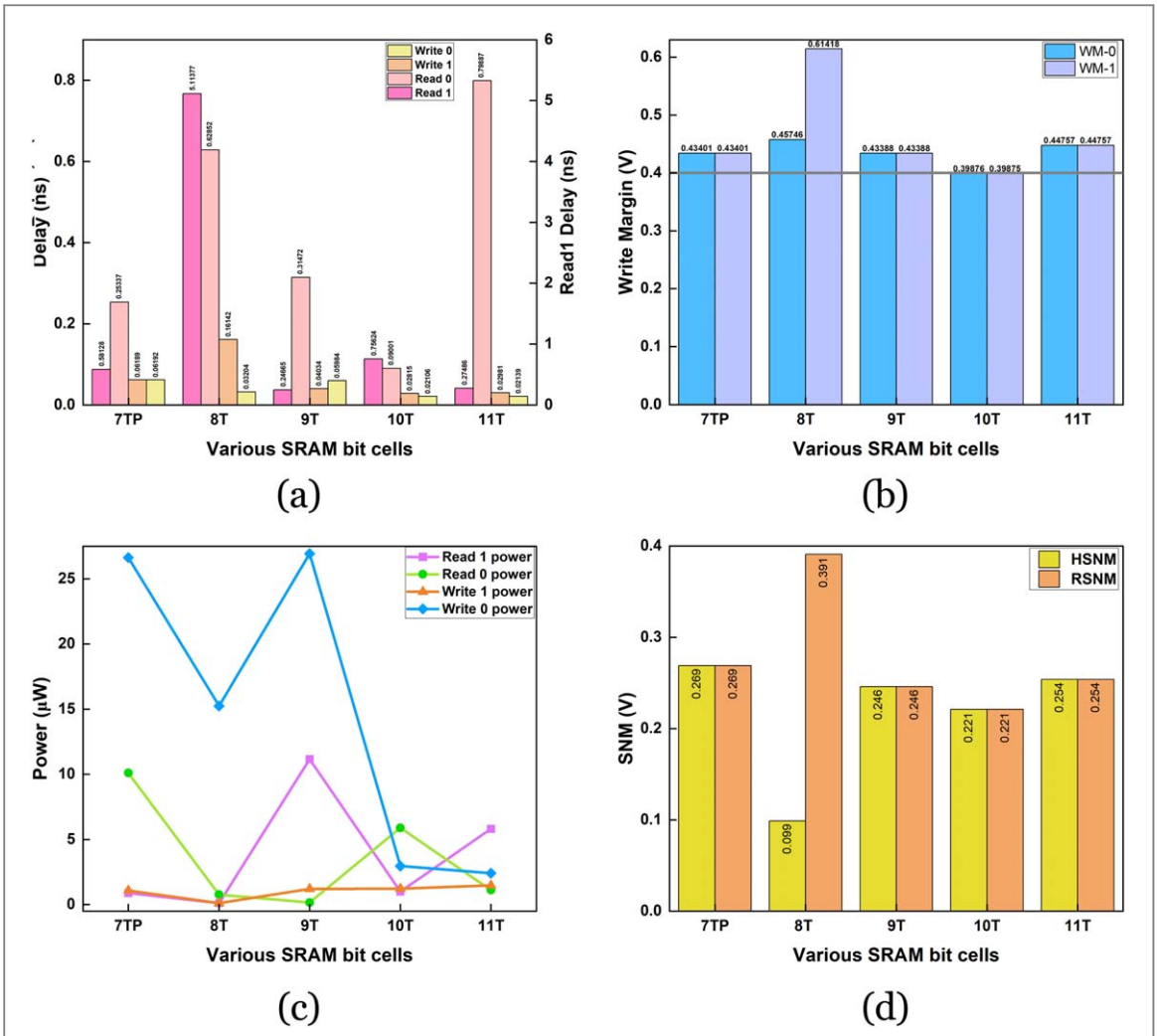


Figure 12. Performance analysis of 7TP with existing SRAM bitcells of varied transistor count (a) read/write delay, (b) WM, (c) power consumption, and (d) SNM.

has one of the lowest read-0 delays of 0.2534 ns. The 10 T cell has the smallest read-0 and write-0 delay of 90.01 ps and 21.06 ps respectively.

The WM-1 and WM-0 values of various SRAM bit-cells under comparison are shown in figure 12(b). The WM of the proposed cell is one of the lowest WM amongst other cells, while the 10 T SRAM bit-cell has the lowest WM (0.398 V). The WM-1 and WM-0 of the 8 T SRAM cell vary due to different control signals given to FCS1 and FCS2 transistors.

The read and write power of various SRAM bit-cells are shown in figure 12(c). The proposed cell has one of the lowest write-1 and read-1 power of 1.09 μW and 0.887 μW respectively, while the lowest write-1 and read-1 power is of the 8 T SRAM bit-cell (0.113 μW and 0.112 μW respectively). The write-0 and read-0 power of the 7TP cell is 26.6 μW and 10.1 μW respectively, but it is fairly compensated for due to the large delay value of the 8 T SRAM cell. The 7TP cell has slightly larger write-0 and read-0 power, compared to other counterparts, but due to the higher transistor count and the larger area occupied by the 10 T and 11 T cells, the difference can be neglected.

The hold SNM (HSNM) and read SNM (RSNM) values of the proposed 7TP SRAM cell are compared with the HSNM and RSNM values of other SRAM cells with varying transistor counts. Figure 12(d) depicts the obtained SNM values. The maximum HSNM is obtained for 7TP cell at 0.2686 V at $V_{DD} = 0.8$ V. The HSNM of 8 T, 9 T, 10 T, and 11 T cells are 0.098 V, 0.2463 V, 0.221 V and 0.2542 V respectively. The 7TP cell has one of the largest RSNM value of 0.2686 V after the 8 T cell whose RSNM value is 0.3913 V. The RSNM of 9 T, 10 T, and 11 T cells are 0.2463 V, 0.221 V, and 0.254 V respectively. Thus, at 0.8 V V_{DD}, the proposed 7TP cell outperforms all the other cells in terms of SNM.

The normalised area occupancy of 8 T, 9 T, 10 T and 11 T cells is shown in table 2. The area occupied by 11 T SRAM is the largest as it has the highest number of transistors among bit-cells under comparison. Its area occupancy is 0.196 sq. μm. The area occupied by 8 T, 9 T, and 10T bit-cells are 0.132 sq. μm, 0.158 sq. μm, and

Table 2. Comparison of 7TP cell with other SRAM bit-cells.

SRAM cell	Half-select resilience	Control signals	Overhead	Area	WM	SNM
7TP	Yes	2	LV_{TH}	1	0.43	0.27
7Tn	Yes	3	No	1.22	0.58	0.25
7To	No	3	No	1.23	0.43	0.26
7Ti	No	2	HV_{TH}	1.21	0.55	0.21
7Tj	No	2	No	1.22	0.61	0.37
7Tr	Yes	3	No	1.34	0.43	0.26
8 T	Yes	4	No	1.27	0.61	0.10
9 T	No	2	No	1.52	0.43	0.25
10 T	No	3	No	1.60	0.39	0.22
11 T	Yes	5	No	1.88	0.45	0.25

0.167 sq. μm respectively. These cells have a high area penalty due to a higher number of transistors. The layout of each of these cells is shown in figure 10.

5. Conclusion

A half-select free single-ended, single port 7 T SRAM bit-cell with a low voltage threshold (LV_{TH}) MOS as a feedback transistor is presented in this work. The proposed cell is designed at 32 nm CMOS technology node. The 7TP bit-cell has one of the highest static noise margin (SNM) with a magnitude of 0.27 V at 0.8 V, the largest SNM is of the 7Tj SRAM bit-cell at 0.37 V. Monte Carlo analysis for calculation of SNM encompassing variations due to PVT variations and temperature variations was also simulated and is presented here. Write margin (WM) of the 7TP cell is 0.434 V which is near to its ideal value of 0.4 V at 0.8 V V_{DD}. WM of the proposed cell was calculated at different process corners under supply voltage and threshold voltage variations. The impact of temperature was also modelled in WM Monte Carlo simulations to check the cell's robustness. Power consumption and leakage power consumption of the 7TP cell are also calculated along with the simulation of the effect of temperature on power consumption of the bit-cell. The read-1(read-0) and write-1(write-0) power consumption of the 7TP bit-cell is 0.89 μW(10.10 μW) and 1.08 μ(26.61 μW) respectively. Delay values are also calculated at different temperatures to check its robustness. 7TP SRAM bit-cell consistently exhibits 6-σ deviation for SNM, WM and delay values when analysed using 1000-point Monte Carlo simulations. The proposed cell is also resilient to half-select error. The comparison of 7TP cell with other state-of-the-art SRAM bit-cells is also stated. The area analysis of the proposed cell was carried out and it occupies an area of 0.104 sq. μm. It occupies the least area when compared to other SRAM bit-cells where the layout of each bit-cell was designed on a grid of 16 nm. Area consideration and other important metrics of each of the cells under comparison are summarised in a table which ultimately points towards the usefulness and robustness of the proposed cell.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

ORCID iDs

Poornima Mittal https://orcid.org/0000-0002-9479-8628

References

[1] Garofalo A, Rusci M, Conti F, Rossi D and Benini L 2020 Pulp-nn: accelerating quantized neural networks on parallel ultra-low-power risc-v processors *Phil. Trans. R. Soc.* A **378** 20190155
[2] Akgun O and Mei J 2020 An energy efficient time-mode digit classification neural network implementation *Philos. Trans. R. Soc.* A **378** 20190163
[3] Fossum J, Yang J-W and Trivedi V 2003 Suppression of corner effects in triple-gate mosfets *IEEE Electron Device Lett.* **24** 745–7
[4] Jena B, Dash S and Mishra G P 2018 Improved switching speed of a cmos inverter using work-function modulation engineering *IEEE Trans. Electron Devices* **65** 2422–9
[5] Tawfik S A and Kursun V 2008 Low power and robust 7t dual-v t sram circuit *2008 IEEE Int. Symp. on Circuits and Systems* 1452–5 (IEEE)
[6] Gupta S, Gupta K and Pandey N 2017 A 32-nm subthreshold 7t sram bit cell with read assist *IEEE Trans. Very Large Scale Integr. VLSI Syst.* **25** 3473–83

[7] Kushwah C, Vishvakarma S K and Dwivedi D 2016 Single-ended boost-less (se-bl) 7t process tolerant sram design in sub-threshold regime for ultra-low-power applications *Circuits Syst. Signal Process.* **35** 385–407
[8] Rawat B and Mittal P 2021 Single bit line accessed high-performance ultra-low voltage operating 7t static random access memory cell with improved read stability *Int. J. Circuit Theory Appl.* **49** 1435–49
[9] Raj B, Saxena A and Dasgupta S 2011 Nanoscale finfet based sram cell design: Analysis of performance metric, process variation, underlapped finfet, and temperature effect *IEEE Circuits Syst. Mag.* **11** 38–50
[10] Ionescu A M and Riel H 2011 Tunnel field-effect transistors as energy-efficient electronic switches *Nature* **479** 329–37
[11] Raikwal P, Kumar P, Panchore M, Dwivedi P and Cecil K 2023 Soft-error-aware radiation- hardened ge-dltfet-based sram cell design *Electronics* **12** 3198
[12] Horowitz M 2014 1.1 computing's energy problem (and what we can do about it) *2014 IEEE international solid-state circuits conference digest of technical papers (ISSCC)* (IEEE) 10–4
[13] Gupta V and Anis M 2009 Statistical design of the 6t sram bit cell *IEEE Trans. Circuits Syst. I* **57** 93–104
[14] Praveen P and Singh R 2023 Design of enhanced reversible 9t sram design for the reduction in sub-threshold leakage current with14nm finfet technology *ACM Trans. Des. Autom. Electron. Syst.* **28** 1–29
[15] Rawat B and Mittal P 2022 A comprehensive analysis of different 7t sram topologies to design a 1r1 w bit interleaving enabled and half select free cell for 32 nm technology node *Proc. R. Soc. A* **478** 20210745
[16] Kushwah C and Vishvakarma S K 2015 A single-ended with dynamic feedback control 8t subthreshold sram cell *IEEE Trans. Very Large Scale Integr. VLSI Syst.* **24** 373–7
[17] Roy C and Islam A 2020 Characterization of single-ended 9t sram cell *Microsyst. Technol.* **26** 1591–604
[18] Eslami N, Ebrahimi B, Shakouri E and Najafi D 2020 A single-ended low leakage and low voltage 10t sram cell with high yield *Analog Integr. Circuits Signal Process.* **105** 263–74
[19] Lorenzo R and Pailly R 2020 Single bit-line 11t sram cell for low power and improved stability *IET Computers & Digital Techniques* **14** 114–21
[20] Dahiya A, Mittal P and Rohilla R 2023 Modified decoupled sense amplifier with improved sensing speed for low-voltage differential sram *ACM Trans. Des. Autom. Electron. Syst.* **28** 1–15
[21] Mittal P, Rawat B and Kumar N 2022 Tetra-variate scrutiny of diverse multiplexer techniques for designing a barrel shifter for low power digital circuits *Microprocess. Microsyst.* **90** 104491
[22] Dahiya A, Mittal P and Rohilla R 2023 Variation-tolerant sense amplifier using decoupling transistors for enhanced sram read performance *Circuits Syst. Signal Process.* **42** 5799–810
[23] Rawat B and Mittal P 2023 A switching nmos based single ended sense amplifier for high density sram applications *ACM Trans. Des. Autom. Electron. Syst.* **28** 1–14
[24] Gupta V, Khandelwal S, Raj B and Gupta R D 2015 Leakage current reduction in finfet based 6t sram cell for minimizing power dissipation in nanoscale memories *2015 5th Nirma University Int. Conf. on Engineering (NUiCONE)* 1–5
[25] Gupta V, Sribhuvaneshwari H, Das T K, Khandelwal S and Akashe S 2024 Innovative circuit level methodology for finfet based low-power 6t sram cell design *2024 28th Int. Symp. on VLSI Design and Test (VDAT)* 1–5
[26] Jeong H, Kim T, Kang K, Song T, Kim G, Won H-S and Jung S-O 2015 Switching pmos sense amplifier for high-density low-voltage single-ended sram *IEEE Trans. Circuits Syst. I* **62** 1555–63
[27] Kulkarni J P, Kim K and Roy K 2007 A 160 mv robust schmitt trigger based subthreshold sram *IEEE J. Solid-State Circuits* **42** 2303–13
[28] Singh J, Mohanty S P and Pradhan D K 2012 *Robust SRAM Designs and Analysis* (Springer Science & Business Media)
[29] Seevinck E, List F J and Lohstroh J 1987 Static-noise margin analysis of mos sram cells *IEEE J. Solid-State Circuits* **22** 748–54
[30] Gupta R and Dasgupta S 2019 Process corners analysis of data retention voltage (drv) for 6t, 8t, and 10t sram cells at 45 nm *IETE Journal of Research* **65** 114–9
[31] Gupta V, Khandelwal S, Mathew J and Ottavi M 2018 45 nm bit-interleaving differential 10t low leakage finfet based sram with column-wise write access control *2018 IEEE Int. Symp. on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)* 1–6
[32] Islam A and Hasan M 2012 Leakage characterization of 10t sram cell *IEEE Trans. Electron Devices* **59** 631–8
[33] Zhao W and Cao Y 2007 Predictive technology model for nano-CMOS design exploration *J. Emerg. Technol. Comput. Syst.* **3** 1–es
[34] Rawat B and Mittal P 2022 A comprehensive analysis of different 7t sram topologies to design a 1r1 w bit interleaving enabled and half select free cell for 32 nm technology node *Proceedings of the Royal Society A* **478** 20210745

Electrical Engineering
https://doi.org/10.1007/s00202-024-02930-7

ORIGINAL PAPER

A novel deep learning framework for PV module thermal condition monitoring

Rahma Aman[1] · Mohammad Rizwan[1,3] · Astitva Kumar[2]

Received: 30 October 2024 / Accepted: 15 December 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
The increasing consumption of solar energy has generated a requirement for efficient techniques to monitor and evaluate the condition of photovoltaic modules. This research approaches the difficulty by developing a novel transfer learning framework that employs thermographic images and deep convolutional neural networks (DCNNs) for non-intrusive and reliable photovoltaic module monitoring. The framework analyzes variations in temperatures over time with the help of a thermal imager, further leading to the categorization of photovoltaic modules as either functional or faulty. Three fault scenarios, namely shadowing causing less than 15% power loss, dust hotspots causing 15–20% power loss and cracking causing above 20% power loss, lead to the classification of faults based on power loss. These criteria were determined by empirical data analysis where the power was recorded by the PV analyzer 9018BT. Furthermore, the images captured under photovoltaic module condition monitoring were used to train the different pretrained DCNN transfer learning models, such as ResNet-50, GoogLeNet and VGG-19, for photovoltaic module health classification, whether healthy or not, and compare the model's efficacy with the proposed VGG-16-based architecture. The findings indicate that the suggested framework reaches an accuracy of 95.5%, surpassing ResNet-50 (83.3%), GoogLeNet (73.5%) and VGG-19 (90.9%). The results underscore the efficacy of the suggested method in precisely identifying and categorizing photovoltaic module defects. This study provides a significant contribution by presenting an economical, precise and scalable approach for the monitoring of photovoltaic module conditions. The novel use of transfer learning for thermographic data offers a powerful tool for early problem diagnosis, enhancing maintenance strategies and prolonging the lifetime of photovoltaic systems.

Keywords Deep learning · Thermal management · Hotspot · Solar photovoltaic · Transfer learning

Abbreviations

Adam optimizer	Adaptive moment estimation optimizer
CNN	Convolutional neural network
DCNN	Deep convolutional neural network
DNN	Deep neural network
EL	Electroluminescence
F1-Score	F1 measure score (harmonic mean of precision and recall)
FDC	Fault detection and classification
FDD	Fault detection and diagnosis
IR	Infrared
I–V Curve	Current–voltage curve
MCC	Matthew's correlation coefficient
MATLAB	Matrix laboratory
PPL	Percentage power loss
PR	Performance ratio
PV	Photovoltaic
P–V Curve	Power–voltage curve
ResNet	Residual network
Sn	Sensitivity (recall or true positive rate)
SoftMax	Soft maximum function

✉ Rahma Aman
rahmaaman_2k21phdee08@dtu.ac.in

Mohammad Rizwan
rizwan@tsue.uz

Astitva Kumar
astitva.kumar@nsut.ac.in

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

[2] Department of Electrical Engineering, Netaji Subhas University of Technology, Delhi 110078, India

[3] Department of Information Systems and Technologies, Tashkent State University of Economics, Tashkent, Uzbekistan

Published online: 23 December 2024

Sp	Specificity (true negative rate)
Sgdm	Stochastic gradient descent with momentum
STC	Standard test conditions
TiS75 +	Thermal imager series 75 +
VGG	Visual geometry group

List of symbols

MW	Megawatt
W_p	Watt peak
V_o	Open circuit voltage
V_m	Voltage at maximum power
I_m	Current at maximum power
I_s	Short circuit current
I_{pv}	Photovoltaic current
T_c	Cell temperature
G	Irradiance
I_o	Reverse saturation current of the diode
n	Ideality factor
T	Temperature in Kelvin
k	Boltzmann constant
R_s	Series resistance
R_{sh}	Shunt resistance
t_p	True positives
t_n	True negatives
f_p	False positives
f_n	False negatives

1 Introduction

As of September 2024, India has an installed generation capacity of a total renewable energy of 2,09,638 MW, with solar power contributing 90,762 MW, as per the Ministry of New and Renewable Energy (MNRE) [1]. Despite growing commodity prices, utility-scale solar photovoltaic (PV) remains the least expensive choice for new energy production in the vast majority of nations throughout the globe. Distributed solar PV, such as rooftop solar on buildings, is also expected to increase quicker as retail power costs rise and regulatory support grows. Clean power is produced by PV systems without the release of harmful emissions. Despite these advantages, PV modules in powerhouses face humid and high-temperature environments that might cause damage or act as a hotspot owing to the partial shading of objects on the PV modules [2]. Further, there are grid stability issues with integration of PV systems, [3] analyzes energy efficiency measures (EEMs) in high-renewable energy source (RES) power systems, evaluating cost-effectiveness using Levelized Avoided Cost of Solar Electricity (LACoSE) calculations, using Brazil as a case study for wider relevance in tackling emissions and stability issues, and for the above limitation certain researchers have tried to overcome the drawbacks. The use of perovskite solar cells (PSCs) in vertical building-integrated photovoltaics (BIPV) applications increases energy output, hence enhancing grid stability using the decentralization of power production, minimization of transmission losses and optimization of overall energy efficiency and management [4–6]. The author in [7] assesses the environmental efficacy of organic and perovskite solar cells using graphene transparent electrodes (GTEs), underscoring their reduced energy payback time and stressing the need for extended lifespans to ensure sustainability. [8] illustrate that textile envelope-integrated flexible photovoltaic (TE-FPV) systems, especially those using perovskite cells, provide superior sustainability, recyclability and efficiency, contributing to diminished grid instability and improved resilience while supplying essential environmental data for building-integrated energy solutions. Decentralizing power production with TE-FPV systems generates electricity locally on building surfaces, reducing transmission losses, minimizing grid congestion and improving grid resilience by alleviating stress on central power plants during peak demand [9]. Further, a solar cell or component of a solar cell that is overheated in relation to adjacent cells is said to have a hotspot. Hotspot cells do not contribute to the generation of electricity [10–12]; this issue needs to be addressed for better PV system performance. Due to shading of panels, solar power is reduced which is also a concern while setting up of PV system; the author in [13] examines the effects of partial shadowing on BIPV systems, presenting a simpler model for performance prediction, emphasizing the importance of shadow distribution and suggesting optimum module configurations for enhanced efficiency. High PV solar cell temperature due to hotspotting can damage the cell and lead to secondary breakdown. Therefore, in [14] the evaluation of the percentage of power loss (PPL) due to hotspots affecting PV modules is assessed. The lowest disparity between the performance ratio (PR) of healthy and hotspotted PV modules is 0.83%, while the most significant difference is 15.47%. The detection and categorization of faults in solar PV modules are imperative for optimizing module power output. These faults can broadly be classified into permanent, encompassing electrical disconnection, wiring losses and aging, and transient, involving factors like dust, shadow and bird droppings [15]. Early detection of these faults is essential for timely intervention and corrective measures. This research emphasizes the significance of not only identifying the presence of defects but also categorizing them based on fault types. By distinguishing between permanent and transient faults, this study aims to provide a comprehensive understanding of the nature of faults in solar

PV modules, facilitating swift and targeted actions for maintenance and performance improvement. In [16] a data-driven method is presented to identify and categorize various solar PV system anomalies.

Deep learning surpasses machine learning (ML) methodologies in terms of feature extraction and capability of learning, which contribute to enhanced classification performance that is both accurate and resilient. In addition to providing generalization and nonlinear representation capabilities for massive data sets, deep learning models may also do so [17] a deep learning-based PV module defect detection method using electroluminescence images. It addresses two technical challenges: generating high-quality EL images for limited samples and creating an efficient model for automatic defect classification using the generated EL image. A deep hybrid learning model discussed in [18] identifies the kind and location of hot areas. Categories of experiments are done on the acquired dataset to benchmark the suggested framework. It has limited fault detection capability, is expensive and time-consuming and is incapable of quickly identifying the actual site of the defect [19]. This research in [20] describes a novel method for identifying modules that detect hot areas and locating them. To provide a robust detection framework, two new region-based convolutional neural networks are combined. The key contribution is the use of thermography and telemetry data to offer a module condition monitoring response. CNN model automatically classifies thermographic pictures into hotspot and operational classes. Various preprocessing approaches were assessed to decrease picture noise. Using a dataset with various acquisition procedures, the model achieved 99% accuracy [21]. The suggested technique beats benchmarked alternatives in effectiveness and efficiency; [22] proposes an end-to-end deep learning pipeline that uses EL pictures to identify, localize and segment cell-level irregularities in solar PV modules. The modular pipeline uses modified Faster-RNN for object recognition, Efficient Net for image classification and autoencoder for weakly supervised segmentation. The modular pipeline enables upgrading deep learning models to state-of-the-art upgrades and adding new functions. The suggested method in [23] automatically classifies thermographic pictures from the system's convolutional neural network (CNN) with 98% accuracy in two-minute testing. This method is faster than others in the literature, with lower expenses, diagnostic time and power production losses. This study develops a computer vision method to semi-automatically extract PV modules from thermographic UAV footage. Deep two-dimensional convolutional neural networks (CNN) are utilized in a novel approach to extract features from 2-D scalograms of PV system data, thereby facilitating the detection and classification of faults. An exhaustive quantitative evaluation of the proposed method is juxtaposed with previous approaches to PV array failure classification, such as traditional ML and deep learning [24]. [25] presents two improved RF classifiers for FDD: the Euclidean distance-based reduced kernel RF (RK-RFED) and the K-means clustering-based reduced kernel RF. [26] uses two sophisticated convolutional neural network models to categorize the modules' defect type and identify its area of interest. The proposed method compares classification models using the F1-score, with the ResNet-50 transfer learning model scoring 85.37%. [27] investigates the use of artificial intelligence (AI) to monitor the health of solar PV panels without invasive procedures. Solar PV panels must be properly maintained to ensure that they operate safely, reliably and efficiently. An AI-based, noninvasive condition monitoring technique is used to diagnose the health state of solar PV panels using thermal pictures of the panels. The performance of PV systems is examined in [28–30]; the author uses support vector machine (SVM) classification. The method streamlines data processing by transforming observed data into multiclass variables, allowing for effective investigation of complex elements influencing PV system performance over time.

- Despite the increasing number of research papers in the field of fault detection in solar PV modules, a noticeable research gap exists, with the majority focusing on the efficient detection of abnormalities in just two categories. This paper fills in that gap by presenting a new method that uses a pretrained convolutional neural network (CNN) that can sort infrared (IR) picture patterns into up to four different types of abnormalities. Unlike existing research, which predominantly explores the efficiency of binary classification, our study extends the scope to include four classes: hotspot modules, shadow modules, cracked modules and healthy modules. The transfer learning models implemented in this study leverage thermal images from a thermal imager in the MATLAB environment. Our work employs visual fault detection and classification (FDC) methods, specifically utilizing deep learning approaches for segmenting binary mask images. This aids in the identification of hotspots indicative of PV module defects. The research compares the performance of pretrained deep CNNs (ResNet-50, GoogLeNet, VGG-16 and VGG-19) based on key performance metrics, including Accuracy, Precision, Sensitivity, Specificity, F1-score and Matthew's correlation coefficient. The Utilization Electrical Engineering (UEE) Laboratory at Delhi Technological University (DTU), New Delhi, India, provided the thermal images for model training. The dataset comprises diverse anomalies, such as hotspots induced by dust deposition, fractures and shadow-affected modules. Notably, limited research papers explore the dimension of classifying faulty and healthy modules based on power loss percentage. The motivation behind the proposed work is outlined in Table 1. Shading, dust and cell faults generate PV system hotspots

and mismatch losses, which are common issues that cause lowering efficiency and perhaps damaging the system. Advanced condition monitoring technologies, including thermal imaging, electrical analysis and AI-driven image processing, have greatly improved defect detection and mitigation. These innovations enhance PV system performance, reliability and early detection. Additionally, the paper investigates crucial parameters impacting simulation accuracy and highlights literature gaps in the context of faulty image classification based on power loss percentage. The findings contribute to the enhancement of fault diagnosis precision in PV systems, addressing the identified research gap and promoting a more comprehensive understanding of solar PV module defects for improved efficiency and performance assessment. The research objectives of the paper:

- This research is to demonstrate a clear relationship between the thermal properties of PV modules and their overall health.
- Based on the power loss percentage criteria, the faults are categorized, and further thermal images have been captured to train the deep CNN architectures such as ResNet-50, GoogLeNet and VGG-19.
- A comprehensive comparative study is performed to assess the proposed framework in comparison to prominent deep CNN architectures such as ResNet-50, GoogLeNet and VGG-19. The results of this evaluation illustrate the better performance of the proposed framework.
- The framework accurately determines the health state of PV modules by comparing power loss percentages to module conditions, providing useful information for maintenance and improvement.

The manuscript has six sections: Section 1 presents the introduction, while Section 2 addresses the condition monitoring of PV modules. Section 3 outlines the proposed methodology. Section 4 defines the transfer learning methodology and associated performance criteria. Section 5 has two segments: The first segment refers to the distribution of classes based on power loss percentages, while the subsequent segment focuses on the classification of these classes using various transfer learning models. Ultimately, Section 6 concludes the manuscript.

2 Condition monitoring of PV modules

A PV cell is very sensitive to the ambient meteorological conditions, and its output current [36, 37] can be represented as below

$$I_{\mathrm{pv}}(T,\,G) = \left(I_{\mathrm{pv(stc)}} + K_i(T-25)\right)G \times 10^{-3} \tag{1}$$

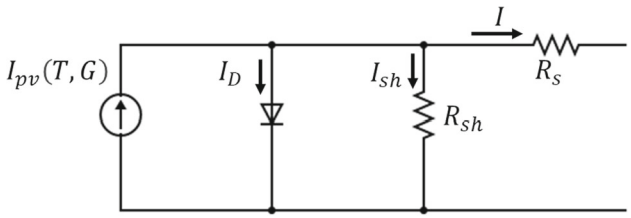


Fig. 1 Single-diode model PV cell

Here, $I_{pv(stc)}$, K_i, T and G are PV current at standard test conditions, temperature coefficient, cell temperature and irradiance, respectively. A single-diode PV cell model can be considered for easy understanding of I-V characteristics and its variation with changing temperatures. This single model is given in Fig. 1.

$$I = I_{\mathrm{pv}} - I_s\left[\exp\left(\frac{q(V + R_s I}{\mathrm{akT}}\right) - 1\right] - \frac{V + R_s I}{R_{\mathrm{sh}}} \tag{2}$$

Here, I_s is reverse saturation current of diode, a is ideality factor, T is temperature in Kelvin, k is Boltzmann constant, and R_s and R_{sh} are series resistance and shunt, resistances respectively.

The series parallel configuration of the PV cell creates a PV module; this is the fundamental block for PV system analysis. The impact of changing temperature and irradiance on performance of PV cell can be studied from Eq. (1) and (2). Furthermore, this impacts the performance of the PV module. This paper explores the effect of thermal stress on the PV module and monitors PV module condition for safe and effective operations. PV module condition monitoring refers to the ongoing process of assessing and analyzing the performance, health and operational status of solar PV modules. The current–voltage (I-V) curve and power–voltage (P–V) curve for varying temperature are shown in Fig. 2. This clearly shows the decrease in performance of PV modules with increasing temperature. As the modules are continuously subjected to thermal stress the overall performance, lifetime and modules condition are to be studied.

The objective is to ensure the reliable and efficient operation of these modules over their operational lifespan. Condition monitoring involves the use of various techniques, technologies and tools to detect, diagnose and mitigate potential issues or faults in PV systems. This proactive approach helps to prevent unexpected failures, optimize energy production and extend the overall lifespan of the PV installation. Solar PV modules can be affected by a variety of factors that have an impact on their performance. Dirt and dust collection on the modules, resulting in lower sunlight absorption, and shading, which can dramatically diminish power output, are common issues. Figure 3 explains the various stages of PV module condition monitoring.

Table 1 Relevant work

Ref	Classification	detection	Faults categories	Remarks
[31]	SVM, CNN	Electroluminescence images	Degradation	EL imaging enables automated PV defect detection via SVM and CNN. where CNN is better
[32]	Multi-Attention U-net	Electroluminescence images	Degradation	MAU-Net improves fault identification in EL images by the use of attention processes
[33]	CNN and encoder-decoder	Aerial images from multi-copters	Shading	This work describes a precise deep network for identifying and prioritizing bird droppings on PV modules
[34]	CNN and SVM	Thermographic images from ground-based and UAVs	Degradation and shading	YOLOv8-GD improves PV defect detection with higher accuracy, efficiency, reduced model size and adaptability
[35]	CNN and VGG-16	Thermographic images from ground-based and UAVs	Degradation	The hybrid system improves PV flaw detection efficiency and accuracy
[21]	CNN	Electroluminescence images	Thermographic images from ground-based and UAVs	The proposed method employs CNN to identify solar failures accurately and remotely

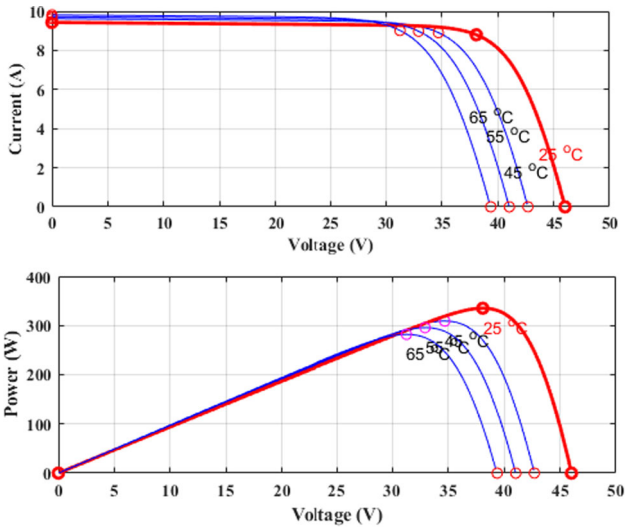


Fig. 2 I–V and P–V characteristics for the PV module for thermal variations

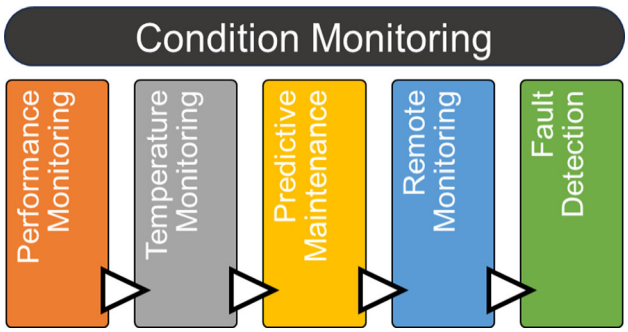
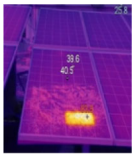
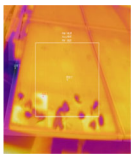
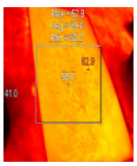
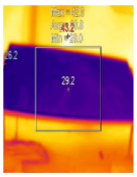


Fig. 3 PV module condition monitoring

Manufacturing flaws can cause cell or module failures, reducing overall system efficiency. Hailstorms, for example, can cause physical damage to modules. Reduced performance is also caused by age-related degeneration, bird droppings and nesting. Regular cleaning, correct installation, monitoring and timely maintenance are critical for addressing these difficulties and ensuring PV module systems long-term efficacy. Considering the power-reducing faults, different classes are defined and thermal images have been captured.

In the proposed work, the thermal IR images are collected with the help of thermal imager TiS75 + . These images are classified into four different categories of PV module condition. This condition classification of PV modules is shown in Table 2. It summarizes different types of faults and healthy images that may occur and estimates potential losses due to such faults. It includes dust hotspots, module shading caused by nearby objects and cracking failures.

Table 2 Classification of condition of PV modules

Thermographic IR Images	Class	Definition
	Hotspots and multi-hotspots	The rate increases or decreases by 1.5% to 6.2%, based upon the location of the PV facility. Severe dust deposition can result in as much as 33% power loss [38, 39]
	Shadowing/bird dropping	Shaded solar PV modules reduce power output by 20% [40]
	Cracked	Studies indicate that microfractures may cause a power loss of over 2.5% in output power and lesser than 21% for solar PV modules with many microcracks [14, 41]. Large cracks can have power loss up to 30%
	Healthy module	No power loss
Discoloration of cell		This type of faults results from prolonged UV exposure or defects in manufacturing, diminishing light absorption. Power loss fluctuates; prolonged discoloration may decrease efficiency by 10%–15% [42]
Material aging		Degradation of encapsulant material (ethylene vinyl acetate) resulting in less protection of PV cells. Results in an annual power decrease of 2% to 5% [43]
SVM degradation due to voltage stress		Occurs when excessive voltage stress leads to power decline particularly in humid conditions. May result in a power loss of up to 30% [44]

3 Proposed methodology

Effective PV module condition monitoring is crucial for maximizing the return on investment in solar energy systems. It provides insights into the system's health, aids in optimizing energy production, reduces maintenance costs and ensures the long-term reliability of the PV installation. Advanced technologies, such as AI and ML, are increasingly being integrated into condition monitoring systems to enhance accuracy and efficiency in fault detection and analysis.

In the proposed work, condition monitoring is performed for four different classes of PV modules. These classes have been defined after calculation of power loss for each of the various conditions with respect to the standard test condition rated value. The four classes are healthy, hotspot, shadow and cracked. Specific shadow, dust and cleaned PV modules were deployed for mimicking a real-time scenario so that a thermal infrared image can be obtained with the help of thermal imager. For another class 'cracked', the dataset was generated considering a faulty cracked module.

The primary goal of this work is to evaluate hotspot-related flaws, which are thought to be the primary factor influencing the overall efficiency of solar PV modules. A significant finding of the research is that pinpointing hotspots is essential to

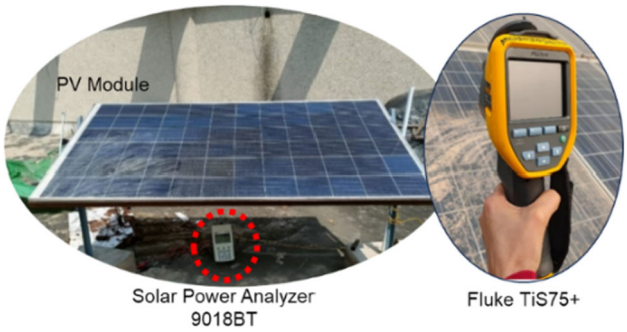


Fig. 4 PV module condition monitoring testbed

Table 3 Specification of PV module testbed

Modules parameters	Values
Maximum power, P_m	335 W_p
Voltage at maximum power, V_m	38.8 V
Current at maximum power, I_m	8.80 A
Open circuit voltage, V_o	46.02 V
Short circuit current, I_s	9.43 A
Tolerance	± 3%

precisely evaluate their deleterious influence on PV cell efficiency. A defect detection algorithm that can preprocess data more accurately and discern between defective and healthy modules is needed to enhance PV module performance. In a real-time setting, the power loss in relation to the rated power is calculated to analyze problematic modules. The errors are classified using deep learning CNN image classifiers. Images for fault identification are captured using a fluke thermal imager TiS75 + , as illustrated in Fig. 4, and its specifications are: resolution: 384 × 288 (110,592 pixels), field of view: 42° × 30°, thermal sensitivity: 40 mK, temperature range: − 20 °C–550 °C, battery life: ≥ 3.5 h (without WiFi), and weight: 0.885 kg (1.95 lbs); here the hotspots are formed by dust deposition and shading is done on the module on purpose to explore the effect of power loss on different type of PV modules. IR thermography uses midwave (MWIR, 3–5 m) or longwave (LWIR, 7–14 m) infrared sensors to produce thermal images or thermograms of the objects under inquiry. IR thermography can measure PV module surface temperature and temperature trend. The hotspot is formed by introducing dust to solar PV modules, and the dust is spread over half of the module to create a real-time scenario to get hotspot. The PV module testbed specification is mentioned in Table 3.

The proposed work is described below in three stages as shown in Fig. 5.

Stage 1: Data collection and filtering involving steps such as thermal picture is generated; based on power loss, the module is categorized into four different classes; and data splitting between training (80%) and testing (20%), batch normalization and 3 × 3 convolution filter are used to divide images into groups. A detailed preprocessing steps are: Image rescaling—adjust photos to a uniform resolution (224 × 224 pixels) to ensure compatibility with deep learning models (ResNet-50, GoogLeNet, VGG-19 and VGG-16).

Standardization—normalize pixel intensity values to [0,1] or [− 1,1] for uniform input and expedited model convergence with the formula:

$$I' = I - \min(I)/(\max(I) - \min(I))$$

Noise reduction—employ filters such as median filtering, which eliminates salt-and-pepper noise, and Gaussian

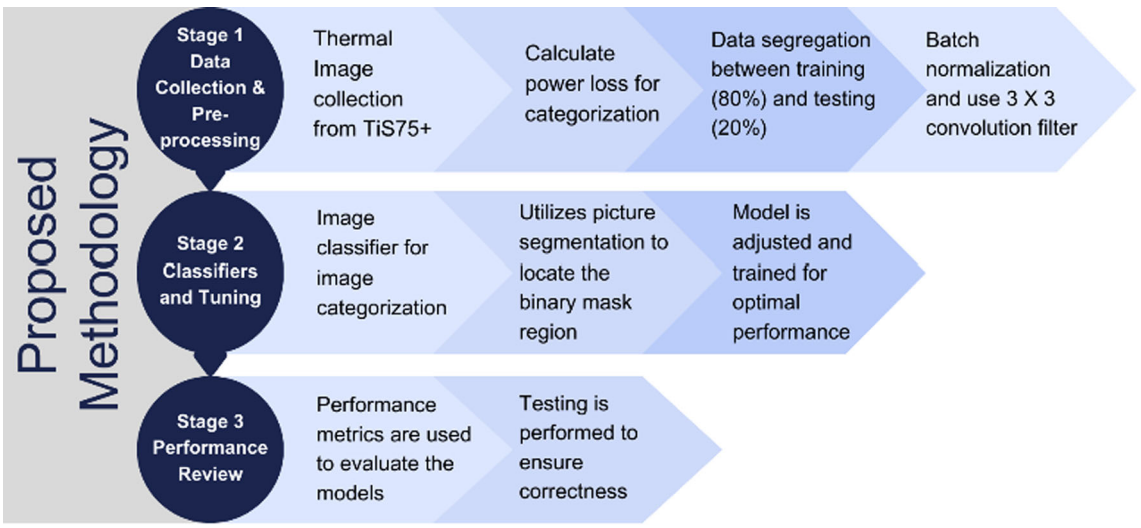


Fig. 5 Proposed methodology for PV module condition monitoring

smoothing, which maintains edges while reducing high-frequency noise.

Contrast augmentation—utilize histogram equalization or adaptive histogram equalization to enhance visibility in photos with little temperature variance.

Segmentation—identify areas of interest (e.g., thermal anomalies) with techniques such as thresholding.

Temperature calibration—translate pixel intensity into real temperature values using the calibration data of the thermal camera.

Stage 2: Classifiers and tuning involves a step for image categorization; a deep learning model (ResNet-50, GoogLeNet, VGG-19, VGG-16) was constructed. The model then utilizes picture segmentation to locate the binary mask region and determine the incorrect class. The model is adjusted and trained for optimal performance.

Stage 3: Performance review stage evaluates the model using performance metrics. The captured pictures are used to categorize faults on the SoftMax layer that are completely related. Furthermore, further testing is performed to ensure correctness (Fig. 6).

3.1 Feature extraction and SoftMax layer

This proposed works uses transfer learning model as feature extractor. The proposed framework uses key extraction features and softmax layer from VGG-16-based CNN architecture. This architecture creates classification layers for 660 image datasets. Some of the images from the different classes of dataset are shown in Fig. 7. Feature extraction and categorization from input training photos is performed using a pretrained CNN model. CNN learns to extract high-level features which helps in discriminating of four class labels, i.e., hotspot, cracked, shadowed and healthy modules. The first is the feature extraction phase, in which a series of strategically arranged convolution and pooling layers reduce the input image to a set of feature maps. The second is the classifier phase, in which the features are fed into a series of fully connected or hidden layers, and last connected to an output layer. The SoftMax layer helps the CNN to make predictions for classification tasks based on the features extracted from the input image it converts the raw data into probability so that the classification can be further done. The proposed framework for PV module conditioning is shown in Fig. 6. The mathematical expression for SoftMax is as follows.

$$\text{SoftMax}(Z) = \frac{e^{Z_i}}{\sum_{j=1}^{k} e^{z_j}} \tag{3}$$

The function notation Z represents the input vector elements.

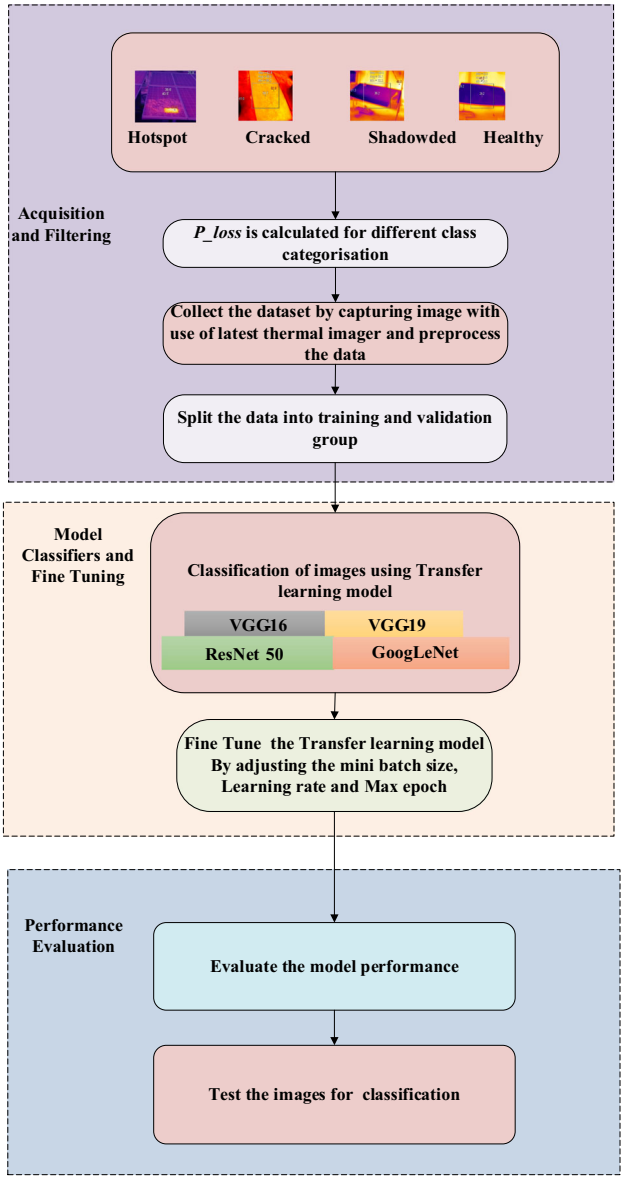


Fig. 6 Framework of the proposed PV module thermal condition monitoring

4 Transfer learning method and performance metrics

In AI, specifically in ML, it is a common practice to leverage the knowledge and experience acquired while solving a particular problem to analyze and tackle other related (or potentially unrelated) problems and challenges. This is particularly true in the domain of ML, which is a subset of AI. Transfer learning is a kind of learning that combines AI and ML. Efficiency is the driving force behind such a tendency, which, in many cases, has restricted research hindered by computer resources. Using models that have been trained in the past may assist bridge certain gaps, but one

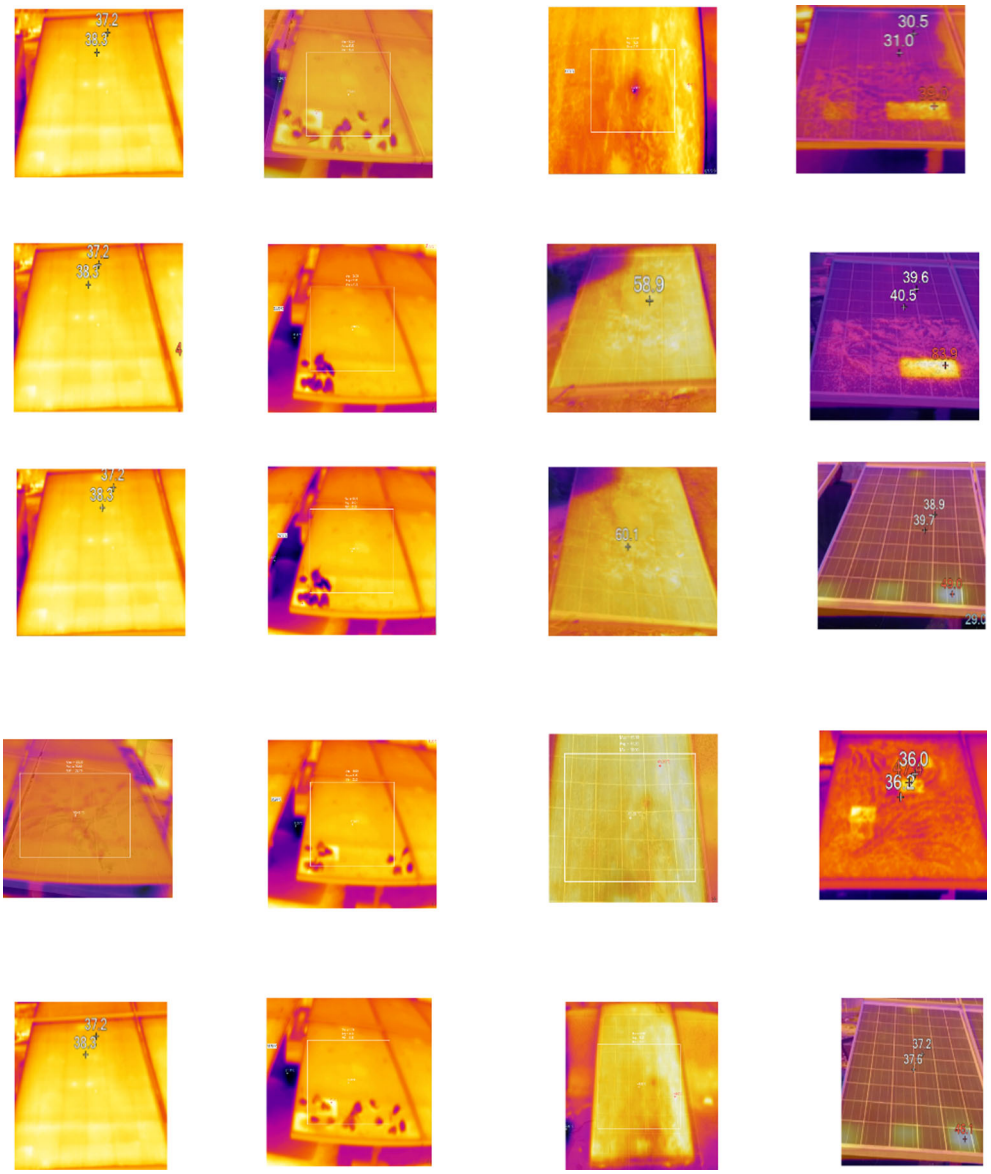


Fig. 7 Infrared (IR) images illustrating PV module conditions across various fault classifications

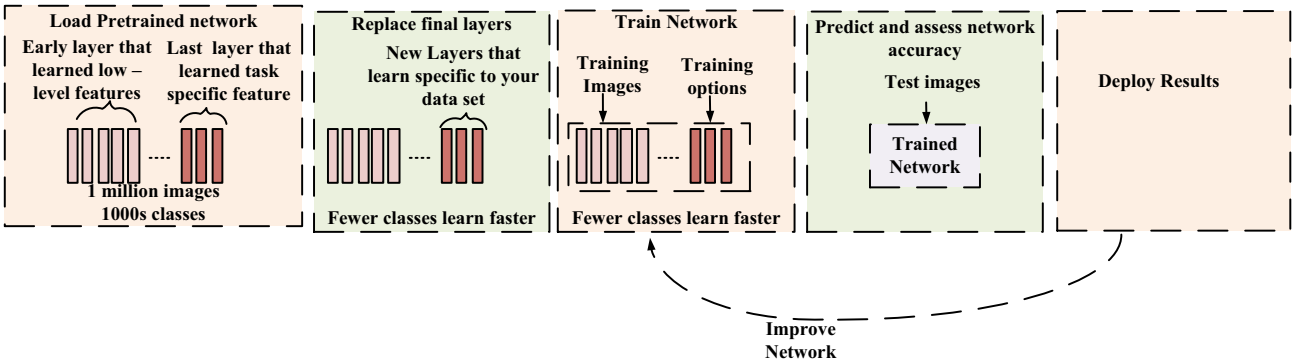


Fig. 8 Transfer learning process of a deep neural network

must be cautious. It is vital to use a model that has already been pretrained in order to guarantee fine adjustment of its performance toward the new domain. Outcomes are more realistic than starting the whole process of model training from the beginning. This procedure is explained in Fig. 8.

In this work, four different transfer learning methods have been compared for PV module condition monitoring. These methods are ResNet 50, GoogLeNet, VGG-19 and proposed VGG-16. In the context of PV module condition monitoring, a well-constructed confusion matrix and associated performance metrics help in assessing how effectively the transfer learning model can distinguish between healthy and faulty PV modules. This evaluation is essential for fine-tuning the model, optimizing its performance and ensuring accurate fault detection in real-world applications. To evaluate the effectiveness of the aforementioned methods, six performance evaluation metrics are employed: Accuracy (A), Precision (P), Sensitivity (S_n), Specificity (S_p), F1-score ($F1$) and Matthew's correlation coefficient (M_{CC}). M_{cc}, is a statistical measure used to assess the quality of binary classification models. The Accuracy provides an evaluation of network performance for different PV modules, while the concept of P refers to the ratio of truly positive results divided by the anticipated outcomes are expected to be favorable. S_n represents the proportion of true positive results in relation to all pertinent samples. S_p statistic quantifies the ratio of true negatives to the total number of negatives that are anticipated to have negative outcomes. $F1$ computes the harmonic average in a given context. The combination of the elements P and S_n. The MCC algorithm also calculates the disparities between the observed and anticipated PV module condition [45]. The key components of a confusion matrix are:

- True Positive (t_p): Instances where the model correctly predicts the positive class. In PV module condition monitoring, this would correspond to correctly identifying faulty modules. True Negative (t_n): Instances where the model correctly predicts the negative class. In PV module monitoring, this would be correctly identifying healthy modules.
- False Positive (f_p): Instances in which the model forecasts the positive class inaccurately. In PV module monitoring, this would imply wrongly categorizing a healthy module as defective.
- False Negative (f_n): Instances in which the model predicts the negative class incorrectly. In PV module monitoring, this would be incorrectly classifying a faulty module as healthy.

Based on the confusion matrix that was acquired, the performance metrics can be defined mathematically from Eq. (4) to (9)

$$A = \frac{t_p + t_n}{t_p + f_p + t_n + f_n} \tag{4}$$

$$P = \frac{t_p}{t_p + f_p} \tag{5}$$

$$S_n = \frac{t_p}{t_p + f_n} \tag{6}$$

$$S_p = \frac{t_n}{t_n + f_p} \tag{7}$$

$$F1 = \frac{2 * t_p}{2 * t_p + f_p + f_n} \tag{8}$$

$$M_{\mathrm{cc}} = \frac{\left(t_p \times t_n\right) - \left(f_p \times f_n\right)}{\sqrt{\left(t_p + f_p\right)\left(t_p + f_n\right)\left(t_n + f_p\right)\left(t_n + f_n\right)}} \tag{9}$$

5 Results and discussions

This study undertook a meticulous investigation and assessed the discriminative capabilities of leading deep learning architectures—ResNet, GoogLeNet, VGG-19 and proposed VGG-16 based CNN framework. By harnessing the power of transfer learning and initializing models with pretrained weights sourced from extensive datasets, the focal point is the adaptation of preexisting features to the unique nuances inherent in PV systems for module condition monitoring. Rigorous evaluation was performed on a dedicated testing set, using comprehensive performance metrics, such as Precision, Sensitivity, Accuracy, Specificity, Matthew's correlation coefficient and F1-score.

The results and discussion is divided into two major sections. In the first section, the PV module condition is analyzed using P_{loss}. This loss in power is calculated using the data collected by solar PV analyzer 9018BT; these data are the power output of SPV modules. P_{loss} is computed by distinguishing between defective and healthy modules. Additionally, the section divides the data into four classes and constructs an image dataset for the purpose of training the transfer learning model, where out of 660 images total 251, 136, 168 and 105 images are of hotspot and multi-hotspot, shadowing, cracked and healthy modules, respectively. The percentage of the four classified solar modules are shown in Fig. 9. Further in the second section, this dataset is given as input to various transfer learning model, and the classification performance of the various transfer learning models discussed in Sect. 4 is evaluated on the datasets obtained from thermal imager captured for four days. The classes used for classifications are divided into two parts: faulty and healthy modules; faulty modules

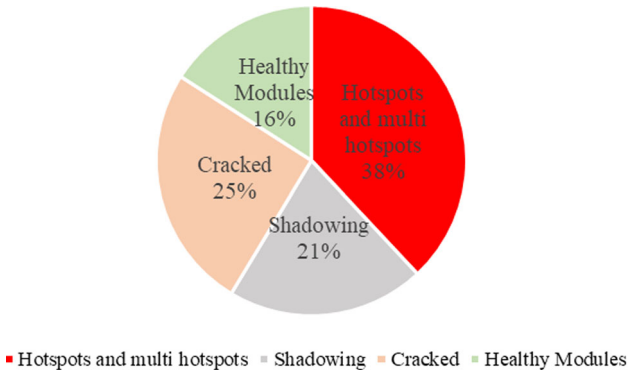


Fig. 9 PV module classification according to P_{loss}

are divided into three classes which are hotspot and multi-hotspot, shadowing by nearby objects and cracked modules and healthy modules.

5.1 Comparative performance for different classes on PV modules

Different faulty conditions performance is considered, where the power is recorded on three different real-time PV module fault scenarios; i.e., shadowing, cracked and hotspots which are created by dust and further compared the power recorded against healthy module. The magnitude of the power loss is determined by a variety of parameters, including the severity of the hotspot, the degree of shadowing and fracture damage.

The power is recorded on a 335 W PV module for 10 min interval for different conditions by the solar system analyzer 9018 B for 4 h. The power recorded is shown in Fig. 10. Different classes of PV modules are discussed below.

Healthy module—The module is cleaned and then power is recorded.

Shadowing module—The shadowed condition is been mimicked as short wide and short narrow shading condition, discussed in [46].

Cracked module datasets—A same 335 W cracked module is considered to record power, and the cracked module considered is having severe cracks.

Hotspot modules—Dust was spread for mimicking a real-time scenario over half of the panel surfaces about 100 to 200 g/m^2. The power loss in different conditions is shown in Table 4, where it is observed that the maximum power loss is due to cracked module.

Based on this power loss, it can be defined that the solar PV modules are faulty. Here less than 20% power loss is considered to be recoverable, and more than 20% power loss the module is stated as cracked and not recoverable and needs replacement. Table 4 discusses the power loss percentage for different faulty conditions. The power loss calculations

Table 4 Power loss in different conditions of PV modules

Conditions	Healthy power (W)	Faulty power (W)	P_{loss}(%)
Hotspots and multi-hotspots	93.1672 W	77.9319 W	16.35%
Shadowing	93.1672 W	91.9368 W	2.22%
Cracked	93.1672 W	72.0247 W	22.69%

presented in Table 4 represent typical performance degradation owing to various faults in PV modules, with references supporting the amount of these losses. In hotspots and multi-hotspots (16.35% power loss) where hotspots are portions of the PV module that experience increased heating owing to localized damage or shading, these locations can cause a significant drop in the module's power output. A power loss of around 16.35% is consistent with findings in various studies that indicate hotspots can cause power losses ranging from 10 to 30% depending on the severity of the thermal damage [47]. Hotspot generation is associated with difficulties such as mismatched or damaged cells, which can increase resistance and generate extra heat, lowering efficiency. In shadowing conditions, there is a power loss of 2.22%; this power loss is due to the shadows occurring on PV systems that reduce the irradiance reaching the solar cells. This loss achieved shows mild shadowing, which is consistent with research stating that partial shading of PV systems can produce power decreases ranging from 2 to 20%, depending on the size of the shadow and the module's [48]. Partial shading affects efficiency, but it has a less severe impact than other faults such as cracks or hotspots. Further cracks in solar modules in this work have power loss of 22.69%; these cracks in PV modules damage internal electrical channels, resulting in significant power loss. [49] shows that cracks can impair power production by up to 20% or more, which supports the 22.69% power loss reported in fractured modules. Research suggests that physical damage, such as cracks, affects cell connection and significantly reduces performance. Here the choice of power loss threshold is validated through real-time monitoring of SPV systems. From Table 4 different faulty modules can be classified and thermal images are being captured for different stated classes.

5.2 Classification using transfer learning methods

The captured thermal infrared images are used to train the deep learning CNN transfer learning models GoogLeNet, ResNet-50, VGG-19 and proposed VGG-16 to classify faulty and healthy solar PV modules. The "GoogLeNet" alters the last three layers of the network; these three layers of the network—"loss3-classifier," "prob" and "output"—are

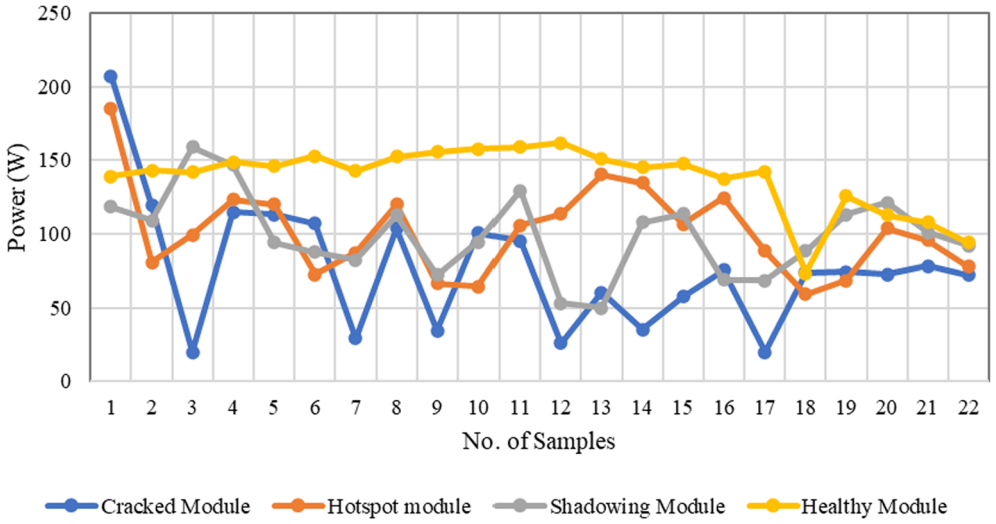


Fig. 10 Comparative performance of different classifications

Table 5 Deep neural network (DNN) parameters

DNN architecture	GoogLeNet	ResNet-50	VGG-19	Proposed VGG-16
Learning rate	0.001	0.001	0.001	0.001
Solver	Sgdm	Sgdm	Sgdm	Sgdm
Max epochs	2	2	2	2
Mini batch size	1	1	1	1
Validation frequency	50	50	50	50
Iteration	1056	1056	1056	1056
Iteration per epoch	528	528	528	528

replaced with "fully connected," "softmax" and "classification output" layers. Connect the last transferred layer ("pool5-drop_7 × 7_s1") to the new layers. The data are divided into eight. The fluke thermal imager TiS75 + is used to create hotspots and capture thermal pictures. The photos are acquired over a span of 4 days in order to gather a dataset encompassing four distinct classes. The purpose of these photos is to train deep learning CNN transfer learning models, such as GoogLeNet, ResNet-50, VGG-19 and proposed architecture, to accurately detect defective and non-defective solar PV modules. The "GoogLeNet" modifies the last three levels of the network. The last three layers of the network, namely "loss3-classifier," "prob" and "output," are substituted by "fully connected," "softmax" and "classification output" layers, respectively. Establish a connection between the most recently transferred layer, namely the "pool5-drop_7 × 7_s1" layer, and the newly introduced layers. The models—ResNet, GoogLeNet, VGG-19 and the proposed VGG-16-based CNN framework —were built and trained in MATLAB R2022b. This version of MATLAB provides extensive deep learning support, including pre-trained architectures, transfer learning capabilities and tools for tracking and analyzing training progress.

Further MATLAB's code was designed for different transfer learning models which were used to import pre-trained networks such as ResNet, GoogLeNet, VGG-19 and proposed VGG-16, which were then fine-tuned via transfer learning. This technique enabled the use of pretrained weights on the ImageNet dataset, which accelerated convergence and improved accuracy.

The training datasets trained with network parameters are shown in Table 5. The solver used for transfer learning models is stochastic gradient descent with momentum (SGDM). The data are split into training and testing as 80% and 20%, respectively, i.e., 528 training images and 132 testing images. The training plots of the input images are shown in Fig. 11. The training accuracy and loss plots of different transfer learning model, i.e., ResNet-50, GoogLeNet, VGG-16 and VGG-19, are shown in Fig. 11a and b. The models developed are optimized for better results, VGG-16 comes out to be 95.5% with highest computational time of approx. 50 min, and ResNet-50 performs with lowest computational time of 2 min 19 s with accuracy of 83.3% shown in Table 6. From Fig. 11 it is observed that the VGG-16 transfer learning model gives the best performances for classifying the various classes of dataset with highest accuracy.

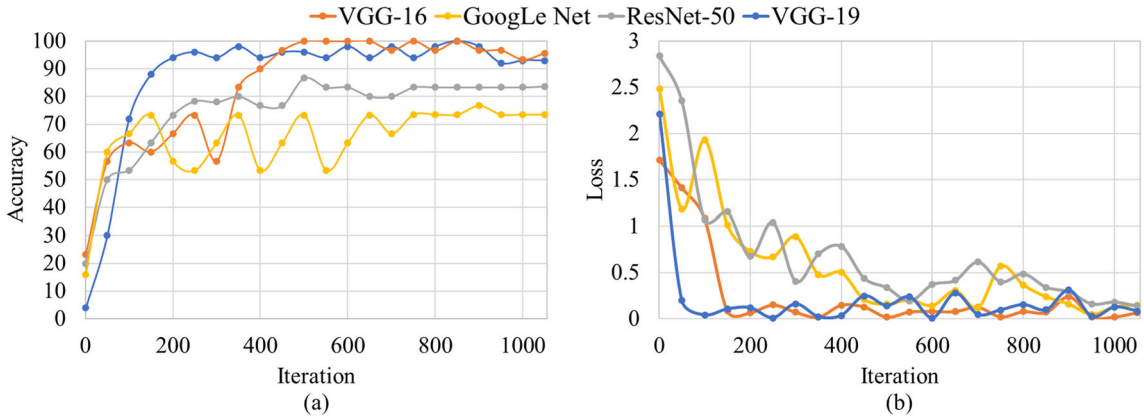


Fig. 11 Comparative performance of various DNN architectures during training phase: **a** Accuracy performance and **b** loss analysis

Table 6 Performance comparison of various DNN architecture during training phase

Transfer learning models	Accuracy	Computational time
ResNet-50	**83.3%**	**2 min 19 s**
GoogLeNet	73.5%	5 min 17 s
VGG-19	90.9%	52 min 12 s
Proposed VGG-16	**95.5%**	5 min 2 s

Performance comparison for complete dataset consisting of different cases is shown in Table 7, where 660 images are used to train the transfer learning models. Further, after training the entire data set is tested and confusion matrix is made so as to find various performance matrices, i.e., accuracy, sensitivity and specificity. The confusion matrix for each of the four DNN architectures for PV condition monitoring is presented in Fig. 12. The four DNN architectures (ResNet-50, GoogLeNet, VGG-19 and VGG-16) symbolize transfer learning method and their respective confusion matrix to evaluate the performance of each model. The accuracy plot for different classes is shown on the transfer learning models. A systematic approach is necessary to derive accuracy (sensitivity), precision or specificity values from a confusion matrix. Perform the necessary computations to determine the True Positives (*tp*), False Positives (*fp*), True Negatives (*tn*) and False Negatives (*fn*) associated with a particular class of interest. As demonstrated by the confusion matrix provided in Fig. 12c, the analysis is centered on shadowing. *tp* represents the number of instances that were accurately classified as shadowing, which is 14 in this instance. On the other hand, (*fp*) represents 13 irrelevant instances that were erroneously classified as shadowing, excluding the first column. This total is calculated by adding the values in the first row. Indicative of correctly identified irrelevant test instances (*tn*) is obtained by deducting (*fp*) from the total number of irrelevant test instances. (*tn*) is calculated in this context using three irrelevant classes and fourteen test instances per class as ($3 \times 14 - 13$), which equals 29. In order to identify (*fn*), which denotes shadowing instances that were incorrectly classified as other defects, the sum of the values in the first column (excluding the first row) yields 0. The sensitivity or accuracy of different classes is represented in the bottom row of the confusion matrix, whereas the precision of different classes is illustrated in the far-right corner. Subsequently, power measurements for both functional and malfunctioning modules are discussed. Thermal images are captured and categorized into four distinct classes, and a performance analysis is conducted on the aforementioned classes. The performance metrices can be calculated by the confusion matrix with the help of *fp, fn, tp, tn*; the testing data set is used to plot the confusion matrix; Fig. 12 describes the different classes sensitivity and precision; and rest parameters are shown in Table 7. This table concluded that VGG-16 transfer learning models perform best among the other transfer learning models discussed in this paper. Further comparison of the proposed work with other studies is given in Table 4. In [24] authors presents a unique deep learning technique that uses 2-D CNNs to identify PV system problems, such as partial shading, MPPT and others, with a 73.53% accuracy. It compares this approach to traditional machine learning and other deep learning methods, exhibiting higher performance. [50] provides a thermography and machine learning-based technique for detecting PV faults, which achieves 93.4% training efficiency and 91.7% testing accuracy by using texture features and an artificial neural network. When compared to different studies, the proposed framework is better.

6 Conclusion

The comprehensive evaluation of different DNN architectures (ResNet, GoogLeNet, VGG-19 and proposed VGG-16)

Table 7 Performance evaluation for different classes of PV module condition

Models	ResNet-50				GoogLeNet				VGG-19				Proposed VGG-16			
Performance metrics	Healthy module	Shadowing module	Cracked module	Hotspot module	Healthy module	Shadowing module	Cracked module	Hotspot module	Healthy module	Shadowing module	Cracked module	Hotspot module	Healthy module	Shadowing module	Cracked module	Hotspot module
A (%)	100	88.8	89.6	94	85.5	76.8	61.1	100	100	88.9	100	94	100	92.9	100	97.3
P (%)	100	77.7	70.6	88	100	51.8	35.3	100	100	77.7	100	88	100	77.7	100	100
Sn (%)	100	77.7	100	88	42	100	66.6	100	100	77.7	100	88	100	100	100	89.3
Sp (%)	100	92.6	86.1	96	100	69	59.3	100	100	92.6	100	96	100	90.5	100	100
F1 (%)	100	77.7	82.8	88	59.2	68.3	46.2	100	100	77.7	100	88	100	87.6	100	94.3
MCC (%)	100	70.4	77.9	84	59.3	59.8	22.5	100	100	70.3	100	84	100	83.9	100	92.8

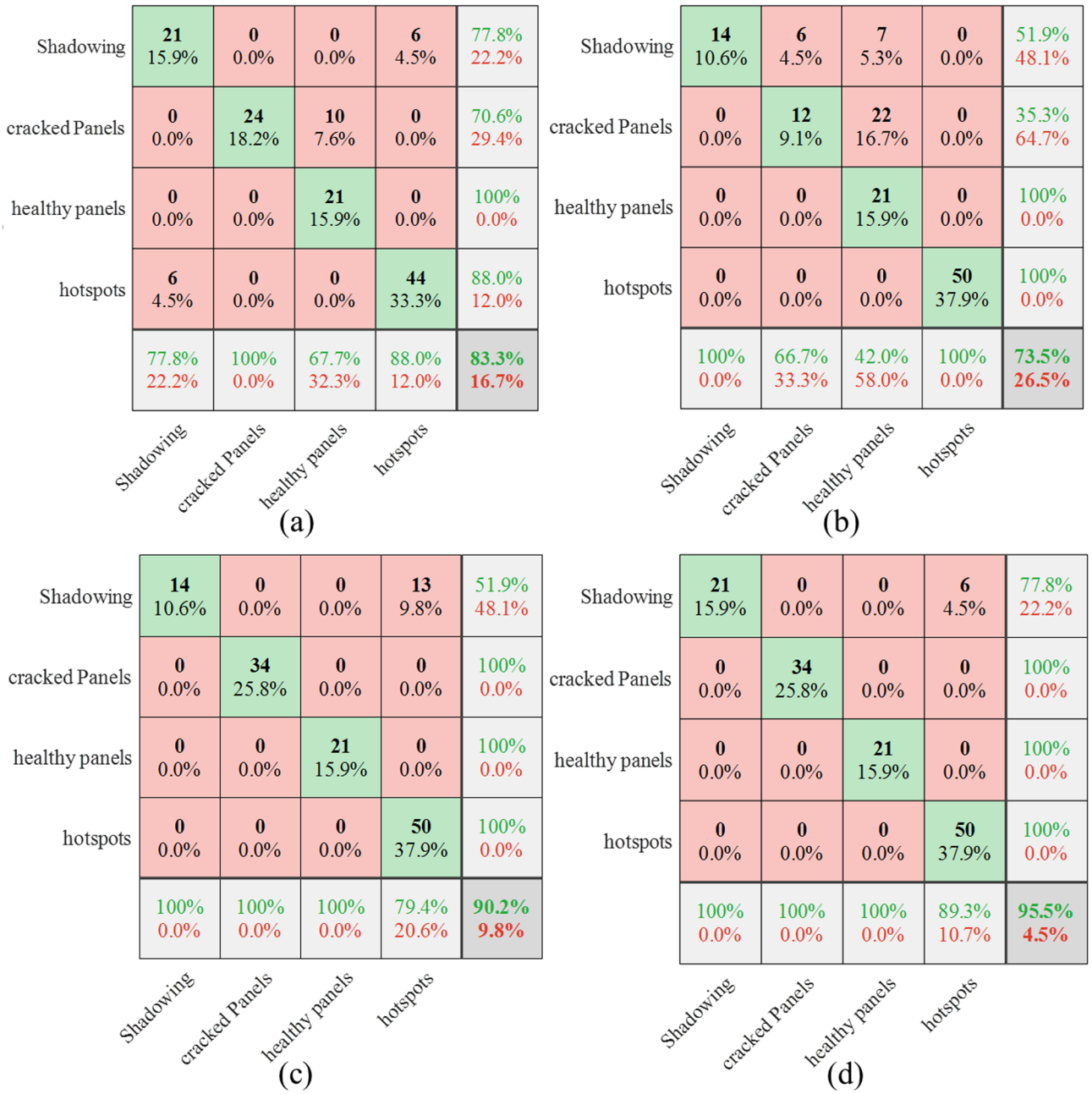


Fig. 12 Confusion matrix for various DNN architectures **a** ResNet-50, **b** GoogLeNet, **c** VGG-19 and **d** VGG-16

in context to PV module condition monitoring has resulted in categorization of the PV module health. This work categorized the module condition into faulty and healthy modules and furthermore was able to identify the faulty performance. The performance analysis presented in the paper helps addressing the key factors for faulty performance of the PV modules. The major causes for faulty performance are dust, shadowing and cracks in PV modules. Moreover, this work has proposed an in-depth performance assessment of a novel framework for PV condition monitoring. The strategy of employing DCNN models with transfer learning proved effective in fault classification. For fault classification, images are obtained through thermal imager by creating a real-time scenario such as shadowing, accumulation of dust on solar PV modules and further using a cracked solar PV module. The proposed PV module conditioning model proved to be the best of the assessed models, with classification values 95.5%. The findings outperformed existing traditional and deep learning approaches for healthy and solar PV faulty image categorization. In comparison to traditional approaches, it is preferable in terms of performance metrics such as Accuracy (*A*), Precision (*P*), Sensitivity (*Sn*),

Specificity (*Sp*), F1-score (*F1*) and Matthew's correlation coefficient (M_{CC}). Furthermore, this paper helps in detecting faulty modules; based on their loss of power, above 20% the module is stated as unrecoverable fault and less than 20% loss fault is recoverable. Future study will concentrate on executing the models on computers with GPU-enabled capability, which is anticipated to minimize calculation time. This research work underscores the pivotal role of model architecture in optimizing PV module condition monitoring, with VGG-16 establishing itself as a robust choice for PV system engineers and researchers. Further, the suggested model for module-level monitoring may be expanded to monitor and optimize the performance of grid-connected PV systems, resulting in improved fault detection, performance optimization and maintenance methods. This would help to improve the efficiency and reliability of large-scale PV systems.

Acknowledgements The authors would like to thank Anusandhan National Research Foundation (ANRF) for carrying out this research under CRG scheme (CRG/2022/003902).

Author contributions Rahma Aman performed writing original draft. M. Rizwan and Astitva Kumar reviewed original draft.

Data availability No datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. (2023) Ministry of New and Renewable Energy (MNRE). (2023). Renewable energy installed capacity report, https://mnre.gov.in/. https://mnre.gov.in/
2. Ebrahim MA, Afify Badawy Afify E, Saied Elzawawy A, Abdelqawee IM (2024) Techno-economic strategy for mitigating Hot-Spot/Partial shading of photovoltaic systems. Sol Energy 279:112813. https://doi.org/10.1016/j.solener.2024.112813
3. Li Q, Zhu L, Sun Y et al (2020) Performance prediction of Building Integrated Photovoltaics under no-shading, shading and masking conditions using a multi-physics model. Energy 213:118795. https://doi.org/10.1016/j.energy.2020.118795
4. Li Q, Chen Z, Li X et al (2024) Vertical perovskite solar cell envelope for the circular economy: a case study using life cycle cost analysis in Europe. J Clean Prod 467:143017. https://doi.org/10.1016/j.jclepro.2024.143017
5. Li Q, Li T, Kutlu A, Zanelli A (2024) Life cycle cost analysis and life cycle assessment of ETFE cushion integrated transparent organic/perovskite solar cells: Comparison with PV glazing skylight. J Build Eng 87:109140. https://doi.org/10.1016/j.jobe.2024.109140
6. Li Q, Yang G, Bian C et al (2025) Autonomous design framework for deploying building integrated photovoltaics. Appl Energy 377:124760. https://doi.org/10.1016/j.apenergy.2024.124760
7. Li Q, Monticelli C, Zanelli A (2022) Life cycle assessment of organic solar cells and perovskite solar cells with graphene transparent electrodes. Renew Energy 195:906–917. https://doi.org/10.1016/j.renene.2022.06.075
8. Li Q, Monticelli C, Kutlu A, Zanelli A (2024) Environmental performance analysis of textile envelope integrated flexible photovoltaic using life cycle assessment approach. J Build Eng 89:109348. https://doi.org/10.1016/j.jobe.2024.109348
9. Li Q, Monticelli C, Kutlu A, Zanelli A (2023) Feasibility of textile envelope integrated flexible photovoltaic in Europe: carbon footprint assessment and life cycle cost analysis. J Clean Prod 430:139716. https://doi.org/10.1016/j.jclepro.2023.139716
10. Rossi D, Omana M, Giaffreda D, Metra C (2015) Modeling and detection of hotspot in shaded photovoltaic cells. IEEE Trans Very Large Scale Integr Syst 23:1031–1039. https://doi.org/10.1109/TVLSI.2014.2333064
11. Akram MN, Lotfifard S (2015) Modeling and health monitoring of DC side of photovoltaic array. IEEE Trans Sustain Energy 6:1245–1253. https://doi.org/10.1109/TSTE.2015.2425791
12. Aman R, Rizwan M, Kumar A (2024) A novel hybrid intelligent approach for solar photovoltaic power prediction considering UV index and cloud cover. Electr Eng. https://doi.org/10.1007/s00202-024-02577-4
13. Zhu L, Li Q, Chen M et al (2019) A simplified mathematical model for power output predicting of building integrated photovoltaic under partial shading conditions. Energy Convers Manag 180:831–843. https://doi.org/10.1016/j.enconman.2018.11.036
14. Dhimish M, Mather P, Holmes V (2018) Evaluating power loss and performance ratio of hot-spotted photovoltaic modules. IEEE Trans Electron Devices 65:5419–5427. https://doi.org/10.1109/TED.2018.2877806
15. Kurukuru VSB, Haque A, Khan MA, Tripathy AK (2019) Fault classification for Photovoltaic Modules Using Thermography and Machine Learning Techniques. In: 2019 International Conference on Computer and Information Sciences (ICCIS). IEEE, pp 1–6
16. Zhao Y, Liu Q, Li D et al (2019) Hierarchical anomaly detection and multimodal classification in large-scale photovoltaic systems. IEEE Trans Sustain Energy 10:1351–1361. https://doi.org/10.1109/TSTE.2018.2867009
17. Tang W, Yang Q, Xiong K, Yan W (2020) Deep learning based automatic defect identification of photovoltaic module using electroluminescence images. Sol Energy 201:453–460. https://doi.org/10.1016/j.solener.2020.03.049
18. Su Y, Tao F, Jin J, Zhang C (2021) Automated overheated region object detection of photovoltaic module with thermography image. IEEE J Photovoltaics 11:535–544. https://doi.org/10.1109/JPHOTOV.2020.3045680
19. Fernández A, Usamentiaga R, de Arquer P et al (2020) Robust detection, classification and localization of defects in large photovoltaic plants based on unmanned aerial vehicles and infrared thermography. Appl Sci 10:5948. https://doi.org/10.3390/app10175948
20. Huerta Herraiz Á, Pliego Marugán A, García Márquez FP (2020) Photovoltaic plant condition monitoring using thermal images analysis by convolutional neural network-based structure. Renew Energy 153:334–348. https://doi.org/10.1016/j.renene.2020.01.148
21. Manno D, Cipriani G, Ciulla G et al (2021) Deep learning strategies for automatic fault diagnosis in photovoltaic systems by thermographic images. Energy Convers Manag 241:114315. https://doi.org/10.1016/j.enconman.2021.114315
22. Otamendi U, Martinez I, Quartulli M et al (2021) Segmentation of cell-level anomalies in electroluminescence images of photovoltaic modules. Sol Energy 220:914–926. https://doi.org/10.1016/j.solener.2021.03.058

23. Cipriani G, D'Amico A, Guarino S et al (2020) Convolutional neural network for dust and hotspot classification in PV modules. Energies 13:6357. https://doi.org/10.3390/en13236357
24. Aziz F, Ul Haq A, Ahmad S et al (2020) A novel convolutional neural network-based approach for fault classification in photovoltaic arrays. IEEE Access 8:41889–41904. https://doi.org/10.1109/ACCESS.2020.2977116
25. Dhibi K, Fezai R, Mansouri M et al (2020) Reduced kernel random forest technique for fault detection and classification in grid-tied PV systems. IEEE J Photovoltaics 10:1864–1871. https://doi.org/10.1109/JPHOTOV.2020.3011068
26. Pathak SP, Patil DS, Patel S (2022) Solar panel hotspot localization and fault classification using deep learning approach. Procedia Comput Sci 204:698–705. https://doi.org/10.1016/j.procs.2022.08.084
27. Krishna Veni KS, Senthil Kumar N, Gnanavadivel J (2024) Non-invasive health status diagnosis of solar PV panel using ensemble classifier. Electr Eng. https://doi.org/10.1007/s00202-024-02760-7
28. Hafdaoui H, Boudjelthia EAK, Chahtou A et al (2022) Analyzing the performance of photovoltaic systems using support vector machine classifier. Sustain Energy, Grids Networks 29:100592. https://doi.org/10.1016/j.segan.2021.100592
29. Hafdaoui H, Boudjelthia EAK, Bouchakour S, Belhaouas N (2022) Employing machine learning by classification for analysis of a monitoring database from a photovoltaic module. Desalin Water Treat 279:147–151. https://doi.org/10.5004/dwt.2022.29100
30. Hafdaoui H, Bouchakour S, Belhaouas N (2022) Use an artificial intelligence method (Machine Learning) for analysis of the performance of photovoltaic systems. J Renew Energies 25(2):199–210
31. Deitsch S, Christlein V, Berger S et al (2019) Automatic classification of defective photovoltaic module cells in electroluminescence images. Sol Energy 185:455–468. https://doi.org/10.1016/j.solener.2019.02.067
32. Rahman MRU, Chen H (2020) Defects inspection in polycrystalline solar cells electroluminescence images using deep learning. IEEE Access 8:40547–40558. https://doi.org/10.1109/ACCESS.2020.2976843
33. Moradi Sizkouhi A, Aghaei M, Esmailifar SM (2021) A deep convolutional encoder-decoder architecture for autonomous fault detection of PV plants using multi-copters. Sol Energy 223:217–228. https://doi.org/10.1016/j.solener.2021.05.029
34. Cao Y, Pang D, Zhao Q et al (2024) Improved YOLOv8-GD deep learning model for defect detection in electroluminescence images of solar photovoltaic modules. Eng Appl Artif Intell 131:107866. https://doi.org/10.1016/j.engappai.2024.107866
35. Hwang HP-C, Ku CC-Y, Chan JC-C (2021) Detection of malfunctioning photovoltaic modules based on machine learning algorithms. IEEE Access 9:37210–37219. https://doi.org/10.1109/ACCESS.2021.3063461
36. Shinong W, Qianlong M, Jie X et al (2020) An improved mathematical model of photovoltaic cells based on datasheet information. Sol Energy 199:437–446. https://doi.org/10.1016/j.solener.2020.02.046
37. Vinod KR, Singh SK (2018) Solar photovoltaic modeling and simulation: As a renewable energy solution. Energy Rep 4:701–712. https://doi.org/10.1016/j.egyr.2018.09.008
38. Laarabi B, El Baqqal Y, Rajasekar N, Barhdadi A (2021) Updated review on soiling of solar photovoltaic systems Morocco and India contributions. J Clean Prod 311:127608. https://doi.org/10.1016/j.jclepro.2021.127608
39. Pathak SP, Patil SA (2023) Evaluation of effect of preprocessing techniques in solar panel fault detection. IEEE Access 11:72848–72860. https://doi.org/10.1109/ACCESS.2023.3293756
40. Bamisile O, Acen C, Cai D et al (2025) The environmental factors affecting solar photovoltaic output. Renew Sustain Energy Rev 208:115073. https://doi.org/10.1016/j.rser.2024.115073
41. Winston DP, Murugan MS, Elavarasan RM et al (2021) Solar PV's micro crack and hotspots detection technique using NN and SVM. IEEE Access 9:127259–127269. https://doi.org/10.1109/ACCESS.2021.3111904
42. Alimi OA, Meyer EL, Olayiwola OI (2022) Solar photovoltaic modules' performance reliability and degradation analysis—A review. Energies 15:5964. https://doi.org/10.3390/en15165964
43. Rahman T, Mansur A, Hossain Lipu M et al (2023) Investigation of degradation of solar photovoltaics: a review of aging factors, impacts, and future directions toward sustainable energy management. Energies 16:3706. https://doi.org/10.3390/en16093706
44. Badran G, Dhimish M (2023) Potential induced degradation in photovoltaic modules: a review of the latest research and developments. Solar 3:322–346. https://doi.org/10.3390/solar3020019
45. Reddy BH, R KP (2022) Classification of fire and smoke images using decision tree algorithm in comparison with logistic regression to measure accuracy, precision, recall, F-score. In: 2022 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS). IEEE, pp 1–5
46. Kumar A, Rizwan M, Nangia U, Alaraj M (2021) Grey wolf optimizer-based array reconfiguration to enhance power production from solar photovoltaic plants under different scenarios. Sustainability 13:13627. https://doi.org/10.3390/su132413627
47. Wang A, Xuan Y (2021) Close examination of localized hot spots within photovoltaic modules. Energy Convers Manag 234:113959. https://doi.org/10.1016/j.enconman.2021.113959
48. Bhallamudi R, Kumarasamy S, Karuppaiyah Sundarabalan C (2021) Effect of dust and shadow on performance of solar photovoltaic modules: experimental analysis. Energy Eng 118:1827–1838
49. Köntges M, Kunze I, Kajari-Schröder S et al (2011) The risk of power loss in crystalline silicon based photovoltaic modules due to micro-cracks. Sol Energy Mater Sol Cells 95:1131–1137. https://doi.org/10.1016/j.solmat.2010.10.034
50. Fan S, Wang Y, Cao S et al (2022) A deep residual neural network identification method for uneven dust accumulation on photovoltaic (PV) panels. Energy 239:122302. https://doi.org/10.1016/j.energy.2021.122302

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

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

Astrophysics and Space Science (2024) 369:120
https://doi.org/10.1007/s10509-024-04386-4

RESEARCH

Accelerating cosmological $f(R, L_m, T)$ gravity model along with bulk viscous fluid

Shivani Sharma[1] · **Pratik V. Lepse**[2] · **Mehul Raj Sharma**[3]

Received: 20 October 2024 / Accepted: 4 December 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract
This article explores the late-time acceleration phase of the universe through a novel $f(R, L_m, T)$ gravity model, particularly, $f(R, L_m, T) = R + \alpha T + 2\beta L_m$, where α and β are free parameters of the model, in the presence of viscous fluid. We obtain the corresponding analytical solution and then we establish the constrain on arbitrary parameters of the solution by considering the Cosmic chronometers and Panthoen+SH0ES data. Further, we analyze the behavior of the obtained constrained solution through the deceleration, effective equation of state, and the Om diagnostic test. We find that the present $f(R, L_m, T)$ gravity model in the presence of viscous cosmic fluid successfully describe the late-time evolution phase of the universe with proper transition from the decelerated epoch to the accelerated one.

Keywords $f(R, L_m, T)$ gravity · Deceleration parameter · Equation of state parameter · Om diagnostic parameter

1 Introduction

The surprising revelation that the Universe's expansion began accelerating (Riess et al. 1998; Perlmutter et al. 1999) at a redshift of approximately $z \approx 0.6$ challenged the core principles of general relativity, the most successful gravitational theory to date. The most straightforward explanation for this late-time acceleration involves revisiting the cosmological constant Λ, originally adopted by Einstein at the time of proposing the first cosmological solution of the general relativity to obtain a static universe model (Einstein 1917). Building on this approach, and strongly backed by the excellent agreement between the cosmological constant and observational data, the standard cosmological model known as the ΛCDM model was proposed. Moreover, the standard model requires the presence of another undetected mysterious component of the Universe, called dark matter (Salucci et al. 2020). Although the ΛCDM model successfully explains many observations, it faces a significant challenge: no fundamental physical theory can account for it. The key issue lies in the difficulties surrounding the explanation of the cosmological constant's nature and origin (Weinberg 1989; Copeland et al. 2006). The first alternative approach to describing cosmological evolution is referred to as the dark constituents model. This model extends Einstein's gravitational field equations by adding two new terms in the stress-energy-momentum tensor, representing dark matter and dark energy (Harko and Lobo 2020).

The second approach, known as dark gravity, moves beyond Einstein's framework. It offers a purely geometrical explanation of gravitational interactions, attributing their dynamics of evolution to modifications in the geometric structure of Einstein's equations. A notable case of dark gravity is the $f(R)$ gravity, first introduced in (Buchdahl 1970). Some interesting consequences of $f(R)$ theory are listed in the references (Nojiri and Odintsov 2007, 2008; Cognola et al. 2008; Santos et al. 2007; Capozziello et al. 2008; Debnath et al. 2023; Dixit et al. 2023). Another approach to understanding gravitation, known as the dark coupling approach, has also been developed. This method replaces the traditional Einstein-Hilbert Lagrangian with a more general algebraic structure having simple additive relationship between

✉ S. Sharma
shivani.sharma14122001@gmail.com

P.V. Lepse
pratiklepse124@gmail.com

M.R. Sharma
mehul.rajsharma.min21@itbhu.ac.in

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

[2] Department of Mathematics, Ramdeobaba University, Nagpur, India

[3] Department of Physics, Indian Institute of Technology, Varanasi, India

Published online: 06 December 2024

matter and curvature term. This dark coupling approach introduced in (Bertolami et al. 2007), where the gravitational action $S = \int [f_1(R) + (1 + \lambda f_2(R))L_m]\sqrt{-g}d^4x$ was considered. This idea was generalized in the reference (Harko and Lobo 2010), eventually resulting in the $f(R, L_m)$ gravity, where the generic action consists an arbitrary function of the Ricci scalar curvatur and the Lagrangian density of matter. Various cosmological aspects of the $f(R, L_m)$ gravity were discussed in (Wang and Liao 2012; Jaybhaye et al. 2022a, 2023a,b; Koussour et al. 2024c; Jaybhaye et al. 2022b; Tangphati et al. 2023; Zeyauddin et al. 2024; Dixit et al. 2024; Pradhan et al. 2023; Shukla et al. 2024; Pradhan et al. 2024; Maurya 2023; Myrzakulov et al. 2024a; Myrzakulova et al. 2024; Koussour et al. 2024d). Another potential interaction between matter and curvature is proposed in the form of $f(R, T)$ gravity (Harko et al. 2011), where the energy-momentum scalar is coupled to the Ricci curvature term. Recently, $f(R, L_m, T)$ modified gravity theory has been proposed (Haghani and Harko 2021) which is the generalization of two aforementioned matter-geometry coupling theories, particularly, $f(R, T)$ and $f(R, L_m)$ gravity. Some recent notable works of $f(R, L_m, T)$ gravity are listed in references (Tangphati et al. 2024; Mota et al. 2024). In this work, we explore this novel $f(R, L_m, T)$ gravity theory by means of a FLRW viscous cosmological model in order to probe the accelerating nature of the expansion phase of the universe. In Sect. 2, we present the gravitational action of the theory. In Sect. 3, we present Friedmann equations of the theory under the flat FLRW background. Further, we calculate the analytical solution of a particular matter-geometry coupling model. In Sect. 4, we constrained the free parameters of the solution utilizing the MCMC approach. Further in Sect. 5, we present the corresponding results. In Sect. 6, we discuss the details of the investigation.

2 Field equation

The dynamics of the $f(R, L_m, T)$ modified gravity theory describing the significant coupling between the spacetime curvature and matter part is ruled by the following action (Haghani and Harko 2021),

$$S = \frac{1}{16\pi}\int f(R, L_m, T)\sqrt{-g}\, d^4x + \int L_m\sqrt{-g}\, d^4x \quad (1)$$

where $f(R, L_m, T)$ is an arbitrary function of the Ricci scalar R, matter Lagrangian density L_m, and the trace T of the matter energy-momentum tensor $T_{\mu\nu}$, which is defined as follows,

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta\left(\sqrt{-g}L_m\right)}{\delta g^{\mu\nu}}. \quad (2)$$

On varying the action (1), one can get the following field equation of the $f(R, L_m, T)$ gravity,

$$f_R R_{\mu\nu} - \frac{1}{2}[f - (f_L + 2f_T)L_m]\, g_{\mu\nu} + \left(g_{\mu\nu}\Box - \nabla_\mu\nabla_\nu\right) f_R = \left[8\pi + \frac{1}{2}(f_L + 2f_T)\right] T_{\mu\nu} + f_T\tau_{\mu\nu}. \quad (3)$$

where,

$$\tau_{\mu\nu} = 2g^{\alpha\beta}\frac{\partial^2 L_m}{\partial g^{\mu\nu}\partial g^{\alpha\beta}}. \quad (4)$$

The trace of the field equation (3) given as,

$$f_R R - 2\left[f - (f_L + 2f_T)L_m\right] + 3\Box f_R = \left[8\pi + \frac{1}{2}(f_L + 2f_T)\right] T + f_T\tau. \quad (5)$$

On utilizing the above equation one can rewrite the field equation of the $f(R, T, L_m)$ gravity as follows,

$$R_{\mu\nu} - \frac{1}{4}Rg_{\mu\nu} = \frac{1}{f_R}\left[8\pi + \frac{1}{2}(f_L + 2f_T)\right] \times \left(T_{\mu\nu} - \frac{1}{4}g_{\mu\nu}T\right) - \frac{1}{f_R}\left(\frac{1}{4}g_{\mu\nu}\Box - \nabla_\mu\nabla_\nu\right) f_R + \frac{f_T}{f_R}\left(\tau_{\mu\nu} - \frac{1}{4}\tau g_{\mu\nu}\right). \quad (6)$$

This form of the gravitational field equations is known as the traceless representation of the $f(R, L_m, T)$ gravity theory.

3 The $f(R, L_m, T)$ gravity model

We begin with the following flat, isotropic and homogeneous FLRW line element,

$$ds^2 = -dt^2 + a(t)^2(dx^2 + dy^2 + dz^2), \quad (7)$$

where $a(t)$ denotes the scale factor.

Note that, the spatial isotropy of the universe implies that the bulk viscous pressure is identical in each spatial directions and is therefore directly proportional to the volume expansion, denoted as $\theta = 3H$ (Eckart 1940; Israel and Stewart 1976; Israel 1976). Bulk viscosity introduces damping effects associated with volumetric straining. We consider a scaling law for the coefficient of bulk viscosity ζ, which reduces the Einstein case to a form proportional to the Hubble parameter. This scaling law has been demonstrated to be

highly beneficial (Koussour et al. 2024a,b; Myrzakulov et al. 2024b). Thus, the corresponding energy-momentum tensor becomes,

$$T_{\mu\nu} = (\rho + \bar{p})u_\mu u_\nu + \bar{p} g_{\mu\nu} \tag{8}$$

where $\bar{p} = p - 3\zeta H$ and p is the usual pressure term. Here $\zeta > 0$ is the constant of proportionality and $u^\mu = (1, 0, 0, 0)$ denotes the four velocity vector of the comoving cosmic fluid.

The Friedmann equations describing $f(R, L_m, T)$ gravity dynamics for the pressureless universe are given as follows (Haghani and Harko 2021),

$$-\dot{H} = \frac{1}{2f_R}(8\pi + f_m)(\rho + p) + \frac{1}{2f_R}\left(\ddot{f}_R - H\dot{f}_R\right), \tag{9}$$

and

$$\dot{H} + 2H^2 = \frac{1}{3f_R}(f - 2L_m f_m) - \frac{1}{6f_R}(8\pi + f_m)(\rho - 3p) + \frac{1}{2f_R}\left(\ddot{f}_R + 3H\dot{f}_R\right). \tag{10}$$

where $f_m = f_T + \frac{1}{2}f_L$.

We consider the following $f(R, L_m, T)$ gravity model,

$$f(R, L_m, T) = R + \alpha T + 2\beta L_m \tag{11}$$

Here α and β are arbitrary parameters of the assumed model. The model under consideration is the straight forward minimal coupling class model of the $f(R, L_m, T)$ gravity that is motivated from the interesting work of Haghani and Harko (Haghani and Harko 2021).

The Friedmann equations (9) and (10) for the functional form (11) reduces to following equations,

$$\frac{1}{(8\pi + \alpha + \beta)}\left[-2\dot{H} + 3(8\pi + \alpha + \beta)\zeta H\right] = \rho \tag{12}$$

and

$$3(\dot{H} + 2H^2) = \frac{1}{2}(8\pi + \alpha + \beta)(\rho + 9\zeta H) + 9\alpha\zeta H - \alpha\rho \tag{13}$$

Here we consider the matter Lagrangian density to be $L_m = -\rho$. Now, on solving the equations (12) and (13), one can obtain the following first order differential equation in the variable $N = ln(a)$,

$$\frac{dH}{dN} + \frac{3(8\pi + \alpha + \beta)}{(16\pi + \alpha + 2\beta)}H = \frac{3\zeta(8\pi + \alpha + \beta)(8\pi + 2\alpha + \beta)}{(16\pi + \alpha + 2\beta)} \tag{14}$$

On integrating the above equation, we obtain the following solution,

$$H(z) = H_0(1+z)^{\frac{3(8\pi+\alpha+\beta)}{(16\pi+\alpha+2\beta)}} + (8\pi + 2\alpha + \beta)\zeta\left[1 - (1+z)^{\frac{3(8\pi+\alpha+\beta)}{(16\pi+\alpha+2\beta)}}\right] \tag{15}$$

where $H(0) = H_0$, the Hubble function value at the present epoch $z = 0$. In a more simplified form one can rewrite the above expression as follows,

$$H(z) = H_0(1+z)^n + \bar{\zeta}\left[1 - (1+z)^n\right] \tag{16}$$

where $\frac{3(8\pi+\alpha+\beta)}{(16\pi+\alpha+2\beta)} = n$ and $(8\pi + 2\alpha + \beta)\zeta = \bar{\zeta}$.

4 Data and methodology

In this section, we constraint the free parameters of the obtained analytical solution by utilizing Markov Chain Monte Carlo method (Foreman-Mackey et al. 2013). In order to perform observational analysis we consider two data sets, first one is the Cosmic Chronometers data consisting of 31 H(z) measurements, and the second one is the Panthoen+SH0ES data consisting of 1701 light curve. To calculate the posterior probability we utilize the Bayesian statistical method and calculate the required likelihood and prior function.

4.1 Cosmic chronometers

Cosmic chronometers refer to a set of mainly old, passive galaxies that have stopped forming new stars and can be recognized by specific characteristics in their spectra and color profiles. The cosmic chronometers data is derived by calculating the age of these galaxies across different redshifts. In this research, we analyzed a dataset containing 31 H(z) samples having the redshift in the range $0.07 \leq z \leq 2.41$ (Sharov and Vasiliev 2018). The χ^2 function for the cosmic chronometer datasets is given as follows,

$$\chi^2_{CC} = \sum_{k=1}^{31} \frac{[H_{th}(z_k, \theta) - H_{obs}(z_k)]^2}{\sigma^2_{H(z_k)}} \tag{17}$$

where $H_{obs}(z_k)$ denotes the observed H(z) value at the observed redshift $z = z_k$, while $H_{th}(z_k, \theta)$ is the H(z) value obtained from the assumed model with the parameter space θ and that is estimated at the observed redshift $z = z_k$, and $\sigma_{H(z_k)}$ is the standard observed error. The contour plot describing the parameter correlation with 1σ-3σ confidence interval for the CC samples is presented in Fig. 1. Moreover the corresponding constraints obtained for the parameter space are listed in Table 1.

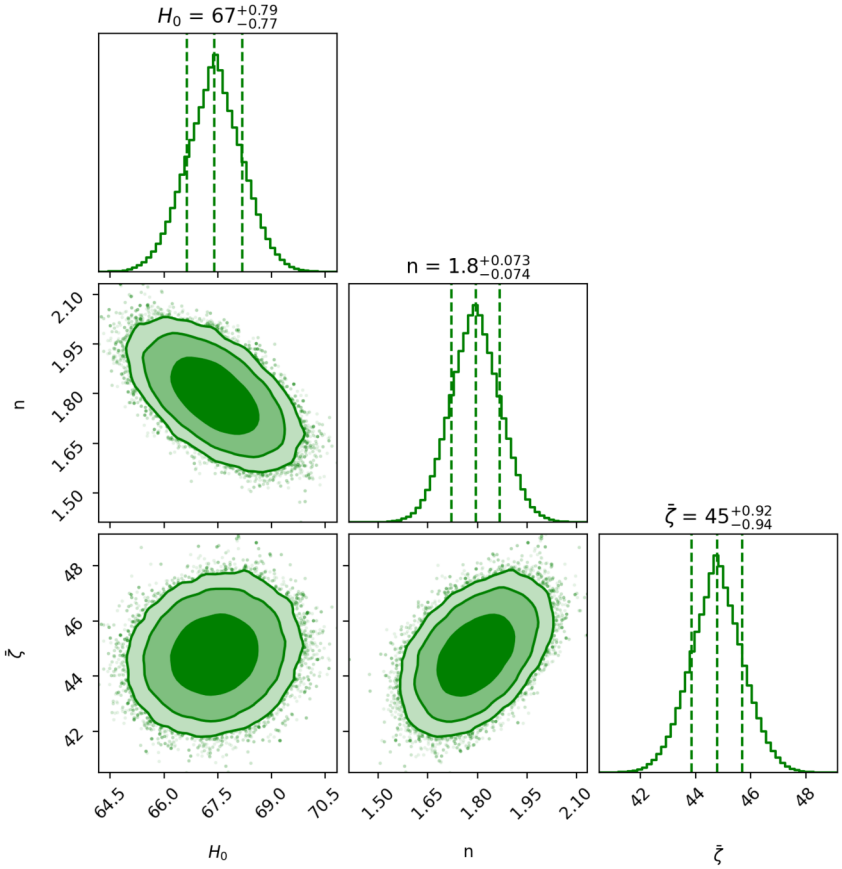


Fig. 1 Contour plot describing the parameter correlation for the CC samples

4.2 Pantheon+SH0ES

Type Ia supernovae are recognized for their uniform brightness, making them reliable standard candles for calculating distances through the distance modulus method. The Pantheon+SH0ES dataset cover a broad range of redshifts, surpassing the previous Pantheon samples by including the most up-to-date observational data (Brout et al. 2022). The χ^2 function for the SNIa datasets is given as follows,

$$\chi^2_{SN} = D^T C^{-1}_{SN} D, \tag{18}$$

where C_{SN} (Brout et al. 2022) is the corresponding covariance matrix for the Pantheon+SH0ES data. Further, the vector D is defined as $D = m_{Bi} - M - \mu^{th}(z_i)$, where m_{Bi} and M are the apparent and absolute magnitude, respectively. Further, $\mu^{th}(z_i)$ gives the distance modulus for the obtained analytical expression for the Hubble function as follows,

$$\mu^{th}(z_i) = 5log_{10}\left[\frac{D_L(z_i)}{1Mpc}\right] + 25, \tag{19}$$

Here $D_L(z)$ is the luminosity distance and that is acquired as follows,

$$D_L(z) = c(1+z)\int_0^z \frac{dx}{H(x,\theta)} \tag{20}$$

The Pantheon+SH0ES samples resolves the degeneracy of parameters H_0 and M by redefining the vector D as follows (Brout et al. 2022),

$$\bar{D} = \begin{cases} m_{Bi} - M - \mu_i^{Ceph} & i \in \text{Cepheid hosts} \\ m_{Bi} - M - \mu^{th}(z_i) & \text{otherwise} \end{cases} \tag{21}$$

Where, μ_i^{Ceph} is independently estimated using Cepheid calibrators. Hence, the corresponding chi-sqaured function becomes $\chi^2_{SN} = \bar{D}^T C^{-1}_{SN} \bar{D}$. The contour plot describing the parameter correlation with 1σ-3σ confidence interval for the Pantheon+SH0ES samples is presented in Fig. 2. Moreover the corresponding constraints obtained for the arbitrary parameter space are listed in Table 1.

4.3 Combined samples

Now, we estimated the free parameters constraint by considering the combined H(z) and supernovae light curve samples. We minimize the total chi-squared function i.e. $\chi^2 = \chi^2_{CC} + \chi^2_{SN}$ to acquire the constraint for the combined samples, along with the priors $H_0 \in [50, 100]$, $n \in [0, 5]$, and $\bar{\zeta} \in [50, 100]$. The contour plot describing the parameter correlation with 1σ-3σ confidence interval for the combined CC+Pantheon+SH0ES samples is presented in Fig. 3. More-

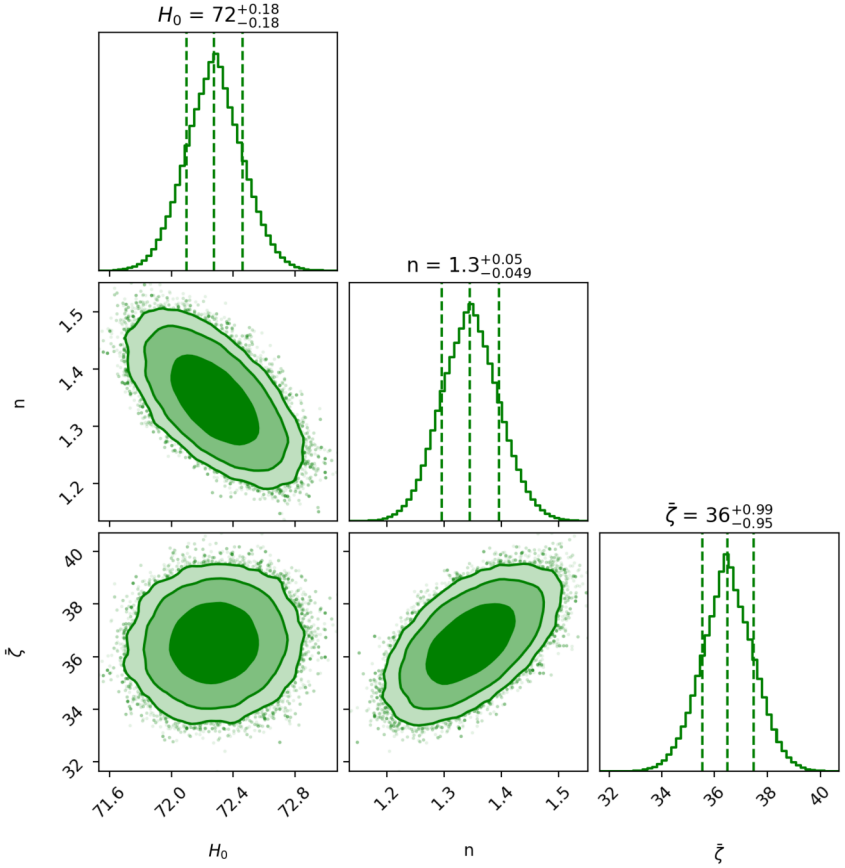


Fig. 2 Contour plot describing the parameter correlation for the Pantheon+SH0ES samples

Table 1 Obtained constraint on different parameters corresponding to the different data sets

Parameters	CC	Pantheon+SH0ES	CC+Pantheon+SH0ES
H_0	$67^{+0.79}_{-0.77}$	72 ± 0.18	72 ± 0.15
n	$1.8^{+0.073}_{-0.074}$	$1.3^{+0.05}_{-0.049}$	$1.5^{+0.034}_{-0.035}$
$\bar{\zeta}$	$45^{+0.92}_{-0.94}$	$36^{+0.99}_{-0.95}$	41 ± 0.48
q_0	$-0.408^{+0.013}_{-0.014}$	-0.35 ± 0.009	-0.354 ± 0.007
ω_0	$-0.605^{+0.008}_{-0.009}$	$-0.566^{+0.005}_{-0.007}$	$-0.569^{+0.004}_{-0.0055}$

over the corresponding constraints obtained for the arbitrary parameter space are listed in Table 1.

4.4 Information criterion

To testify the viability of the MCMC analysis performed, we utilize a statistical diagnostic that is well known as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The first one is formulated using the following expression,

$$AIC = \chi^2_{min} + 2d \tag{22}$$

Here, d is the number of independent model parameters in the assumed cosmological model. In order to compare the assumed model with the ΛCDM one, we calculate $\Delta AIC = |AIC_{Model} - AIC_{\Lambda CDM}|$. The value $\Delta AIC < 2$ indicates high evidence for the considered model over the standard ΛCDM, whereas the value $4 < \Delta AIC \leq 7$ shows a mild evidence. However, the value $\Delta AIC > 10$ indicates disagreement for the assumed model. The second formulation is given as follows,

$$BIC = \chi^2_{min} + dln(N) \tag{23}$$

where, N is the number of observed data points utilized in the MCMC process. As similar to the previous case, $\Delta BIC < 2$ indicates high evidence, whereas the value $2 < \Delta BIC \leq 6$ shows a mild evidence. Utilizing the obtained $\chi^2_{min} = 1640.525$ value, we acquired $AIC_{Model} = 1646.525$ and $BIC_{Model} = 1662.896$ and hence we obtained $\Delta AIC = 2.377$ and $\Delta BIC = 7.784$, where ΛCDM value are estimated as $AIC_{\Lambda CDM} = 1644.148$ and $BIC_{\Lambda CDM} = 1655.112$. Hence, using this information criterion, we conclude that the considered model has significant evidence over the standard ΛCDM.

5 Results

Now, we present the results obtained corresponding to the obtained observational constraints for the different datasets.

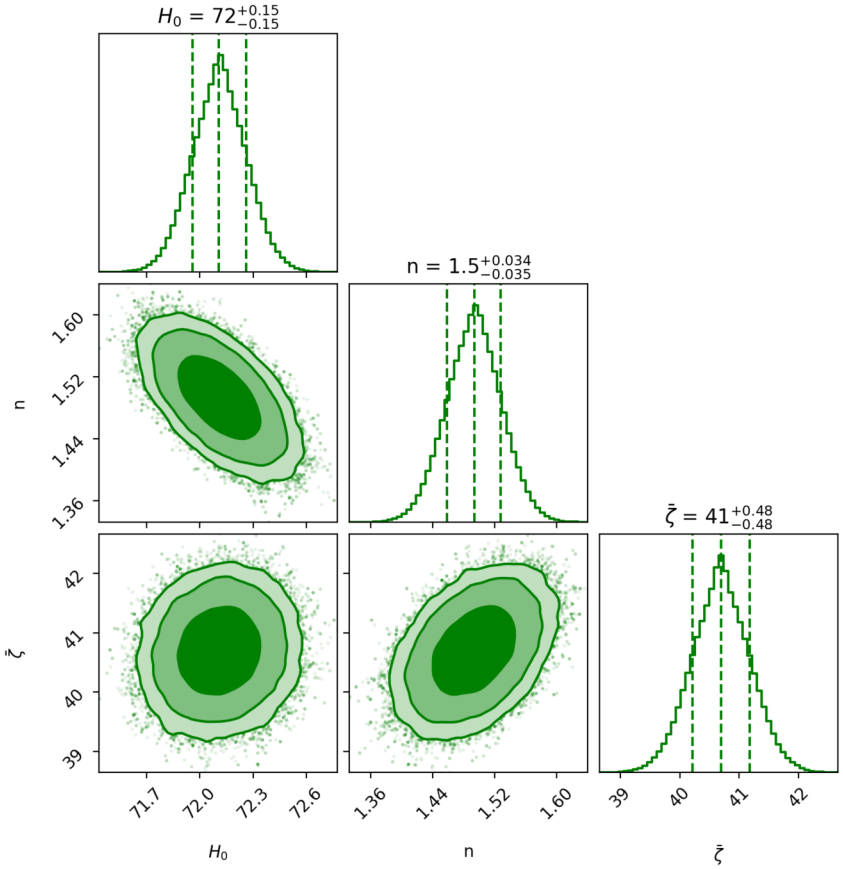


Fig. 3 Contour plot describing the parameter correlation for the CC+Pantheon+SH0ES samples

The deceleration parameter and the effective EoS parameter are key ingredients in order to probe the history of the universe's accelerating phase. These parameters are defined as follows,

$$q = -1 - \frac{\dot{H}}{H^2} \quad \text{and} \quad \omega_{eff} = -1 - \frac{2\dot{H}}{3H^2} \tag{24}$$

The evolutionary profile of the deceleration parameter and the effective EoS parameter are presented in Fig. 4 and 5 respectively, corresponding to the parameter estimation obtained for the CC, Pantheon+SH0ES, and combined samples.

From Fig. 4 and 5, we found that our theoretical $f(R, L_m, T)$ model shows a evolution from decelerated phase of expansion to the accelerated phase. Moreover, the present value of the deceleration parameter and effective EoS parameter are listed in Table 1. Further to assess the type of dark energy favored by the assumed model we employ the Om diagnostic test. The Om diagnostic parameter is defined as follows (Sahni et al. 2008),

$$Om(z) = \frac{\left(\frac{H(z)}{H_0}\right)^2 - 1}{(1+z)^3 - 1} \tag{25}$$

were H_0 is the Hubble function value at the present epoch. The Fig. 6 shows the behavior of Om diagnostic parameter for the values of parameter set that is acquired from the observational data. Note that, the negative slope of $Om(z)$ curve indicates the quintessence dark energy behavior of the assumed theoretical model whereas the positive one corresponds to the phantom type behavior. The constant $Om(z)$ behavior depicts the ΛCDM type behavior. From Fig. 6 it is clear that the considered theoretical $f(R, L_m, T)$ model shows quintessence behavior.

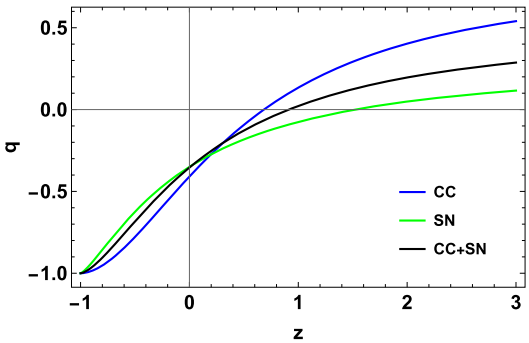


Fig. 4 Evolutionary profile of the deceleration parameter vs redshift

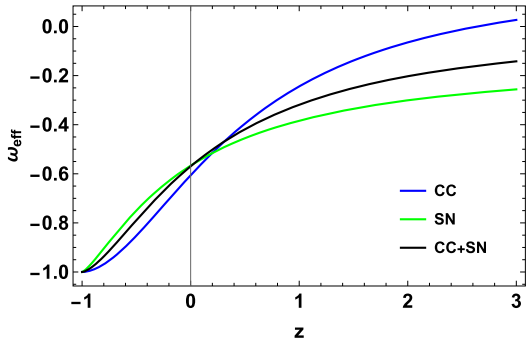


Fig. 5 Evolutionary profile of the effective equation of state parameter vs redshift

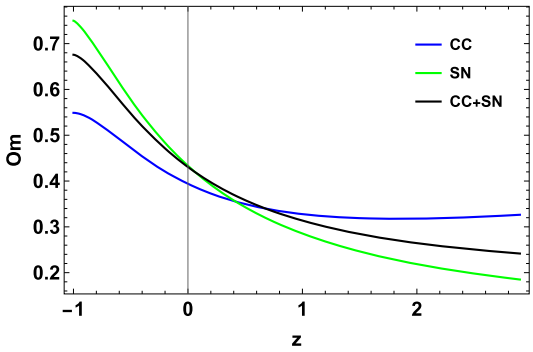


Fig. 6 Evolutionary profile of the Om diagnostic parameter vs redshift

6 Discussion

In this study, we attempted to investigate the late-time acceleration phase of the universe through a novel $f(R, L_m, T)$ gravity model. We considered the functional form $f(R, L_m, T) = R + \alpha T + 2\beta L_m$, where α and β are arbitrary constants. We have calculated the exact analytical solution of the assumed $f(R, L_m, T)$ model in the presence of viscous cosmological fluid having the viscosity coefficient $\zeta > 0$. The obtained analytical solution i.e. the explicit expression of the Hubble function is presented in the equation (16). Then to obtained the constraints on model parameters we invoke MCMC method along with the Bayesian statistical method to estimate the posterior probability with the help of prior and likelihood function. The obtained parameter values for the CC, Pantheon+SH0ES, and the combined CC+Pantheon+SH0ES datasets are listed in Table 1. Moreover, to examine the robustness of the MCMC analysis performed, we employed the information criterion such as AIC and BIC. We have obtained $\Delta AIC = 2.377$ and $\Delta BIC = 7.784$ which shows the significant favor of the assumed cosmological scenario. Further, we have presented the profiles of the deceleration and the effective EoS parameter in Fig. 4 and 5 respectively. We found that our theoretical $f(R, L_m, T)$ model shows a evolution from decelerated phase of expansion to the accelerated phase. Moreover, the present value of the deceleration parameter and effective EoS parameter are listed in Table 1. Further, we employed the Om diagnostic test to predict the model behavior. We found that our assumed model shows the quintessence type dark energy behavior (see Fig. 6). From outcomes of the investigation, we conclude that our considered $f(R, L_m, T)$ gravity model in the presence of viscous cosmic fluid successfully describe the late-time evolution phase of the universe with proper transition from the decelerated epoch to the accelerated one.

Author contributions Shivani Sharma and Pratik V Lepse: Writing - original draft, methodology, graph plotting, review, and editing. Mehul Raj Sharma: data analysis, review, and editing.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

Bertolami, O., Böhmer, C.G., Harko, T., Lobo, F.S.N.: Extra force in $f(R)$ modified theories of gravity. Phys. Rev. D **75**, 104016 (2007)

Brout, D., Scolnic, D., Popovic, B., Riess, A.G., Carr, A., Zuntz, J., Kessler, R., Davis, T.M., Hinton, S., Jones, D., Kenworthy, W.D., Peterson, E.R., Said, K., Taylor, G., Ali, N., Armstrong, P., Charvu, P., Dwomoh, A., Meldorf, C., Palmese, A., Qu, H., Rose, B.M., Sanchez, B., Stubbs, C.W., Vincenzi, M., Wood, C.M., Brown, P.J., Chen, R., Chambers, K., Coulter, D.A., Dai, M., Dimitriadis, G., Filippenko, A.V., Foley, R.J., Jha, S.W., Kelsey, L., Kirshner, R.P., Möller, A., Muir, J., Nadathur, S., Pan, Y.-C., Rest, A., Rojas-Bravo, C., Sako, M., Siebert, M.R., Smith, M., Stahl, B.E., Wiseman, P.: The pantheon+ analysis: cosmological constraints. Astrophys. J. **938**(2), 110 (2022)

Buchdahl, H.A.: Non-linear Lagrangians and cosmological theory. Mon. Not. R. Astron. Soc. **150**(1) (1970)

Capozziello, S., Cardone, V.F., Salzano, V.: Cosmography of $f(R)$ gravity. Phys. Rev. D **78**, 063504 (2008)

Cognola, G., Elizalde, E., Nojiri, S., Odintsov, S.D., Sebastiani, L., Zerbini, S.: Class of viable modified $f(R)$ gravities describing inflation and the onset of accelerated expansion. Phys. Rev. D **77**, 046009 (2008)

Copeland, E.J., Sami, M., Tsujikawa, S.: Dynamics of dark energy. Int. J. Mod. Phys. D **15**, 1753–1935 (2006)

Debnath, U., Molla, N.U., Pradhan, A.: Non-commutative wormhole in non-minimal curvature-matter coupling of $f(R)$ gravity with Gaussian and Lorentzian distributions. Int. J. Geom. Methods Mod. Phys. **20**(12), 2350214 (2023)

Dixit, A., Bhardwaj, V.K., Pradhan, A., Krishnannair, S.: Observational constraint in Kantowski-Sachs $f(R)$ gravity model with strange quark matter. Indian J. Phys. **97**(12), 3695–3705 (2023)

Dixit, A., Pradhan, A., Zeyauddin, M., Singh, J.: Dark energy nature of logarithmic $f(R, L_m)$-gravity models with observational constraints. Int. J. Mod. Phys. A **39**, Article ID 2450043 (2024)

Eckart, C.: The thermodynamics of irreversible processes iii relativistic theory of the simple fluid. Phys. Rev. **58**, 919–924 (1940)

Einstein, A.: Cosmological considerations in the general theory of relativity. Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **1917**, 142–152 (1917)

Foreman-Mackey, D., Hogg, D.W., Lang, D., Goodman, J.: emcee: the MCMC hammer. Publ. Astron. Soc. Pac. **125**(925), 306 (2013)

Haghani, Z., Harko, T.: Generalizing the coupling between geometry and matter: $f(R, L_m, T)$ gravity. Eur. Phys. J. C **81**(7), 615 (2021)

Harko, T., Lobo, F.S.N.: $f(R, L_m)$ gravity. Eur. Phys. J. C **70**, 373 (2010)

Harko, T., Lobo, F.S.N.: Beyond Einstein's general relativity: hybrid metric-Palatini gravity and curvature-matter couplings. Int. J. Mod. Phys. D **29**(13), 2030008 (2020)

Harko, T., Lobo, F.S.N., Nojiri, S., Odintsov, S.D.: $f(R, T)$ gravity. Phys. Rev. D **84**, 024020 (2011)

Israel, W.: Nonstationary irreversible thermodynamics: a causal relativistic theory. Ann. Phys. **100**(1), 310–331 (1976)

Israel, W., Stewart, J.: Thermodynamics of nonstationary and transient effects in a relativistic gas. Phys. Lett. A **58**(4), 213–215 (1976)

Jaybhaye, L.V., Solanki, R., Mandal, S., Sahoo, P.: Cosmology in $f(R, L_m)$ gravity. Phys. Lett. B **831**, 137148 (2022a)

Jaybhaye, L.V., Solanki, R., Mandal, S., Sahoo, P.: Cosmology in $f(R, L_m)$ gravity. Phys. Lett. B **831**, 137148 (2022b)

Jaybhaye, L.V., Bhattacharjee, S., Sahoo, P.: Baryogenesis in $f(R, L_m)$ gravity. Phys. Dark Universe **40**, 101223 (2023a)

Jaybhaye, L.V., Solanki, R., Mandal, S., Sahoo, P.K.: Constraining the viscous dark energy equation of state in $f(R, L_m)$ gravity. Universe **9**(4) (2023b)

Koussour, M., Altaibayeva, A., Bekov, S., Donmez, O., Muminov, S., Rayimbaev, J.: Observational constraints on the equation of state of viscous fluid in $f(R, T)$ gravity. Phys. Dark Universe **46**, 101577 (2024a)

Koussour, M., Errehymy, A., Donmez, O., Myrzakulov, K., Khan, M.A., Çil, B., Güdekli, E.: Constraints on bulk viscosity in $f(Q, T)$ gravity from H(z)/Pantheon+ data. Phys. Dark Universe **45**, 101527 (2024b)

Koussour, M., Myrzakulov, N., Rayimbaev, J., Alfedeel, A.H.A., Elkhair, H.M.: Bouncing behavior in $f(R, L_m)$ gravity: phantom crossing and energy conditions. Int. J. Geom. Methods Mod. Phys. **21**(11), 2450184 (2024c)

Koussour, M., Myrzakulov, N., Rayimbaev, J., Alfedeel, A.H.A., Elkhair, H.M.: Bouncing behavior in $f(R, L_m)$ gravity: phantom crossing and energy conditions. Int. J. Geom. Methods Mod. Phys. **21**(11), 2450184 (2024d)

Maurya, D.C.: Constrained ΛCDM dark energy models in higher derivative $f(R, L_m)$ gravity theory. Phys. Dark Universe **42**, 101373 (2023)

Mota, C.E., Pretel, J.M.Z., Flores, C.O.V.: Neutron stars in $f(R, L_m, T)$ gravity. Eur. Phys. J. C **84**(7), 673 (2024)

Myrzakulov, Y., Donmez, O., Yildiz, G.D.A., Güdekli, E., Muminov, S., Rayimbaev, J.: Linear redshift parametrization of deceleration parameter in $f(R, L_m)$ gravity. Phys. Dark Universe **45**, 101545 (2024a)

Myrzakulov, Y., Koussour, M., Davletov, I.Y., Rayimbaev, J.: Constraining $f(Q, L_m)$ gravity with bulk viscosity (2024b)

Myrzakulova, S., Koussour, M., Myrzakulov, N.: Investigating the dark energy phenomenon in $f(R, L_m)$ cosmological models with observational constraints. Phys. Dark Universe **43**, 101399 (2024)

Nojiri, S., Odintsov, S.D.: Unifying inflation with Λcdm epoch in modified $f(R)$ gravity consistent with solar system tests. Phys. Lett. B **657**(4), 238–245 (2007)

Nojiri, S., Odintsov, S.D.: Modified $f(R)$ gravity unifying R^m inflation with the ΛCDM epoch. Phys. Rev. D **77**, 026007 (2008)

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R.A., Nugent, P., Castro, P.G., Deustua, S., Fabbro, S., Goobar, A., Groom, D.E., Hook, I.M., Kim, A.G., Kim, M.Y., Lee, J.C., Nunes, N.J., Pain, R., Pennypacker, C.R., Quimby, R., Lidman, C., Ellis, R.S., Irwin, M., McMahon, R.G., Ruiz-Lapuente, P., Walton, N., Schaefer, B., Boyle, B.J., Filippenko, A.V., Matheson, T., Fruchter, A.S., Panagia, N., Newberg, H.J.M., Couch, W.J., Project, T.S.C.: Measurements of Ω and Λ from 42 high-redshift supernovae. Astrophys. J. **517**(2), 565–586 (1999)

Pradhan, A., Goswami, G., Beesham, A.: Reconstruction of an observationally constrained $f(R, T)$ gravity model. Int. J. Geom. Methods Mod. Phys. **20**(10), 2350169 (2023)

Pradhan, A., Dixit, A., Zeyauddin, M.: Reconstruction of Λcdm model from $f(T)$ gravity in viscous fluid universe with observational constraints. Int. J. Geom. Methods Mod. Phys. **21**(01), 2450027 (2024)

Riess, A.G., Filippenko, A.V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P.M., Gilliland, R.L., Hogan, C.J., Jha, S., Kirshner, R.P., Leibundgut, B., Phillips, M.M., Reiss, D., Schmidt, B.P., Schommer, R.A., Smith, R.C., Spyromilio, J., Stubbs, C., Suntzeff, N.B., Tonry, J.: Observational evidence from supernovae for an accelerating universe and a cosmological constant. Astron. J. **116**(3), 1009 (1998)

Sahni, V., Shafieloo, A., Starobinsky, A.A.: Two new diagnostics of dark energy. Phys. Rev. D **78**, 103502 (2008)

Salucci, P., Turini, N., Paolo, C.D.: Paradigms and scenarios for the dark matter phenomenon. In: Particle Physics at the Year of 150th Anniversary of the Mendeleev's Periodic Table of Chemical Elements (2020)

Santos, J., Alcaniz, J.S., Rebouças, M.J., Carvalho, F.C.: Energy conditions in $f(R)$ gravity. Phys. Rev. D **76**, 083513 (2007)

Sharov, G., Vasiliev, V.: How predictions of cosmological models depend on Hubble parameter data sets. Math. Modell. Geom. **6**(1) (2018)

Shukla, A., Raushan, R., Chaubey, R.: Dynamical systems analysis of $f(R, L_m)$ gravity model (2024)

Tangphati, T., Panotopoulos, G., Banerjee, A., Pradhan, A.: Charged compact stars with color-flavor-locked strange quark matter in $f(R, T)$ gravity. Chin. J. Phys. **82**, 62–74 (2023)

Tangphati, T., Sakalli, I., Banerjee, A., Pradhan, A.: Anisotropic quark stars in $f(R, L_m, T)$ gravity (2024)

Wang, J., Liao, K.: Energy conditions in $f(R, L_m)$ gravity. Class. Quantum Gravity **29**(21), 215016 (2012)

Weinberg, S.: The cosmological constant problem. Rev. Mod. Phys. **61**, 1–23 (1989)

Zeyauddin, M., Dixit, A., Pradhan, A.: Anisotropic dark energy models in $f(R, L_m)$-gravity. Int. J. Geom. Methods Mod. Phys. **21**(02), 2450038 (2024)

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

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

Electrical Engineering
https://doi.org/10.1007/s00202-024-02795-w

ORIGINAL PAPER

An energy storage system with SOA-based FONPID controller for AGC study in restructured power systems

Ram Naresh Mishra[1,2] · **Narendra Kumar**[2]

Received: 28 December 2023 / Accepted: 7 October 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

The growing integration of renewable energy sources (RESs) into the power grid to tackle climate change is making the network design of the present electrical system more complex every day. Thus, the inertia of the power system is gradually decreasing. Therefore, a minor load perturbation or dynamic system disturbance is the cause of the power imbalance. The control architecture of the proposed systems focuses on minimizing the frequency deviations and tie-line power fluctuations. Thus, novel controller techniques with new optimization algorithms for real-world issues are always desirable. Under this situation, a skill optimization algorithm (SOA)-based fractional-order nonlinear PID (SOA-FONPID) controller is proposed with an ultra-capacitor (UC) and electric vehicle (EV) for frequency management of multi-area restructured power system with governor dead band (GDB). The suggested controller's settings are optimized to minimize the objective function, which is the integral of time multiplied by absolute error (ITAE). The performance of the SOA-FONPID controller with UC and EV is evaluated by comparing it with several controllers modified using the skill optimization algorithm (SOA) and genetic algorithm (GA). A comparison shows that considering all time-domain behaviors, the SOA-FONPID controller with ultra-capacitor (UC) and electric vehicle (EV) participation is viable, effective, and immune to optimal frequency stabilization issues. The objective of a polar fuzzy controller is to restore the frequency and tie-line power to its nominal value with minimum computational efforts in the shortest possible time. A fuzzy logic controller's performance depends on having sufficient and accurate information. A knowledge base's complexity rises with its rule count, influencing computation time and memory. Another study proposes a GA-polar fuzzy controller (GA-PFC) considering various input membership functions to stabilize a reorganized thermal-gas power system frequency. A comparison shows that the GA-PFC (Dsig., i.e., difference of two sigmoidal functions) controller with RFB is practical and effective for frequency stability because it quickly restores frequency and tie-line power to nominal values in case of load disturbances. The outcomes of the simulation are shown to highlight the unique qualities and benefits of the present work.

Keywords Automatic generation control (AGC) · Electric vehicle (EV) · Energy storage system · Fractional-order nonlinear PID controller (FONPID) · Nonlinear PID controller (NPID) · Polar fuzzy controller (PFC) · Skill optimization algorithm (SOA)

Abbreviations

ABC	Artificial bee colony
ACSA	Artificial cooperative search algorithm
ALO	Ant lion optimizer
BFO	Bacterial foraging optimization
CRAZYPSO	Craziness-based particle swarm optimization
DES	Distributed energy source
DISCOs	Distribution companies
ESS	Energy storage system
EV	Electric vehicle
FA	Firefly algorithm

Ram Naresh Mishra and Narendra Kumar contributed equally to this work.

✉ Ram Naresh Mishra
ramnaresh.mishra@gla.ac.in; rnaresh2007@gmail.com

Narendra Kumar
dnk_1963@yahoo.com

[1] Department of Electrical Engineering, GLA University, Mathura, UP 281406, India

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

Published online: 19 November 2024

FOOA	Fifth-order Oustaloup's approximation
FLC	Fuzzy logic controller
FONPID	Fractional-order nonlinear PID
FPA	Flower pollination algorithm
GA	Genetic algorithm
GAMSF	GA-based multi-stage fuzzy
GDB	Governor dead band
GENCOs	Generation companies
GNA	Gradient Newton algorithm
GRC	Generation rate constraint
GWO	Gray wolf optimization
hBFOA-PSO	Hybrid BFO algorithm-particle swarm optimization
HES	Hydrogen energy storage
hFA-PS	Hybrid firefly algorithm-pattern search
ICCL	Improved cone complementarity linearization
ISOs	Independent system operators
MARL	Multi-agent reinforcement learning
MBA	Mine blast algorithm
MPC	Model predictive controller
MSFC	Multi-stage fuzzy controller
OAR	Optimal AGC regulator
PFC	Polar fuzzy controller
PID	Proportional–integral–derivative
PIDF	PID controller with derivative filter
PSOMSF	PSO-based multi-stage fuzzy
RFB	Redox flow batteries
RPS	Restructured power system
RPI	Robust performance index
RLD	Random load disturbance
SLP	Step load perturbations
SMES	Superconducting magnetic energy storage
SSA	Salp swarm algorithm
STPSO	Statistical tracked PSO
TLBO	Teaching learning-based optimization
TCPS	Thyristor controlled phase shifter
TD	Time delay
TRANSCOs	Transmission companies
UPFC	Unified power flow controller
VIU	Vertically integrated utility
XCSR	Continuous-valued input, extended classifier system
ZN	Ziegler–Nichols

1 Introduction

1.1 Literature review

The International Energy Agency says that the world's energy demand has grown fastest in recent years due to a growing economy, intense heat waves, and more people using electric technologies such as heat pumps and electric vehicles. Renewable energy sources, such as solar PV, are also increasing rapidly. They are on track to set new scores. The power quality issues are linked to frequency deviation, which illustrates the imbalance between load demand and generation. A growing frequency deviation might result in power interruptions. It is crucial to manage the system frequency within proper limits and have an acceptable tie-line power profile to ensure a steady energy supply. The primary reason for these changes is a disproportion between the suddenly needed power and the currently insufficient generation. The frequency difference between the control areas and the tie-line profile is kept in check by automated generation control (AGC). AGC works by changing the generation to meet the power needs and losses related to those needs. As a consequence of this, AGC is essential to the provision of high-quality power in power systems that have been rebuilt. Energy storage systems (ESSs) are used in RPS to improve AGC's work because ESSs react quickly and perfectly when absorbing excess power and compensate for insufficient power in dynamic settings. In addition to this, they reduce fluctuations in the frequency of the grid, which helps to improve power quality.

In a liberalized environment with several GENCOs and DISCOs, a DISCO can establish a contract for selling electrical power with any GENCO in its own or another pool. As a result, customers are anticipated to incur lower costs for power in an open and competitive market. The security and stability of the electricity system are the responsibility of ISO, an impartial organization. Frequency control, often known as AGC, is one of its many auxiliary functions [1–3]. The operation of the reformed power system becomes more challenging to manage due to the emergence of new entities. AGC is increasingly important to provide reliable, high-quality power in the reorganized electrical market. Over the past two decades, various researchers have contributed to the field of AGC in the deregulated power system.

Practically speaking, there are no linear systems. Practical systems are all somewhat nonlinear. The authors devoted close attention to considering the impact of system nonlinearities [2, 3]. AGC in the power system may also be supported by the ESS, which has a lot of promise for preserving the power balance and sustaining the grid frequency during abrupt disruptions [4]. Hota et al. [5] discussed the effectiveness of DE and GA-based I, PI, ID, and PID controllers for a deregulated thermal-gas system, this study shows that the DE

algorithm is more effective in solving AGC issues. Yogendra et al. [6] discussed the simulation results of CRAZYPSO and hBFOA-PSO-based PI controllers for the AGC of multisource hydrothermal DPS in order to show the optimal PI regulators' advantage. The promising results of the statistically tracked PSO (STPSO) algorithm-based controller are shown in [7]. In another study, BFOA optimizes the integral gains of controllers for LFC of two-area thermal reheat RPSs. According to simulation studies, the system's dynamic performance is further improved when the RFB is coordinated with IPFC units [8]. A new BFOA-based PI^2 controller is proposed and implemented in the AGC loop to ensure load variation stability for a reheat DPS with HES and UPFC coordinated action [9]. The integration of SMES units and RFB into the thermal deregulated LFC system is reported in [10]. Moreover, a multi-stage fuzzy RPS controller is proposed in ref. [11]. IAE, ITAE, and ITSE performance indices compare their performance to FPID and PID controllers to enhance system performance. Another study suggests a novel FA-hybrid fuzzy PIDF controller for DPS for LFC by addressing GRC and GDB nonlinearities. Comparing FA results to tuned GA shows its superiority. Simulation findings show that the developed controllers are resilient to diverse loading situations and system parameters. Finally, the proposed control mechanism is tested under RSLD [12]. The LFC of the two areas of the deregulated thermal system is addressed using a GA-FLC [13]. It outperforms integral and non-tuned fuzzy logic controllers. An MBA-FPID controller is proposed for the LFC of multi-area reheat thermal power systems. The controller takes into account the GDB and GRC nonlinearity. The proposed controller is compared with an ant lion optimizer, ABC, hybrid DE-PSO, and hybrid PSO-PS algorithms. The results of the FPID controller were found to be better [14]. In [15], a DE-PIDF controller is proposed for the AGC of a multi-area DPS using the ITAE criterion. The controller is shown to be superior to FLC. To further improve system performance, an IPFC is considered with RFB. To gain a better understanding of the AGC problem, TD and GRC are taken into account. Furthermore, the study is extended to a hydrothermal system to achieve greater accuracy. The control performance of the BFOA-based FOFPID controller for AGC is compared against FOPID, FPID, and PI/PID controllers based on hFA-PS and GWO algorithm for a two-area hydrothermal power system. The study also includes the hydrothermal and thermal-gas systems. The results show that the BFOA-optimized FOFPID controller performs better than the BFOA-optimized FPID/FOPID/PID and DE/GA-optimized PID controllers in power transactions with/without GRC, TD, and GDB while meeting AGC requirements. Additionally, the proposed approach is tested for its robustness [16]. Nisha et al. [17] customized a three-area hybrid DPS's LFC with a novel type-2 fuzzy-based PD with a dual degree-of-freedom PID controller utilizing the arithmetic optimized African vulture's optimization approach. A particular emphasis is placed on the polar fuzzy-based LFC in [18]. The stabilization of frequency in PV-thermal system with JAYA optimization considering UC and RFB is discussed in [19].

The use of plug-in electric vehicles (EVs) is anticipated to become widespread in the near future, with the intention of controlling the frequency of the system. Dutta et al. examined the regulation of frequency in a deregulated market, specifically focusing on the use of vehicle-to-grid services in residential distribution networks [20]. Reference [21] highlighted optimal LFC of non-reheat thermal DPS with EVs using TLBO. Dutta et al. [22] designed and implemented emotional controller for LFC of renewable energy sources-based DPS with EVs. Control of load frequency based on mixed H_2/H_∞ with EVs discussed in [23]. While considering the CPDN-FOPIDN controller for the three regions' hybrid power systems in the presence of an HVDC connection, Dekaraja addresses the influence that different energy storage units have on the combined ALFC-AVR loop [24]. The first time that the authors have attempted to apply the Harris hawks optimization (HHO)-based 2DOFTID (two-degree-of-freedom tilt–integral–derivative) controller for the ALFC and AVR loops in a multi-area thermal-CCGT included electric vehicle system [25]. In another work, they constructed a fuzzy two-degree-of-freedom tilt–integral–derivative controller, which they referred to as F2DOFTID [26]. The use of Harris' hawks optimization-based fuzzy fractional-order tilt–integral–derivative (FFOTID) controller in the unified voltage–frequency management of a two-area power system that comprises thermal, solar thermal, biodiesel engine generator, and electric vehicles was one of the first attempts that were undertaken [27]. A Coot algorithm-based FO PID-FO PD cascade controller is proposed to suppress oscillation in AGC and AVR systems, including thermal units, renewable energy sources (RES), and electric vehicles. Additionally, real-time validation takes place on the OPAL-RT OP 4510 [28]. A novel master–slave (MS) control is designed for a hybrid power system in which an SSA-based FOPID controller acts as a master, and MPC acts as a slave controller [29]. A CBOA-based cascaded PI-PD controller is proposed for the LFC of power systems with electric vehicles (EVs). It discussed its superiority over ARA: 1+PD, CGA: 1+PI, DSA: FOPID, and FA: PID [30]. In another study, an optimal controller is designed to mitigate frequency and voltage variations by considering the effect of communication time delays in the multisource power system. Area 1 comprises thermal, hydro, and gas units, whereas Area 2 includes wind, solar, and diesel [31]. The cascaded fractional model predictive controller coupled with the FOPID controller (CFMPC-FOPID) is designed and discussed in [32] to justify its superiority.

For a nominal process, linear controllers are usually enough. However, the process often misbehaves when environmental conditions or under operation alter or become unclear. This paper presents a nonlinear PID controller (NPID) to circumvent this restriction. Furthermore, since FONPID controllers have six parameters to adjust (as opposed to FOPID's five), they are more adaptable and perform better than FOPID.

However, tweaking the FONPID controller can be quite challenging [33, 34]. This is a significant aspect of the research, highlighting the complexities and nuances involved in developing and implementing advanced control strategies in power systems and control engineering.

1.2 Research gap and motivation

The literature review provides insight into the research done by the researchers in this domain, who have offered novel approaches to frequency management to assure the continuous power systems operation in the presence of load perturbation in other words the mostly researchers presented linear controllers for frequency management in deregulated AGC systems. Due to the limitations of linear controllers to convey the uniqueness of the suggested investigations compared to the existing body of the literature, nonlinear controllers are presented for frequency management in deregulated AGC systems including the study of sensitivity and stability. The following are the observations that are the most important as follows:

(a) The novelty of our research lies in implementing the SOA-based FO nonlinear PID controller with ultra-capacitor (UC) and electric vehicle (EV) in the anticipated reorganized power system. The effectiveness of electric vehicles (EVs) and ultra-capacitors in this context is a topic that has yet to be thoroughly investigated.
(b) Moreover, the suggested reorganized power system with various input membership functions, such as trapezoidal, difference of two sigmoidal, and triangular, does not implement polar fuzzy controllers (PFCs).

Following a thorough examination of the existing body of research, it has been shown that there are still many unanswered questions that need further investigation. Therefore, it is of utmost importance to conduct a comprehensive study of numerous sources that use a variety of energy storage technologies, taking into consideration as discussed above.

1.3 Key challenges and contributions

1.3.1 Key challenges

In reorganized power system, when the power demand of customers is unpredictable. The task of concerned utility becomes challenging to meet the equilibrium between generation and demand. It can lead to changes in the system's frequency and the expected flow of electricity. The main challenges in addressing this issue are as follows:

(a) Insufficient dampening may cause frequency instability. Using an intelligent AGC controller reduces variations in system frequency and tie-line flows so that the system responds with the shortest settling time, minimal overshoot, and undershoot.
(b) Explore a suitable optimization method to fine-tune the controller design parameters.
(c) To ensure superior and dependable performance of the suggested control strategy in comparison with the current control techniques for advanced power system models.

1.3.2 Contributions

The main contributions of the present research work are as follows:

(a) To recommend a novel SOA-tuned fractional-order nonlinear PID (SOA-FONPID) controller with ultra-capacitor (UC) and electric vehicle (EV) to solve the frequency management issues in a two-area proposed nonlinear thermal reorganized power system subjected to a 10% step load disturbance in each control area.
(b) To determine the advantages of the SOA-FONPID controller with ultra-capacitor (UC) and electric vehicle (EV) as compared to the SOA-FONPID with RFB, GA-FONPID with RFB, GA-FONPID, GA-NPID, and GA-PID controllers.
(c) To assess how resilient the recommended controllers are to random load disturbances (RLD) in each control region.
(d) Another study implements GA-PFCs in a two-area thermal-gas reorganized system with different input membership functions such as trapezoidal, difference of two sigmoidal (Dsigmoidal), and triangular.
(e) To determine the advantages of the GA-PFC (Dsig.) with an energy storage system (RFB) in comparison with GA-PFC (Dsig.) without RFB, GA-PFC (Trap.), GA-PFC (Triang.), and GA-PI controllers.

This paper explores the design and analysis of SOA-FONPID controller with ultra-capacitor (UC) and electric vehicle (EV) to minimize frequency deviation in a two-area reorganized power system considering nonlinearity. The performance of this controller is compared with various controllers tuned by skill optimization algorithm (SOA) as well as genetic algorithm (GA). In another study, an adaptive polar fuzzy controller (PFC) is proposed considering various input membership functions to stabilize a reorganized thermal-gas power system frequency since a fuzzy logic controller's performance depends on a sufficient and accurate knowledge base.

As the number of rules in a knowledge base increases, its complexity increases, affecting the computation time and memory requirements. The remainder of the research is structured in the following manner: The model that was studied is described in Sect. 2, while Sect. 3 discusses energy storage batteries. Section 4 discusses the methodology used in the manuscript. Section 5 depicts describes the performance of various nonlinear controllers such as SOA-FONPID with UC and EV, SOA-FONPID with RFB, GA-FONPID with RFB, GA-FONPID, GA-NPID, and GA-PID including its robustness; Sect. 6 describes the performance of GA-PFCs considering trapezoidal, dsigmoidal, and triangular membership functions with RFB for thermal-gas AGC systems; and the last section summarizes the study's findings.

2 System under study

In the proposed two-area system as shown in Fig. 1, region 1 comprises two generation organizations, particularly GENCO-1 and -2 (reheat-thermal generation units), with two distribution organizations, specifically DISCO-1 and DISCO-2. Each GENCO has a reheat thermal unit. Additionally, region-two likewise incorporates two generation companies, namely, GENCO-3 and -4 (reheat-thermal generation units), and two DISCOs, namely, DISCO-3 and DISCO-4. The AC transmission link interconnects these two control areas. The thermal region has reheat turbines with 3% per-minute generating constraints. DPM simplifies contract visualization. This illustrates DISCO's deal with GENCO. DPM has four rows for GENCOs and four columns for DISCOs. Consequently, each "contract participation factor" (cpf) entry in the matrix represents the proportion of a total load contracted by a DISCO-j from a GENCO-i. Matrix columns add to unity. The $\text{apf}_{11} = \text{apf}_{12} = \text{apf}_{21} = \text{apf}_{22} = 0.5$.

The transfer function (TF) of the power system is given below.

$$G_{ps}(s) = \frac{K_{ps}}{1 + sT_{ps}}$$

The TF of the thermal speed governor with dead band

$$G_{Tg}(s) = \frac{-(0.2/\pi)s + 0.8}{sT_{\text{g}} + 1}$$

The TF of steam turbine

$$G_{Tt}(s) = \frac{1}{1 + sT_t}$$

The TF of the reheater

$$G_{Tr}(s) = \frac{1 + sK_rT_r}{1 + sT_r}$$

The stable-state tie-line flow for an EHVAC-connected system is as follows:

$$\Delta P_{tie}^{actual} = \frac{2\pi T_{12}}{S}(\Delta F_1 - \Delta F_2) \tag{1}$$

$$\Delta P_{tie}^{Scheduled} = \sum_{i=1}^{3}\sum_{j=3}^{4} Cpf_{ij}\Delta P_{Lj} - \sum_{i=4}^{6}\sum_{j=1}^{2} Cpf_{ij}\ \Delta P_{Lj} \tag{2}$$

$$\Delta P_{tie}^{Error} = \Delta P_{tie}^{Actual} - \Delta P_{tie}^{Scheduled} \tag{3}$$

In the stable zone, condition (3) becomes zero. The area control errors (ACEs) in two regions are as follows:

$$ACE_1 = B_1\Delta F_1 + \Delta P_{tie}^{Error} \tag{4}$$

$$ACE_2 = B_2\Delta F_2 + \alpha_{12}\Delta P_{tie}^{Error} \tag{5}$$

3 Energy storage system

3.1 Redox flow batteries (RFB)

Figure 2 shows RFB blocks. In mild sulfuric acid electrolytes, RFB stores energy using redox couplings V^{2+}/V^{3+} in the − ve half cells and V^{4+}/V^{5+} in the + ve half cells. During charging and discharging, positive H + ions travel between electrolyte tanks. When load demand spikes, RFBs rapidly return energy to the system. A pair of converters convert AC/DC or DC/AC. When power generation and unanticipated load demand are out of sync, RFBs help. Thus, they may increase AGC performance and electrical quality [4]. RFB's first-order lag transfer function is presented here. Redox flow batteries have anolyte and catholyte at their negative and positive electrodes. Active materials dissolved in tanks are continually delivered to the electrodes during discharge and returned to the tank. A battery-powered device's

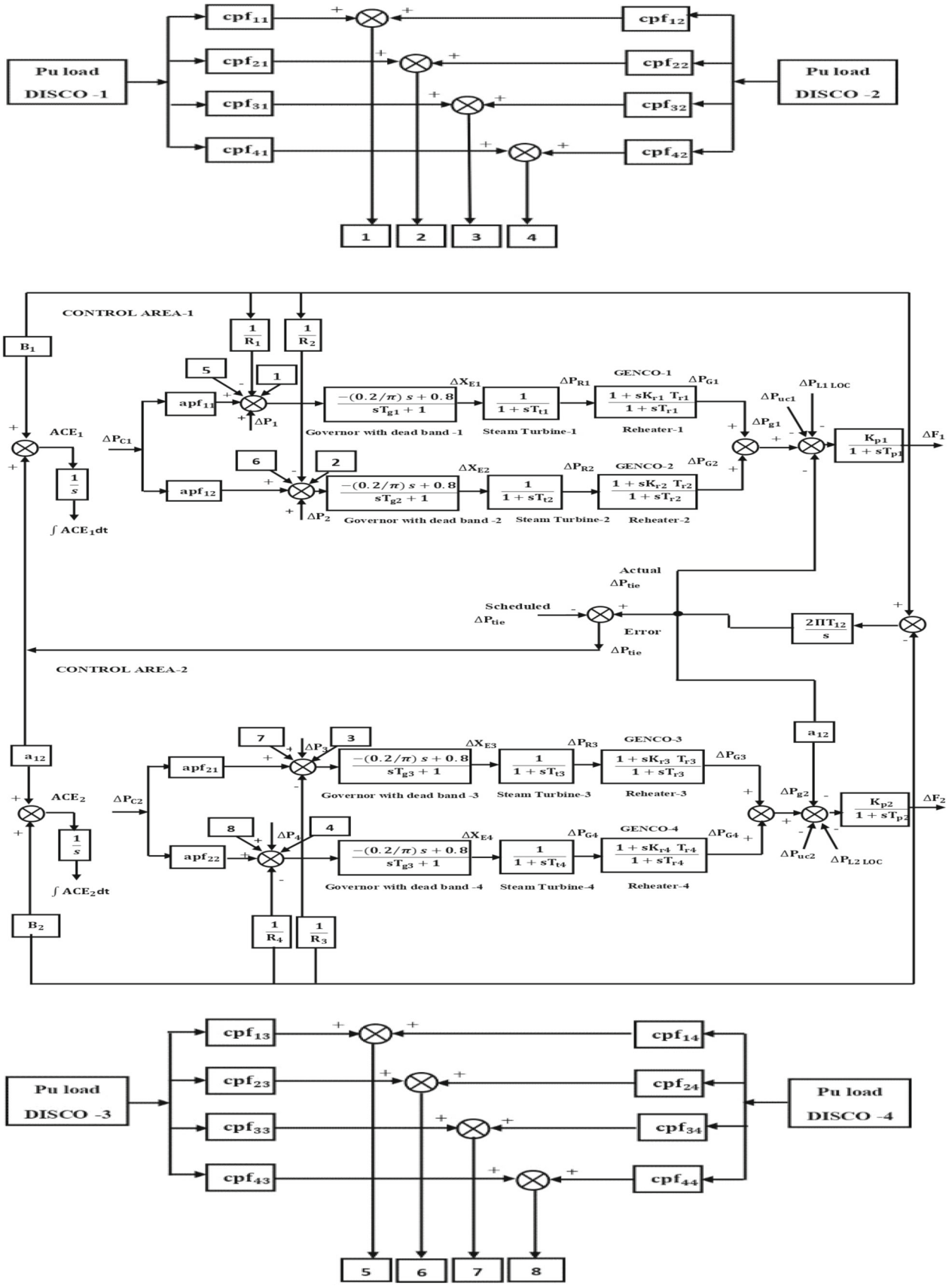


Fig. 1 Stucture of proposed reorganized system

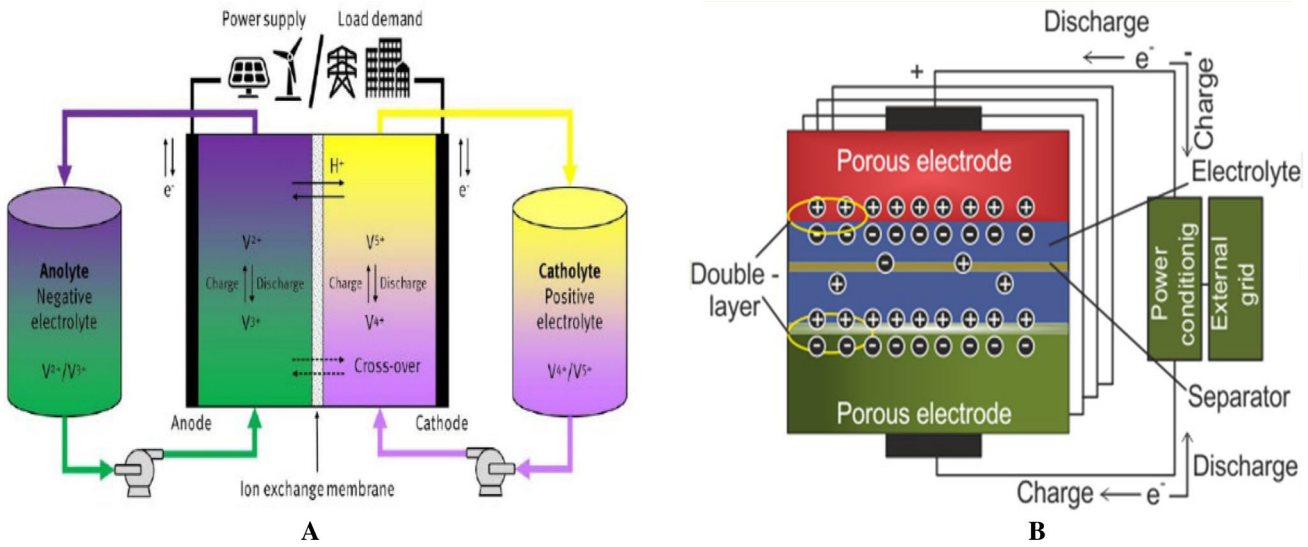


Fig. 2 **A** RFB in frequency management. **B** Ultra-capacitor (UC) in frequency management

current flows between electrodes during charge exchange. Though electrolyte energy density is too low, RFBs are being studied for electric cars. An RFB may be "refilled" in minutes by draining and recharging the electrolyte. Redox couples such as poly S–Br and Fe–Ti have been studied and tested using RFBs.

3.2 Ultra-capacitor (UC)

A novel electrochemical energy storage device known as UC, a supercapacitor, or electric double-layer capacitor, has emerged. A potassium hydroxide electrolyte, two permeable anodes, and an ion exchange layer are employed in this system. A UC's surface area is significantly larger than a capacitor because of the presence of a fluid electrolyte and permeable terminals. The incorporation of a thin double layer into UC results in an increase in its capacitance. These two attributes guarantee that a UC has a 100–1000 times greater capacitance than that of an electrolytic capacitor. UC exhibits superior power density compared to batteries. The specific energy ranges from 1 to 10 Wh/kg, while the power ranges from 1000 to 5000 W/kg. It exhibits a faster rate of charging and discharging than a battery. Therefore, UC is reliable and well-suited for AGC. The control zone's frequency variation will work as the input to UC, and its TF is:

$$\Delta P_{UC}(s) = \left\{ \frac{K_{UC}}{1 + sT_{UC}} \right\} \Delta F(s) \tag{6}$$

3.3 Distributed energy resource (electric vehicle)

Intelligent charging technology known as vehicle-to-grid allows EV batteries to contribute energy to the power grid. It functions as distributed energy resources (DER) during periods of high demand to provide additional capacity to support and stabilize the grid. Incorporating V2G-enabled EVs presents additional obstacles to preserving this equilibrium. It is extremely complex to connect V2G-enabled EVs to the grid; therefore, developers of both EVs and EV supply equipment must assure safety, dependability, and power quality. Each EV is equipped with a dead band function featuring droop characteristics to mitigate the risk of disconnection from the grid, which could cause an undesirable frequency response. The dead band's upper limit (Δf_{ul}) and lower limit value (Δf_{ll}) are set at 10 mHz and − 10 mHz, respectively. 2.4 Hz/p.u.MW is the value of the aggregate model droop coefficient (RAG). ΔP_{EV} stands for the incremental generation change of EV. $\Delta P_{AG,\max}$ and $\Delta P_{AG,\min}$ represent EV fleets' highest and lowest power output. Its calculation is as follows:

$$\Delta P_{AG,\max} = +\left(\frac{1}{N_{EV}}\right) \Delta P_{EV} \tag{7}$$

$$\Delta P_{AG,\min} = -\left(\frac{1}{N_{EV}}\right) \Delta P_{EV} \tag{8}$$

The value of K_{EV} can be determined with the help of state of charge (SOC) of EV.

$$K_{EV} = K_{\max} \left\{ 1 - \left(\frac{SOC - SOC_{low(high)}}{SOC_{\max(\min)} - SOC_{low(high)}} \right)^{n} \right\} \tag{9}$$

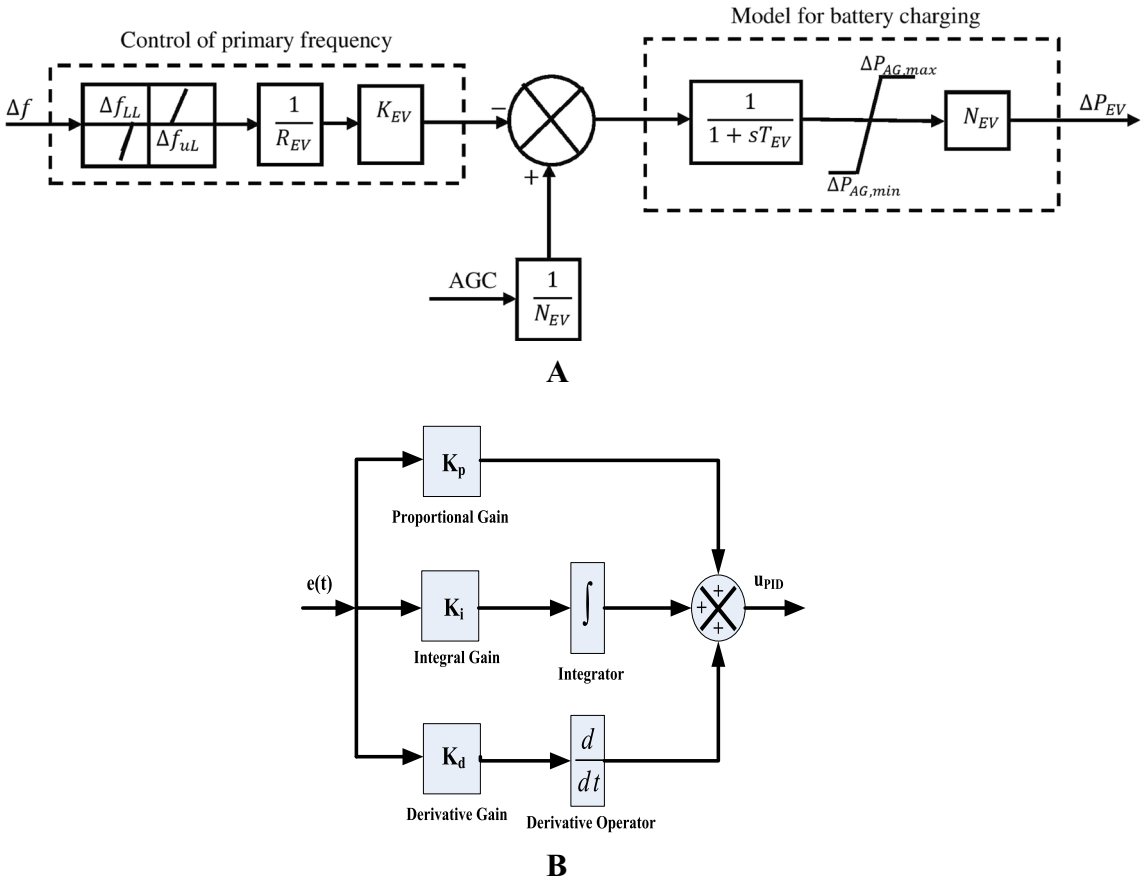


Fig. 3 **A** Structure of EV and **B** structure of PID controller

The number of electric vehicles that are linked is denoted by the symbol N_{EV}. The equation that describes the transfer function of EV is as follows:

$$G_{EV} = \left\{ \frac{K_{EV}}{1 + sT_{EV}} \right\} \tag{10}$$

The K_{EV} and T_{EV} are abbreviations that stand for the gain of electric vehicles and the time constant, respectively.

4 Methodology

This section describes the design of PID, NPID, FONPID, and polar fuzzy controllers. A genetic algorithm (GA) and skill optimization algorithm (SOA) are used to optimize the gain of controllers.

4.1 The design of controllers

This research uses the PID, integer-order NPID, and FO-NPID controllers for frequency management in a reorganized power system. These controllers are tuned by using GA and SOA considering the performance metric ITAE. The control strategies are described in the following segments:

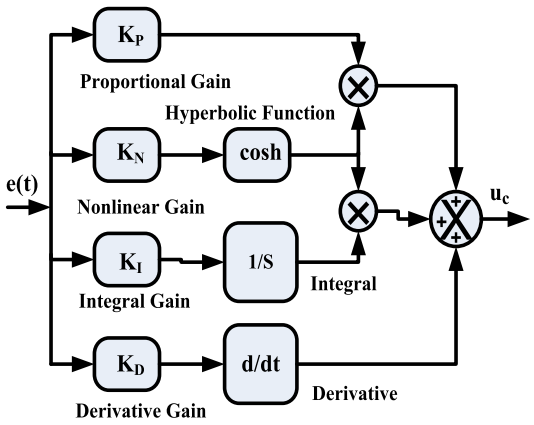
4.1.1 The PID controller

The PID controller is the kind of feedback controller most commonly used in process industries. Proportional controller shortens the rising time. Integral controls remove steady-state inaccuracy. Derivative controls increase overshoot, system stability, and transient response. Its design includes three parameters: K_P, K_I, and K_d. Figure 3B shows that it is basic setup. It can be represented by Eq. (11) in time-domain.

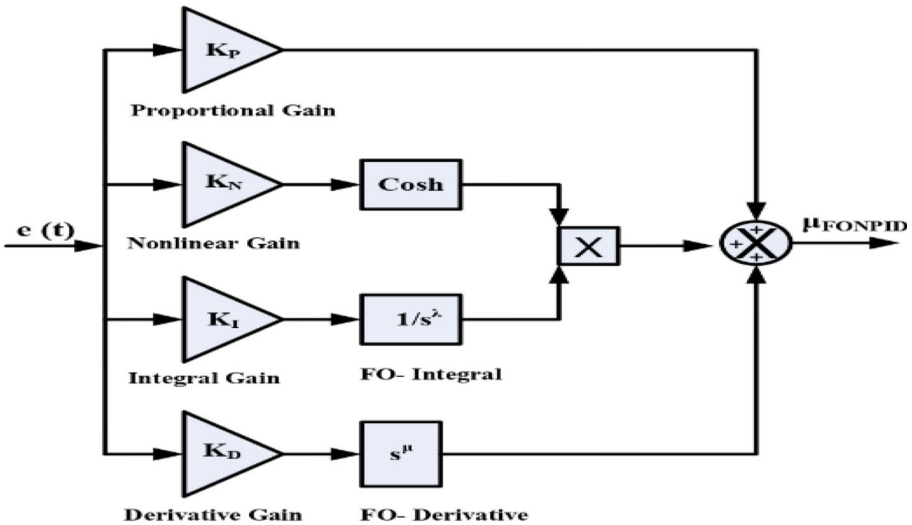
$$u_{PID}(t) = \left[k_p e(t) + k_i \int e(t)dt + k_d \frac{de(t)}{dt} \right] \tag{11}$$

Its s-domain representation is given by Eq. (12).

$$u_{PID}(s) = \left[k_p + k_i * \frac{1}{s} + k_d * s \right] E(s) \tag{12}$$



A



B

Fig. 4 **A** Structure of NPID controller. **B** Structure of FONPID controller

4.1.2 The nonlinear PID (NPID) controller

Linear systems do not exist in practice. Nonlinearity is common in practical systems. PID controllers function poorly when operational or environmental circumstances change or are unavailable. In this research, an NPID controller is designed to automatically modify controller gain and integral time based on control errors. K_p and K_i gains are represented by one nonlinear hyperbolic function. The structure of the NPID controller is shown in Fig. 4A. The system's nonlinearity makes the NPID controller adaptive by permitting real-time gain modifications to the proportional and integral controllers. Equation (13) describes a time-domain NPID controller.

$$u_{NPID}(t) = \left[K_P e(t) f(e) + K_I \int f(e) e(t) dt + K_D \frac{de(t)}{dt} \right] \tag{13}$$

where f (e) is a nonlinear proportional gain factor that is also integral, defined as a function of the error e, written as follows:

$$f(e) = \cosh(K_N\ e) \tag{14}$$

$$\text{or, } f(e) = \frac{\exp(K_N e) + \exp(-K_N e)}{2} \tag{15}$$

$$\text{where, } e = \begin{cases} e : |e| \leq e_{\max} \\ e_{\max} * \operatorname{sgn}(e) : |e| \succ e_{\max} \end{cases} \tag{16}$$

K_N and e_{max} are defined as the positive values taken into account by the user in this study. If e was equal to zero, the minimum value of f (e) would be 1. In this case, GA optimizes K_N, while $e_{max} = 30$ is assumed [35].

4.1.3 The fractional-order nonlinear PID (FONPID) controller

FONPID controllers use fractional-order and nonlinear terms to accumulate integral gain. The structure of the FONPID controller is shown in Fig. 4B. FO operators increase DOF and provides accurate control solutions. The best scenarios for incorporating these design aspects are those that occur in real-world settings. For optimizing such FONPID controllers, various advanced techniques have been put forth. Investigators are aware of numerous methods for fractional-order items, i.e., for the purpose of integrating the FO—derivative and integral, the FOOA approach is used. This approximation method keeps these updated throughout the optimization process [36]. According to this method, the predicted transfer function is provided by

$$O_f(s) = S^{\alpha} = K \prod_{-N}^{N} \frac{s + \omega'_k}{s + \omega_k} \tag{17}$$

where α is the differ-integration order, and (2N + 1) is the approximation filter order. The Oustaloup filter's zeros and poles are ω'_k and ω_k, where K is the gain. This study uses a 5th-order Oustaloup filter as described in [36].

$$\omega_k = \omega_b \left(\frac{\omega_h}{\omega_b}\right)^{\frac{k+N+\frac{1}{2}(1+\alpha)}{2N+1}},$$
$$\omega'_k = \omega_b \left(\frac{\omega_h}{\omega_b}\right)^{\frac{k+N+\frac{1}{2}(1-\alpha)}{2N+1}}, \quad K = \omega_h^{\alpha} \tag{18}$$

4.1.4 Polar fuzzy controller (PFC)

A fuzzy controller needs accurate knowledge to perform. Knowledge bases with more rules are more complicated, requiring more processing and memory. In Fig. 5, fuzzy controllers use the variance in system frequency (ΔF) and its integral ($\Delta F'$), translated into polar coordinates for PFC, to calculate the control action. Due to its simplicity, polar fuzzy controller (PFC) has only two rules. PFC only utilizes one input angle, which depends on the ratio of suitably scaled inputs. Thus, system frequency and integral changes do not require two input gains. The ΔF or $\Delta F'$ is calculated using only one parameter. Figure 6's origin "0" is the ideal equilibrium position. Thus, all control attempts should advance vector R in that direction. The fuzzy input immediately generates the PFC's output angle. For zone A (0°–90°), FLC should be considerably positive as both ΔF and $\Delta F'$ are positive. Zone B (315°–360°) requires a low-positive FLC signal due to notable positive scaled ΔF and somewhat negative $\Delta F'$. Zone C (270°–315°) requires a low negative control signal from FLC, which leads to a little positive scaled ΔF and a significant negative $\Delta F'$. While zones D, E, and F differ substantially from A, B, and C. Trapezoidal, dsigmoidal, and triangular are considered as input membership functions. While triangular is considered as output membership function, and two simple rules are considered [18]:

Rule1: If (angle is LP), then (output is P); Rule2: If (angle is LN), then (output is N).

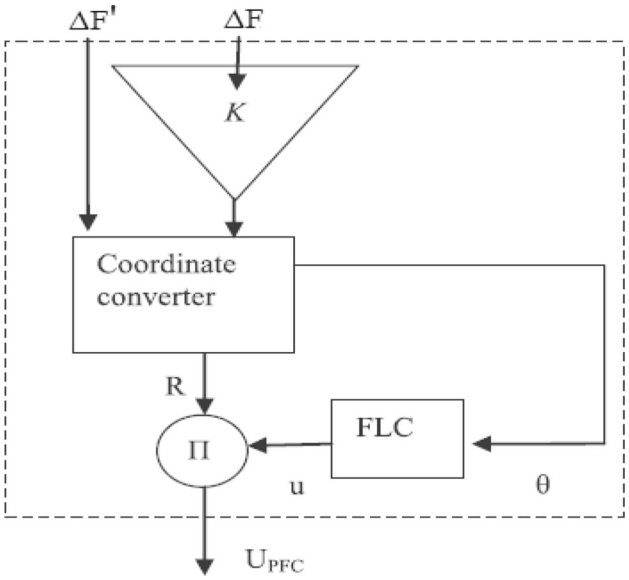


Fig. 5 Structure of PFC

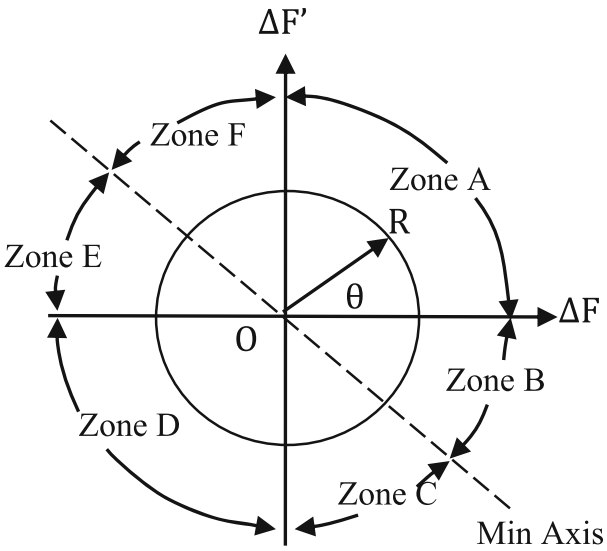


Fig. 6 Plan for six zones

In the context of this study, the integral time and absolute error (ITAE) serves as the objective function, and its expression is as follows:

$$J : ITAE = \int_0^{tsim} t.(|\Delta F_1 + \Delta F_2 + \Delta P_{tie}|).dt \tag{19}$$

4.1.5 Genetic algorithm (GA)

The gain settings for the suggested controllers are optimized using the GA optimization approach. The Darwinian principle of evolution, which represents the maxim "survival of the fittest," is the basis for the flowchart of GA, as shown in Fig. 7. It belongs to the class of evolutionary algorithms as a random probability distributive optimizer. Bit strings, called solution vectors or chromosomes, store GA solution variables. Over generations, crossing, reproduction, and mutation build these chromosomes, duplicating natural evolution. Optimization requires an objective function. Notable GA markers include mutation fraction and crossover percent. Mutation criteria determine the population's mutation rate outside the elite. The crossover criterion shows the number of non-elite group members who advanced via crossover. Carefully choose crossover and mutation criteria [35, 37]. Table 1 shows the suggested controllers' tuned gain parameters. Figure 8B compares performance indexes.

4.1.6 Skill optimization algorithm [38]

Skill optimization algorithm (SOA) is a novel metaheuristic optimization tool that draws inspiration from human skill development and enhancement. Participants in population-based SOA are real people who are always growing and learning. Members of an SOA can resolve the optimization problem. The search space placements of these members define the values of the problem decision variables. SOA member assignments are initially random. A matrix may be used to represent the SOA population (Eq. 20).

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} x_{1,1} & \cdots & x_{1d} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i,1} & \cdots & x_{i,d} & \cdots & x_{i,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,d} & \cdots & x_{N,m} \end{bmatrix} \tag{20}$$

X is the population matrix for the SOA. Each member is inserted into the problem's variable to compute the objective function value. Hence, a vector in accordance with Eq. (21). F is a vector of all objective function values, and F_i is the value based on the ith candidate solution.

The objective function values identify the best and poorest members. The best and worst population members vary with each iteration because the objective function values and population members are modified.

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{NX1} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{NX1} \tag{21}$$

SOA population updates include exploration as well as exploitation. The exploration phase simulates learning a skill from an expert. You'll imitate users' skill improvement throughout exploitation. In SOA design, the update process comprises two phases: exploration and exploitation. Exploration searches the problem-solving space globally, while exploitation searches locally. During the investigation phase, SOA members followed each other's instructions rather than following the best individual. This increases the algorithm's exploration capability, helping it better scan the search space and find the optimum spot. Local search near each population member helps the algorithm converge on better likely solutions during the exploitation phase.

In the first step, SOA members learn a skill from a community expert. A person's contribution to the population is directly proportional to their objective function. The objective function value determines if a SOA member is an expert. The "expert set" for a SOA member includes those with a higher objective function value. One of these people will randomly mentor the person in question. So, the professional picked to lead the SOA member may not be the best choice. The optimum solution is a non-rotating expert set member applicable to all SOA members. The population is guided to various search spaces by learning the algorithm's global search and exploration skills. Experts handle this. If the new location estimated for each population member increases the goal function, it may be acceptable.

Hence, equations may define the initial phase of the update in line with the concepts given.

$$X_i^{P1}{:}X_{i,d}^{P1} = x_{i,d} + r\ X\ (E_{i,d} - 1\ X\ x_{i,d}),\ E_i = X \tag{22}$$

Where $F_k \prec F_i$ and k is randomly selected from{1, 2, ...N},

$$k \neq i \quad X_i = \begin{Bmatrix} X_i^{p1}, & F_i^{p1} \prec F_i \\ X_i, & \text{else} \end{Bmatrix} \tag{23}$$

On the basis of the first phase, X_i^{P1} represents the ith candidate solution's newly estimated status, and so on. Everyone studies and practices alone in the second phase to improve. Local search increases exploitation in SOA by having each

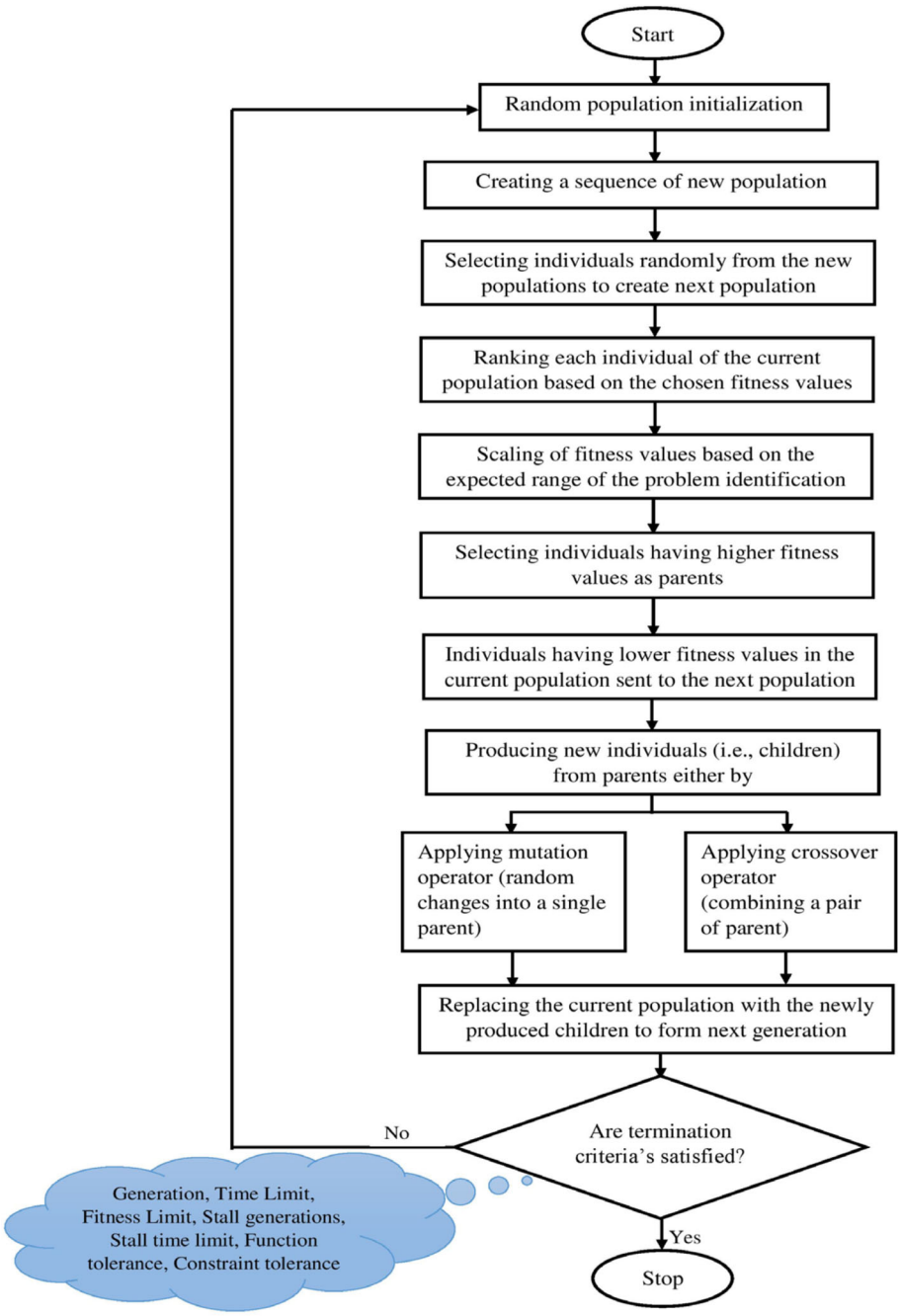


Fig. 7 The framework of the GA flowchart

member hunt for better conditions near its position to raise its objective function value. This encourages exploitation (shows expertise). In the second phase, everyone studies and practices alone to improve their skills. The newly estimated position is acceptable, as with the prior step, if it raises the aim function. Equations describe SOA updating features mathematically.

$$X_i^{p2}{:}X_{i,d}^{p2} = \begin{Bmatrix} x_{i,d} + \frac{1-2r}{t} X\ x_{i,d}, & r \prec 0.5 \\ x_{i,d} + \frac{lb_{i,d}(ub_j - lb_j)}{t} X\ x_{i,d}, & else \end{Bmatrix} \tag{24}$$

$$X_i = \begin{Bmatrix} X_i^{p1}, & F_i^{p1} \prec F_i \\ X_i, & else \end{Bmatrix} \tag{25}$$

Table 1 The proposed controllers' tuned gain parameter

Sl. No.	Controller/gain	K_{p1}	K_{i1}	K_{d1}	K_{N1}	λ_1	μ_1	K_{p2}	K_{i2}	K_{d2}	K_{N2}	λ_2	μ_2
1	GA:PID	0.427	2.997	0.662	–	–	–	0.658	2.998	0.355	–	–	–
2	GA:NPID	0.107	0.899	0.129	0.629	–	–	0.103	0.898	0.589	0.677	–	–
3	GA:FONPID	0.436	2.999	1.038	0.538	0.99	0.96	1.566	2.998	1.849	0.849	0.98	0.99
4	GA:FONPID with RFB	0.445	3.019	1.136	0.639	0.98	0.96	1.667	2.907	1.849	0.857	0.99	0.98
5	SOA:FONPID with RFB	0.438	3.998	1.039	0.539	0.989	0.957	1.568	4.907	1.846	0.858	0.988	0.977
6	SOA:FONPID with UC and EV	0.429	5.019	1.138	0.594	0.968	0.968	1.577	5.907	1.889	0.849	0.991	0.996

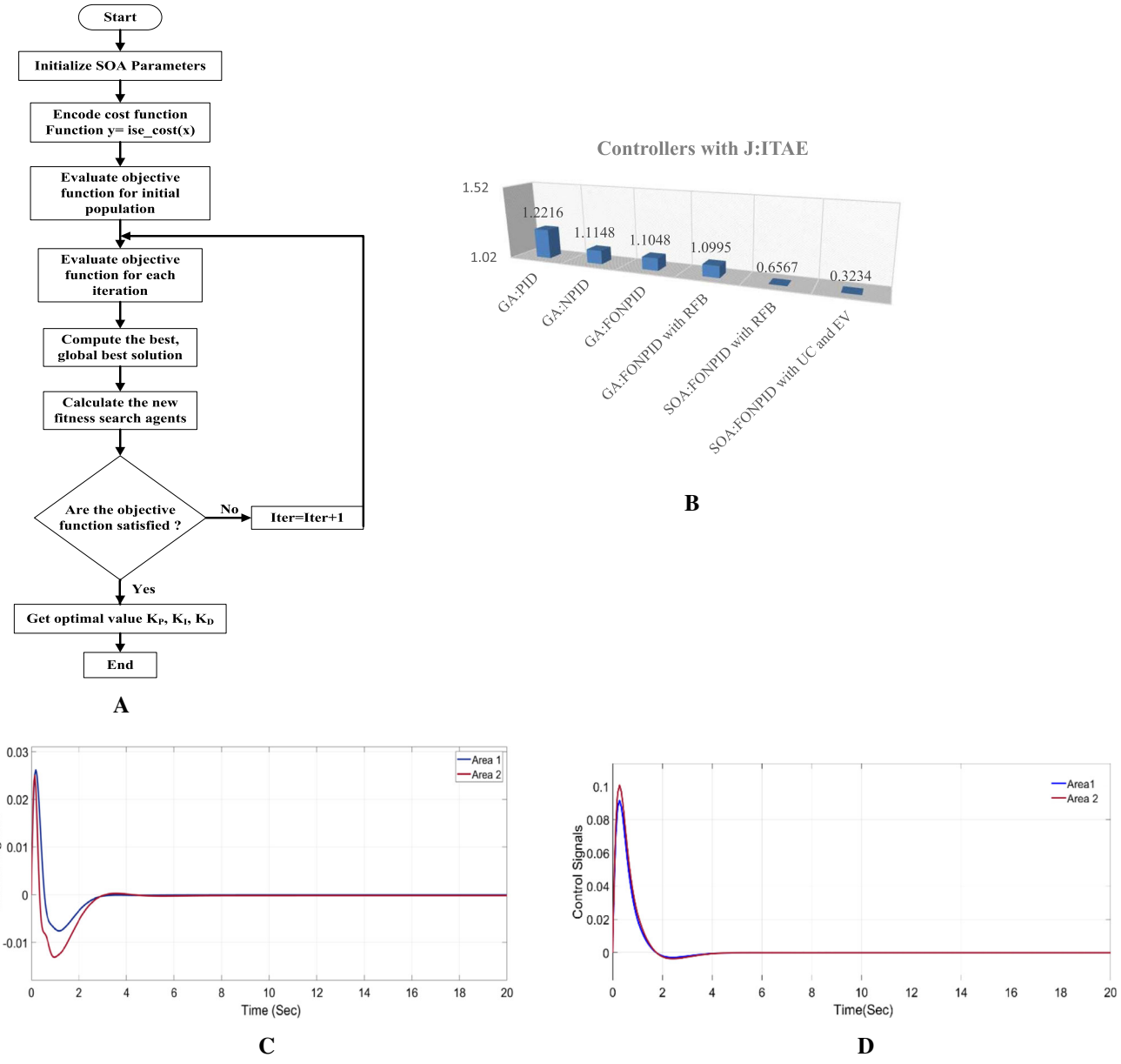


Fig. 8 **A** The flowchart of SOA, **B** performance index (J_{ITAE}) comparison, **C** control signals for CV [SOA:FONPID with UC and EV], and **D** control signals for CV[GA:PFC (Dsig with RFB)]

Based on the second phase, X_i^{P2} reflects the ith candidate solution's most recent calculation of its status, and so on. The initialization of the population and the initial assessment of the objective function, which is equal to O (Nm), where N is the number of population members, and m is the number of issue variables, account for the computational complexity of SOA. The SOA member update process consists of two stages, and each phase and iteration evaluates the updated member's location relative to the goal function. Because T is the maximum number of repetitions, the computational complexity of SOA throughout the upgrading process is thus equal to O (2NmT). As a result, O (Nm (1 + 2 T)) equals the entire complexity of SOA [38].

5 Simulation results and analysis

This section investigates the performance of the two-area thermal–thermal reorganized power system with a governor that exhibits dead band nonlinearity using a few nonlinear controllers adjusted by GA and SOA for various kinds of power transactions. Figure 8C shows the SOA:FONPID with UC and EV controller signals for control areas 1 and 2 while considering contract violation (CV).

Multiple simulations have been carried out inside the MATLAB/Simulink environment. The GA and SOA programs are implemented using m files, and this research considers a total of one hundred repetitions (N). Each GENCO's participation in AGC may be seen via the following apfs: $apf_1 = apf_2 = apf_3 = apf_4 = 0.5$;

5.1 Test I—10% SLP in each zone

5.1.1 Poolco transaction (PT)

DISCO1 and DISCO2 necessitate load only. DISCO3 and DISCO4 do not require load from GENCOs; hence, DPM matrix cpfs are 0. The DPM matrix governs DISCO-GENCO contracts, as given in Eq. (26). An algorithm, namely, GA, is used to optimize PID, NPID, and FONPID controllers gain parameters for simulation and is depicted in Table 1. The growth in settling time for ΔF_1 is found to be 68.48%, 34.55%, and 18.26% by GA-FONPID controller with RFB as compared to GA-PID, GA-NPID, and GA-FONPID controllers, respectively. Similarly, for ΔF_2, these are 68.03%, 42.93%, and 29.72% by GA-FONPID controller with RFB as compared to GA-PID, GA-NPID, and GA-FONPID controllers, respectively. The growth in settling time for ΔP_{tie} is found to be 53.50%, 33.52%, and 18.87% by GA-FONPID controller with RFB. Figures 9, 10, and 11 depict the suggested system reactions. The improvement in settling time for ΔF_1, ΔF_2, and ΔP_{tie} is 17.99%, 27.09%, and 31.03%, respectively, by SOA:FONPID with UC and EV as compared to GA:FONPID with RFB.

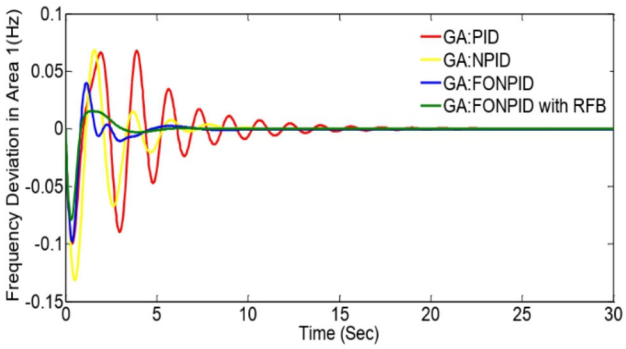


Fig. 9 ΔF_1 responses for PT

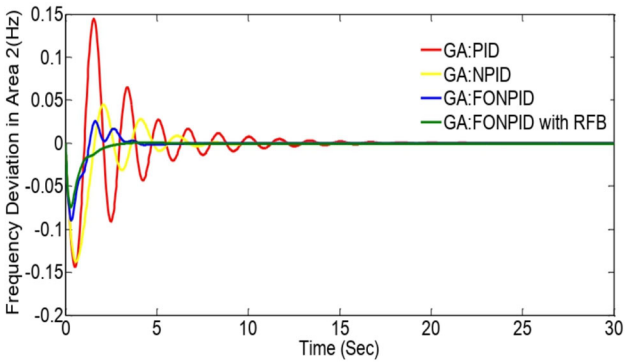


Fig. 10 ΔF_2 responses for PT

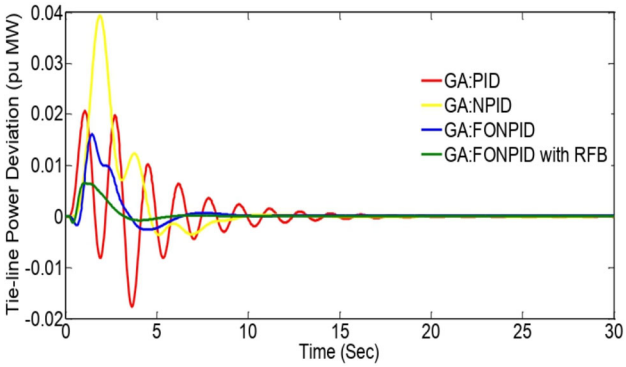


Fig. 11 ΔP_{tie} responses for PT

$$\text{DPM for PT} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{26}$$

$$\text{DPM for BT} = \begin{bmatrix} 0.5 & 0.25 & 0 & 0.3 \\ 0.2 & 0.25 & 0 & 0 \\ 0 & 0.25 & 1 & 0.7 \\ 0.3 & 0.25 & 0 & 0 \end{bmatrix} \tag{27}$$

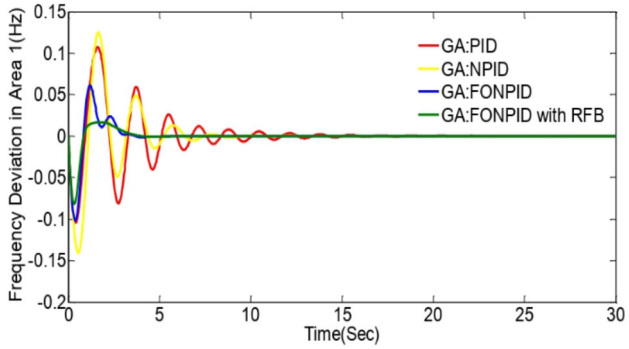


Fig. 12 ΔF_1 responses for BT

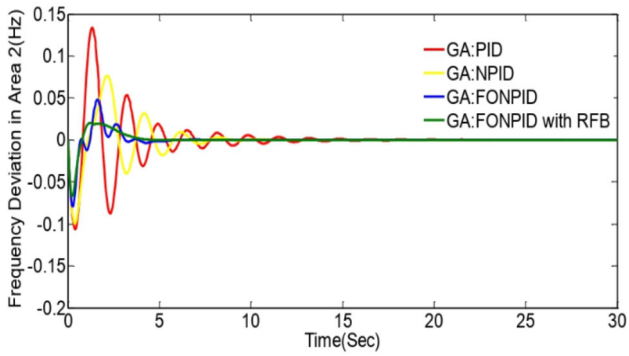


Fig. 13 ΔF_2 responses for BT

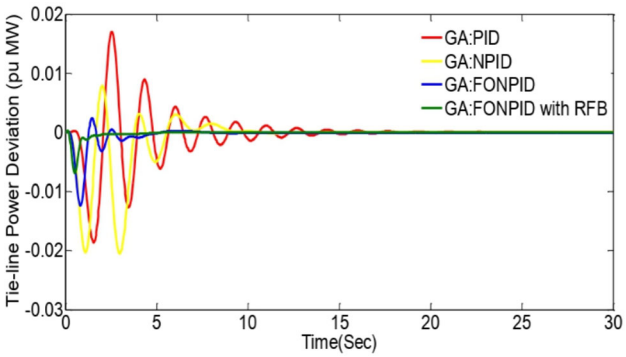


Fig. 14 ΔP_{tie} responses for BT

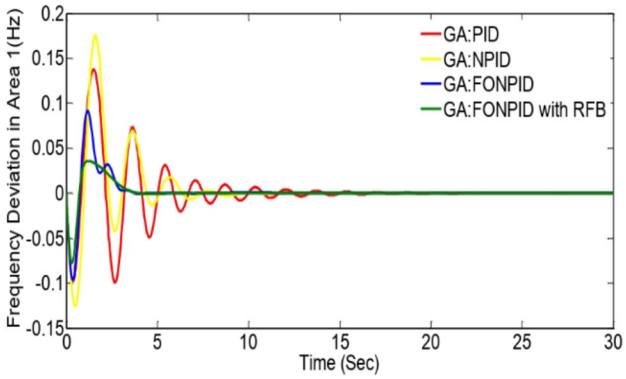


Fig. 15 ΔF_1 responses for CV

5.1.2 Bilateral transaction (BT)

GENCOs may have power agreements with any DISCOs in this transaction. The DPM matrix governs DISCO-GENCO contracts, as shown above. An algorithm, namely, GA, is used to optimize PID, NPID, and FONPID controllers gain parameters for simulation and is depicted in Table 1. The growth in settling time for ΔF_1 is found to be 79.56%, 57.93%, and 23.75% by GA-FONPID controller with RFB as compared to GA-PID, GA-NPID, and GA-FONPID controllers, respectively. Similarly, for ΔF_2, these are 75.71%, 57.19%, and 20.24% by GA-FONPID controller with RFB as compared to GA-PID, GA-NPID, and GA-FONPID controllers, respectively. The growth in settling time for ΔP_{tie} is found to be 67.17%, 58.08%, and 12.39% by GA-FONPID controller with RFB. Figures 12, 13, and 14 depict the suggested system reactions. The improvement in settling time for ΔF_1, ΔF_2, and ΔP_{tie} is 18.74%, 35.29%, and 43.4%, respectively, by SOA:FONPID with UC and EV as compared to GA:FONPID with RFB.

5.1.3 Contract violation (CV)

DISCO may need more electricity than the contract. The DISCOs' local GENCOs must provide extra electricity. DISCO1 requires 0.005 p.u. MW of surplus electricity, which is the local load. Area-1's local load is Delta $P_{L1,\text{loc}}$ = load of DISCO1 (0.005) + load of DISCO2 (0.005) + 0.005 = 0.015 p.u. MW**.** Moreover, DISCO3 requires 0.005 p.u. MW of surplus electricity, which is the local load. Area 2's local load is Delta $P_{L2,\text{loc}}$ = load of DISCO3 (0.005) + load of DISCO4 (0.005) + 0.005 = 0.015 p.u. MW**.** The bilateral case's DPM matrix persists here. DISCO1's un-contracted load tends to generate GENCO1 and GENCO2. At the same time, DISCO3's un-contracted load tends to generate GENCO3 and GENCO4. The growth in settling time for ΔF_1 is found to be 73.71%, 47.88%, and 18.50% by GA-FONPID controller with RFB as compared to GA-PID, GA-NPID, and GA-FONPID controllers, respectively. Similarly, for ΔF_2, these are 69.6%, 46.29%, and 7.79% by GA-FONPID controller with RFB as compared to GA-PID, GA-NPID, and GA-FONPID controllers, respectively. The settling time has improved for ΔP_{tie} and was found to be 64.41%, 53.11%, and 5.04% by GA-FONPID controller with RFB. Figures 15, 16, and 17 depict the suggested system reactions. Figures 18, 19, 20, 21, 22, 23, 24, 25, and 26 illustrate the dynamic reactions that occur for a variety of power transactions when the proposed system is subjected to random load disturbance (RLD).

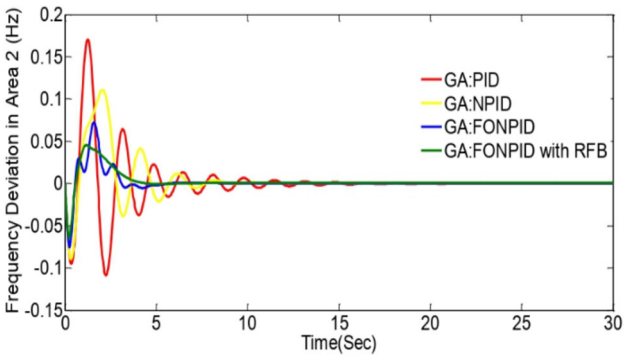


Fig. 16 ΔF_2 responses for CV

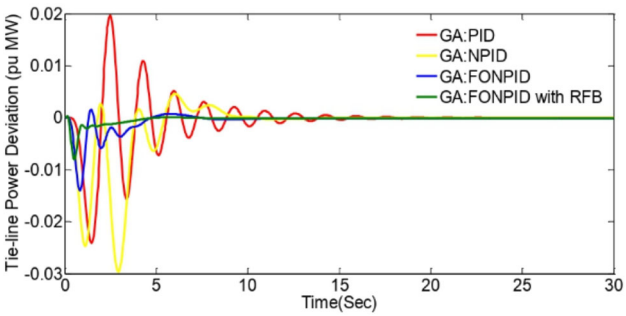


Fig. 17 ΔP_{tie} responses for CV

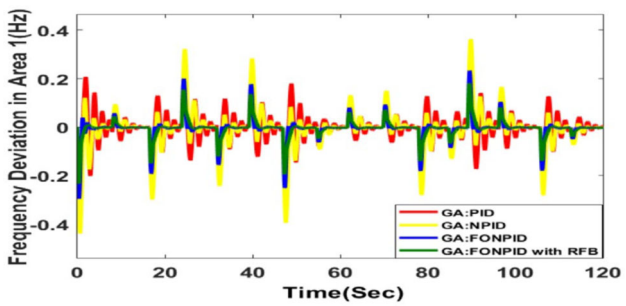


Fig. 18 ΔF_1 responses for PT considering RLD

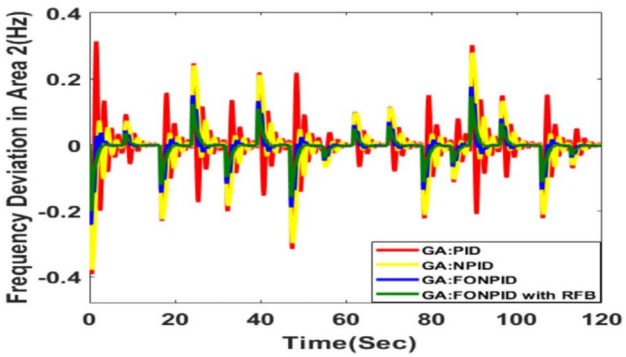


Fig. 19 ΔF_2 responses for PT considering RLD

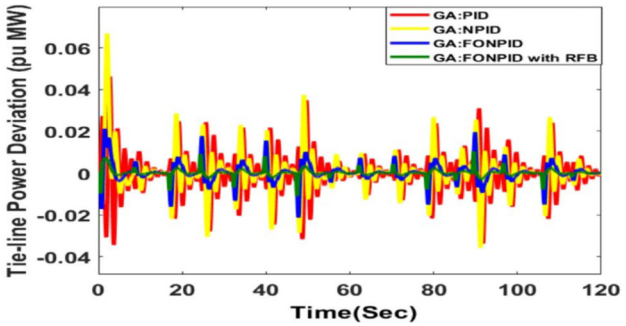


Fig. 20 ΔP_{tie} responses for PT considering RLD

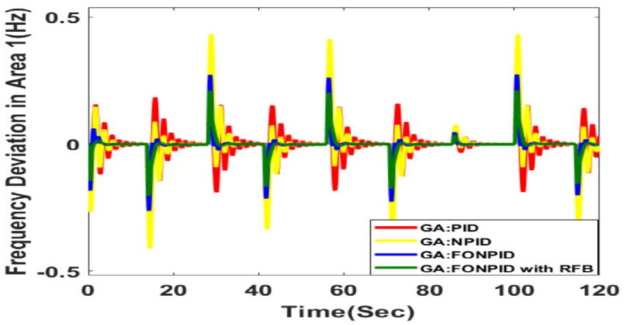


Fig. 21 ΔF_1 responses for BT considering RLD

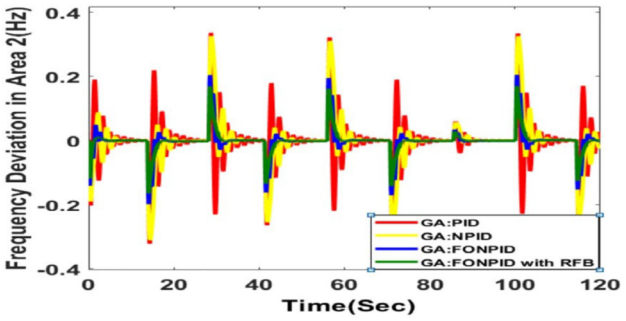


Fig. 22 ΔF_2 responses for BT considering RLD

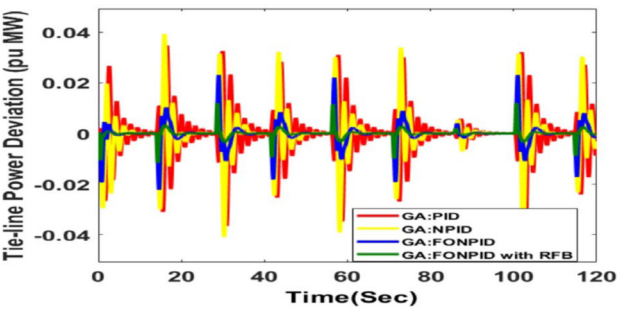


Fig. 23 ΔP_{tie} responses for BT considering RLD

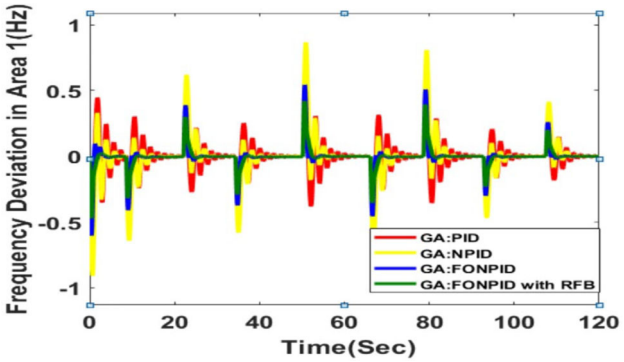


Fig. 24 ΔF_1 responses for CV considering RLD

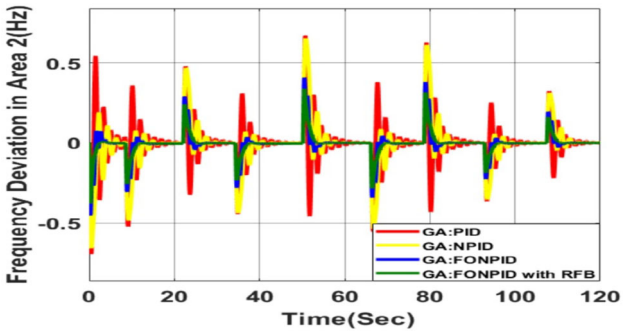


Fig. 25 ΔF_2 responses for CV considering RLD

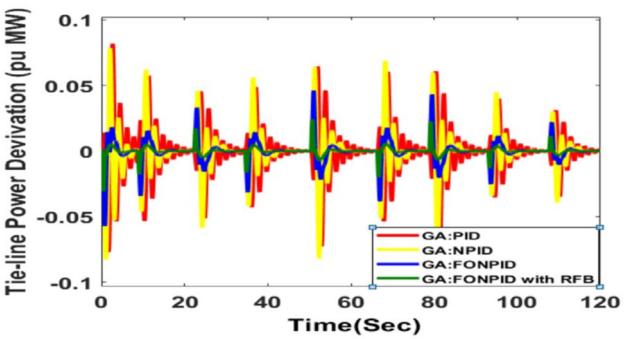


Fig. 26 ΔP_{tie} responses for CV considering RLD

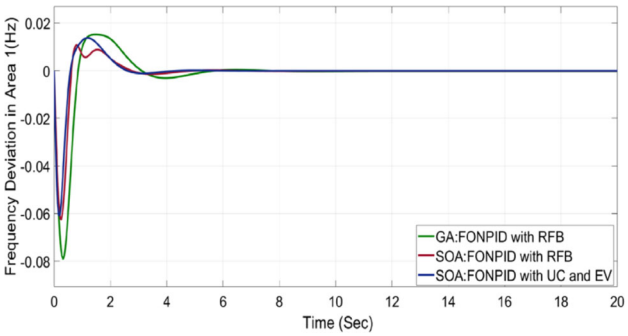


Fig. 27 ΔF_1 responses for PT

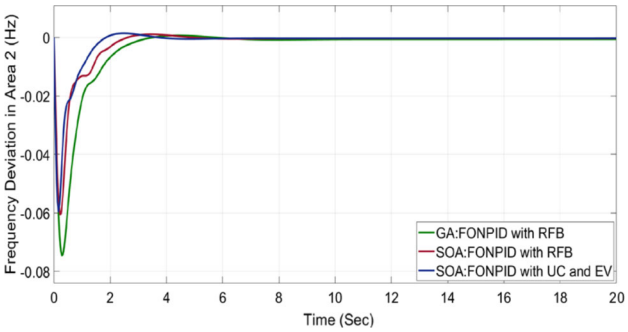


Fig. 28 ΔF_2 responses for PT

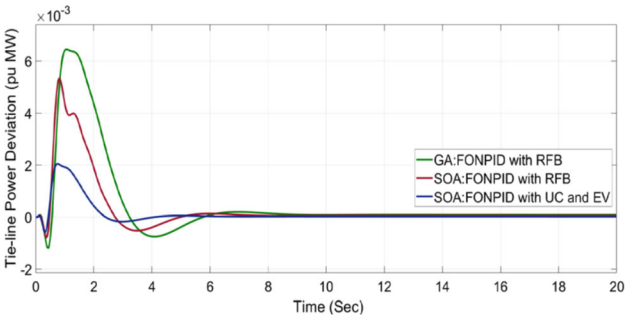


Fig. 29 ΔP_{tie} responses for PT

In order to extend this study, a novel metaheuristic optimization tool that draws inspiration from human skill development and enhancement, i.e., SOA is used [38]. The SOA-FONPID controller with UC and EV is designed and implemented for the system as shown in Fig. (1). The performance of this controller is compared with SOA-FONPID controller with RFB, GA-FONPID controller with RFB controllers for various power transactions. As shown in Table 2, the values of O_{sh}, U_{Sh}, and T_S for the ΔF_2, and ΔP_{tie} responses are obtained for various controllers, considering many transactions. The numbers that are given in the very last row of Table 2 indicate that the values of O_{sh}, U_{sh}, and T_s for SOA-FONPID with UC and EV are the lowest when compared to SOA-FONPID with RFB, GA-FONPID with RFB, GA-FONPID, GA-NPID, and GA-PID controllers. This is the case since these values are the minimum values. The results reveal that the SOA-FONPID controller with UC and EV produces the least amount of U_S ($\Delta F_1 = -0.0609$ Hz, $\Delta F_2 = -0.0577$ Hz, $\Delta P_{tie} = -0.0024$ pu MW) and least J_{ITAE} (0.3234), and quickly reach zero in minimal T_S ($\Delta F_1 = 3.468$ s, $\Delta F_2 = 4.094$ s, $\Delta P_{tie} = 3.873$ s) as compared to other controllers for the burst case, i.e., contract violation. As a result, the SOA-FONPID controller with UC and EV produces superior results. ΔF_1, ΔF_2, and ΔP_{tie} responses for various power transactions are shown in Figs. 27, 28, 29, 30, 31, 32, 33, 34, and 35. Figure 47A–C compares the

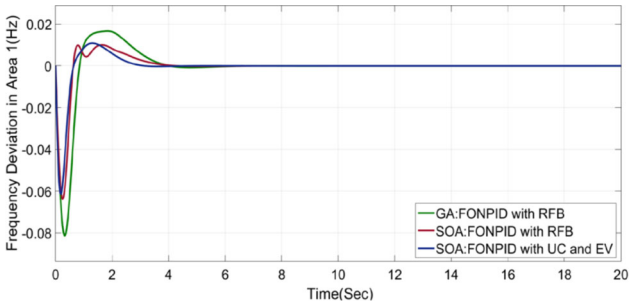


Fig. 30 ΔF_1 responses for BT

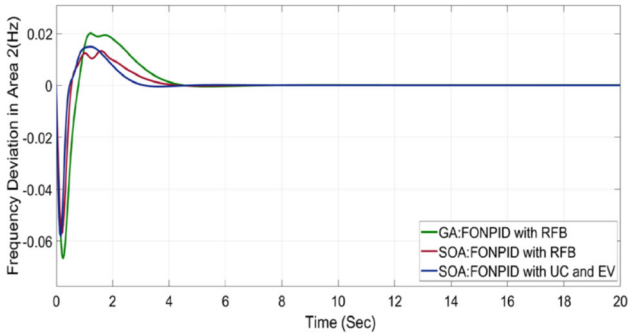


Fig. 31 ΔF_2 responses for BT

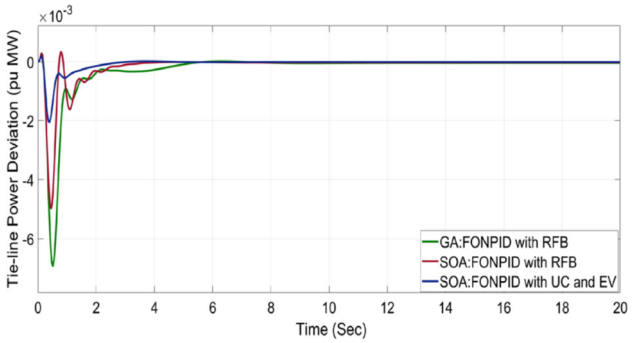


Fig. 32 ΔP_{tie} responses for BT

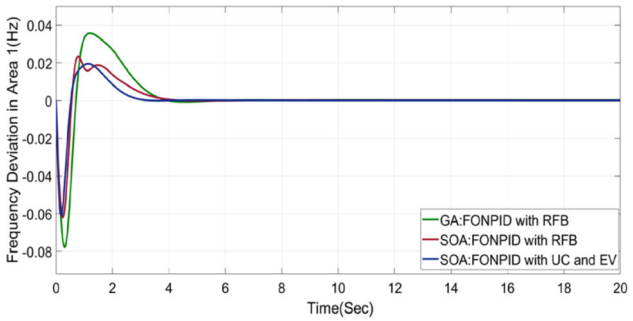


Fig. 33 ΔF_1 responses for CV

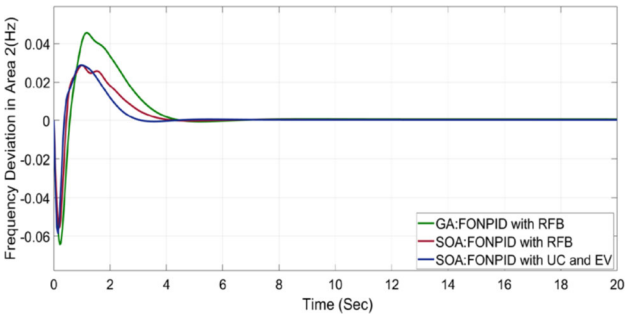


Fig. 34 ΔF_2 responses for CV

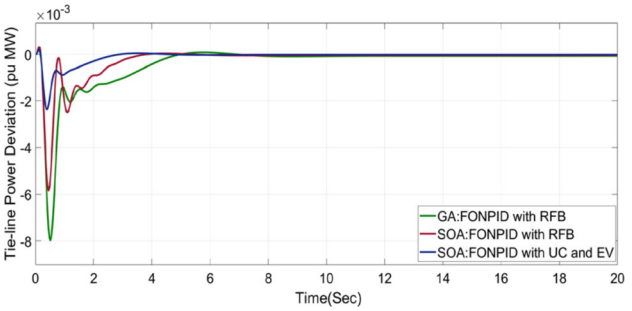


Fig. 35 ΔP_{tie} responses for CV

settling times of ΔF_1, ΔF_2, and ΔP_{tie} for different types of power transactions. The improvement in settling time for ΔF_1, ΔF_2, and ΔP_{tie} is 37.01%, 32.89%, and 44.71%, respectively, by SOA:FONPID with UC and EV as compared to GA:FONPID with RFB.

5.2 Test II: an examination of the sensitivity for the proposed controllers

Since the GDB denotes a speed shift that is continuous up to the point when valve position has no change. The GDB is responsible for oscillations in the outcomes produced by the system. In this scenario, the suggested control technique's robustness and exceptional performance are demonstrated by using a random load disturbance (RLD) in every portion of the provided RPS. Figures 18, 19, 20, 21, 22, 23, 24 25, and 26 provide a comparative study of the dynamic responses for the proposed system. The analysis includes the use of GA-FONPID with RFB, GA-FONPID, GA-NPID, and GA-PID controllers. The analysis is conducted using various power transactions. The responses indicate that the GA-based FONPID controller with RFB yields superior and more dependable outcomes. In contrast with GA-PID, GA-NPID, GA-FONPID, and GA-FONPID with RFB controllers, the findings demonstrate that a GA-FONPID controller with RFB rapidly establishes its optimal value. Moreover, sensitivity analysis is carried in order to test robustness of the SOA-FONPID controller with UC and EV, it is discovered

Table 2 Measures of performance for the thermal-thermal reorganized system

Controller	Parameter	Overshoot (O_{sh}) PT	Undershoot (U_{sh})	Settling time (T_s)	Overshoot (O_{sh}) BT	Undershoot (U_{sh})	Settling time (T_s)	Overshoot (O_{sh}) CV	Undershoot (U_{sh})	Settling time (T_s)
GA:PID	ΔF_1	0.0678	− 0.0995	18.601	0.1068	− 0.1042	19.95	0.1374	− 0.0993	20.942
	ΔF_2	0.1443	− 0.1437	17.907	0.1336	− 0.1059	19.315	0.1703	− 0.1087	20.069
	ΔP_{tie}	0.0206	− 0.0177	18.369	0.01699	− 0.0186	18.947	0.0196	− 0.0241	19.684
GA:NPID	ΔF_1	0.0684	− 0.1314	8.958	0.1248	− 0.1403	9.692	0.1756	− 0.126	10.565
	ΔF_2	0.0447	− 0.1377	10.03	0.0764	− 0.0992	10.962	0.1106	− 0.0893	11.359
	ΔP_{tie}	0.0394	− 0.0036	12.807	0.0079	− 0.02045	14.841	0.00457	− 0.0296	14.94
GA:FONPID	ΔF_1	0.0395	− 0.0976	7.173	0.06089	− 0.1017	5.347	0.0914	− 0.0958	6.756
	ΔF_2	0.0255	− 0.0898	8.145	0.048	− 0.07879	5.883	0.0725	− 0.0754	6.617
	ΔP_{tie}	0.0161	− 0.0026	10.495	0.00234	− 0.0123	7.101	0.001579	− 0.01395	7.377
GA:FONPID with RFB	ΔF_1	0.0152	− 0.0788	5.863	0.01667	− 0.08137	4.077	0.0357	− 0.0778	5.506
	ΔF_2	0.0006	− 0.0746	5.724	0.02012	− 0.0667	4.692	0.0456	− 0.064	6.101
	ΔP_{tie}	0.00645	− 0.00119	8.514	0.00029	− 0.0069	6.221	0.000329	− 0.00798	7.005
SOA:FONPID with RFB	ΔF_1	0.0107	− 0.0621	5.218	0.0098	− 0.0636	3.948	0.0234	− 0.0616	4.31
	ΔF_2	0.001	− 0.0602	5.02	0.0132	− 0.0568	4.107	0.0286	− 0.0586	4.358
	ΔP_{tie}	0.0053	− 0.0007	7.714	0.0003	− 0.0049	4.187	0.00029	− 0.0057	4.984
SOA:FONPID with UC and EV	ΔF_1	0.0137	− 0.0607	4.808	0.0108	− 0.0613	3.313	0.0194	− 0.0609	3.468
	ΔF_2	0.0014	− 0.0592	4.173	0.0149	− 0.0575	3.036	0.0285	− 0.0577	4.094
	ΔP_{tie}	0.00205	− 0.00058	5.872	0.00017	− 0.002	3.521	0.0003	− 0.0024	3.873

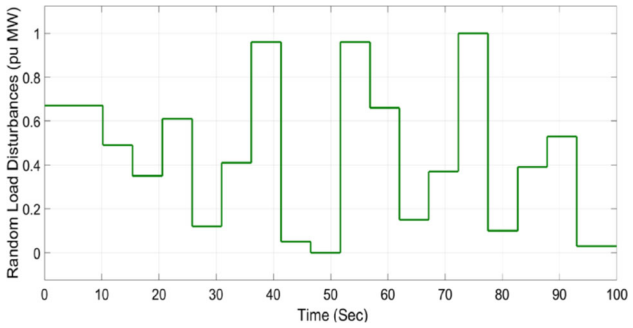


Fig. 36 Load pattern for PT

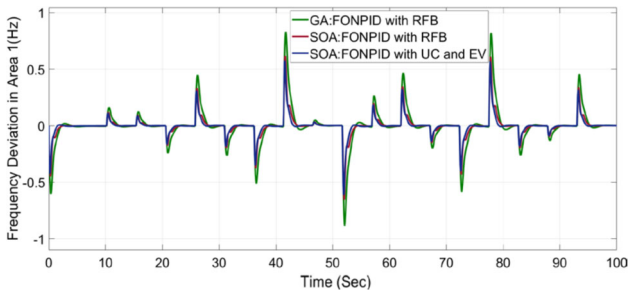


Fig. 37 ΔF_1 responses for PT with RLD

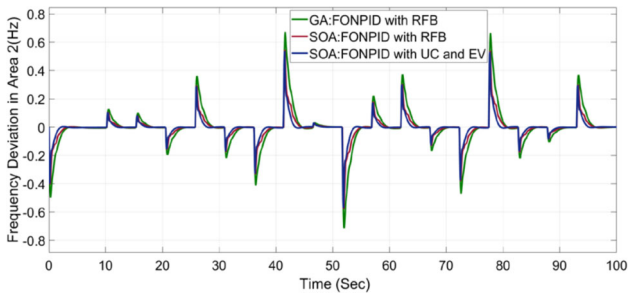


Fig. 38 ΔF_2 responses for PT with RLD

that the performance of SOA-FONPID with UC and EV is superior for the proposed system as compared to other controllers for various power transactions such as PT, BT, and CV. ΔF_1, ΔF_2, and ΔP_{tie} responses for various power transactions are shown in Figs. 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47. If a dependable controller is not present, the system will inevitably experience degradation.

6 Thermal-gas reorganized system

There are two GENCOs and two DISCOs assigned to each area in this system. In the framework that is now being investigated, Region 1 is comprised of two-generation organizations, notably GENCO-1 and -2, as well as two distribution organizations, especially DISCO-1 and DISCO-2.

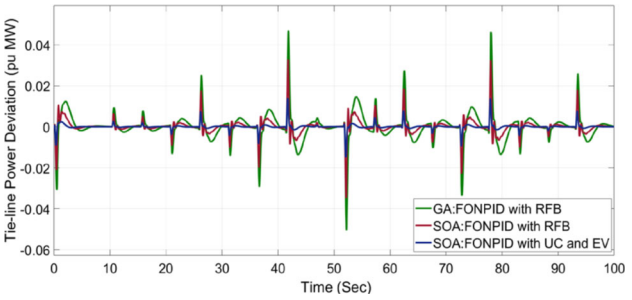


Fig. 39 ΔP_{tie} responses for PT with RLD

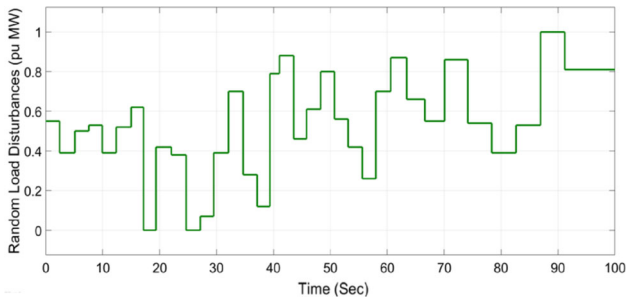


Fig. 40 Load pattern for BT

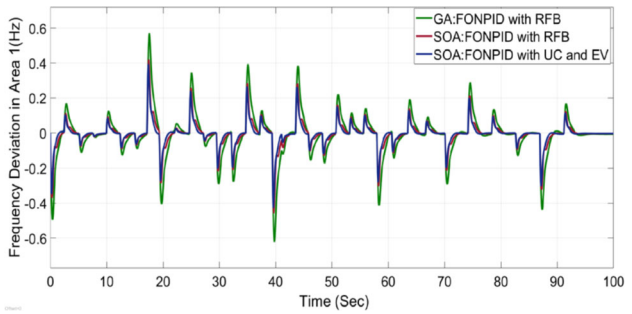


Fig. 41 ΔF_1 responses for BT with RLD

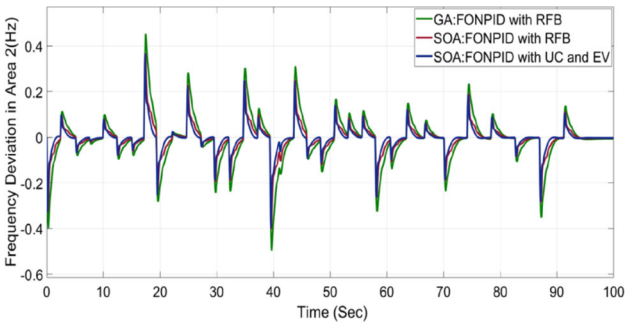


Fig. 42 ΔF_2 responses for BT with RLD

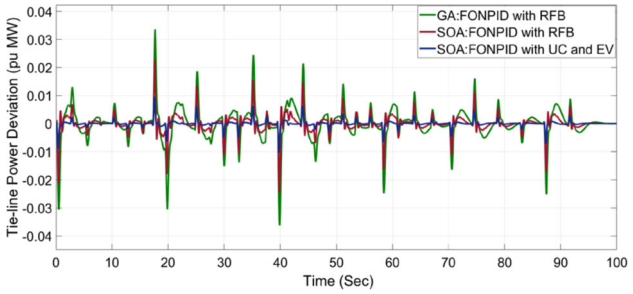


Fig. 43 ΔP_{tie} responses for BT with RLD

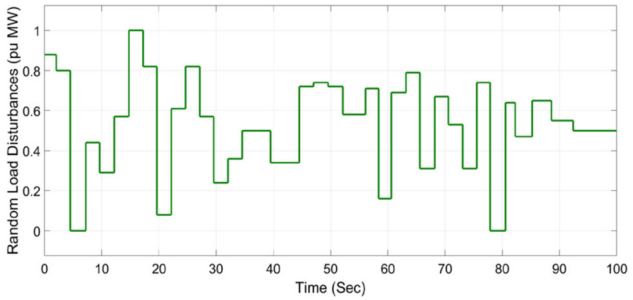


Fig. 44 Load pattern for CV

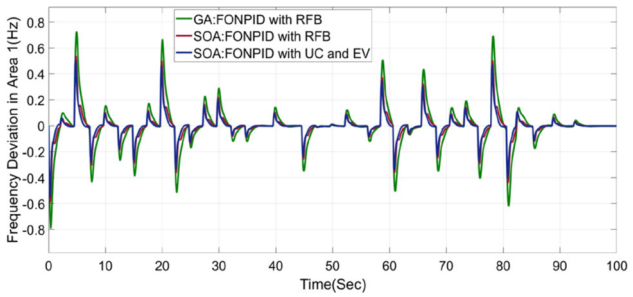


Fig. 45 ΔF_1 responses for CV with RLD

In GENCO-1, there is a reheat thermal unit, and in GENCO-2, there is a gas unit. In addition, area 2 is comprised of two generating units, namely, GENCO-3 (thermal unit) and GENCO-4 (gas unit), as well as two distribution companies, namely, DISCO-3 and DISCO-4. AC/DC transmission links that are parallel to one another connect these two regions. The thermal generation has reheat turbines with 3% per-minute generating constraints. The $\text{apf}_{11} = \text{apf}_{21} = 0.75$; $\text{apf}_{12} = \text{apf}_{22} = 0.25$. The transfer function (TF) of the power system is given below.

$$G_{ps}(s) = \frac{K_{ps}}{1 + sT_{ps}}$$

The TF of the thermal speed governor

$$G_{Tg}(s) = \frac{1}{1 + sT_g}$$

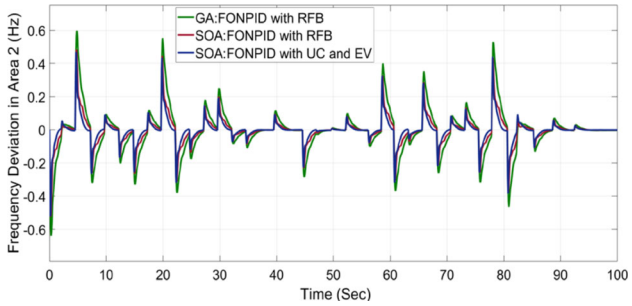


Fig. 46 ΔF_2 responses for CV with RLD

The TF of steam turbine

$$G_{Tt}(s) = \frac{1}{1 + sT_t}$$

The TF of the reheater

$$G_{Tr}(s) = \frac{1 + sK_rT_r}{1 + sT_r}$$

The TF of the gas turbine speed governor

$$G_{Gg}(s) = \frac{1 + \text{sX}}{1 + \text{sY}}$$

The TF of gas governor positioner

$$G_{Gp}(s) = \frac{\text{a}}{\text{c} + \text{sb}}$$

The TF of fuel system and combustor

$$G_{Gf}(s) = \frac{1 - \text{sT}_{\text{CR}}}{1 + \text{sT}_{\text{F}}}$$

The TF of gas turbine

$$G_{Gt}(s) = \frac{1}{1 + \text{sT}_{\text{CD}}}$$

This part analyzes the performance of a two-area thermal-gas restructured power system using genetic algorithm-based polar fuzzy controllers, considering diverse input membership functions for different power transactions. The GA:PFC (Dsig with RFB) controller signals are shown in Fig. 8D for control areas 1 and 2, considering contract violation (CV).

6.1 Poolco transaction (PT)

DISCO1 and DISCO2 necessitate load only. DISCO3 and DISCO4 do not require load from GENCOs; hence, DPM matrix cpfs are 0. The DPM matrix governs DISCO-GENCO contracts, as shown in Eq. (26). An algorithm, namely, GA,

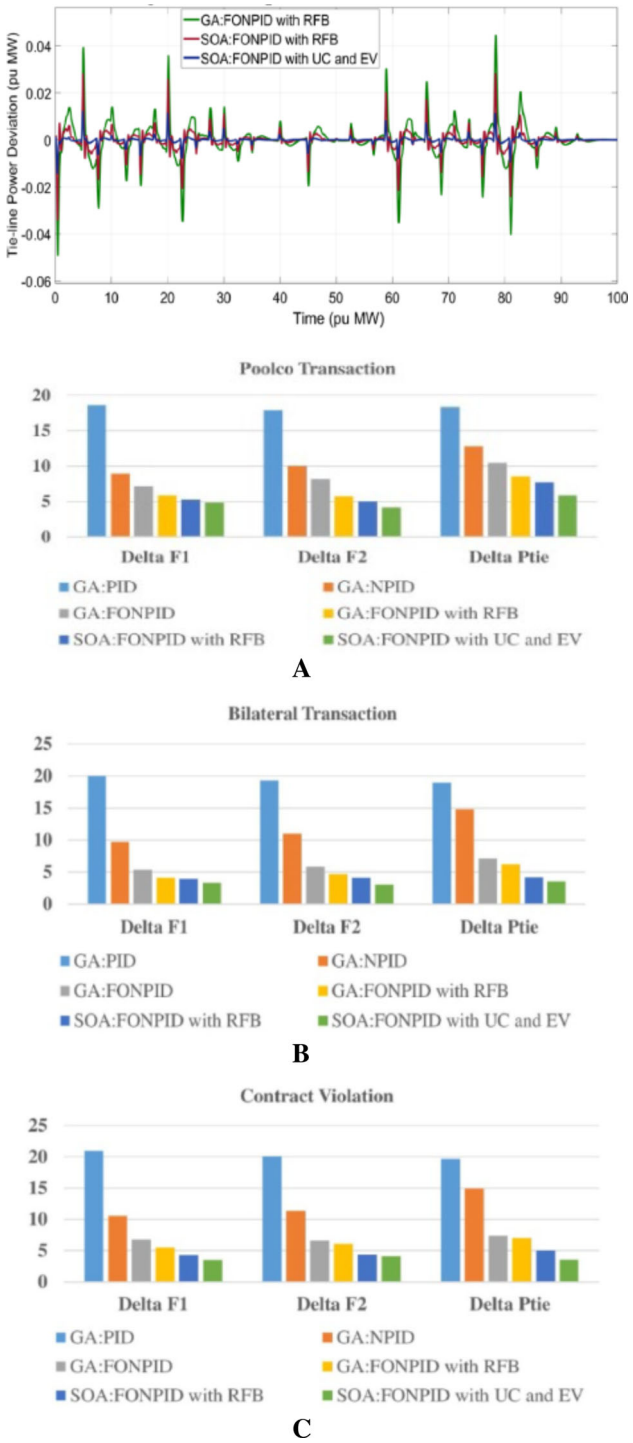


Fig. 47 ΔP_{tie} responses for CV with RLD. **A** Settling time comparision of $\Delta F_{1,}$ $\Delta F_{2,\text{ and}}$ ΔP_{tie}. **B** Settling time comparision of $\Delta F_{1,}$ $\Delta F_{2,\text{ and}}$ ΔP_{tie}. **C** Settling time comparision of $\Delta F_{1,}$ $\Delta P_{2,\text{ and}}$ ΔP_{tie}

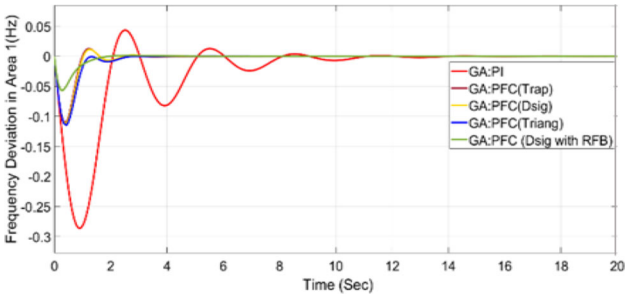


Fig. 48 ΔF_1 responses for PT

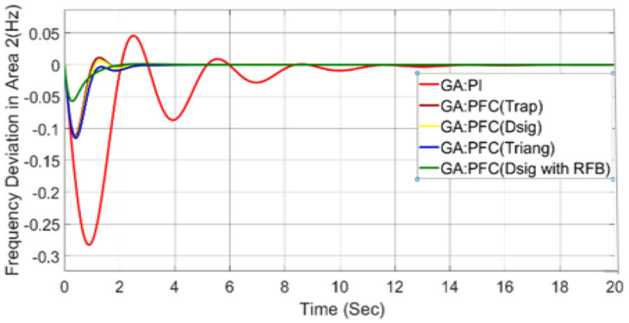


Fig. 49 ΔF_2 responses for PT

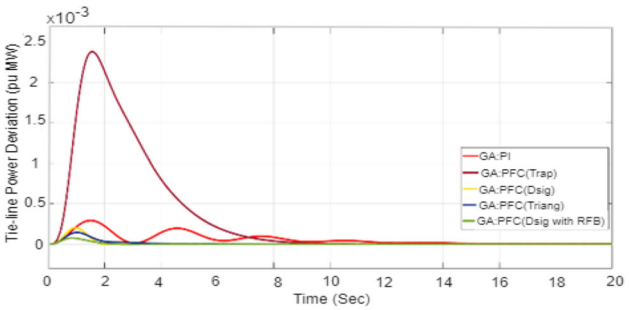


Fig. 50 ΔP_{tie} responses for PT

is used to optimize PI, PFC controllers (considering various membership functions) gain parameters for simulation. The improvement in settling time for $\Delta F_{1,}\Delta F_2$, and ΔP_{tie} is found by GA-PFC (Dsig.) controller with RFB as compared to GA-PI, GA-PFC (trap.), GA-PFC (Dsig.), and GA-PFC (Triang.) without RFB. Figures 48, 49, and 50 depict the suggested system responses. The improvement in settling time for $\Delta F_{1,}$ ΔF_2, and ΔP_{tie} is 27.29%, 28.54%, and 54.28%, respectively, by GA:PFC (Dsig with RFB) controller as compared to GA:PFC (Trap).

6.2 Bilateral transaction (BT)

GENCOs may have power agreements with any DISCOs in this transaction. The DPM matrix governs DISCO-GENCO contracts, as shown in Eq. (27). An algorithm, namely, GA,

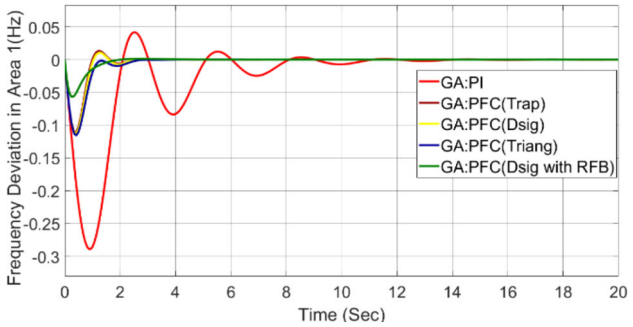


Fig. 51 ΔF_1 responses for BT

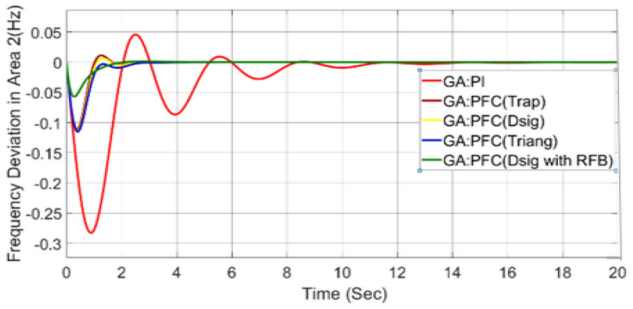


Fig. 52 ΔF_2 responses for BT

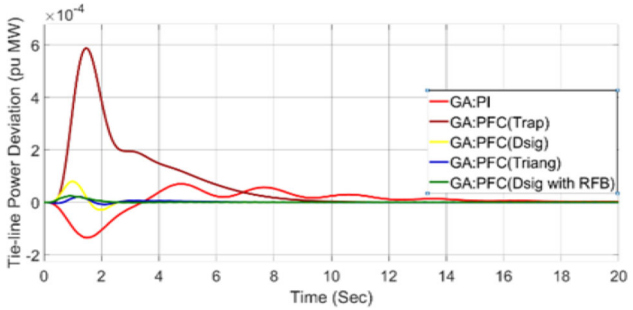


Fig. 53 ΔP_{tie} responses for BT

is used to optimize PI, PFC controllers (considering various membership functions) gain parameters for simulation. The improvement in settling time for $\Delta F_1, \Delta F_2$, and ΔP_{tie} is found by GA-PFC (Dsig.) controller with RFB as compared to GA-PI, GA-PFC (Trap), GA-PFC (Dsig.), and GA-PFC (Triang.) without RFB. Figures 51, 52, and 53 depict the suggested system responses. The improvement in settling time for ΔF_1, ΔF_2, and ΔP_{tie} is 10.18%, 4.25%, and 57.03%, respectively, by GA:PFC (Dsig with RFB) controller as compared to GA:PFC (Trap).

6.3 Contract violation (CV)

DISCO may need more electricity than the contract. The DISCOs' local GENCOs must provide extra electricity. DISCO1

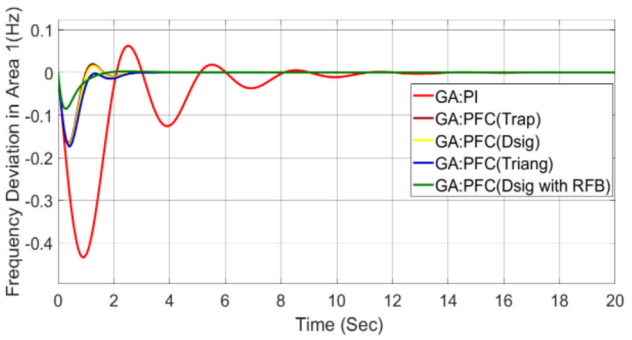


Fig. 54 ΔF_1 responses for CV

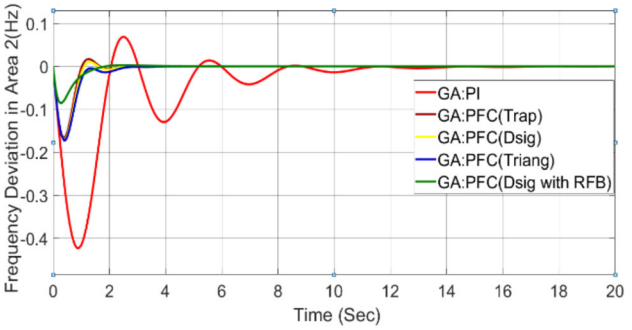


Fig. 55 ΔF_2 responses for CV

requires 0.005 p.u. MW of surplus electricity, which is the local load. Area-1's local load is Delta $P_{L1,loc}$ = load of DISCO1 (0.005) + load of DISCO2 (0.005) + 0.005 = 0.015 p.u. MW. Moreover, DISCO3 requires 0.005 p.u. MW of surplus electricity, which is the local load. Area 2's local load is Delta $P_{L2,loc}$ = load of DISCO3 (0.005) + load of DISCO4 (0.005) + 0.005 = 0.015 p.u. MW. The bilateral case's DPM matrix persists here. DISCO1's un-contracted load tends to generate GENCO1 and GENCO2. At the same time, DISCO3's un-contracted load tends to generate GENCO3 and GENCO4. The improvement in settling time for $\Delta F_1, \Delta F_2$, and ΔP_{tie} is found by GA-PFC (Dsig.) controller with RFB as compared to GA-PI, GA-PFC (Trap), GA-PFC (Dsig.), and GA-PFC (Triang.) without RFB. Figures 54, 55, and 56 depict the suggested system responses. The improvement in settling time for ΔF_1, ΔF_2, and ΔP_{tie} is 10.64%, 8.12%, and 45.06%, respectively, by GA:PFC (Dsig with RFB) controller as compared to GA:PFC (Trap).

As shown in Table 3, the values of O_{sh}, U_S, and T_S for the ΔF_1, ΔF_2, and ΔP_{tie} responses are obtained for various controllers, considering many transactions. The numbers that are given in the very last row of Table 3 indicate that the values of O_{sh}, U_{Sh}, and T_S for GA-PFC (Dsig.) with RFB are the lowest when compared to GA-PFC (Triang.), GA-PFC (Trap), and GA-PI controllers that do not have RFB. This is the case since these values are the minimum values. The results reveal that the GA-PFC (Dsig.) controller with RFB

Table 3 Measures of performance for the thermal-gas reorganized system

Sl. No.	Para	Controller	Overshoot (O_{sh})	Undershoot (U_{sh})	Settling time (T_s)	Overshoot (O_{sh})	Undershoot (U_{sh})	Settling time (T_s)	Overshoot (O_{sh})	Undershoot (U_{sh})	Settling time (T_s)
			Poolco transaction			Bilateral transaction			Contract violation		
1	ΔF_1	GA PI	0.0435	− 0.2864	13.962	0.0418	− 0.2894	14.04	0.0611	− 0.4341	14.819
2		GA PFC (Trap)	0.1264	− 0.1094	6.144	0.01307	− 0.1104	4.782	0.0196	− 0.1674	4.955
3		GA PFC (Dsig)	0.0112	− 0.1117	4.705	0.01036	− 0.1129	4.395	0.0154	− 0.1695	4.795
4		GA PFC (Triang)	0.0002	− 0.1150	4.683	0	− 0.1147	4.573	0	− 0.1737	4.861
5		GA PFC (Dsig with RFB)	0.0012	− 0.0565	4.467	0.00116	− 0.0566	4.295	0.0025	− 0.0853	4.428
1	ΔF_2	GA PI	0.0379	− 0.3020	14.689	0.0457	− 0.2823	14.94	0.0680	− 0.04230	17.691
2		GA PFC (Trap)	0.0005	− 0.1134	6.173	0.01127	− 0.1112	4.398	0.0171	− 0.1678	4.874
3		GA PFC (Dsig)	0.0004	− 0.1167	4.644	0.0068	− 0.1140	4.226	0.0102	− 0.1701	4.517
4		GA PFC (Triang)	0.0005	− 0.1184	4.645	0	− 0.1154	4.279	0	− 0.1726	4.729
5		GA PFC (Dsig with RFB)	0.00005	− 0.05793	4.411	0.00123	− 0.0569	4.211	0.0027	− 0.0863	4.478
1	ΔP_{tie}	GA PI	0.00029	0	18.221	0.00007	− 0.00013	18.17	0.0001	− 0.00021	18.678
2		GA PFC (Trap)	0.00238	0	9.587	0.00058	0	10.54	0.0008	− 0.000003	10.814
3		GA PFC (Dsig)	0.00020	− 0.000002	6.052	0.000079	− 0.000028	5.976	0.0001	− 0.00004	6.007
4		GA PFC (Triang)	0.00014	0	6.957	0.000021	− 0.000003	7.033	0.00003	− 0.00001	7.375
5		GA PFC (Dsig with RFB)	0.00007	− 0.0000007	4.383	0.00002	− 0.00000008	4.529	0.00003	− 0.000003	5.941

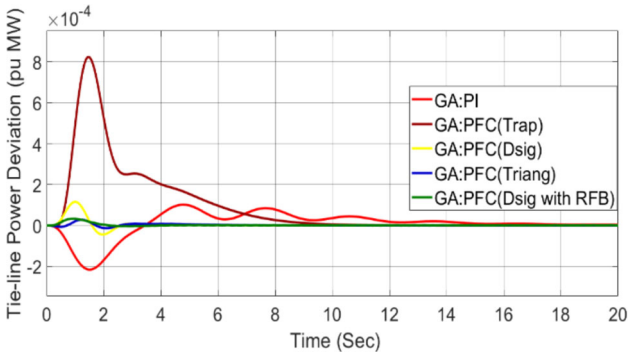


Fig. 56 ΔP_{tie} responses for CV

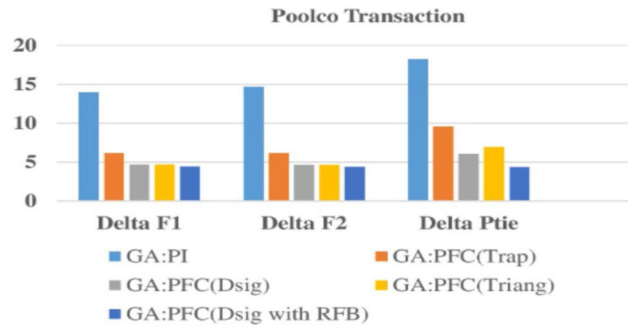


Fig. 57 Settling time comparision of ΔF_1, $\Delta F_{2,\text{ and}}$ ΔP_{tie}

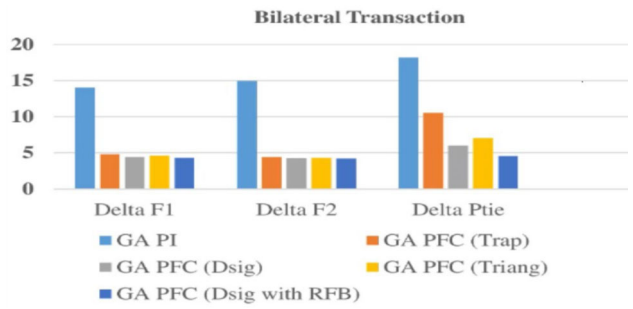


Fig. 58 Settling time comparision of ΔF_1, $\Delta F_{2,\text{ and}}$ ΔP_{tie}

produces the least amount of U_S ($\Delta F_1 = -0.0565$ Hz, $\Delta F_2 = -0.0566$ Hz, $\Delta P_{\text{tie}} = -0.0853$ p.u. MW), and quickly reach zero in minimal T_S ($\Delta F_1 = 4.428$ s, $\Delta F_2 = 4.478$ s, $\Delta P_{\text{tie}} = 5.941$ s) as compared to other controllers without RFB in contract violation case as highlighted in Table 3. Figures 57, 58, and 59 compare the settling times of ΔF_1, ΔF_2, and ΔP_{tie} for different types of power transactions. As a result, the GA-PFC (Dsig.) controller with RFB produces results that are vastly superior to GA-PFC (Dsig.) without RFB.

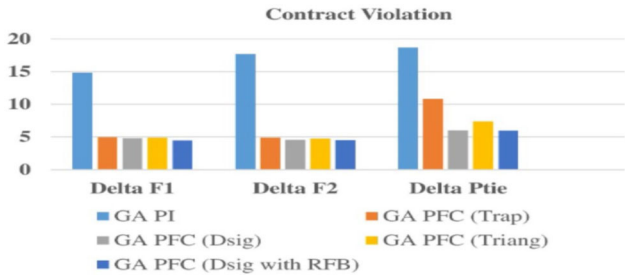


Fig. 59 Settling time comparision of ΔF_1, $\Delta F_{2,\text{ and}}$ ΔP_{tie}

7 Conclusions

The authors present a novel implementation of the SOA-GONPID controller with UC and EVs in a two-area reorganized power system with 10% step load perturbations in both zones. When the system dynamic responses are compared, the SOA-GONPID controller with UC and EVs outperforms the SOA-FONPID with RFB, GA-FONPID with RFB, GA-FONPID, GA-NPID, and GA-PID in all time-domain indices. When comparing SOA:FONPID with UC and EV to GA:FONPID with RFB, the improvements in settling time for ΔF_1, ΔF_2, and ΔP_{tie} in the poolco transaction are 17.99%, 27.09%, and 31.93%, respectively. Whereas the improvements in settling time for ΔF_1, ΔF_2, and ΔP_{tie} are found to be 18.74%, 35.29%, and 43.4%, respectively, by SOA:FONPID with UC and EV as compared to GA:FONPID with RFB under bilateral transaction. In the event of contract violation, the improvements in settling time for ΔF_1, ΔF_2, and ΔP_{tie} are found to be 37.01%, 32.89%, and 44.71%, respectively, by SOA:FONPID with UC and EV as compared to GA:FONPID with RFB. The findings support the superiority of the suggested control (SOA:FONPID with UC and EV) strategy under examination. Moreover, sensitivity analysis is carried out to test the robustness of the proposed controller; it is discovered that the performance of SOA-FONPID with UC and EV is superior for the proposed system as compared to rival controllers. In another study, the frequency management of a two-area system (thermal-gas) with parallel transmission links was controlled using GA-PI and GA-based polar fuzzy controller (PFC), considering various input membership functions. At the same time, both areas are concurrently exposed to 10% step load perturbations. In poolco transaction, the improvements in settling time for ΔF_1, ΔF_2, and ΔP_{tie} are found to be 27.29%, 28.54%, and 54.28%, respectively, by GA:PFC (Dsig with RFB) controller as compared to GA:PFC (Trap). Whereas the improvement in settling time for ΔF_1, ΔF_2, and ΔP_{tie} is found to be 10.18%, 4.25%, and 57.03%, respectively, GA:PFC (Dsig with RFB) controller as compared to GA:PFC (Trap) under bilateral transaction. In the event of contract violation, the improvement in settling time for ΔF_1, ΔF_2, and

ΔP_{tie} is found to be 10.64%, 8.12%, and 45.06%, respectively, by GA:PFC (Dsig with RFB) controller as compared to GA:PFC (Trap). This research work could be extended for frequency management in hybrid reorganized power systems using cutting-edge optimization algorithms.

Appendix

The system under study has the following nominal parameters [6, 9, and 22]:

Parameters	Values
Area participation factor (apf)	0.5 p.u
Rated capacity of each control area (P_{ri})	2000 MW
Nominal frequency (f^0)	60 Hz
Base power (P_{base})	2000 MVA
Frequency bias constants (Bi)	0.425 p. u MW/Hz
Regulation constants ($R_1 = R_3, R_2 = R_4$)	(0.417,0.1227)Hz/p.u
Steam turbine governor time constant (T_G)	0.06 s
Steam turbine time constant (T_t)	0.3 s
Coefficient of reheater steam turbine (K_r)	0.3
Steam turbine reheater time constant (T_r)	10 s
Gain constant of power system (K_{PS1}, K_{PS2})	120 Hz/p.u. MW
The time constant of the power system (T_{P1},T_{P2})	20 s
Synchronizing coefficient of AC link (T_{12})	0.2712
Area size ratio coefficient (a_{12})	− 1
Tie-line power capacity (P_{tie})	200 MW
Gas turbine speed governor lead time constant (X_G)	0.6 s
Gas turbine speed governor lag time constant (Y_G)	1.1 s
Gas turbine constant of valve positioner (b_g,c_g)	0.049 s,1
Gas turbine fuel time constant (T_F)	0.239 s
Gas turbine combustion reaction time delay (T_{CR})	0.01 s
Gas turbine compressor discharge volume time constant (T_{CD})	0.2 s
Gain of redox flow battery (K_{RFB})	0.2787
Time constant of redox flow battery (T_{RFB})	0 s
Gain of ultra-capacitor (K_{UC})	− 0.7
Time constant of ultra-capacitor (T_{UC})	0.9 s

Author contribution Ram Naresh Mishra helped in methodology, software, formal analysis, investigation, resources, and writing—original draft. Narendra Kumar helped in conceptualization, visualization, and supervision.

Funding The authors did not receive support from any organization for the submitted work. No funding was received for conducting this study and preparation of the manuscript.

Availability of data and materials Not applicable. The details about the references have been presented in references section. No datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The authors declare no competing interests.

Ethical approval This research does not involve any human participants, human material, or human data.

Consent for publication This research does not involve any individual person's data in any form (including individual details, images, or videos).

Consent to participate Not applicable.

References

1. Fosha CE, Elgerd OI (1970) The megawatt-frequency control problem: a new approach via optimal control theory. IEEE Trans Power Appar Syst 4:563–577
2. Mishra RN, Chaturvedi DK, Kumar P (2020) Recent philosophies of AGC techniques in deregulated power environment. J Inst Eng India Ser B. https://doi.org/10.1007/s40031-020-00463-8
3. Esmail M, Tzoneva R, Krishnamurthy S (2017) Review of automatic generation control in deregulated environment. IFAC 50(2):88–93
4. Chakraborty MR, Dawn S, Saha PK, Basu JB, Ustun TS (2022) A comparative review on energy storage systems and their application in deregulated systems. Batteries 8:124. https://doi.org/10.3390/batteries8090124
5. Hota PK, Mohanty B (2016) Automatic generation control of multi-source power generation under deregulated environment. Electr Power Energy Syst 75:205–214
6. Arya Y, Kumar N (2016) AGC of a multi-area multi-source hydrothermal power system interconnected via AC/DC parallel links under deregulated environment. Electr Power Energy Syst 75:127–138
7. Jain C, Verma HK, Arya LD (2014) A novel statistically tracked particle swarm optimization method for AGC. J Mod Power Syst Clean Energy 2(4):396–410. https://doi.org/10.1007/s40565-014-0083-x
8. Chidambaram IA, Paramasivam B (2013) Optimized load–frequency simulation in restructured power system with redox flow batteries and interline power flow controller. Int J Electr Power Energy Syst 50:9–24
9. Thirunavukarasu R, Chidambaram IA (2016) Enhanced restoration in a restructured power system with hydrogen energy storage and unified power flow controller. IOSR-JEEE 11(1):54–73

10. Selvaraju RK, Somaskandan G (2016) Impact of energy storage units on load frequency control of deregulated power systems. Energy 97:214–228
11. Malla Reddy CH, Shravan Kumar A (2014) Load frequency control of deregulated power system using different fuzzy controllers. Int J Sci Res (IJSR) 4:731–736
12. Chandra Sekhar GT, Sahu RK, Baliarsingh AK, Panda S (2016) Load frequency control of power system under deregulated environment using optimal firefly algorithm. Electr Power Energy Syst 74:195–211
13. Ajithapriyadarsini S, Melba Mary P, Willjuice Iruthayarajan M (2017) Automatic generation control in deregulated power system using genetic algorithm optimized fuzzy controller. Appl Math Inf Sci 11(2):479–487
14. Fathy A, Kassem AM, Abdelaziz AY (2020) Optimal design of fuzzy PID controller for deregulated LFC of multi-area power system via mine blast algorithm. Neural Comput Appl 32:4531–4551. https://doi.org/10.1007/s00521-018-3720-x
15. Gorripotu TS, Sahu RK, Panda S (2015) AGC of a multi-area power system under deregulated environment using redox flow batteries and interline power flow controller. Eng Sci Technol Int J 18(4):555–578
16. Arya Y, Kumar N (2017) BFOA-scaled fractional order fuzzy PID controller applied to AGC of multi-area multi-source electric power generating systems. Swarm Evol Comput 32:202–218
17. Kumari N, Pulakraj Aryan G, Raja L, Arya Y (2023) Dual degree branched type-2 fuzzy controller optimized with a hybrid algorithm for frequency regulation in a triple-area power system integrating renewable sources. Prot Control Mod Power Syst. https://doi.org/10.1186/s41601-023-00317-7
18. Chaturvedi DK, Umrao R, Malik OP (2015) Adaptive polar fuzzy logic-based load frequency controller. Electr Power Energy Syst 66:154–159
19. Sharma G, Krishnan N, Arya Y, Panwar A (2021) Impact of ultra-capacitor and redox flow battery with JAYA optimization for frequency stabilization in linked photovoltaic-thermal system. Int Trans Electr Energ Syst. https://doi.org/10.1002/2050-7038.12883
20. Dutta A, Debbarma S (2018) Frequency regulation in deregulated market using vehicle-to-grid services in residential distribution network. IEEE Syst J 12(3):2812–2820. https://doi.org/10.1109/JSYST.2017.2743779
21. Kumar R, Prasad LB (2023) Optimal load frequency control of multi-area multi-source deregulated power system with electric vehicles using teaching learning-based optimization: a comparative efficacy analysis. Electr Eng. https://doi.org/10.1007/s00202-023-02027-7
22. Dutta A, Prakash S (2019) Utilizing electric vehicles and renewable energy sources for load frequency control in deregulated power system using emotional controller. IETE J Res. https://doi.org/10.1080/03772063.2019.1654936
23. Khan M, Sun H, Xiang Y, Shi Di (2021) Electric vehicles participation in load frequency control based on mixed H2/H∞. Electr Power Energy Syst. https://doi.org/10.1016/j.ijepes.2020.106420
24. Dekaraja B, Saikia LC, Ramoji SK, Behera MK, Bhagat SK (2022) Performance analysis of diverse energy storage on combined ALFC and AVR Control of multi-area multiunit system with AC/HVDC interconnection. IFAC-Papers On Line 55(1):479–485. https://doi.org/10.1016/j.ifacol.2022.04.079
25. Ramoji SK, Saikia LC (2021) Utilization of electric vehicles in combined voltage frequency control of multi-area thermal-combined cycle gas turbine system using two degree of freedom tilt-integral derivative controller. Energy Storage 3(4):e234. https://doi.org/10.1002/est2.234
26. Ramoji SK, Saikia LC (2021) Maiden application of fuzzy-2DOFTID controller in unified voltage-frequency control of power system. IETE J Res. https://doi.org/10.1080/03772063.2021
27. Ramoji SK, Saikia LC (2022) Comparative performance of multiple energy storage systems in unified voltage and frequency regulation of power system including electric vehicles. Energy Storage. https://doi.org/10.1002/est2.360
28. Shukla H, Raju M (2024) Combined frequency and voltage regulation in an interconnected power system using fractional order cascade controller considering renewable energy sources, electric vehicles and ultra-capacitor. J Energy Storage. https://doi.org/10.1016/j.est.2024.110875(2024)
29. Iqbal M, Gulzar MM (2022) Master-slave design for frequency regulation in hybrid power system under complex environment. IET Renew Power Gener 16(14):3041–3057. https://doi.org/10.1049/rpg2.12553
30. Jameel A, Gulzar MM (2023) Load frequency regulation of interconnected muli-source multi-area power system with penetration of electric vehicles aggregator. Res Square. https://doi.org/10.21203/rs.3.rs-2961150/v1
31. Gulzar MM, Irfan R, Shahid K, Raza MT, Javed I (2023) Frequency and voltage stabilization of multi-area hybrid power system by considering the impact of communication time delay. Electr Power Compon Syst 1–15. https://doi.org/10.1080/15325008.2023.2251489
32. Gulzar MM, Sibtain D, Khalid M (2023) Cascaded fractional model predictive controller for load frequency control in multi-area hybrid renewable energy system with uncertainties. Int J Energy Res 2023:1–25
33. Guerrero J, Torres J, Creuze V, Chemori A, Campos E (2019) Saturation based nonlinear PID control for underwater vehicles: design, stability analysis and experiments. Mechatronics 61:96–105. https://doi.org/10.1016/j.mechatronics.2019.06.006
34. Petras I (2024) Non-linear fractional-order PID controller. (https://www.mathworks.com/matlabcentral/file exchange/51190-non-linear-fractional-order-pid-controller), MATLAB Central File Exchange. Retrieved 5 Aug 2024
35. Gupta D, Goyal V, Kumar J (2022) Design of fractional-order NPID controller for the NPK model of advanced nuclear reactor. Prog Nucl Energy. https://doi.org/10.1016/j.pnucene.2022.104319
36. Farshad Merrikh-Bayat (2012) Rules for selecting the parameters of Oustaloup recursive approximation for the simulation of linear feedback systems containing $PI^{\lambda}D^{\mu}$ controller. Commun Nonlinear Sci Numer Simu 17(4):1852–1861. https://doi.org/10.1016/j.cnsns.2011.08.042
37. Sivandam SN, Deepa SN (2008) Genetic algorithm optimization problems. Springer, pp 165–209
38. Givi H, Hubalovska M (2023) Skill optimization algorithm: a new human-based metaheuristic technique. Comput Mater Contin 74(1):179–202

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

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

Wireless Personal Communications
https://doi.org/10.1007/s11277-024-11669-0

An Optimized Multi-kernel Based Extreme Learning Machine for Authentication Threat Detection with Feature Reduction Scheme in IoT

Prachi Dahiya[1] · **Vinod Kumar**[1]

Accepted: 1 December 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
Internet of Things (IoT) refers to a network of interconnected objects, computer systems, and physical or mechanical devices that are capable of sending and receiving data over a network without the intervention of a human. A growing number of undesirable security issues are occurring as a result of the rapid advancements in IoT infrastructures. IoT devices are therefore becoming more vulnerable to security threats and network attacks. IoT services and innovative ecosystem devices can be severely damaged as a result of these attacks and issues. So we propose a multi-kernel-based intensive learning machine model to improve security and overcome the above challenges. To enhance the performance of the insider threat detection model, significant feature sets are generated using correlation coefficient, random forest mean (RFM) reduction accuracy, and gain ratio. To obtain an optimal feature set, the features are combined using an appropriate mechanism (AND function). Afterward, the combined feature set is fed into a machine-learning model based on multi-kernel optimization (MKELM). MKELM is optimized using the Adaptive Crocodile Optimization Algorithm (ACOA) to increase its performance in this detection model. Java is used for the implementation of this proposed approach. In order to analyze performance, NSL-KDD 99 dataset and UNSW-NB15 dataset have been used. The performance of the proposed scheme is evaluated in terms of detection rate, precision, accuracy, recall, F-measure, and detection rate.

Keywords IoT · Network security · Threat detection · MKELM · Kernel · ACOA

1 Introduction

Internet of Things (IoT) applications are increasingly important in smart city applications, including smart homes, smart buildings, smart health, and smart transportation [1]. Urban cities are benefiting greatly from IoT platforms that are intelligently enhancing the quality of life of the residents. There will be a rapid increase in the number of Internet of Things

✉ Prachi Dahiya
prachidahiya15@gmail.com

[1] Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India

devices worldwide from 20.35 billion in 2017 to 75.44 billion in 2025 [2]. Because of the large amount of data generated by cloud computing, intrusion detection systems (IDS) based on this technology are more complex and challenging [3]. The vulnerability of centralized data storage makes cloud computing less suitable for the support of IoT needs [4] (e.g., low latency, resource limitations, distribution, mobility). Traditional cloud-based intrusion detection has the disadvantage of having limited functionality and resources, for example, smart lights and smartwatches are not supported [5].

With the sophistication of attacks increasing every day, IoT security challenges are increasing [6]. One of the most widespread applications of IDS is detecting intrusions into IT infrastructure. Network IDS guarantees network security by detecting attacks and monitoring intrusions [7]. Zero-day attacks (unknown) are efficiently detected by anomaly-based IDS approaches. It is defined as a series of unexpected activities within a network that affect its confidentiality, integrity and availability (CIA triad) [8]. Packet header fields contain information about network traffic. Contradictions are detected by analyzing the features of those events. A network intrusion detection system (NIDS) is designed to detect and/or prevent abnormal behavior (e.g. unauthorized access), and passive and active network intrusion activities, thereby improving information security [9].

There have been numerous ML and DL-based solutions proposed by researchers to improve the effectiveness of NIDS in detecting malicious attacks over the last decade. However, the massive increase in network traffic and the resulting security threats have presented many challenges for NIDS systems to efficiently detect malicious intrusions [10]. So to improve security, the following contributions are presented in this paper.

- In order to improve the performance of authentication threat detection model, significant features are generated using correlation coefficient, random forest average reduction accuracy, and gain ratio.
- An optimal feature set is then obtained by combining the features using an AND function.
- After that, the combined feature set is fed as input into a multi-kernel intensive learning machine detection model that uses multi-kernel learning.
- The performance of multi-kernel based intensive learning machine is improved by adaptive crocodile optimization algorithm.
- The performance of the proposed scheme is evaluated in terms of detection rate, precision, accuracy, recall, F-Measure and G-Mean.

2 Literature Survey

Jayaselvi et al. [11] presented secure IDS targeting the IoT environment that maintained data confidentiality and integrity. As a first step, the data is pre-processed. Using Pearson Correlation Coefficients (IPCC) and Feature Extraction (FE) techniques, this work generated data knowledge by optimizing Pearson Correlation Coefficients (IPCC). Particle Swarm Optimization (PSO) Centered Sea Turtle Foraging Algorithm (XPSO-STFA) extracted and analyzed features while maintaining computation efficiency. An artificial neural network training and analysis was conducted to detect selected feature attacks by training and analyzing a Look Ahead Artificial Neural Network (LAANN). LAANN also offers low error rates. ADR and empirical work accuracy reached 98.16 and 95.65%, respectively.

Zhang et al. [12] proposed a hierarchical intrusion detection model for WSNs that would reduce their energy consumption. As part of this research, multiple kernel functions were combined to improve the accuracy of detecting anomalous behavior in the WSN intrusion detection system and reduce false alarms by using a kernel-intensive learning machine classification algorithm based upon the Mercer property. By testing and utilizing the multi-kernel function, the optimal linear combination was achieved for WSN intrusion detection systems by developing a multi-kernel intensive learning machine. A simulation study demonstrated that the system was not only capable of providing high detection accuracy, but also significantly reduced detection times.

Zheng et al. [13] developed an extreme learning machine (ELM) classifier for intrusion detection algorithms (ILECA). The first step in this process was to develop and employ a linear discriminant analysis (LDA) in order to reduce the number of features in a feature set. A single hidden layer neural network was employed as part of the extreme learning machine (ELM) algorithm used by the authors. Due to the high demand for detection performance from IoT devices, this project ensured the accuracy of intrusion detection as well as improved the processing capability to detect intrusions more quickly. A dataset from NSL-KDD was used for the experiment. The evaluation results indicate that this ILECA exhibits good real-time characteristics and good generalization characteristics, with a detection accuracy of 92.35%.

Li et al. [14] proposed a two-stage intrusion detection system based on artificial intelligence and enhanced by software-defined technology. Using AI algorithms, it intelligently detects attacks by capturing network flows with a global view. Initially, common features were selected using swarm segmentation and differential mutation using the BAT algorithm. A weighted voting algorithm was used to modify the weights of the samples and classify the flows using random forests. In terms of flow classification, it was found that modified intelligent algorithms perform better by selecting the most important features for classification. A comparative analysis of intelligent intrusion detection and existing solutions showed that it is more accurate with less overhead than existing solutions.

Li et al. [15] developed an algorithm combining deep learning with intrusion detection technologies for an intelligent city based on deep migration learning. It was proposed in this paper to model migration learning models and extract data features based on the existing literature and methodology. Test data were selected from KDD CUP 99, and training data were selected from 10% of the test data. Furthermore, a comparison was made between this algorithm and existing algorithms. Using this algorithm, it was found that it was both faster and more efficient at detecting targets.

Lin et al. [16] implemented a multi-feature extreme learning machine (MFE-ELM) algorithm for cloud computing, adding a multi-feature extraction process to cloud servers and discovering the network by applying the MFE-ELM algorithm to cloud nodes. In the simulation experiments, the classical dataset for intrusion detection was pre-processed, features were extracted, a model was trained, and the results were analysed. It has been demonstrated that this algorithm is capable of detecting and identifying most network data packets with good sampling performance, as well as detecting IoT intrusions using heterogeneous data from cloud nodes.

Ilyas et al. [17] proposed using artificial neural networks to detect threats in the IoT for authentication purposes. After detecting the attacks, controller commands were rejected based on supervised learning algorithms. A three-layer ANN was constructed, including an input layer, a hidden layer, and an output layer. To convert the input data into an output signal, the input layer sent signals to the hidden layer, where they were computed using weights and activation functions. A timely decision is made to deal

with attacks when this technique is used to detect them effectively. In tenfold cross-validation, this ANN achieved an average accuracy of 84% and a false positive rate of under 8%.

Annadurai et al. [18] presented a technique for securing data transmission and detecting intrusions in biometric authentication systems. Intruders can be detected with the help of biometric databases that are based on the Internet of Things. These biometric data have been de-noised, smoothed, and normalized. To extract features from the processed data, kernel-based principal component analysis (KPCA) was utilized. To classify the processed features, we used a modified version of the VGG16 network structure. To ensure the security of the entire network, a Strong Trust Transfer Protocol (DTTP) is used. In the test, the results were 96% precision, 85% F-score, 92% precision, 80% recall, and 46% RMSE.

3 Proposed Authentication Threat Detection Model

3.1 Overview

This section proposes a multikernel-based ELM threat detection technique. Figure 1 represents the proposed threat detection scheme. To enhance the performance of the threat detection model, significant feature sets are initially generated utilizing correlation coefficients, random forest mean accuracy loss, and gain ratios. Feature sets can be combined using the AND operator to obtain a feature set. The combined single feature set is then fed as input to MKELM for threat detection. The performance of MKELM is improved using adaptive crocodile optimization algorithms.

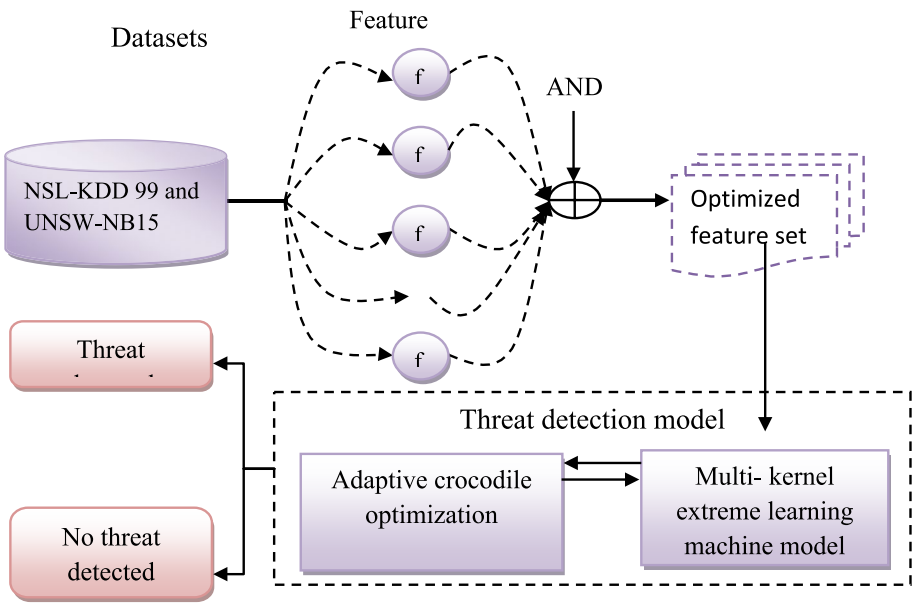


Fig. 1 Proposed authentication threat detection and feature reduction model

3.1.1 Feature Selection Model

A subset of features is selected as part of the feature selection process. With the help of this strategy, the classifier is able to identify and select the most appropriate features. Classifiers built using smaller features may yield more accurate results than using full attributes. Network traffic covers a wide range of features. This facilitates the reduction of the training dataset, the speeding up of the processing, and the accuracy of the test as a result of feature selection.

Based on a novel hybrid combination of filter- and wrapper-based feature selection algorithms, the proposed model extracts the most effective intrusion detection features from IoT network data. A hybrid model is capable of combining gain ratio, mean reduction accuracy, and correlation coefficient. After performing these procedures, the three subsets of features are ranked based on their scores. The most important properties are obtained using the intersection technique (AND function). The following are a brief description of the methods for selecting features in the proposed models:

i Correlation Coefficient

This study selects the features based on the correlation coefficient. In most cases, incoming network communication contains multiple features that serve no purpose and have the same set of values. An independent feature that does not have any correlation with the target variable is one of the most likely candidates for removal. Further, if two independent features have a high correlation one with another, using them together will not only make the detection system more complex, but will also prolong the computation time. Therefore, the proposed approach eliminates all of these variables from the raw information and provides a more straightforward model for identifying threats quickly and accurately. The purpose of this study is to determine the linear dependence of two measurements using the person's correlation coefficient (PCC). When the two traits are interdependent, the PCC will be 1, and when they are independent, it will be 0. As a result, PCC seeks to ascertain how much improving one trait will improve another and vice versa. PCC (ρ) can be calculated for two features $Fe \in \{fe_1, fe_2, \ldots fe_K\}$ using Eq. (1), where V stands for the target variable.

$$\rho(Fe, V) = \frac{\text{cov}(Fe, V)}{\sigma_{Fe}\sigma_V} \tag{1}$$

In this case, the standard deviation is used to indicate the covariance (cov) between the characteristics Fe and V, σ_{Fe} *and* σ_V are represented as the standard deviation (S_d) will be computed with Eqs. (2) and (3)

$$S_d(\sigma_{Fe}) = \sqrt{\text{var}_{Fe}} \tag{2}$$

$$S_d(\sigma_V) = \sqrt{\text{var}_V} \tag{3}$$

Equations (4) and (5) can be used to get the mean and variance for each feature, where K is the total number of values in each feature and μ_{Fe} and μ_V are the mean of the corresponding features.

$$\text{var}(\sigma_{Fe}^2) = \frac{1}{K}\sum_{j=1}^{K}(F_j - \mu_{Fe})^2 \tag{4}$$

$$\text{var}(\sigma_{V}^2) = \frac{1}{K}\sum_{j=1}^{K}(V_j - \mu_{V})^2 \tag{5}$$

By using Eqs. (6) and (7), we can calculate the mean for both features, where *valu* represents the values of the corresponding features in the dataset.

$$\mu_{Fe} = \frac{\sum_{j=1}^{K} valu_j}{K} = \frac{valu_1 + valu_2 + valu_3 + \cdots + valu_n}{K} \tag{6}$$

$$\mu_{V} = \frac{\sum_{j=1}^{K} valu_j}{K} = \frac{valu_1 + valu_2 + valu_3 + \cdots + valu_n}{K} \tag{7}$$

Equation (8) can be used to get the covariance between the two features.

$$\text{cov}(F, V) = \frac{1}{K}\sum_{j=1}^{K}(F_j - \mu_{Fe})(V_j - \mu_{V}) \tag{8}$$

All of the previous equations can be used to create an optimum feature set.

ii Random Forestmean (RFM) reduce accuracy and gain ratio

During this stage, RFM is used to reduce the gain ratio and improve accuracy. It is first necessary to rank the features one by one to determine the best feature from the subset. According to this method, a permuted independent variable Y_j is used in conjunction with an unpermuted independent variable T to forecast the target variable T. In bootstrapping, a large sample is divided into a number of small samples that are routinely taken from the larger sample. It is necessary to use the Out-Of-Bag (OOB) score in order to calculate the drop in accuracy. The Out-Of-Bag (OOB) score is calculated before and after the permutation of the variable fei, and then the difference is calculated to provide a more accurate estimate of the Out-Of-Bag value. The averaged value of the variation between the predicted class in assured explanation i before permuting the feature fe_i, $X_l = F(f_l)$, and after swapping the feature fe_i, $X_l = F(f_l^i)$, is the feature relevance of f_i in the tree $t \in \{1, 2, 3, \ldots, k_{\text{tree}}\}$. Equation (9) can be used to determine RFMAR.

$$fe_i = \frac{1}{k_{tree}}\sum_{V=1}^{k_{tree}}\left[\frac{\sum_{l\in ooB} I(X_l = Fe(fe_l)) - \sum_{l\in ooB} I(X_l = Fe(X_l^i))}{|ooB|}\right] \tag{9}$$

iii Gain Ratio (GR)

In order to select optimal features, the gain ratio technique is applied. In contrast to information gain (I_g), this strategy has the advantage of not favouring attributes with high magnitude values. I_g may choose features that are not optimal for making predictions, leading to overfitting. When ranking a feature, GR takes numbers and branches into account. When the data are evenly distributed, GR has a high value; however, when the data are all

concentrated in one branch, it is low. By considering the intrinsic information (*Ii*) of the split, GR corrects Ig. *Ii* is used to identify the branch of a particular instance. Use Eq. (10) to calculate the GR for the F feature with values $Fe \in \{fe_1, fe_2, \ldots, fe_K\}$

$$GR(Fe) = \frac{Ig_{Fe}}{Ii_{Fe}} \tag{10}$$

It is necessary to compute Ig using Eq. (11) before calculating GR for each feature f.

$$I_g(Fe) = (I_q - I_p) \tag{11}$$

where I_p is feature information after the split and I_q is feature information before the split. Equation (12) can be used to calculate information about the feature prior to the split.

$$I_q(T) = Entropy(T) = -\sum_{j=1}^{n} P(l_i) \log_2 P(l_i) \tag{12}$$

where T stands for the dataset's target variable, l for its value, and P for probability. Then utilize Eq. (13) to compute information for each feature.

$$I_p(T, Fe) = -\sum_{y} P(l)\, Entropy(l) \tag{13}$$

where l stands for the feature's potential values. *Ii* can be determined using Eq. (14).

$$Ii(T, Fe) = -\sum_{j-l}^{n} \frac{|l_i|}{|l|} * \log_2 \frac{|l_i|}{|l|} \tag{14}$$

As a result of the normalisation described above, the value of GR is always between [0, 1]. In addition, if the GR score for the feature Fe is 1, then Fe has full predictive power over T. On the other hand, if the GR score for Fe is 0, then Fe has no predictive power over T.

The obtained optimal feature sets are combined using Eq. (15)

$$O_{FeS} = (\rho(Fe, V)) \,\&\, (fe_i) \,\&\, GR(Fe) \tag{15}$$

Algorithm 1 An optimal feature selection technique

Input: feature set $Fe \in \{fe_1, fe_2, \cdots fe_K\}$

Output: O_{FeS} (Optimized feature set)

1. Read original feature set $Fe \in \{fe_1, fe_2, \cdots fe_K\}$
2. Using PCC, determine how similar all features are.
3. For j= 1, 2,..., K do// j is the feature column index
4. Compute PCC (ρ) using Eq.1
5. $O_{FS1} \leftarrow Fe \in \{fe_1, fe_2, \cdots fe_a\}$ //where a represents the set O_{FeS1}'s most important attributes.
6. End for
7. Calculate OOB score using RFMAR before and after permutation.
8. For j= 1, 2,..., K
9. Calculate RFMAR score using Eq.(9)
10. $O_{FeS2} \leftarrow Fe \in \{fe_1, fe_2, \cdots fe_b\}$ //where b represents the set O_{FeS2}'s most important attributes.
11. End for
12. Compute normalized GR score
13. For j=1, 2,...,c do
14. By using Eq. (10) and Ii for each split, calculate the GR score.
15. $O_{FeS3} \leftarrow Fe \in \{fe_1, fe_2, \cdots fe_c\}$ /where c represents the set O_{FeS3}'s most important attributes.
16. End for
17. $O_{FeS} \leftarrow (O_{FeS1}) \& (O_{FeS2}) \& (O_{FeS3})$
18. Return O_{FeS}

The above extracted features are given as input to multi-kernel ELM for authentication threat detection.

3.2 Multi-kernel Based Extreme Learning Machine Model (MKELM)

In the early days of neural network learning, ELM was implemented only for 1-layer feedforward architectures. Due to its fast-learning speed and excellent generalization performance, the ELM learning algorithm has recently attracted the interest of researchers. It is possible to use most nonlinear piecewise continuous functions as hidden

neurons in ELM since the initial parameters do not need to be tuned. Hence, for S arbitrary discrete samples $\{(y_j, t_j)|y_j \in R^a, t_j \in R^b, j = 1, 2, \ldots, S\}$, H is the output function of ELM with hidden neurons

$$o_H(y) = \sum_{j=1}^{H} \beta_j m_j(y) = m(y)\beta, \tag{16}$$

where $\beta = [\beta_1, \beta_2, \ldots, \beta_H]$ is the hidden layer of H neurons and the output neuron and $m(y) = [m_1(y), m_2(y), \ldots, m_H(y)]$ is the output vector of the hidden layer concerning the input y that maps the data from the input space to the ELM feature space.

To reduce the training error and improve the generalization performance of neural networks, the output weights and the training error must be minimized simultaneously.

$$minimize = \|M\beta - T\|, \|\beta\|. \tag{17}$$

A neural network's generalization performance is improved if the training error and the output weights are minimized simultaneously, as follows:

$$\beta = M^T \left(\frac{1}{c} + MM^T\right)^{-1} \tag{18}$$

An output matrix of the hidden layer is represented by M, a regularization coefficient is represented by C, and an output matrix of the predicted layer is represented by T. This results in the following output function for the ELM learning algorithm.

$$o(y) = m(y)\, M^T \left(\frac{1}{c} + MM^T\right)^{-1} T. \tag{19}$$

It is possible to define a kernel matrix of ELM based on Mercer's conditions (G) if the feature mapping $o(y)$ is unknown:$G = MM^T : g_{ji} = m(y_j)m(y_i) = n(y_j, y_i)$,

The output function $o(y)$ of a multi kernel-based extreme learning machine (MKELM) can be formulated as follows:

$$o(y) = [n(y, y_1), \ldots, n(y, y_S)]\left(\frac{1}{c} + G\right)^{-1} T \tag{20}$$

A single hidden layer feed-forward neural network has $G = MM^T$, and n(y, z) refers to the kernel function of hidden neurons. Several kernel functions can satisfy Mercer's condition, including polynomial kernels, linear kernels, exponential kernels, and Gaussian kernels. For simulation and performance analysis, we use three common kernel functions, which are listed below.

I Gaussian Kernel:

$$n(y, z) = \exp(-s\|y - z\|); \tag{21}$$

II Hyperbolic tangent (sigmoid) kernel:

$$n(y, z) = \tanh(uy^T z + v); \tag{22}$$

III Wavelet kernel:

$$n(y,x) = \cos\left(d\frac{\|y-z\|}{w}\right)\exp\left(-\frac{\|y-z\|^2}{e}\right), \tag{23}$$

In general, Gaussian kernel functions are generalized local kernel functions and tangent kernel functions are generalized global atomic functions. Moreover, the performance of the algorithms is tested using a complex wavelet kernel function. A neural network's performance is greatly influenced by the adjustable parameters s, u, v, w, and e of the three kernel functions discussed above, which should be adapted according to the type of problem being addressed. In contrast to the ELM learning algorithm, the MKELM algorithm does not require knowledge of the hidden layer feature mapping or a choice of hidden neuron number. Utilizing an adaptive crocodile optimization algorithm, the kernel parameters of MKELM are optimized to improve its performance.

3.2.1 Adaptive Crocodile Optimization Algorithm (ACOA)

Crocodile hunting behaviour is used to develop a meta-heuristic optimizer called Crocodile optimization Algorithm (COA). In a recent study, Abualigah et al. [15] developed the RSA metaheuristic algorithm. COA replicates crocodile hunting behaviour in its natural habitat. It is generally considered that crocodiles belong to the family Crocoodylinae, and they prefer a climate with plenty of water and food. Crocodiles belong to the amphibian family and can hunt in the water. They also live outside of water.

i. *Vision* Crocodiles are one of the few animals with night vision. Because of their poor night vision, they hunt at night, unlike other animals.
ii. *Eating* They eat fish, cows, deer, baby elephants, zebras and small crocodiles as part of their diet in the environment around their habitat. Also, large crocodiles will eat cats and sharks in addition to other predators. Without food sources in the environment, it is capable of surviving long periods without food. Sources indicate that some of them consume fruits.
iii. *Physical action* Swimming, walking and running are all abilities of crocodiles. Swimmers steer with their tails, ignoring their legs. Their legs facilitate movement and carry their bodies when they walk, while their tails act as a balance and steering mechanism. The tail transmits energy to the body for forward acceleration, allowing crocodiles to run out of water to attack prey.
iv. *Cognition* Crocodiles can identify their prey to determine which animals frequent the water.
v. *Hunting* In order to catch animals that dive or swim up from the shore to drink, snucking into the water is a common practice. Once their prey approaches the water, crocodiles attack stealthily. Once crocodiles catch their prey, they drag them into the water and drown them. A crocodile eventually eats its prey whole after cutting it up into large pieces. Prey is often shared between crocodiles through fighting.
vi. It is common for crocodiles to live in groups. In preparation for predation ambushes, crocodiles use this pattern. Each member of the group plays a part in predatory behaviour. Crocodiles attack animals that drink from riverbanks to push them to the water, where they attack the prey from beneath the water. As illustrated in Fig. 2, the ACOA procedure steps are described in detail.

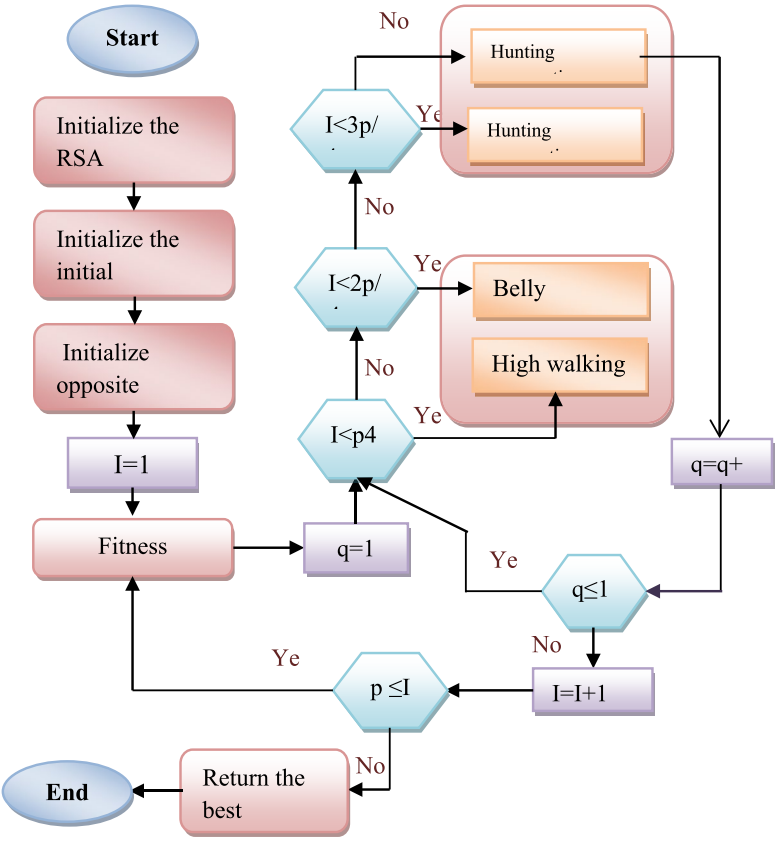


Fig. 2 Steps for ACOA

i *Initialization* ACOA should be executed with both kernel parameters and algorithmic parameters initialized. In this work, we use three kernel functions with a set of adjustable parameters (s, u, v, w, and e) that are represented by C, and the number of iterations are represented by I. Furthermore, ACOA uses two algorithmic parameters, α and β. This algorithmic parameter controls the exploration and exploitation abilities in order to achieve the right balance between exploration and exploitation.

ii *Population Initialization* Using the following equation, an initial set of solutions is generated randomly

$$Y_{p,q} = Y_q^{\min} + rc\, D \times (Y_q^{\max} - Y_q^{\min}), \quad \forall p = 1, 2, \ldots, C, \quad \forall q = 1, 2, \ldots, D, \tag{24}$$

Here Y represents the kernel parameter. The decision variable $Y_{p,q}$ represents the pth solution at the qth position. There are two upper and lower bounds to the decision variable at the qth position: $Y_q^{\max}$ *and* $Y_q^{\min}$. Rand indicates the random number between 0 and 1, while D indicates the total number of decision variables at each solution. In Y, a set of C solutions is generated and is stored as follows:

$$Y = \begin{bmatrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,D-1} & Y_{1,D} \\ Y_{2,1} & Y_{2,2} & \cdots & Y_{2,D-1} & Y_{2,D} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ Y_{C,1} & Y_{C,2} & \cdots & Y_{C,D-1} & Y_{C,D} \end{bmatrix} \tag{25}$$

$Y_{1,D} = \{s, u, v, w, e\}_{1,D}$, where $Y_{1,D}$ indicates the kernel parameter. In each row $Y_p = (Y_{p,1}, Y_{p,2}, \ldots, Y_{p,D-1}, Y_{p,D})$ represents the solution of position p.

iii *Opposite solution* In order to create candidate solutions (the initial position of the crocodile) without prior knowledge, the random initialization method is applied using Eq. (5). The distance between the starting point and the optimum solution determines convergence speed and performance. Generally, the approach performs better when the randomly generated solutions have a lower target function value. An adaptive version of the classic RSA approach is presented in this study in order to increase convergence and the likelihood of finding global optima. In contrast to the previous ARSA, the new method creates the opposite positions for each solution after first generating the initial random solutions (i.e., the positions of the rats) based on the opposite number principle. It is necessary to define the opposite number in order to describe the initialization of the new population. A vector Y with a dimension of C is shown below:

$$Y = (Y_1, Y_2, \ldots, Y_c) \tag{26}$$

where $Y_q$is $Y_q^{\min}$, $Y_q^{\max}$. As a result, the opposite point of Y_q, represented by $\overline{Y_q}$

$$\overline{Y_q} = (Y_q^{\min} - Y_q^{\max}) - Y_q,\ q = 1, 2, \ldots, c \tag{27}$$

By assuming that X_p is a random generated solution in C-dimensional space, the ARSA population is initialized. As a result of solving inverse Eq. (27), $\overline{Y_q}$ represents a stochastic solution. As the objective function of the evaluation, fn (.) is used to evaluate both solutions (i.e., Y_p and $\overline{Y}_q$. If $f_c(Y_q)$ is superior to $f_c(\overline{Y_q})$ (i.e.,$f_c(\overline{y}_q) < f_c(y_q)$ then replaces agent $\overline{Y_q}$; otherwise, continue with agent q. The best initial solution is selected in the first iteration. When dealing with more complex tasks, the ARSA method can become trapped in local optima. Although it is more efficient than traditional methods, it tends to trap in local optima. The search process may require many iterations for certain agents to remain stationary. It will address these shortcomings and improve exploration and search capabilities by replacing the poorest solution with the best fitness value.

iv *Fitness Function* A fitness calculation is conducted at this stage to compare each solution with its opposite. Each node along the routing path has an energy and confidence value, which are used to calculate a fitness function. By summing all intermediates along the routing path, the fitness of the kernel is calculated. For this study, the fitness function is the maximum accuracy value. A fitness function is calculated for each solution based on the classification accuracy function

$$f(Y_p)\ \ \forall_p = \max[accuracy] \tag{28}$$

v *Encircling phase* In the COA, crocodiles explore the water. According to Eq. (29), we introduce two strategies for locating new regions in the search space of a problem: high walking and belly walking. Using Eq. (29) we can see that high walking is controlled by $i \leq I/4$, belly walking is controlled by $I/4 < i \leq 2I/4$.

$$Y_{p,q}(i+1) = \begin{cases} Y_q^{best}(i) - \eta_{p,q}(i) \times \beta - r_{p,q}(i) \times rc\,D, & i \le \frac{I}{4}, \\ Y_q^{best}(i) \times Y_{R1,q}(i) \times ES(i) \times rcD, & \frac{I}{4} < i \le \frac{2I}{4}, \end{cases} \tag{29}$$

The decision variable $Y_{p,q}$ corresponds to the solution at position q. The best solution obtained at i iterations is $Y_q^{best}(i)$. Iteration i + 1 represents the new iteration, while iteration i represents the previous iteration. Using (30), we calculate the hunting operator for the qth position in the ith solution. In the high-walking strategy, β controls its exploration capability. The value of β is set to 0.1. A random value between zero and one is generated by rcD. The decision variable $Y_{r1,q}(i)$ is at the qth position in the r1th solution, where $r1 \in [1, C]$. It is calculated that $\eta_{p,q}(i)$, $F_{p,q}$ and $Avg(Y_p)$ are, respectively,

$$\eta_{p,q} = Y_q^{best} \times F_{p,q}, \tag{30}$$

$$F_{p,q} = \alpha + \frac{Y_{p,q} - Avg(Y_p)}{Y_q^{best} \times (Y_q^{\max} - Y_q^{\min}) + \varepsilon}, \tag{31}$$

$$Avg(y_p) = \frac{1}{D} \sum_{q=1}^{D} Y_{p,q}, \tag{32}$$

Y^{best} at the qth position and Y_p at its corresponding position in this example constitute a percentage difference between the decision variables. α controls RSA exploration ability during cooperative hunting by setting it to 0.1. The value should be random between 0 and 2. In the current solution, Y_p, all decision variables have a combined average value of $Avg(Y_p)$. Pth solution positions q are reduced by $R_{p,q}(i)$, and ES(i) factors assign their values.

$$R_{p,q} = \frac{Y_q^{best} - Y_{r2,q}}{Y_q^{best} + \varepsilon} \tag{33}$$

$$ES(i) = 2 \times r3 \times \left(1 - \frac{1}{I}\right), \tag{34}$$

In a population of solutions, r2 represents the probability of selecting one solution among them at random. The value of r3 can be represented as an integer between 0 and 1.

vi *Hunting phase* This is how crocodiles exploit their habitat in the ACOA. To find optimal solutions based on the existing research regions, ACOA adopts two strategies in this phase: coordination and cooperation. Coordinating hunting is controlled by $i \le 3I/4$,

$$Y_{p,q}(i+1) = \begin{cases} Y_q^{best}(i) \times F_{p,q}(i) \times rcD, & \frac{2I}{4} < i \le \frac{3I}{4}, \\ Y_q^{best}(i) - \eta_{p,q}(i) \times \varepsilon - R_{p,q}(i) \times rcD, & \frac{3I}{4} < i \le I. \end{cases} \tag{35}$$

vii *Stop Criterion* The process is repeated from Step 3 to Step 5 until I is reached.

Algorithm 2 Adaptive crocodile optimization (ACO)

Input: C, I, α and β

Output: Optimal solution

Initialize the **ACO** parameters α **,** β **, etc.**

Generate the initial set of solutions using equation (24)

Evaluate the opposite solution using equation (27)

While fitness value (! = stopping criteria) do

Calculate the fitness value (Y) using equation (28)

Obtain the best solution

Update $R_{p,q}$, ES using equation (33) and (34)

End for

I=i+1

End while

Return to the best position and fitness

4 Result and Discussion

The purpose of this section is to provide an analysis of the proposed intrusion detection model. On a Windows 10 test computer with an Intel Core i5 processor and 6 GB of RAM, the method is implemented in Java. The purpose of this section is to provide an analysis of the proposed intrusion detection model. On a Windows 10 test computer with an Intel Core i5 processor and 6 GB of RAM, the method is implemented in Java. The proposed work is evaluated based on precision (p), recall (R), F-measure, precision, sensitivity specificity, detection rate, training time and fitness.

4.1 Dataset Description

For this study, two datasets were used: the NSL KDD dataset and the UNSW-NB15 dataset. NSL KDD data consists of five classes: normal, DoS, U2R, Probe, and R2L. A total of 148,517 records are contained in this dataset. A total of 115,973 records are used for training, and 22,544 records are used for testing. There are nine classes contained in the UNSW-NB15 dataset, including Normal, Backdoor, Fuzzers, Exploit, Analysis, Worm, Shellcode, and Reconnaissance. There are 2,360,854 records in this dataset. A total of 1,914,519 records are available for training, while a total of 446,335 records are available for testing.

Table 1 Metrics evaluation for the NSL KDD dataset

Methods	P	R	F-score	Accuracy	Sensitivity	Specificity	Detection rate
Proposed ACOA-MKELM	98.84	98.76	98.79	98.98	98.86	98.95	98.97
Existing RF	95.15	95.28	95.21	95.41	95.35	95.46	94.89
Existing KNN	91.12	92.82	91.46	92.84	92.82	92.86	91.65
Existing XGBoost	90.54	89.84	90.18	89.92	90.84	88.79	88.45

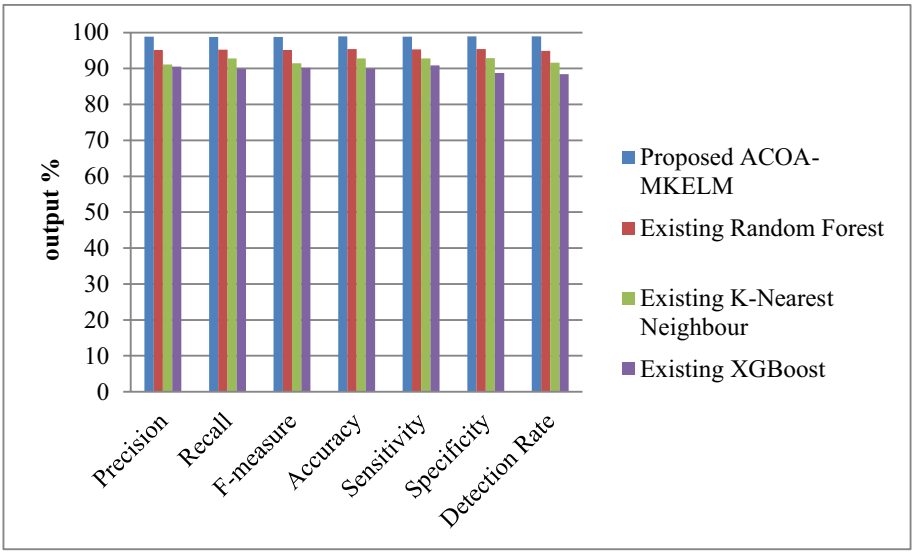


Fig. 3 Metrics evaluation for the NSL KDD dataset

4.1.1 Results for NSL KDD Dataset

The experimental results for the NSL KDD dataset are presented in this subsection. The proposed work is evaluated quantitatively through Table 1 and Fig. 3. Using the NSL KDD dataset, we propose a comparison between three existing techniques: Random Forest (RF), K-Nearest Neighbor (KNN) and XGBoost. According to our proposed work, the precision rate of the proposed method is 98.94%, which is higher than the RF rate of 3.69%, the KNN rate of 7.72%, and the XGBoost rate of 8.3%. The recall rate of our proposed work is 98.76%, our proposed work is improved by 3.48, 5.94 and 8.92% from the above three techniques, respectively. Our proposal has an F-measure of 98.79%, and our proposed work has improved by 3.58, 7.33 and 8.61% over the existing techniques. The accuracy of our proposed work is 98.98%, which is 3.57% higher than RF, 6.145 higher than KNN, and 9.06% higher than XGBoost, respectively. The sensitivity of our proposed work is improved by 3.51, 6.045 and 8.02% from the existing techniques, respectively. The specificity of our proposed work is improved by 3.49, 6.09 and 10.16% from the existing techniques, and the detection rate of our proposed work is improved by 4.08, 7.32 and 10.52% from the existing techniques. Table 2 represents the training time of the proposed scheme with the

Table 2 Result for Training time

Methods	Training time
Proposed ACOA-MKELM	49,874
Existing Random Forest	52,658
Existing K-Nearest Neighbor	55,487

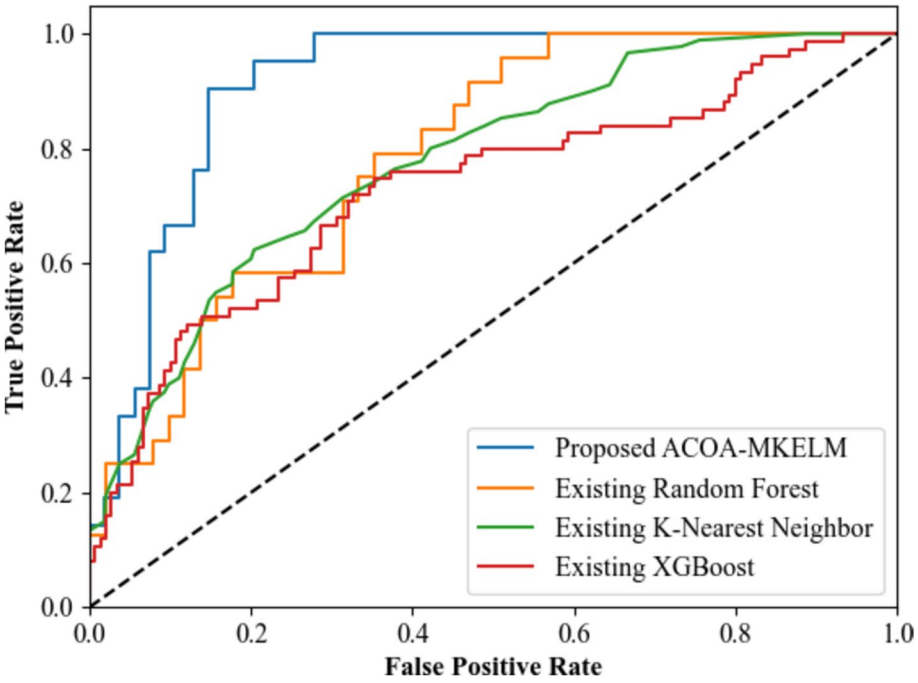


Fig. 4 ROC curve for NSL KDD dataset

existing techniques. From this, it can be seen that our proposed work has 49,874 ms less training time than existing techniques for NSL KDD dataset.

The ROC curve of the NSL KDD dataset is shown in Fig. 4 and the AUC curve is shown in Fig. 5. Figure 6 illustrates the confusion matrix between the proposed and existing techniques. Figure 7 illustrates the fitness validation of the proposed and existing optimization algorithms, comparing our proposed approach to crocodile optimization (COA), whale optimization (WOA), and rain optimization algorithm (ROA). Based on the NSL KDD dataset, it is demonstrated that our proposed work has a high fitness value.

4.1.2 Results for the UNSW-NB15 Dataset

Similar to the previous subsection, this one presents the experimental analysis for the UNSW-NB15 dataset. Based on the UNSW-NB15 dataset, Table 3 and Fig. 8 show the evaluation of the proposed metric for the proposed task. As a result of our proposed work, the precision of RF, KNN, and XGBoost is improved by 5.22, 6.42, and 7.84%, respectively. The recall rate of our proposed task is 97.48%, the recall rate is improved by 2.12,

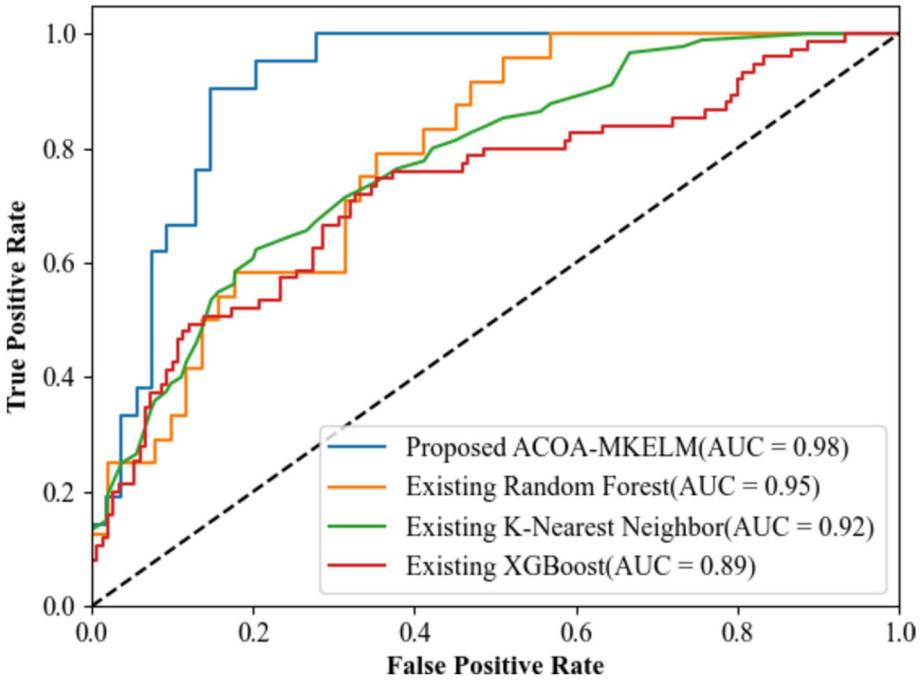


Fig. 5 AUC curve for NSL KDD dataset

6.23 and 7.85%. The F-measure of our proposed work is 97.99%, which is 3.68, 5.72 and 7.87% higher than existing techniques. The accuracy of our proposed work for this dataset is 98.93%, which is 3.53, 5.01 and 9.2% higher than the existing techniques. The sensitivity rate of our proposed work is 98.62%, our proposed work is improved by 4.26, 6.72 and 7.94% from the existing techniques, respectively. The specificity of our proposed work is 98.68%, which is 5.24, 5.74, and 10.13% higher than RF, KNN, and XGBoost, respectively. The detection rate of our proposed technique is 98.89%, our proposed work improves the detection rate from RF, KNN and XGBoost by 6.05, 8.14 and 11.24% respectively. Table 4 represents the training time estimation of the proposed work. From that, it proves that our proposed task has less training time.

AUC and ROC curves are shown in Figs. 9 and 10 for the UNSW-NB15 dataset. Confusion matrices of the proposed and existing models are shown in Fig. 11. Figure 12 shows the fitness validation for the proposed and existing optimization techniques, comparing our proposed work with crocodile optimization (COA), whale optimization (WOA) and rain optimization (ROA) algorithm. It is demonstrated that our proposed work has a high fitness value with UNSW NB15 dataset.

4.2 Comparison with Previous Literature

Table 5 depicts the comparison of previous literature with the proposed scheme. As depicted in the table, Zhang et al. [12]have presented a hierarchical IDS model for intrusion detection and they have achieved 98.34 and 92.10% of accuracy by analyzing their

Fig. 6 Confusion matrics for proposed and the existing techniques

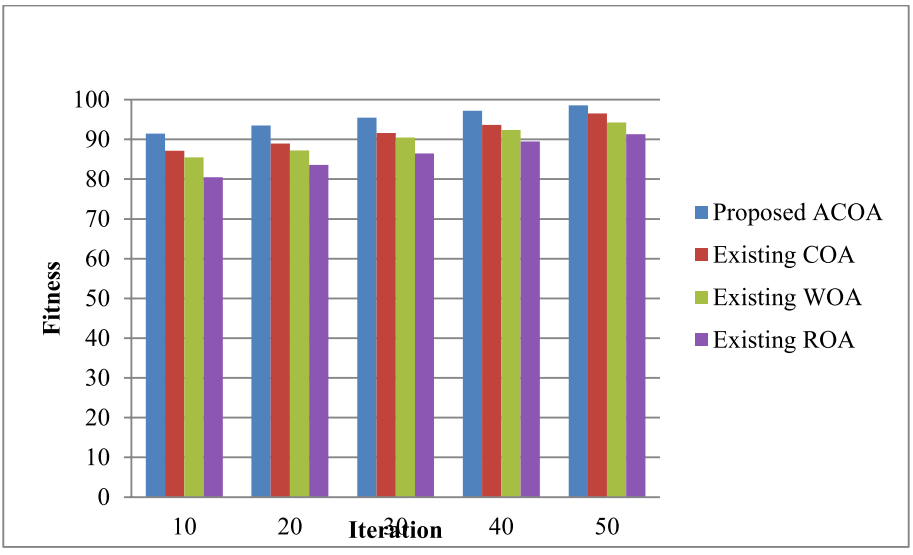


Fig. 7 Comparative result evaluation for the proposed vs. existing techniques

Table 3 Metrics evaluation for the UNSW-NB15 dataset

Methods	Precision	Recall	F-measure	Accuracy	Sensitivity	Specificity	Detection rate
Proposed ACOA-MKELM	98.52	97.48	97.99	98.93	98.62	98.68	98.89
Existing Random Forest	93.30	95.36	94.31	95.40	94.36	93.44	92.84
Existing K-Nearest Neighbor	92.10	91.25	92.27	93.92	91.90	92.94	90.75
Existing XGBoost	90.68	89.63	90.12	89.73	90.68	88.55	87.65

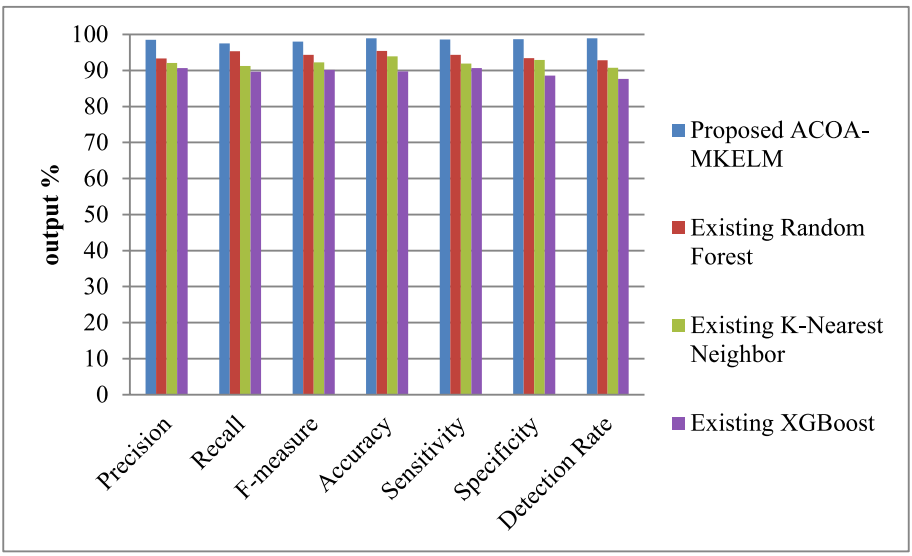


Fig. 8 Metrics evaluation for the UNSW-NB15 dataset

Table 4 Training time result for the UNSW-NB15

Method	Training time
Proposed ACOA-MKELM	46,875
Existing random forest	49,875
Existing K-Nearest Neighbor	52,658
Existing XGBoost	55,689

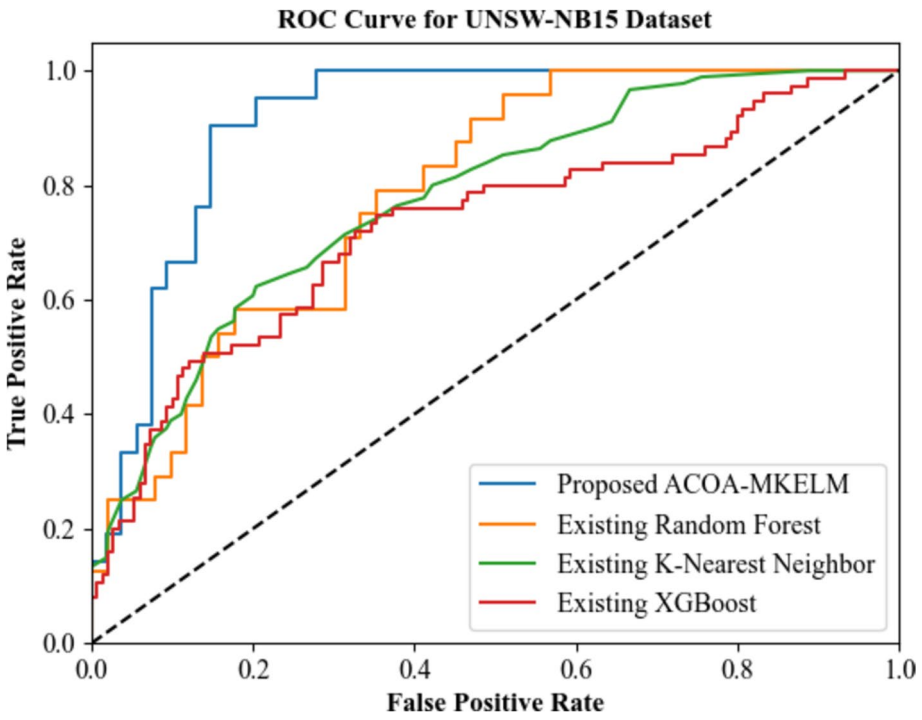


Fig. 9 ROC curve for UNSW- NB15 dataset

model with NSL-KDD and UNSW-NB 15 respectively. Compared to [13–15] and [16], Zhang et al. [12] attained better detection accuracy. Nevertheless, the proposed scheme obtained an accuracy of 98.98 and 98.93% by analyzing the proposed model with the same dataset NSL-KDD and UNSW-NB 15 respectively than that of [12].

5 Conclusion

To enhance the security against threats in IoT, ACOA-MKELM based authentication threat detection model has been presented in this approach. To improve the performance of the insider threat detection model, significant feature sets are generated using correlation coefficients, RFM reduction accuracy, and gain ratio. After that, functionality is used to combine the features to form an optimal feature set. Then the optimal feature set is given as input to the optimized MKELM model, which detects the threats in the network based on the features. Besides, to enhance the performance of MKELM, the ACOA algorithm is used. Using this algorithm, kernel parameters of kernel functions are optimized. Performance analysis was conducted using NSL-KDD 99 and

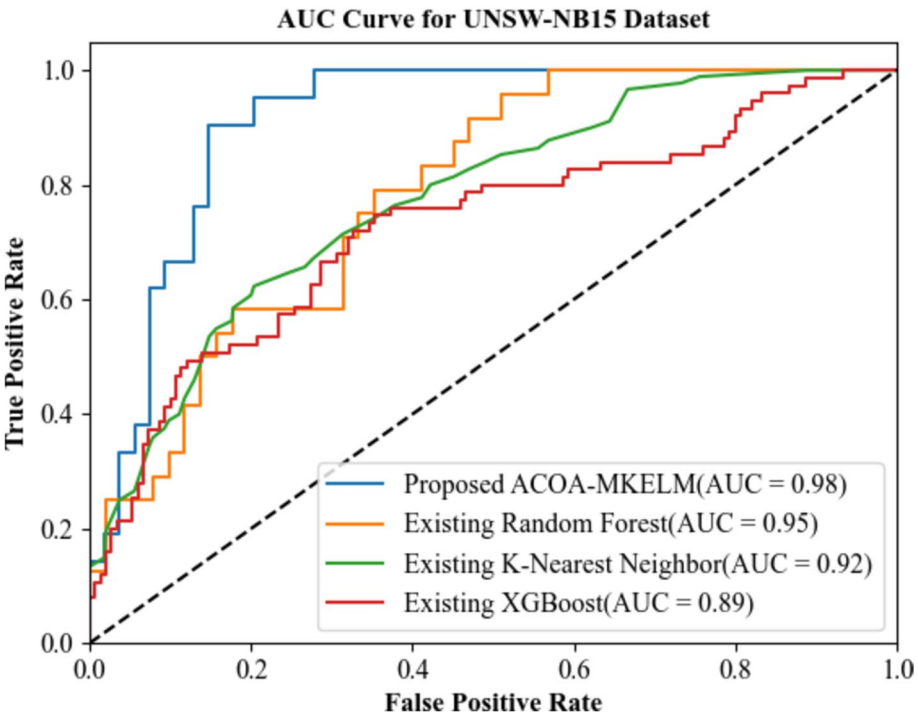


Fig. 10 AUC curve for UNSW- NB15 dataset

Fig. 11 Confusion matrics for the proposed and existing techniques

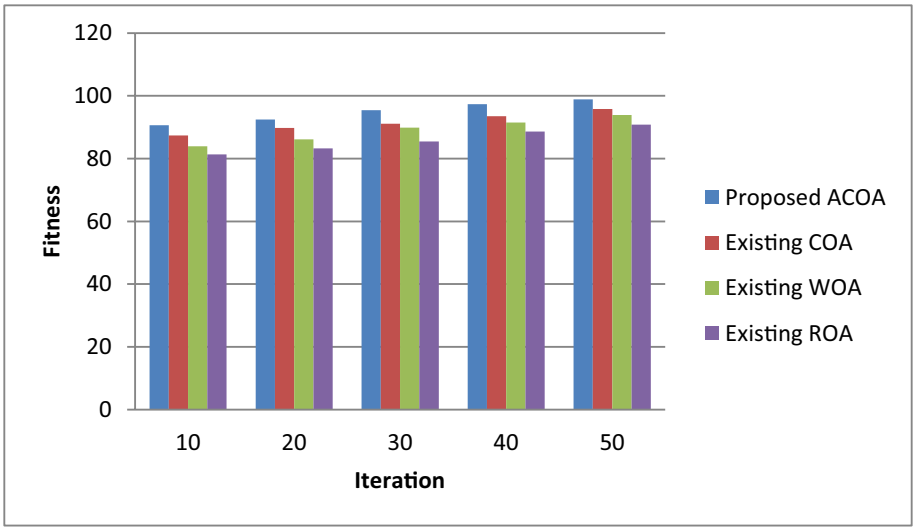


Fig. 12 Comparative result evaluation for the Proposed vs. existing techniques

Table 5 The comparison of previous literature with the proposed scheme

References	Dataset	Methods	Accuracy (%)	Recall (%)	F-measure (%)	Detection rate (%)
Zhang et al. [12]	NSL-KDD and UNSW-NB 15	Hierarchical IDS	98.34 and 92.10	98.04 and 89.42	–	–
Zheng et al. [13]	NSL-KDD	LDA-ELM	92.35	–	–	91.53
Li et al. [14]	KDD Cup 1999	AI-based two-stage intrusion detection	96.42	95.17	97.29	–
Li et al. [15]	KDD Cup 1999	Deep migration learning model	–	–	–	91.05
Lin et al. [16]	NSL-KDD	MFE-ELM	96.53	–	–	–
Proposed	NSL-KDD and UNSW-NB 15	ACOA-MKELM	98.98 and 98.93	98.76 and 97.48	98.79 and 97.99	98.97 and 98.89

UNSW-NB15 datasets. As per the simulation results, the proposed model attained an accuracy of 98.98 and 98.93% for each dataset respectively. However, the computational complexity of the model is to be reduced. Thus, in the future, we focus on presenting an efficient deep learning model as IDS with the evaluation using real-time data.

Authors' Contributions All authors read and approved the final manuscript.

Funding The authors declare that they have competing interests and funding.

Availability of Data and Material 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 Interests On behalf of all authors, the corresponding author states that there is no conflict of interest.

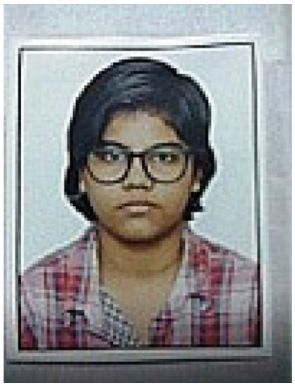
References

1. Minoli, D., Sohraby, K., & Occhiogrosso, B. (2017). IoT considerations, requirements, and architectures for smart buildings—Energy optimization and next-generation building management systems. *IEEE Internet of Things Journal, 4*(1), 269–283.
2. Alrashdi, I., Alqazzaz, A., Alharthi, R., Aloufi, E., Zohdy, M.A., and Ming, H., (2019). FBAD: Fog-based attack detection for IoT healthcare in smart cities. In *2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)* (pp. 0515–0522).
3. Agrawal, S., Sarkar, S., Aouedi, O., Yenduri, G., Piamrat, K., Alazab, M., Bhattacharya, S., Maddikunta, P. K. R., & Gadekallu, T. R. (2022). Federated learning for intrusion detection system: Concepts, challenges and future directions. *Computer Communications, 195*, 346–361.
4. Dang, L. M., Piran, M. J., Han, D., Min, K., & Moon, H. (2019). A survey on internet of things and cloud computing for healthcare. *Electronics, 8*(7), 768.
5. Skarlat, O., Nardelli, M., Schulte, S., Borkowski, M., & Leitner, P. (2017). Optimized IoT service placement in the fog. *Service Oriented Computing and Applications, 11*(4), 427–443.
6. Ustundag, A., Cevikcan, E., Ervural, B.C. and Ervural, B. (2018). Overview of cyber security in the industry 4.0 era. In *Industry 4.0: Managing the digital transformation*, pp. 267–284.
7. Hady, A. A., Ghubaish, A., Salman, T., Unal, D., & Jain, R. (2020). Intrusion detection system for healthcare systems using medical and network data: A comparison study. *IEEE Access, 8*, 106576–106584.
8. Zoppi, T., & Ceccarelli, A. (2021). Prepare for trouble and make it double! Supervised-unsupervised stacking for anomaly-based intrusion detection. *Journal of Network and Computer Applications, 189*, 103106.
9. Anwar, S., Mohamad Zain, J., Zolkipli, M. F., Inayat, Z., Khan, S., Anthony, B., & Chang, V. (2017). From intrusion detection to an intrusion response system: Fundamentals, requirements, and future directions. *Algorithms, 10*(2), 39.
10. Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al-Nemrat, A., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. *IEEE Access, 7*, 41525–41550.
11. Jeyaselvi, M., Dhanaraj, R. K., Sathya, M., Memon, F. H., Krishnasamy, L., Dev, K., Ziyue, W., & Qureshi, N. M. F. (2023). A highly secured intrusion detection system for IoT using EXPSO-STFA feature selection for LAANN to detect attacks. *Cluster Computing, 26*(1), 559–574.
12. Zhang, W., Han, D., Li, K. C., & Massetto, F. I. (2020). Wireless sensor network intrusion detection system based on MK-ELM. *Soft Computing, 24*, 12361–12374.
13. Zheng, D., Hong, Z., Wang, N., & Chen, P. (2020). An improved LDA-based ELM classification for intrusion detection algorithm in IoT application. *Sensors, 20*(6), 1706.
14. Li, J., Zhao, Z., Li, R., & Zhang, H. (2018). Ai-based two-stage intrusion detection for software defined IOT networks. *IEEE Internet of Things Journal, 6*(2), 2093–2102.

15. Li, D., Deng, L., Lee, M., & Wang, H. (2019). IoT data feature extraction and intrusion detection system for smart cities based on deep migration learning. *International Journal of Information Management, 49*, 533–545.
16. Lin, H., Xue, Q., Feng, J., & Bai, D. (2023). Internet of things intrusion detection model and algorithm based on cloud computing and multi-feature extraction extreme learning machine. *Digital Communications and Networks, 9*(1), 111–124.
17. Hanif, S., Ilyas, T., and Zeeshan, M. (2019). Intrusion detection in IoT using artificial neural networks on UNSW-15 dataset. In *2019 IEEE 16th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT and AI (HONET-ICT)* (pp. 152–156).
18. Annadurai, C., Nelson, I., Devi, K. N., Manikandan, R., Jhanjhi, N. Z., Masud, M., & Sheikh, A. (2022). Biometric authentication-based intrusion detection using artificial intelligence internet of things in smart city. *Energies, 15*(19), 7430.

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

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



Prachi Dahiya is currently pursuing Ph.D. from the Department of Computer Science and Engineering at Delhi Technological University in New Delhi, India. She completed her Masters in Technology (M.Tech.) in Computer Engineering (CE) from J.C. Bose University of Science and Technology (JCBUST), YMCA, Faridabad, Haryana, India in 2020, and Bachelors in Computer Science and Engineering (CSE) from Guru Gobind Singh Indraprastha University (GGSIPU), Dwarka, New Delhi, India. She is working as a Guest Faculty in Delhi Technological University for the past 2.5 years alongside pursuing Ph.D. Her research interests and specializations include Data Mining, Web Mining, Webpage Ranking Algorithms, Collaborative Filtering (CF), Recommender Systems, Internet of Things (IoT) and Machine Learning (ML).

Dr. Vinod Kumar is working as Professor in the Department of Computer Science and Engineering at Delhi Technological University in New Delhi, India. He obtained his Ph.D. in Computer Science and Technology from Jawaharlal Nehru University in 2019, his M.Tech. in Computer Science and Technology from Jawaharlal Nehru University in 1998, and his B.E in Electronics and Communication from IIT Roorkee (formerly University of Roorkee) in 1994. Dr. Kumar has over 20 years of teaching experience, having served as a lecturer and senior lecturer in computer science at Jamia Millia Islamia from 1999–2009, and as an associate professor at Delhi Technological University since 2009. In addition, he worked as a research engineer at C-DOT in New Delhi from 1994–1996. Dr. Kumar's research interests and specializations include Information Theory and Coding, Voice and Data Communication, Monte Carlo Simulation, Networking, ISDN, GSM, VoIP, and TCP/IP. He has published numerous research papers, including "ADAS Authentic Data allowed in Internet of Things" in the Turkish Journal of Computer and Mathematics Education and "Grouping and Sponsoring Centric Green Coverage Model for Internet of Things" in Sensors. Dr. Kumar is also Head of the Department of Computer Science and Engineering at Delhi Technological University.

Check for updates

Received: 13 February 2024 | Revised: 9 August 2024 | Accepted: 14 October 2024

DOI: 10.1002/ep.14527

ENVIRONMENTAL PROGRESS & SUSTAINABLE ENERGY

ORIGINAL RESEARCH

Sustainable Energy

Annual energy, exergy, and efficiency analyses for conical solar still combined with N number of evacuated collectors

Abhishek Kumar[1,2] | Rajesh Kumar[1]

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

[2]Department of Mechanical and Allied Engineering, IILM University, Greater Noida, Uttar Pradesh, India

Correspondence
Rajesh Kumar, Department of Mechanical Engineering, Delhi Technological University, Shahbad Daulatpur, Bawana Road, Delhi 110042, India.
Email: rajesh.kumar@dtu.ac.in

Abstract

The utilization of solar energy technology for obtaining clean water for use of society in remote locations will help in fulfilling the sustainable development goals of the United Nations. Also, promotion of solar energy for the use of society will subside dependency on the fossil fuel or conventional energy. This article deals with the annual energy, exergy, and efficiency analyses of N identical ETCs incorporated conical solar still (NETC-CSS). The thermal model for the proposed system has been developed based on equating input and output heats for different elements. The developed fundamental equations are fed to the MATLAB computational code. The four weather situations in each month of year for New Delhi climate is considered for the annual analysis. The annual energy, exergy, thermal efficiency, and exergy efficiency for NETC-CSS are computed to be 1796.61, 170.19 kWh, 38.79%, and 3.94% under optimized values of mass flow rate and number of collectors. Results are compared with earlier published research. Concludingly, the increase in annual energy, exergy, thermal efficiency, and exergy efficiency for NETC-CSS is 68.03% than modified solar still, 74.14% than conventional conical solar still, 61.12% than modified solar still, and 72.59% than solar still with parabolic trough collector.

KEYWORDS
conical solar still, efficiency, energy, evacuated collector, exergy, yield

1 | INTRODUCTION

The examination of active solar still is essential in the present era to address the freshwater scarcity issue. This technology powered by solar energy offers a sustainable solution without harming the environment. Also, it aligns with the United Nations sustainable goal number 6, which focuses on ensuring clean water and sanitation for all. The greenhouse effect is used by solar still for getting freshwater from dirty water with the help of solar energy.

Solar stills can be categorized as either passive or active. Passive solar stills are straightforward in design but typically produce lower output. To enhance performance, heat can be supplied to the basin of a passive solar still using a solar collector. When a passive solar still is combined with a heat supplying element, it becomes an active solar still. This concept was initially explored by Rai and Tiwari in 1983.[1] Since then, many advancements in design have been made, which are detailed in following sections.

Utilizing an evacuated collector increases the heat provided to the basin of a solar still, as it effectively minimizes convective heat loss, leaving only radiative heat loss.[2] In contrast, other collectors experience both radiative and convective heat transfers.[3] Research incorporating evacuated tubes in a natural circulation mode into a solar still concluded that the optimal number of tubes was 10 and resulted in a freshwater output of 3.8 kg per unit area.[4,5] Esen and Esen[6] studied solar water heater with different refrigerants and concluded that the best performance was obtained using R410A because of its higher latent heat value. Esen et al.[7] investigated heat pump with heat exchanger and reported the efficiency of the system as

 © 2024 American Institute of Chemical Engineers.

56.3%. Further, study on the utilization of solar energy for cooking using solar cooker by incorporating heat pipe with refrigerant was done and they concluded that several foods were cooked in 27–70 min period.[8] In another study, solar cooker with collector containing heat pipe was studied and concluded that several foods were cooked in 25–73 min periods using Freon 404A and Freon 410A as refrigerants.[9] Kumar et al.[10] studied solar still by incorporating evacuated tubes and concluded that the daily optimum freshwater output was 3.9 kg which was better than the freshwater output of solar still with evacuated tubes operating in natural circulation mode. Shafii et al.[11] experimented with solar still by incorporating evacuated tube and thermoelectric generator for the utilization of heat of condensation of vapor. They concluded that the freshwater production for the reported system was 3.13 times higher than the passive solar still due to the heat supplied to basin. Sharshir et al.[12] explored the integration of an evacuated heater and the humidification-dehumidification concept in a solar still, finding that the gain output ratio was 39% higher compared to a conventional solar still. Singh[13] conducted exergo-enviro-economic analyses of a dual slope solar still with incorporated photovoltaic thermal (PVT) collectors and concluded that flat plate collectors integrated system performed better than other systems. Singh and Tiwari[14] studied basin type solar still by incorporating evacuated tubular collectors and concluded that the yearly energy output was higher by 6.85% for double slope setup than the corresponding single slope setup due to better distribution of solar intensity in the case of double slope setup.

Singh and Al-Helal[15] and Singh[16] examined solar stills with the inclusion of evacuated collectors and analyzed their energy matrices. Their findings indicated that the energy payback time for systems with evacuated collectors was lower than other collector types due to the reduction of convection heat loss in the evacuated collectors. Singh et al.[17,18] reviewed various solar distillers with and without the use of nanofluids. Sharshir et al.[19] experimented with pyramid solar still by incorporating evacuated collectors and nanofluids. They reported that the modified pyramid solar still produced 4.77% more freshwater compared to the conventional design, attributing the improvement to the combined effects of the evacuated collector and the nanofluid. Singh et al.[20] and Kumar et al.[21] explored the performance of single slope solar stills integrated with evacuated tubular collectors. They found that the life cycle conversion efficiency decreased as the mass flow rate increased for a given number of collectors and water mass in the basin. Additionally, their research indicated that carbon mitigation improved with an increase in the number of collectors. Panchal et al.[22] conducted experiments on a solar still enhanced with calcium stones and evacuated tubes. Their results showed that the setup produced 113.52% more freshwater than a conventional solar still. The increased efficiency was attributed to the combination of evacuated tubes supplying heat to the solar still and calcium stones serving as heat storage material. Sharma et al.[23] developed equations for double slope solar still combined with concentrator integrated evacuated collectors. Mevada et al.[24] experimented with solar still incorporated with fins, evacuated collectors and condenser. They concluded that the freshwater output of the modified solar still was 73.45% higher freshwater output than the conventional solar still due to the presence of fins and condenser. Sadeghi and Nazari[25] experimented with solar still by incorporating concentrator integrated evacuated collector and nanofluid having disinfection property and concluded that the freshwater production from the reported solar still was 117% more than the conventional solar still due to combined effect of collector and nanofluid.

Dawood et al.[26] experimented with solar still combined with parabolic concentrator integrated evacuated tubes and phase change material in receiver tube. They concluded that the cost of freshwater production from the proposed system was 29.87% lower than the conventional solar still due to the presence of phase change material which provided heat to the basin at night, too. Sensitivity analysis[27–33] of solar systems was performed and it was concluded that the output of such analysis gave information to the designer and installer about which parameter affected most to the output of the system. The information would be helpful while installing such systems as per the need of user. Sharshir et al.[34] conducted an experimental study on a pyramid solar still equipped with evacuated tubes, condenser, ultrasonic defogger and nanofluid. Their analysis revealed that the cost of producing freshwater with this enhanced system was 32.04% lower than that of a traditional pyramid solar still. This cost savings was due to the increased rate of evaporation facilitated by the advanced components. Moghadam and Samimi[35] examined a solar still with an evacuated tubular collector basin and found that increasing the specific surface area of the condenser was the key factor in improving freshwater production from the solar still. Gangavathi et al.[36] experimented with evacuated tube solar still containing nano coated condenser and concluded that the freshwater production from the proposed system was 28.53% higher than the conventional solar still because of the provision of heat by collector to the basin. Gopi et al.[37] reported the thermal modeling of multi effect solar still and validated with experimental data. They concluded that there was a reasonable correlation between theoretical end experimental values. The recommended number of effects is seven, as no significant enhancement in freshwater production was observed beyond this point. Kumar et al.[38] studied the performance of a solar still integrated with evacuated collectors focusing on the impact of water mass. They concluded that the optimal water depth for maximum efficiency was 0.56 m.

Farghaly et al.[39] conducted experimental research on a solar still integrated with evacuated collectors and found that incorporating parabolic reflector increased the freshwater output by 30.31% compared to using evacuated collector only. This improvement was due to the concentration of solar energy to receiver surface in parabolic reflector combined with evacuated collector. Dahab et al.[40] investigated hemispherical solar still with evacuated collectors and reported that the efficiency of the proposed system was 47.3% than the efficiency of conventional hemispherical solar still because of the heat supplied to the basin by evacuated collectors. Nagpal and Singh[41] investigated double slope solar still by incorporating evacuated collectors and concluded that the carbon mitigation was 16.57% more for concentrator integrated evacuated collectors solar still than evacuated collectors integrated solar still. Further, a review on different solar stills including solar collectors with their design,

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

output, and cost consideration was reported.[42–45] Their findings indicated that the active solar still exhibited superior performance compared to the passive solar still, however cost of active solar still was higher side than the passive solar still. Arani et al.[46] experimented with solar stills from energy and exergy considerations by incorporating tar coated blue metal stone. They concluded that the freshwater output from the reported system was 34.4% higher than the conventional solar still. Kumar et al.[47] experimented with solar still by incorporating evacuated tubes and concluded that the freshwater output of the reported system was 52.97% higher than the conventional solar still.

The conical solar still experiences less shading compared to basin type solar stills, which enhances its potential for higher output. This reduced shading effect has garnered significant interest from researchers. Gad et al.[48] found through experimentation that the conical solar still produced 42.90% more freshwater than traditional solar stills. Additionally, Kabeel et al.[49] investigated conical solar still with inclination angles of 30°, 45°, and 60° and recommended a 30° angle for getting optimal yield and efficiency. Abdallah and Aldarabseh[50] conducted experiments on a hexagonal pyramid solar still with a conical inner surface varying the mass flow rate. Their findings revealed that the reported system produced 221.5% more freshwater compared to conventional solar stills by reduced shading effect and design improvements. Additionally, Singh et al.[51,52] evaluated the conical solar stills from various perspectives including exergo-enviro-economic analysis, energy metrics and sensitivity. They found that cost of freshwater from the conical solar still was 13.56% lower than that from conventional solar stills. The conical solar still also showed superior performance in terms of energy metrics and efficacy. Further, Kumar et al.[53] reported the development of characteristic equations for conical solar still by incorporating partially/fully covered PVT concentrators. Singh et al.[54] reported the effect of different PVT technologies on the performance of active solar still and they concluded that the crystalline silicon was the most suitable considering environment aspect. In another study, solar still with evacuated collectors was investigated from economic and energy metrics consideration.[55,56] Krishna et al.[57,58] studied the effect of input parameters on the output of active solar still and suggested that the results would help in deciding number of collectors and mass flow rate. Yan et al.[59] investigated hybrid microgrid system for its optimization and concluded that the system consisting of wind turbine/hydrogen energy/ solar configuration has the minimum cost among all the studied systems. In another study, Yan et al.[60] studied solar energy system and liquified natural gas for its integration in a best and most favorable way and they concluded that there was a significant improvement in the capacity of production. Zheng et al.[61] investigated a multigeneration system and concluded that the exergy efficiency was 33.25% under optimized condition. Gao et al.[62] investigated a system consisting of solar and geothermal energies and concluded that the exergy efficiency was 9.34% under optimal condition. Xiao et al.[63] investigated closed loop geothermal system and suggested that the study would work as guideline for engineers in developing system for extracting heat from geothermal resources.

From the literature, it is clear that the investigation on the conical solar still combined with evacuated collectors has not been carried out by any researcher which form a research gap. This research gap has been addressed in the present analysis. The fundamental equations have been developed followed by the conduction of annual energy, exergy and efficiency analyses for a conical solar still integrated with N evacuated collector (NETC-CSS). The main aim of the study includes:

i. To develop thermal model for NETC-CSS followed by the computation of annual energy and exergy.
ii. To evaluate thermal efficiency and exergy efficiency for NETC-CSS.
iii. To compare results of the proposed system with earlier published research.

2 | MATERIALS AND METHODS FOR THE ANALYSIS OF NETC-CSS

The proposed system depicted in three dimensions in Figure 1 and its specifications are detailed in Table 1. In Figure 1, a series connection of N evacuated collectors is shown, designed to achieve the higher water temperatures needed for increased evaporation and higher freshwater yield. When solar radiation hits the condensing surface, part of it is reflected and absorbed, while the rest is transmitted to the water surface. The remaining solar radiation is absorbed by the basin liner, causing its temperature to rise. This elevated temperature of the basin liner then transfers heat to the water in the basin. Additionally, the water in the basin receives heat from the collectors. As the basin water temperature rises, it creates a temperature gradient between the inner surface and the condensing surface. This temperature difference causes evaporation leading to the condensation of vapor on the inner surface of the condensing cover through film wise condensation. The condensate then drips down due to gravity, and it is collected in the beaker.

Thermal model has been developed following the assumptions given in Singh and Tiwari.[14] Following Singh et al.,[51] the energy balance equations (EBEs) for condensing surface are written as follows:

For east facing condensing surface

$$\dot{\alpha_g} I_{SE}(t)\frac{A_g}{4}+h_{1wE}(T_w-T_{gE})\frac{A_b}{4}+h_{EW}(T_{gW}-T_{gE})\frac{A_g}{4}=h_{1gE}(T_{gE}-T_a)\frac{A_g}{4}, \tag{1}$$

For west facing condensing surface

$$\dot{\alpha_g} I_{SW}(t)\frac{A_g}{4}+h_{1wW}(T_w-T_{gW})\frac{A_b}{4}-h_{EW}(T_{gW}-T_{gE})\frac{A_g}{4} = h_{1gW}(T_{gW}-T_a)\frac{A_g}{4}, \tag{2}$$

For south facing condensing surface

$$\dot{\alpha_g} I_{SS}(t)\frac{A_g}{4}+h_{1wS}(T_w-T_{gS})\frac{A_b}{4}+h_{NS}(T_{gS}-T_{gN})\frac{A_g}{4}=h_{1gS}(T_{gS}-T_a)\frac{A_g}{4}, \tag{3}$$

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

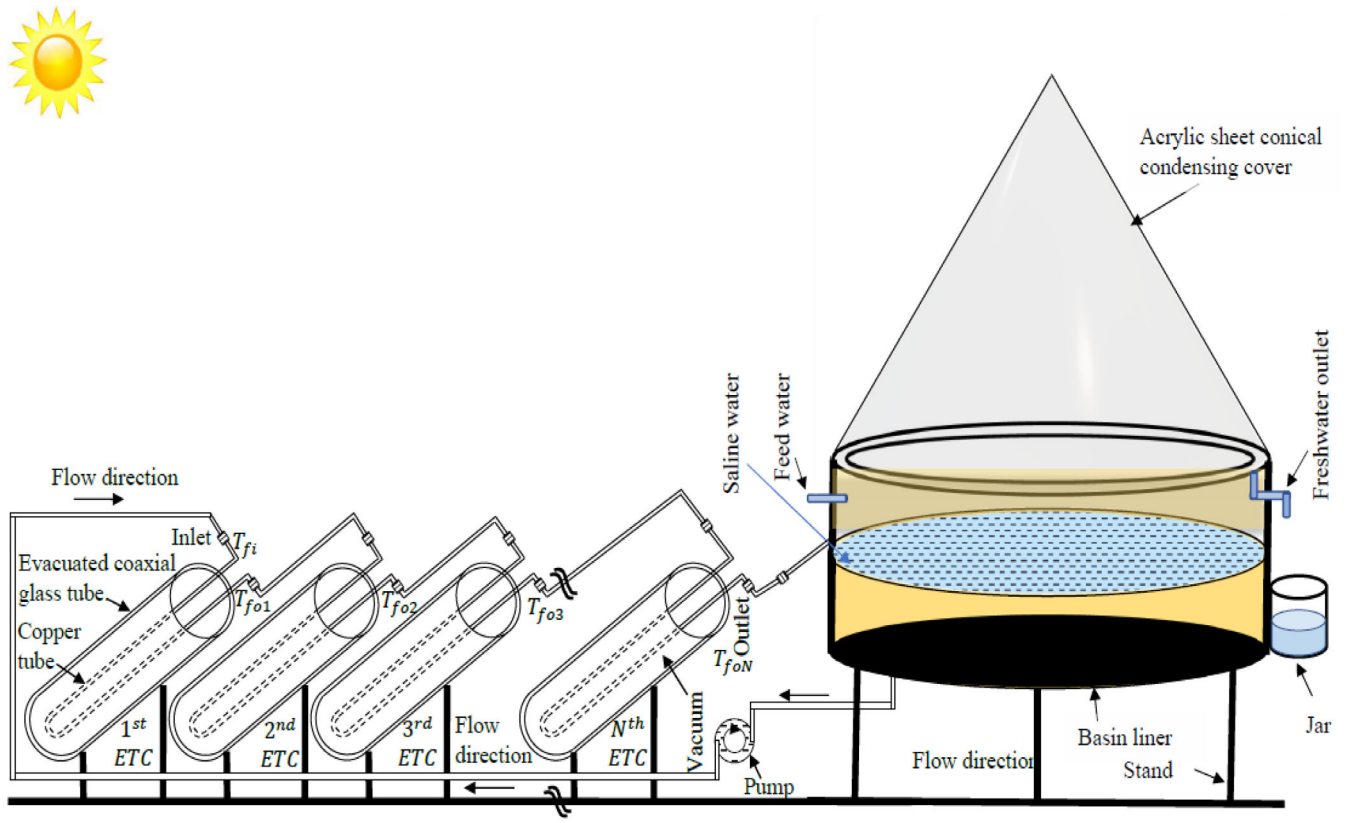


FIGURE 1 Conical solar still combined with *N* alike evacuated tubular collectors.

For north facing condensing surface

$$\dot{\alpha_g} I_{SN}(t)\frac{A_g}{4}+h_{1wN}(T_w-T_{gN})\frac{A_b}{4}-h_{NS}\left(T_{gS}-T_{gN}\right)\frac{A_g}{4}=h_{1g}(T_{gN}-T_a)\frac{A_g}{4}, \tag{4}$$

All unidentified terms in Equations (1) to (4) are given in Appendix A. Following the steps provided in Singh et al.,[53] solution to Equations (1) to (4) are written as follows:

$$T_{gE}=\frac{A_1+A_2T_w}{P_1}, \tag{5}$$

$$T_{gW}=\frac{B_1+B_2T_w}{P_1}, \tag{6}$$

$$T_{gS}=\frac{C_1+C_2T_w}{P_2}, \tag{7}$$

$$T_{gN}=\frac{D_1+D_2T_w}{P_2}. \tag{8}$$

Again, all unidentified terms in Equations (5) to (8) are given in Appendix A. Now, EBEs for basin liner is as follows:

$$\left[\dot{\alpha_b}(I_{SE}(t)+I_{SW}(t)+I(t)+I_{SS}(t))\frac{A_b}{4}\right] =h_{bw}(T_b-T_w)A_b+h_{ba}(T_b-T_a)A_b, \tag{9}$$

Equation (9) can further be simplified as:

$$h_{bw}(T_b-T_w)A_b=\dot{\alpha_b}h_1(I_{SE}(t)+I_{SW}(t)+I_{SN}(t)+I_{SS}(t))A_b -U_bA_b(T_w-T_a), \tag{10}$$

where

$$h_1=\frac{h_{bw}}{4(h_{bw}+h_{ba})}, \tag{11}$$

and

$$U_b=\frac{h_{ba}h_{bw}}{(h_{bw}+h_{ba})}. \tag{12}$$

Further, EBS for water mass can be written as:

$$[\dot{\alpha_w}(I_{SE}(t)+I_{SW}(t)+I_{SN}(t)+I_{SS}(t))\frac{A_b}{4}]+[h_{bw}(T_b-T_w)A_b]+\dot{Q}_{u,N} =h_{1wE}(T_w-T_{gE})\frac{A_b}{4}+h_{1wW}(T_w-T_{gW})\frac{A_b}{4}+h_{1wS}(T_w-T_{gS})\frac{A_b}{4} +h_{1wN}(T_w-T_{gN})\frac{A_b}{4}+M_wC_w\frac{dT_w}{dt}, \tag{13}$$

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Specifications of NETC-CSS and the mean wind speed for different month.

CSS			
System component	**Specification**	**System component**	**Specification**
Surface area of conical passive solar still	1 square meter	α_a	0.01
Configuration of basin	Round	α_w	0.02
Condensing surface material	Acrylic	ε_w	0.96
Acrylic condensing surface depth	0.004 m	α_b	0.82
Insulation material composition	Foam	σ	5.67×10^{-8} W/m^2-K^4
Insulation layer depth	5 cm	ε_g	0.88
Basin material composition	Black-coated GI	Water basin mass	25 kg
Cover incline for condensation	30°	Acrylic reflectivity	0.05
Insulation thermal conductivity	0.166 watts per meter kelvin	Water reflectivity	0.05
Insulation layer depth	0.1 m		

NETC			
System component	**Specification**	**System component**	**Specification**
Type and no. of collectors	ETC, N	α_p	0.8
DC motor rating	12 *V*, 24 *W*	F'	0.968
Diameter of inner copper tube	0.0125 m	τ_g	0.95
Thickness of copper tube	0.0005 m	$K_g\ (Wm^{-1}K^{-1})$	1.09
Differential thickness between external and internal glass elements of evacuated coaxial glass tube	0.002 m	Individual length of each copper tube	2 m
The external radius of the outer glass component in the evacuated coaxial glass tube	0.024 m	Inclination angle of the ETC with respect to the horizontal plane	30°
The internal radius of inner glass element within the evacuated coaxial glass tube	0.0165 m		

Average air velocity variations across different months of the year												
Month	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**June**	**July**	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**
Air velocity (m/s)	2.77	3.13	3.46	3.87	4.02	4.11	3.39	2.91	2.85	2.16	1.83	2.40

where, $\dot{Q}_{u,N}$ is the useful heat obtained from series connected *N* evacuated collectors and can be written as[64]:

$$\dot{Q}_{uN} = \frac{(1-K_k{}^N)}{(1-K_k)}(A F_R(\alpha\tau))_1 I(t) + \frac{(1-K_k{}^N)}{(1-K_k)}(A F_R U_L)_1 (T_{fi} - T_a), \quad (14)$$

where, the value of T_{fi} is equal to T_w. The hot water obtained at the exit of Nth ETC is discharged to the basin. Hence, the value of T_{wo} is equal to the value of T_{foN}. The expression of temperature at the exit of Nth ETC (T_{foN}) can written as:

$$T_{foN} = \frac{(A F_R(\alpha\tau))_1}{\dot{m}_f C_f}\frac{(1-K_k{}^N)}{(1-K_k)} I(t) + \frac{(A F_R U_L)_1}{\dot{m}_f C_f}\frac{(1-K_k{}^N)}{(1-K_k)} T_a + K_k{}^N T_{fi}. \quad (15)$$

All unknown parameters used in Equations (14) and (15) are provided in Appendix A.

Now, one can obtain differential equation using Equations (5)–(8), Equations (10) and (13)–(14) as follows:

$$\frac{dT_w}{dt} + a_1 T_w = f(t), \quad (16)$$

where,

$$f(t) = \frac{1}{M_w C_w}\left[\left(\frac{\dot{\alpha}_w}{4} + \dot{\alpha}_b h_1\right) A_b (I_{SE}(t) + I_{SW}(t) + I_{SN}(t) + I_{SS}(t)) + \frac{(1-K_k^N)}{(1-K_k)}(A F_R(\alpha\tau))_1 I(t) - E_1 + \left\{\frac{(1-K_k^N)}{(1-K_k)}(A F_R U_L)_1 + U_b A_b + F_1\right\} T_a\right], \quad (17)$$

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$a_1 = \left(\frac{\dot{m}_f C_f\left(1 - K_k^N\right) + F_1 + U_b A_b}{M_w C_w}\right). \quad (18)$$

With the help of assumptions that is, taking the interval of time from 0 to Δt, taking the average values of both environment temperature and intensity of solar radiation to keep constant the $f(t)$ and therefore using the average $\bar{f}(t)$ in the expression, Equation (16) will be solved provided, the value of a_1 will remain constant during the time Δt. After carrying out the procedure, the solution of Equation (16) is as follows:

$$T_w = \frac{\bar{f}(t)}{a_1}\left(1 - e^{-a_1 t}\right) + T_{wo} e^{-a_1 t}, \quad (19)$$

where,

$$\bar{f}(t) = \frac{1}{M_w C_w}\left[\left(\frac{\dot{\alpha}_w}{4} + \dot{\alpha}_b h_1\right) A_b\left(\bar{I}_{SE}(t) + \bar{I}_{SW}(t) + \bar{I}_{SN}(t) + \bar{I}_{SS}(t)\right) - E_1 + (U_b A_b + F_1)\bar{T}_a\right]. \quad (20)$$

To clarify the variables, which are not known, Appendix A can be referred.

In the Equation (21), the heat transfer rate due to evaporation has been given.

$$\dot{q}_{ew} = \left[h_{ewgE}(T_w - T_{gE}) + h_{ewgW}(T_w - T_{gW}) + h_{ewgS}(T_w - T_{gS}) + h_{ewgN}(T_w - T_{gN})\right]\left(\frac{A_b}{4}\right). \quad (21)$$

The hourly freshwater output by NETC-CSS is as follows:

$$\text{Hourly freshwater yield} = \frac{\dot{q}_{ew} \times 3600}{(\text{Latent heat})}. \quad (22)$$

2.1 | Experimental Validation

Kumar et al.[53] did experimentation with passive conical solar still for the climatic condition of Greater Noida, India. The conical passive solar still was fabricated and collection of data was done on 13th March 2024 for the atmospheric conditions of Greater Noida, India. A fair agreement was found between experimental and analytical values. Concludingly, values of correlation coefficient for water temperature, condensing surface temperature and freshwater were reported as 0.996, 0.992, and 0.993 respectively. The same code for the conical solar still part has been used in the proposed analysis. Further, Mishra et al.[65] experimented with evacuated tubular collectors arranged in series ($N = 6$) for the climatic condition of Ghaziabad, India. The setup was fabricated, and collection of data was done on 16th March 2016 for the climatic condition of Ghaziabad, India. A good agreement was obtained between analytical and experimental values. Concludingly, the value of correlation coefficient for temperature at the outlet of 6th collector was reported as 0.981. The same code for the NETC part has been used in the proposed analysis.

TABLE 2 Considered climatic situations for the present work.

Type of climatic situation	Description	Remarks
a_1	Clear day (blue sky)	$\gamma \le 0.25$ and $S \ge 9\,h$
b_1	Hazy day (fully)	$0.25 \le \gamma \le 0.50$ and $7\,h \le S \le 9\,h$
c_1	Hazy and cloudy (partially)	$0.50 \le \gamma \le 0.75$ and $5\,h \le S \le 7\,h$
d_1	Cloudy day (fully)	$\gamma \ge 0.75$ and $S \le 5\,h$

Hence, the already validated thermal model for NETC-CSS has been applied for the present analysis.

3 | ANALYSIS

To investigate the efficacy of NETC-CSS, the environmental conditions selected for different month of each calendar year are described in Table 2 which follow the conditions taken by Singh and Tiwari.[66] In this context S denotes the duration of sunlight and γ represents the relative fraction of diffuse radiation compared to the total global radiation received during the day.

3.1 | Exergy analysis

The law of thermodynamics (II) is applied to assess exergy of NETC-CSS. The computation of exergy gained on hourly basis from NETC-CSS is carried out as follows[67]:

$$\dot{E}x_{hourly} = h_{ewgE} \times \frac{A_b}{4} \times \left[(T_w - T_{gE}) - (T_a + 273) \times \ln\frac{(T_w + 273)}{(T_{gE} + 273)}\right] + h_{ewgW} \times \frac{A_b}{4} \times \left[(T_w - T_{gW}) - (T_a + 273) \times \ln\frac{(T_w + 273)}{(T_{gW} + 273)}\right] + h_{ewgS} \times \frac{A_b}{4} \times \left[(T_w - T_{gS}) - (T_a + 273) \times \ln\frac{(T_w + 273)}{(T_{gS} + 273)}\right] + h_{ewgN} \times \frac{A_b}{4} \times \left[(T_w - T_{gN}) - (T_a + 273) \times \ln\frac{(T_w + 273)}{(T_{gN} + 273)}\right]. \quad (23)$$

For the weather condition of type "a_1", the calculation of gain in exergy per day is carried out by adding per hour values of exergy for hours' time of 24. In a similar way, the per day gain in value of exergy of type c_1, b_1, and d_1 weather situation can be calculated. Now, for determining the exergy gain for a whole month for weather condition of type "a_1", the daily gain in exergy for type "a_1" will be added up for a whole month and likewise the per month gain in exergy for type c_1, d_1, and b_1 weather conditions can be calculated. To calculate the monthly gain in exergy in total, the gain in exergy for type a_1, b_1, c_1, and d_1 weather conditions per month will be added. The gain in exergy for a complete year for NETC-CSS will be found by the addition of per month gain in exergy for a total of 12 months.

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.2 | Energy analysis

With the implementation of 1st thermodynamic law, the examination of energy in NETC-CSS will be carried out. To calculate the per hour energy, Equation (24) will be implemented.

$$\text{Hourly energy} = \frac{(\text{Hourly freshwater output}) \times (\text{Latent heat})}{3600}. \quad (24)$$

In order to evaluate the per year energy, the process will be followed as we have done in case of calculation of per year exergy.

3.3 | Efficiency analysis for NETC-CSS

A very important aspect of solar still is the time, material, energy consumed during the operation and before it (during the time of installation) which is also to be reduced. Meanwhile, the yield of fresh water is to be increased. Efficiency also represents how effectively the resources are utilized without leaving any waste that harms the atmosphere. The analysis of thermal efficiency was performed in accordance with the first thermodynamics law while the second thermodynamics law was implemented to examine the exergy efficiency. The various types of efficiencies evaluated for the considered case of NETC-CSS is as follows:

3.3.1 | Thermal efficiency (η_{therm}) for NETC-CSS

The value of η_{therm} on hourly, daily and yearly basis for NETC-CSS can be estimated as follows:

$$\eta_{\text{thermal, hourly, NETC}-\text{CSS}} = \frac{\frac{((\text{Hourly freshwater output}) \times (\text{Latent heat}))}{3600}}{((\text{Basin area}) \times (\text{Hourly solar intensity})) + \text{Hourly pump work}} \times 100, \quad (25)$$

$$\eta_{\text{thermal, daily, NETC}-\text{CSS}} = \frac{\frac{((\text{Daily freshwater output}) \times (\text{Latent heat}))}{3600}}{((\text{Basin area}) \times (\text{Daily solar intensity})) + \text{Daily pump work}} \times 100, \quad (26)$$

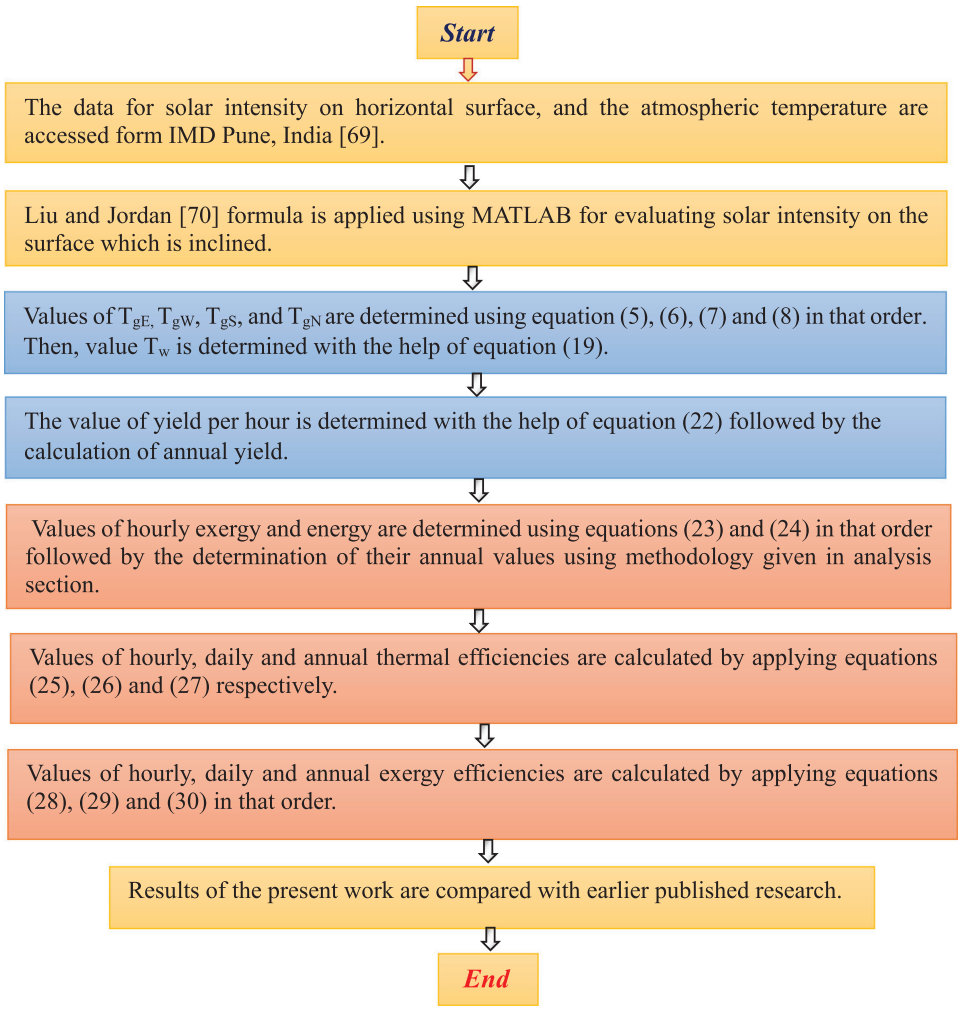


FIGURE 2 Flow chart representing the computation of performance parameters of N-ETC-CSS.

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\eta_{\text{thermal, yearly, NETC-CSS}} = \frac{\frac{((\text{Annual freshwater output}) \times (\text{Latent heat}))}{3600}}{((\text{Basin area}) \times (\text{Annual solar intensity})) + \text{Annual pump work}} \times 100. \quad (27)$$

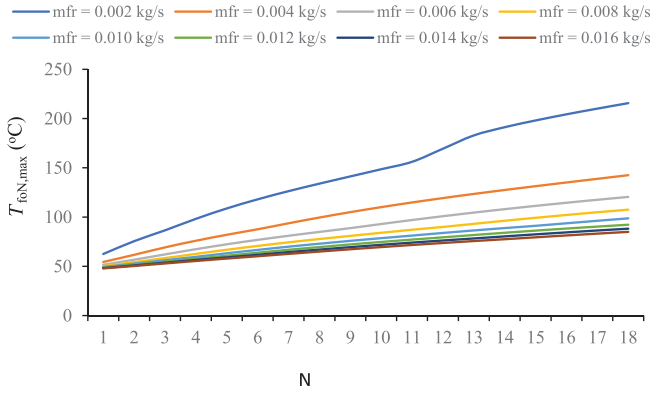


FIGURE 3 Variation of $T_{foN,max}$ with N and mfr for NETC-CSS for a typical day in May.

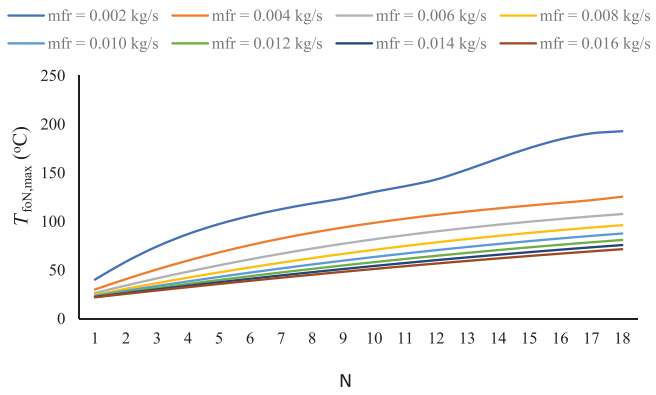


FIGURE 4 Variation of $T_{foN,max}$ with N and mfr for NETC-CSS for a typical day in December.

3.4 | Exergy (η_{exer}) efficiency

The value of η_{exer} on hourly, daily and yearly basis for NETC-CSS can be computed as follows:

$$\eta_{\text{exergy, hourly, NETC-CSS}} = \frac{(\text{Hourly exergy output})}{[(\text{Basin area} \times 0.93 \times \text{Hourly solar intensity}) + \text{Horly pump work}]} \times 100, \quad (28)$$

$$\eta_{\text{exergy, daily, NETC-CSS}} = \frac{(\text{Daily exergy output})}{[(\text{Basin area} \times 0.93 \times \text{Daily solar intensity}) + \text{Daily pump work}]} \times 100, \quad (29)$$

$$\eta_{\text{exergy, yearly, NETC-CSS}} = \frac{(\text{yearly exergy output})}{[(\text{Basin area} \times 0.93 \times \text{Yearly solar intensity}) + \text{Yearly pump work}]} \times 100. \quad (30)$$

The factor 0.933 alters solar energy into the consistent exergy, which can be obtained as.[68]

$$\text{Exergy input} = (\text{Area} \times \text{Solar intensity}) \left[1 - \frac{4}{3}\left(\frac{(\text{Atmospheric temperature})}{(\text{Sun temperature})}\right) + \frac{1}{3}\left(\frac{(\text{Atmospheric temperature})}{(\text{Sun temperature})}\right)^4\right]. \quad (31)$$

4 | METHODOLOGY

The steps included in the procedure to find all the necessary parameters contained in NETC-CSS are described below.

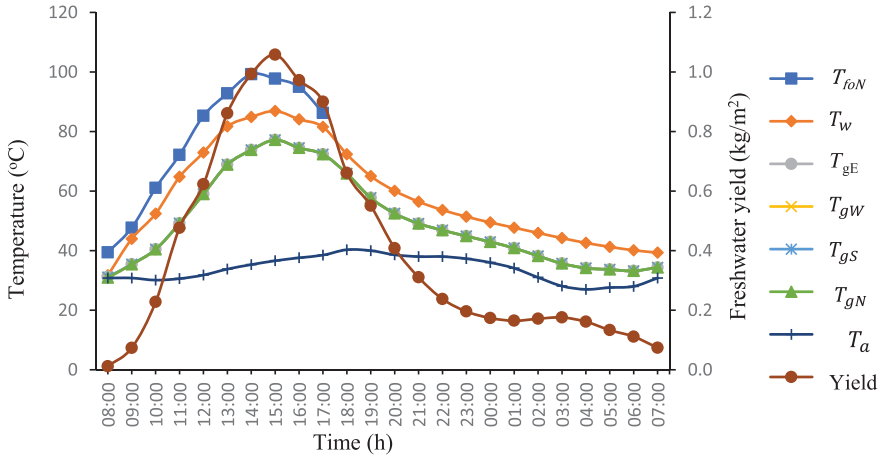


FIGURE 5 The hourly dissimilarities of different temperatures and freshwater yielding for an archetypal day in May for NETC-CSS.

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.1 | Step I

After receiving the data provided by IMD,[69] Pune, India, all the necessary equations are applied to calculate the solar intensity on inclined surface as per Liu and Jordan formula[70] through developing a computation code fed to MATLAB.

4.2 | STEP II

In this step, several temperatures of condensing surface are determined. With the help of Equation (5), T_{gE} is calculated. In a similar manner, T_{gW}, T_{gS}, and T_{gN} are calculated using Equations (6)–(8) respectively. After that, using Equation (19), T_w is determined.

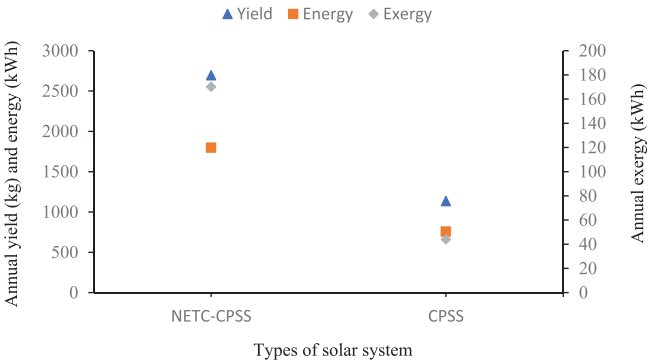


FIGURE 6 Performance comparison of NETC-CSS with CPSS by incorporating yield, energy, and exergy. CPSS, conical passive solar still.

4.3 | Step III

In this step, Equations (22)–(24) are utilized for determining production of fresh water, associated net exergy, and net energy output respectively. This follows the calculation of per day, per month and per year production of fresh water, associated exergy and output energy.

4.4 | Step IV

Here, Equations (25)–(27) are implemented to get various thermal efficiencies. Further, Equations (28)–(30) are implemented for getting various exergy efficiencies.

4.5 | Step V

Results of NETC-CSS are compared with results of earlier published research considering overall energy, overall exergy, and efficiency.

The flow chart representing the computation of performance parameters of NETC-CSS is shown in Figure 2 for better understanding the methodology followed in the present analysis.

5 | RESULTS AND DISCUSSION

The representation shown in Figures 3–6 and data represented in Tables 2–5 are the results of the computations carried out in MATLAB with the input data, equations, and concept. The analysis was carried out for four different types of weather conditions in a given Calander year.

TABLE 3 Computation of yearly yield for NETC-CSS.

	Type a_1			Type b_1			Type c_1			Type d_1			Monthly yield
Month	Daily yield	N	Monthly yield	Daily yield	n	Monthly yield	Daily yield	n	Monthly yield	Daily yield	n	Monthly yield	(kg)
Jan	9.61	3	28.83	9.36	8	74.86	5.89	11	64.74	2.89	9	26.02	194.46
Feb	9.90	3	29.70	10.12	4	40.47	6.00	12	71.97	2.77	9	24.97	167.12
Mar	10.33	5	51.67	11.07	6	66.44	8.16	12	97.91	6.88	8	55.01	271.03
Apr	10.39	4	41.58	10.72	7	75.04	8.26	14	115.60	8.39	5	41.94	274.16
May	9.73	4	38.90	9.72	9	87.44	9.10	12	109.14	8.04	6	48.26	283.75
Jun	10.20	3	30.59	10.34	4	41.37	9.33	14	130.56	7.30	9	65.72	268.24
Jul	9.14	2	18.27	9.22	3	27.67	8.19	10	81.90	6.99	17	118.90	246.73
Aug	8.86	2	17.72	9.05	3	27.16	7.78	7	54.49	6.54	19	124.22	223.59
Sep	8.82	7	61.77	8.84	3	26.53	7.75	10	77.53	6.18	10	61.78	227.61
Oct	7.48	5	37.40	7.07	10	70.74	5.92	13	77.00	4.47	3	13.40	198.54
Nov	7.15	6	42.88	6.40	10	64.03	3.78	12	45.30	3.73	2	7.45	159.67
Dec	8.69	3	26.06	8.11	7	56.78	5.78	13	75.19	2.75	8	21.96	180.00
Yearly yield (kg)													2694.91

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Courtesy to IMD Pune, India, for providing necessary information such as intensity of solar radiation falling on a horizontal plane and temperature of environment. With the valuable information provided by IMD Pune,[69] India, the data for solar radiation falling on surface which is horizontal was converted to solar radiation falling on the surface which is inclined by implementing formula given by Liu and Jordan.[70]

Figures 3 and 4 depict how the highest temperature reached by the fluid at Nth ETC ($T_{foN,max}$) changes in relation to N and mfr for NETC-CSS for a typical day in May and December in that order. From Figure 3, it is noticed that the increase in mfr beyond 0.008 kg/s is not significant because curves are overlapping. The increase in mfr value is considered because mfr is directly proportional to the useful heat hain by ETC. Further, it is noticed that the value of $T_{foN,max}$ becomes higher than 100°C for $N > 15$. Hence, values of N and mfr are taken as 15 and 0.008 kg/s. These values of N and mfr are also true for the month of December shown in Figure 4.

The hourly alteration in yield of freshwater and the temperatures like temperature of fluid at the exit of Nth ETC, temperature of water, temperature of condensing surfaces and environment temperature are shown in Figure 5. Upon giving a clear observation to Figure 5, it has been noticed that the temperature of water is greater than

TABLE 4 Computation of yearly energy for NETC-CSS.

Month	Type a_1			Type b_1			Type c_1			Type d_1			Energy per month
	Daily energy	N	Monthly energy	Daily energy	n	Monthly energy	Daily energy	N	Monthly energy	Daily energy	n	Monthly energy	(kWh)
Jan	6.41	3	19.22	6.24	8	49.91	3.92	11	43.16	1.93	9	17.35	129.64
Feb	6.60	3	19.80	6.75	4	26.98	4.00	12	47.98	1.85	9	16.65	111.41
Mar	6.89	5	34.44	7.38	6	44.29	5.44	12	65.27	4.58	8	36.68	180.69
Apr	6.93	4	27.72	7.15	7	50.03	5.50	14	77.07	5.59	5	27.96	182.78
May	6.48	4	25.93	6.48	9	58.29	6.06	12	72.76	5.36	6	32.18	189.17
Jun	6.80	3	20.39	6.89	4	27.58	6.22	14	87.04	4.87	9	43.82	178.83
Jul	6.09	2	12.18	6.15	3	18.45	5.46	10	54.60	4.66	17	79.26	164.49
Aug	5.91	2	11.81	6.04	3	18.11	5.19	7	36.33	4.36	19	82.82	149.06
Sep	5.88	7	41.18	5.89	3	17.68	5.17	10	51.69	4.12	10	41.19	151.74
Oct	4.99	5	24.94	4.72	10	47.16	3.95	13	51.33	2.98	3	8.93	132.36
Nov	4.76	6	28.58	4.27	10	42.69	2.52	12	30.20	2.48	2	4.97	106.45
Dec	5.79	3	17.37	5.41	7	37.86	3.86	13	50.13	1.83	8	14.64	120.00
Yearly energy (kWh)													1796.61

TABLE 5 Computation of yearly exergy for NETC-CSS.

Month	Type a_1			Type b_1			Type c_1			Type d_1			Exergy per month
	Daily exergy	N	Monthly exergy	Daily exergy	N	Monthly exergy	Daily exergy	N	Monthly exergy	Daily exergy	n	Monthly exergy	(kWh)
Jan	0.79	3	2.37	0.73	8	5.86	0.28	11	3.05	0.08	9	0.76	12.04
Feb	0.86	3	2.57	0.89	4	3.55	0.31	12	3.73	0.09	9	0.81	10.67
Mar	0.86	5	4.29	0.97	6	5.80	0.46	12	5.50	0.35	8	2.76	18.35
Apr	0.81	4	3.24	0.86	7	6.02	0.48	14	6.76	0.50	5	2.50	18.51
May	0.69	4	2.74	0.68	9	6.13	0.60	12	7.22	0.45	6	2.68	18.77
Jun	0.74	3	2.21	0.78	4	3.13	0.62	14	8.75	0.37	9	3.29	17.38
Jul	0.67	2	1.34	0.69	3	2.06	0.53	10	5.31	0.33	17	5.55	14.25
Aug	0.67	2	1.35	0.73	3	2.18	0.51	7	3.59	0.33	19	6.23	13.34
Sep	0.67	7	4.68	0.65	3	1.95	0.51	10	5.11	0.28	10	2.84	14.58
Oct	0.55	5	2.74	0.48	10	4.78	0.32	13	4.11	0.18	3	0.53	12.16
Nov	0.53	6	3.15	0.37	10	3.72	0.15	12	1.81	0.15	2	0.29	8.97
Dec	0.71	3	2.12	0.59	7	4.13	0.32	13	4.15	0.10	8	0.77	11.16
Yearly exergy (kWh)													170.19

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

temperature of condensing surface as expected. This is the difference required to get the freshwater produced from NETC-CSS. In this analysis, useful parameters were evaluated by utilizing mean value of intensity of solar radiation and temperature of atmosphere. Also, the change in temperature of water and change in condensing surface temperature with respect to the change in intensity of solar radiation were also evaluated using useful parameters. Intensity of solar radiation was noted to be highest at 12:00 noon, while the temperature of basin ceased to be highest at around 2′o clock in the afternoon. This is obvious phenomena because first water takes solar radiation and then starts getting heated. It takes some time for water to reach the state of maximum temperature.

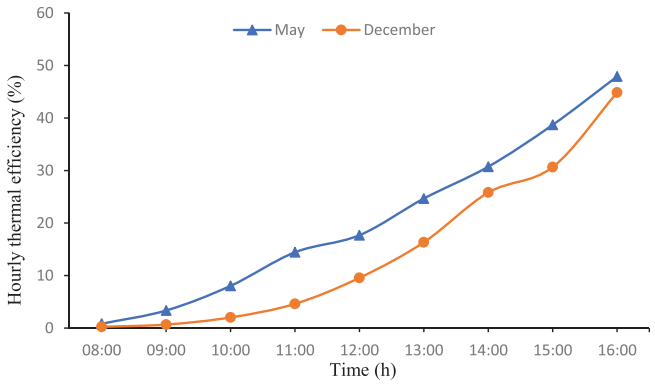


FIGURE 7 Variation of hourly thermal efficiency for NETC-CSS considering a typical day of May.

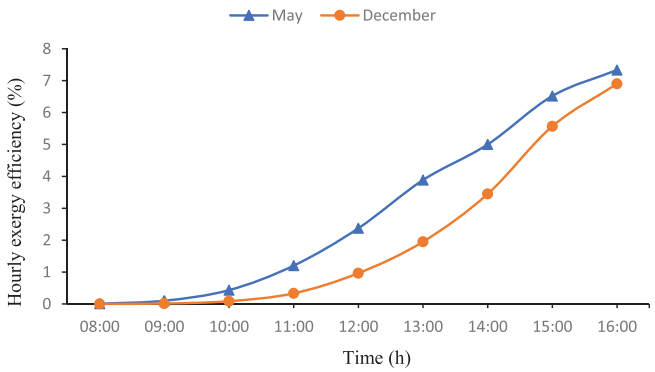


FIGURE 8 Variation of hourly exergy efficiency for NETC-CSS considering a typical day of May.

From the data tabulated in Table 3, the evaluation of freshwater output for the NETC-CSS has been performed. The technique implemented for the analysis has already been represented in the analysis section. It is to be noted that during the month of May, the output of freshwater is highest among all the months whereas it is noted to be lowest during month of December. In May, the weather tends to be warm and humid, with sunlight lingering for longer durations than in other months of the year. Due to the occurrence of winter season in month of December, the presence of sunlight during daytime reduces drastically leading to reduction in supply of heat by the sun to the NETC-CSS resulting in comparatively lower yield of freshwater.

From the data tabulated in Table 4, the evaluation of energy output for the NETC-CSS has been performed. The technique implemented for the analysis has already been represented in the analysis section. It is to be noted that during the month of May, the energy output is highest among all the months whereas it is noted to be lowest in the month of December. This occurs because of the similar variation in value of freshwater yielding. From the data tabulated in Table 5, the evaluation of exergy output for the NETC-CSS has been performed.

TABLE 6 Evaluation of yearly thermal and exergy efficiencies for NETC-CSS.

Parameter	Value
Annual energy output (kWh)	1796.61
Annual solar energy falling on the system (kWh)	4631.42
Annual thermal efficiency (%)	38.79
Annual energy output (kWh)	170.19
Annual solar energy falling on the system (kWh)	4321.12
Annual thermal efficiency (%)	3.94

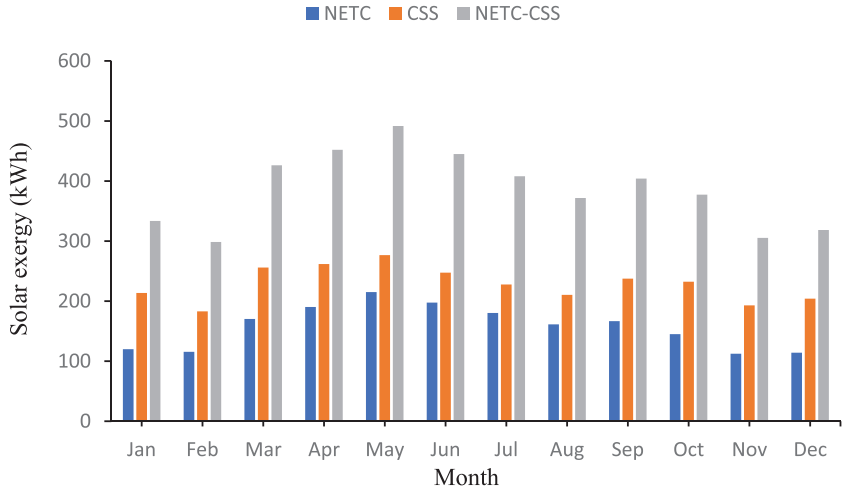


FIGURE 9 Monthly distribution of solar energy falling on the system.

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 7 Comparison of results with earlier published research considering yearly energy exergy and efficiency.

System	Energy (kWh)	Exergy (kWh)	Thermal efficiency (%)	Exergy efficiency (%)
ETC-CSS (Present system)	1796.61	170.19	38.79	3.94
CPSS[51,52]	755.71	44.01	40.02	2.5
Solar still coupled with parabolic trough collector (PTC-SS)[71]	1527.04	233.94	21.53	1.08
Solar still with wire mess in the basin[71]	1602.30	102.00	34.76	1.42
PTC-SS with wire mesh in the basin[71]	861.00	264.57	22.31	1.19
SS with sand in the basin[71]	931.00	111.11	38.27	1.58
PTC-SS with sand in the basin[71]	1728.79	295.47	24.13	1.33
Modified single slope SS[72]	574.37	72.96	15.08	1.76
Modified double slope SS[72]	1034.32	66.32	17.48	1.16
Modified single slope multiwick SS[72]	781.35	100.70	23.93%	2.58
Modified double slope multiwick SS[72]	1722.66	97.36	28.78%	2.23

The technique implemented for the analysis has already been represented in the analysis section. It is to be noted that during the month of May, the exergy output is highest among all the months whereas it is noted to be lowest in the month of December. The weather during month of May is usually hot and humid and the presence of sunlight remains for longer time as compared to other months in a year. Due to the occurrence of winter season in month of December, the presence of sunlight during daytime reduces drastically leading to reduction in supply of heat by the sun to the NETC-CSS resulting in comparatively lower exergy output.

The comparison of results of NETC-CSS with earlier published results of CPSS by incorporating freshwater yield, annual energy, and annual exergy are rendered as Figure 6. It is seen that the value of freshwater yield, annual energy, and annual exergy for NETC-CSS are higher by 57.94%, 57.94%, and 74.14% in that order. The freshwater yield, annual energy and annual exergy are higher for NETC-CSS than CPSS because NETCs present in NETC-CSS provides heat to the basin. Further, the percentage increase in exergy is higher because heat is accessible at elevated temperatures in situations where the case of NETC-CSS is more than the temperature of heat in CPSS because collectors provide heat which raises the temperature.

The variation of hourly thermal and exergy efficiencies of NETC-CSS is shown in Figures 7 and 8 in that order. The values of both thermal and exergy efficiencies are higher for May than their corresponding values in December. It occurs because heat is available at higher temperature in May. Further, exergy efficiency is much lower than thermal efficacies. Due to exergy's representation of energy's quality, its value is often substantially lower than that of energy. Value of daily thermal efficiency is computed using Equation (26). Its value is obtained as 38.48% for an archetypal day in May and 37.69% for an archetypal day in December. Further, Value of daily exergy efficiency is computed using Equation (29). Its value is obtained as 4.09% for an archetypal day in May and 3.76% for an archetypal day in December. Values of thermal and exergy efficiencies are higher for May than their corresponding values in December probably due to the fact that heat is available at higher temperature in May.

The distribution of monthly solar energy falling on the system is shown in Figure 9. The value of corresponding exergy coming to the system can be evaluated as the product of solar energy falling on the system and 0.933.[67] Further, values of annual thermal and exergy efficiencies are computed using Equations (27) and (30) respectively. Table 6 presents the computation of annual thermal and exergy efficiencies for NETC-CSS. Values of annual thermal and exergy efficiencies for NETC-CSS are obtained as 38.79% and 3.94%, respectively.

The comparison of results of present system with earlier published research is shown in Table 7 considering yearly energy, exergy thermal efficiency, and exergy efficiency. It is observed form Table 7 that values of overall energy, exergy, and exergy efficiency are higher for NETC-CSS than corresponding values for CPSS because heat is provided to the basin by NETCs. Further, thermal efficiency is lower because thermal efficiency of NETC-CSS depends on thermal efficiencies of both CSS and NETCs. One can further observe that values of annual energy, thermal efficiency, and exergy efficiency are higher for NETC-CSS than other solar stills under consideration probably because the shading effect is less in the case of NETC-CSS as compared to other solar stills. However, annual exergy output of NETC-CSS is higher as compared to other solar stills under consideration except PTC-SS, PTC-SS with wire mesh in the basin and PTC-SS with sand in the basin. The highest increase in annual energy, annual exergy, thermal efficiency, and exergy efficiency for NETC-CSS is 68.03% than the modified single slope SS, 74.14% than CPSS, 61.12% than the modified single slope SS, and 72.59% than PTC-SS respectively.

6 | CONCLUSIONS

Fundamental equations are derived for NETC-CSS followed by the computation of performance parameters using MATLAB code. Results of NETC-CSS are compared with results of earlier published research considering annual energy, exergy, thermal efficiency and exergy efficiency. Based on the analysis, conclusions are drawn which can be summarized as follows:

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

i. Annual freshwater yield, energy, exergy, thermal efficiency and exergy efficiency for NETC-CSS are computed to be 2694.91 kg, 1796.61, and 170.19 kWh, 38.79%, and 3.94%, respectively.
ii. Values of annual freshwater yield, energy, exergy and exergy efficiency are higher respectively by 57.94%, 57.94%, 74.14% and 36.55% for NETC-CSS than CPSS. However, the thermal efficiency for NETC-CSS is lower by 3.12% than CPSS.
iii. Value of monthly yield is highest in May and the lowest in December. The hourly thermal and exergy efficiencies of the suggested system are greater in May compared to their respective values in December. The hourly exergy efficiency is significantly lower than the hourly thermal efficiency.
iv. A comparison of results of present system with earlier published research represents that the increase in annual exergy, energy, thermal efficiency and exergy efficiency is 68.03% than modified solar still, 74.14% than conventional conical solar still, 61.12% than modified solar still and 72.59% than solar still with parabolic trough collector.

Some limitations are there as revealed in assumptions. These limitations may cause small variation in results which are not going to change conclusions.

Integrating the NETC-CSS system into small businesses can significantly reduce costs and enhance sustainability, especially for those in food processing, laboratories, cosmetics, and electronics applications. For manufacturing setups, the NETC-CSS system can be incorporated into in-house water purification stages, benefiting industries like pharmaceuticals, cosmetics, and automotive manufacturing. This integration supports global sustainability goals and promotes eco-friendly practices, thereby enhancing the industrial relevance and appeal of the technology.

NOMENCLATURE

Symbols

A_g	area of condensing cover, m^2
A_b	area of basin, m^2
a_1	clear day (blue sky) climatic situation
b_1	hazy day (fully) climatic situation
c_1	hazy and cloudy (partially) climatic situation
C_w	specific heat capacity of water, J/kg-K
d_1	cloudy day (fully) climatic situation
EHTC	equivalent heat transfer coefficient
$\dot{E}x_{hourly}$	hourly exergy output from conical PSS, kWh
HTC	heat transfer coefficient
h_{rwgE}	radiative HTC from water surface to condensing cover facing east, W/m^2-K
h_{rwgW}	radiative HTC from water surface to condensing cover facing west, W/m^2-K
h_{rwgS}	radiative HTC from water surface to condensing cover facing south, W/m^2-K
h_{rwgN}	radiative HTC from water surface to condensing cover facing north, W/m^2-K
h_{cwgE}	convective HTC from water surface to condensing cover facing east, W/m^2-K
h_{cwgW}	convective HTC from water surface to condensing cover facing west, W/m^2-K
h_{cwgS}	convective HTC from water surface to condensing cover facing south, W/m^2-K
h_{cwgN}	convective HTC from water surface to condensing cover facing north, W/m^2-K
h_{ewgE}	evaporative HTC from water surface to condensing cover facing east, W/m^2-K
h_{ewgW}	evaporative HTC from water surface to condensing cover facing west, W/m^2-K
h_{ewgS}	evaporative HTC from water surface to condensing cover facing south, W/m^2-K
h_{ewgN}	evaporative HTC from water surface to condensing cover facing north, W/m^2-K
h_{1wE}	equivalent heat transfer coefficient between water and east facing glass surfaces
h_{1wW}	equivalent heat transfer coefficient between water and west facing glass surfaces
h_{1wS}	equivalent heat transfer coefficient between water and south facing glass surfaces
h_{1wN}	equivalent heat transfer coefficient between water and north facing glass surfaces
h_{1gN}	EHTC from north facing condensing surface to the surrounding
h_{1gS}	EHTC from south facing condensing surface to the surrounding
h_{1gE}	EHTC from east facing condensing surface to the surrounding
h_{1gW}	EHTC from west facing condensing surface to the surrounding
h_{bw}	HTC from basin liner to water
h_{ba}	HTC from basin liner to ambient
I_{SE}	solar radiation impinging on glass cover facing east, W/m^2
I_{SW}	solar radiation impinging on glass cover facing west, W/m^2
I_{SN}	solar radiation impinging on glass cover facing north, W/m^2
I_{SS}	solar radiation impinging on glass cover facing south, W/m^2
h_{EW}	radiative HTC between east and west condensing surfaces, W/m^2-K
h_{NS}	radiative HTC between south and north condensing surfaces, W/m^2-K
ln	natural log
M_w	mass of water in the basin, kg
mfr	mass flow rate of fuid/water, kg/s
n	number of clear days
NETC-CSS	conical solar still integrated with N identical evacuated collectors
t	time, s
PF_c	penalty factor due to the glass covers for the glazed portion
PF_1	penalty factor first, dimensionless
PF_2	penalty factor second, dimensionless

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$\dot{Q}_{u,N}$	useful energy gain for N identical collector connected in series, kWh
R_{o1}	inner radius of outer glass tube of evacuated coaxial glass tube, m
R_{o2}	outer radius of outer glass tube of evacuated coaxial glass tube, m
R_{i2}	outer radius of inner glass tube of evacuated coaxial glass tube, m
r	radius of copper tube
R_g	reflectivity of glass
R_w	reflectivity of water
T_w	temperature of water in the basin, °C
T_{wo}	temperature of water in the basin at $t = 0$, °C
T_a	ambient temperature, °C
T_b	temperature of basin liner
T_{gE}	condensing cover temperature facing east, °C
T_{gW}	condensing cover temperature facing west, °C
T_{gS}	condensing cover temperature facing south, °C
T_{gN}	condensing cover temperature facing north, °C
$T_{foN.max}$	maximum collector at the outlet of N^{th} collector
L	latent heat
S	number of sunshine hours
U_L	overall heat transfer coefficient
V	velocity of air

Greek letters

$\dot{\alpha_g}$	effective absorptivity of condensing cover, fraction
α_g	absorptivity of condensing cover, fraction
$\dot{\alpha_w}$	effective absorptivity of water mass, fraction
α_w	absorptivity of water mass, fraction
$\dot{\alpha_b}$	effective absorptivity of basin liner, fraction
α_b	absorptivity of basin liner, fraction
γ	ratio of daily diffuse to daily global irradiation
η	efficiency

Abbreviations

CPSS	conical passive solar still
EHTC	equivalent heat transfer coefficient, W/m^2-K
HTC	heat transfer coefficient, W/m^2-K

Suffix

b	basin
g	glass
E	east
W	west
S	south
N	north

AUTHOR CONTRIBUTIONS

Abhishek Kumar: Writing – review and editing; software; data curation; conceptualization; investigation; methodology. **Rajesh Kumar:** Formal analysis; project administration; writing – review and editing; supervision.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Abhishek Kumar https://orcid.org/0009-0003-4773-3869

REFERENCES

1. Rai SN, Tiwari GN. Single basin solar still coupled with flat plate collector. *Energ Conver Manage*. 1983;23:145-149.
2. Singh DB. Sensitivity investigation of N similar evacuated tubular collectors having series connection by incorporating exergy analysis. *Int J Exergy*. 2022;38(3):367-381.
3. Raturi A, Patel RV, Singh DB. A sensitivity study for n similar partly enclosed with photovoltaic thermal flat plate collectors having series connection. *J Thermal Eng*. 2023;9(1):33-44.
4. Omara ZM, Eltawil MA, ElNashar EA. A new hybrid desalination system using wicks/solar still and evacuated solar water heater. *Desalination*. 2013;325:56-64.
5. Singh RV, Kumar S, Hasan MM, Khan ME, Tiwari GN. Performance of a solar still integrated with evacuated tube collector in natural mode. *Desalination*. 2013;318:25-33.
6. Esen M, Esen H. Experimental investigation of a two-phase closed thermosyphon solar water heater. *Solar Energy*. 2005;79(5):459-468.
7. Esen H, Inalli M, Esen M, Pihtili K. Energy and exergy analysis of a ground-coupled heat pump system with two horizontal ground heat exchangers. *Build Environ*. 2007;42(10):3606-3615.
8. Esen M. Thermal performance of a solar cooker integrated vacuum-tube collector with heat pipes containing different refrigerants. *Solar Energy*. 2004;76(6):751-757.
9. Esen M, Hazar H, Esen H. Experimental investigation of a solar cooker using collector with heat pipes. *Sigma J Eng Nat Sci*. 2005;23(1):114-122.
10. Kumar S, Dubey A, Tiwari GN. A solar still augmented with an evacuated tube collector in forced mode. *Desalination*. 2014;347:15-24.
11. Shafii MB, Shahmohamadi M, Faegh MS. Examination of a novel solar still equipped with evacuated tube collectors and thermoelectric modules. *Desalination*. 2016;382:21-27.
12. Sharshir SW, Peng G, Yang N, El-Samadony MOA, Kabeel AE. A continuous desalination system using humidification—dehumidification and a solar still with an evacuated solar water heater. *Appl Therm Eng*. 2016;104:734-742.
13. Singh DB. Exergoeconomic and enviroeconomic analyses of N identical photovoltaic thermal integrated double slope solar still. *Int J Exergy (IJEX)*. 2017;23(4):347-366.
14. Singh DB, Tiwari GN. Energy, exergy and cost analyses of N identical evacuated tubular collectors integrated basin type solar stills: a comparative study. *Solar Energy*. 2017;155:829-846.
15. Singh DB, Al-Helal IM. Energy metrics analysis of N identical evacuated tubular collectors integrated double slope solar still. *Desalination*. 2018;432:10-22.
16. Singh DB. Energy metrics analysis of N identical evacuated tubular collectors integrated single slope solar still. *Energy*. 2018;148:546-560.
17. Singh AK, Singh DB, Dwivedi VK, Kumar N, Yadav JK. A review of performance enhancement in solar desalination systems with the application of Nanofluids. *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*. IEEE Xplore; 2018:814-819. doi:10.1109/ICACCCN.2018.8748441
18. Singh DB, Singh AK, Navneet K, Dwivedi VK, Yadav JK, Singh G. Performance analysis of special design Single Basin passive solar distillation systems: a comprehensive review. In: Prasad A, Gupta S, Tyagi R, eds. *Advances in Engineering Design*. Lecture Notes in Mechanical Engineering. Springer; 2019. doi:10.1007/978-981-13-6469-3_27

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19. Sharshir SW, Kandeal AW, Ismail M, Abdelaziz GB, Kabeel AE, Yang N. Augmentation of a pyramid solar still performance using evacuated tubes and nanofluid: experimental approach. *Appl Therm Eng*. 2019;160:113997.
20. Singh DB, Raturi A, Kumar N, Nirala A, Singh AK, Tiwari S. Effect of flow of fluid mass per unit time on life cycle conversion efficiency of single slope solar desalination unit coupled with N identical evacuated tubular collectors. *Mater Today: Proceed*. 2020;28(4):2096-2100.
21. Kumar R, Singh DB, Kumar N, Nirala A, Tiwari GN. Effect of number of collectors (N) on the environment due to single slope solar desalination unit coupled with N identical evacuated tubular collectors. *Mater Today: Proceed*. 2020;28(4):2161-2165.
22. Panchal H, Hishan SS, Rahim R, Sadasivuni KK. Solar still with evacuated tubes and calcium stones to enhance the yield: an experimental investigation. *Process Saf Environ Prot*. 2020;142:150-155.
23. Sharma SK, Mallick A, Gupta SK, Kumar N, Singh DB, Tiwari GN. Characteristic equation development for double slope solar distiller unit augmented with n identical parabolic concentrator integrated evacuated tubular collectors. *Desalination Water Treat*. 2020;187:178-194.
24. Mevada D, Panchal H, Sadasivuni KK. Investigation on evacuated tubes coupled solar still with condenser and fins: experimental, exergo-economic and exergo-environment analysis. *Case Stud Therm Eng*. 2021;27:101217.
25. Sadeghi G, Nazari S. Retrofitting a thermoelectric-based solar still integrated with an evacuated tube collector utilizing an antibacterial-magnetic hybrid nanofluid. *Desalination*. 2021;500:114871.
26. Dawood MMK, Nabil T, Kabeel AE, Shehata AI, Abdalla AM, Elnaghi BE. Experimental study of productivity progress for a solar still integrated with parabolic trough collectors with a phase change material in the receiver evacuated tubes and in the still. *J Energy Storage*. 2020;32:102007.
27. Prasad H, Kumar P, Yadav RK, Mallick A, Kumar N, Singh DB. Sensitivity analysis of N identical partially covered (50%) PVT compound parabolic concentrator collectors integrated double slope solar distiller uni. *Desalination Water Treat*. 2019;153:54-64.
28. Singh DB. Sensitivity analysis of N identical evacuated tubular collectors integrated double slope solar distiller unit by incorporating the effect of exergy. *Int J Exergy*. 2021;34(4):424-447.
29. Raturi A, Singh DB, Patil PP, Sharma AK. Sensitivity analysis of a solar still of a single slope type included with N similar evacuated tubular collectors having series connection. *Desalination Water Treat*. 2021;234:309-323.
30. Raturi A, Dobriyal R, Sharma RK, Dwivedi A, Singh SP, Singh DB. Sensitivity investigation for solar still of double slope type included with N similar CPC integrated ETC by incorporating yield and efficiency. *Desalination Water Treat*. 2021;244:12-26.
31. Singh AR, Agarwal AB, Singh DB. Sensitivity analysis of double slope solar still having compound parabolic concentrator integrated ETCs by incorporating heat transfer coefficients, exergy gain and exergy efficiency. *Desalination Water Treat*. 2022;276:13-27.
32. Singh DB, Bansal G, Prasad H, Mallick A, Kumar N, Sharma SK. Sensitivity analysis of N undistinguishable photovoltaic thermal compound-parabolic-concentrator collectors (partly covered, 50%) integrated single-slope solar distiller unit. *J Solar Energy Eng*. 2021;143(2):021003.
33. Bharti K, Manwal M, Kishore C, Yadav RK, Tiwari P, Singh DB. Sensitivity analysis of N alike partly covered PVT flat plate collectors integrated double slope solar distiller unit. *Desalination Water Treat*. 2021;211:45-59.
34. Sharshir SW, Kandeal AW, Algazzar AM, Eldesoukey A, El-Samadony MOA, Hussien AA. 4-E analysis of pyramid solar still augmented with external condenser, evacuated tubes, nanofluid and ultrasonic foggers: a comprehensive study. *Process Saf Environ Protect*. 2022;164:408-417.
35. Moghadam H, Samimi M. Effect of condenser geometrical feature on evacuated tube collector basin solar still performance: productivity optimization using a box-Behnken design model. *Desalination*. 2022;542:116092.
36. Gangavathi P, Ramya G, Karuna MS, Jeyapiriya K, Magade PB, Manirathnam D. Influence of nanomaterial coated condensing surface on the productivity of evacuated tube assisted solar still. *Mater Today: Proceed*. 2022;69(3):853-857.
37. Gopi G, Premalatha M, Arthanareeswaran G. Transient mathematical modelling and investigation of radiation and design parameters on the performance of multi-effect solar still integrated with evacuated tube collector. *Energy Convers Manag: X*. 2022;14:100210.
38. Kumar D, Sharma RK, Singh DB. Effects of dissimilarity of water depth on energy and exergy efficiencies and productivity of single slope solar still coupled to evacuated tubular collectors. *Desalination Water Treat*. 2022;278:301-310.
39. Farghaly MB, Alahmadi RN, Sarhan HH, Abdelghany ES. Experimental study of simultaneous effect of evacuated tube collectors coupled with parabolic reflectors on traditional single slope solar still efficiency. *Case Stud Thermal Eng*. 2023;49:103304.
40. Dahab MA, Omara MA, El-Dafrawy MM, Abdelaziz GB, Sharshir SW. Thermo-economic performance enhancement of the hemispherical solar still integrated with various numbers of evacuated tubes. *Thermal Sci Eng Progress*. 2023;42:101922.
41. Nagpal G, Singh DB. Improving exergo-enviro-economic parameters and annual productivity of double slope solar desalting unit by incorporating concentrator integrated evacuated tubular collectors. *Desalination Water Treat*. 2023;306:9-21.
42. Shelake A, Kumbhar D, Sutar K. Desalination using solar stills: a review. *Environ Progress Sustain Energy*. 2023;42(3):e14025.
43. Kumar R, Singh DB, Dewangan A, Singh VK, Kumar N. Performance of evacuated tube solar collector integrated solar desalination unit—a review. *Desalination Water Treat*. 2021;230:92-115.
44. Patel RV, Bharti K, Singh G, Mittal G, Singh DB, Yadav A. Comparative investigation of double slope solar still by incorporating different types of collectors: a mini review. *Mater Today: Proceed*. 2020;38:300-304.
45. Singh G, Singh DB, Kumar S, Bharti K, Chhabra S. A review of inclusion of nanofluids on the attainment of different types of solar collectors. *Mater Today: Proceed*. 2020;38:153-159.
46. Arani RP, Vaithilingam S, Selvaraj K, Afzal A. Experimental studies of solar still with tar-coated blue metal stones: energy and exergy study. *Environ Prog Sustain Energy*. 2023;42(2):e13997. doi:10.1002/ep.13997
47. Kumar A, Tiwari A, Mehla N. Experimental comparison of solar thermal desalination systems based on energy, exergy and economic (3-E) analysis. *Environ Prog Sustain Energy*. 2023;42(2):e13975. doi:10.1002/ep.13975
48. Gad HE, El-Din SS, Hussien AA, Ramzy K. Thermal analysis of a conical solar still performance: an experimental study. *Solar Energy*. 2015;122:900-909.
49. Kabeel AE, Abdelgaied M, MElH A, Arıcı M, Abdel-Aziz MM. Performance enhancement of a conical solar still by optimizing inclination angle. *Solar Energy*. 2023;264:112001.
50. Abdallah S, Aldarabseh SM. Performance of modified conical solar still integrated with continuous volume flowrate. *J Solar Energy Eng*. 2024;146(1):011001.
51. Singh G, Singh PK, Saxena A, Dobriyal R. Kumar N, Singh DB, Exergo-enviro-economic and yearly productivity analyses of conical passive solar still for sustainable solar distillation. *Environ Sci Pollut Res*. 2023;30:104350-104373.
52. Singh G, Singh PK, Saxena A, Kumar N, Singh DB. Investigation of conical passive solar still by incorporating energy metrics, efficiency, and sensitivity analyses for sustainable solar distillation. *J Clean Prod*. 2024;434:139949.
53. Kumar N, Purnachandrakumar D, Saxena A, Tiwari S, Singh DB, Sinhmar H. Development of characteristic equations of conical solar still by incorporating N identical concentrator collectors with/without photovoltaic panel: a comparative investigation. *Desalination*. 2024;582:117621.
54. Singh V, Kumar R, Saxena A, Dobriyal R, Tiwari S, Singh DB. An analytical study on the effect of different photovoltaic technologies on enviro-economic parameter and energy metrics of active solar desalting unit. *Energy*. 2024;294:130851.
55. Singh AR, Agarwal AB, Singh DB. Comparative exergy and economic analyses of single slope solar desalination unit combined with

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

concentrator integrated evacuated collectors. *Desalination and Water Treatment*. 2024;318:100354.
56. Singh AR, Agarwal AB, Singh DB. Comparative energy metrics and annual efficiency analyses of CPC-ETC integrated single slope solar desalting unit. *Environ Progress Sustain Energy*. 2024;43:e14452. doi: 10.1002/ep.14452. (in press).
57. Krishna H, Singh J, Dwivedi VK, Singh DB. Impact of variation in input parameters on energy matrices of dual slope active solar desalting unit. *Environ Progress Sustain Energy*. 2024;43:e14372. doi:10.1002/ep.14372 (in press).
58. Krishna H, Singh J, Dwivedi VK, Singh DB. Effect of dissimilarity of N and mass flow rate on yearly overall energy, exergy, and efficiency of PVT-FPC included dual-slope solar desalting units. *Water Supply*. 2023;23(8):3156-3178.
59. Yan C, Zou Y, Wu Z, Maleki A. Effect of various design configurations and operating conditions for optimization of a wind/solar/hydrogen/fuel cell hybrid microgrid system by a bio-inspired algorithm. *Int J Hydrogen Energy*. 2024;60:378-391.
60. Yang C, Nutakki TUK, Alghassab MA, et al. Optimized integration of solar energy and liquefied natural gas regasification for sustainable urban development: dynamic modeling, data-driven optimization, and case study. *J Clean Prod*. 2024;447:141405.
61. Zheng S, Hai Q, Zhou X, Stanford RJ. A novel multi-generation system for sustainable power, heating, cooling, freshwater, and methane production: thermodynamic, economic, and environmental analysis. *Energy*. 2024;290:130084.
62. Gao J, Wang Z, Li X, Zhou X. Investigation of a novel scheme utilizing solar and geothermal energies, generating power and ammonia: Exergoeconomic and exergoenvironmental analyses and cuckoo search optimization. *Energy*. 2024;298:131344.
63. Xiao D, Liu M, Li L, et al. Model for economic evaluation of closed-loop geothermal systems based on net present value. *Appl Therm Eng*. 2023;231:121008.
64. Mishra RK, Garg V, Tiwari GN. Thermal modeling and development of characteristic equations of evacuated tubular collector (ETC). *Solar Energy*. 2015;116:165-176.
65. Mishra RK, Garg V, Tiwari GN. Energy matrices of U-shaped evacuated tubular collector (ETC) integrated with compound parabolic concentrator (CPC). *Solar Energy*. 2017;153:531-539.
66. Singh HN, Tiwari GN. Monthly performance of passive and active solar stills for different Indian climatic condition. *Desalination*. 2004;168:145-150.
67. Nag PK. *Basic and Applied Thermodynamics*. ISBN 0-07-047338-2. Tata McGraw-Hill; 2004.
68. Petela R. Exergy of undiluted thermal radiation. *Sol Energy*. 2003;74(6):469-488.
69. https://imdpune.gov.in/
70. Liu BYH, Jordan RC. The interrelationship and characteristic distribution of direct, diffuse, and total solar radiation. *Sol Energy*. 1960;4(3):1-19.
71. Hassan H, Yousef MS, Fathy M, Ahmed MS. Assessment of parabolic trough solar collector assisted solar still at various saline water mediums via energy, exergy, exergoeconomic, and enviroeconomic approaches. *Renew Energy*. 2020;155:604-616.
72. Pal P, Patel SK, Singh D. Energy, exergy, energy matrices, exergoeconomic and enviroeconomic assessment of modified solar stills. *Sustain Energy Technol Assess*. 2021;47:101514.

How to cite this article: Kumar A, Kumar R. Annual energy, exergy, and efficiency analyses for conical solar still combined with N number of evacuated collectors. *Environ Prog Sustainable Energy*. 2024;e14527. doi:10.1002/ep.14527

APPENDIX A

$$(AF_R(\alpha\tau))_1 = PF_1\alpha\tau^2 A_R F_R; \quad (A F_R U_L)_1 = (1-K_k)\dot{m}_f c_f; \quad PF_1 = \frac{h_{pf}}{F' h_{pf} + U_{tpa}};$$

$$U_L = \frac{U_{t,pa}.h_{pf}}{F' h_{pf} + U_{t,pa}}; F_R = \frac{\dot{m}_f C_f}{U_L A_R}\left[1 - \exp\left(-\frac{2\pi r' L' U_L}{\dot{m}_f c_f}\right)\right]; \quad K_K = \left(1 - \frac{A_R F_R U_L}{\dot{m}_f c_f}\right),$$

$$h_{pf} = 100\,\mathrm{W m^2 K^{-1}} \text{ and } U_{t,pa} = \left[\frac{Ro_2}{R_{o1} h_i} + \frac{R_{o2}\ln\left(\frac{Ri_2}{Ri_1}\right)}{K_g} + \frac{1}{C_{ev}} + \frac{R_{o2}\ln\left(\frac{Ro_2}{Ro_1}\right)}{K_g} + \frac{1}{h_o}\right]^{-1},$$

$$A_1 = A_g^2 h_{1g}^2 T_a + A_b A_g h_{1g} h_{1wW} T_a + 2A_g^2 h_{1g} h_{EW} T_a + A_g^2 h_{1g} I_{SE}\dot{\alpha}_g + A_b A_g h_{1wW} I_{SE}\dot{\alpha}_g + A_g^2 h_{EW} I_{SE}\dot{\alpha}_g + A_g^2 h_{EW} I_{SW}\dot{\alpha}_g,$$

$$A_2 = A_b A_g h_{1wE} h_{1g} + A_b^2 h_{1wE} h_{1wW} + A_b A_g h_{1wE} h_{EW} + A_b A_g h_{1wW} h_{EW},$$

$$B_1 = A_g^2 h_{1g}^2 T_a + A_b A_g h_{1g} h_{1wE} T_a + 2A_g{}^2 h_{1g} h_{EW} T_a + A_g^2 h_{1g} I_{SE}\dot{\alpha}_g + A_b A_g h_{1wE} I_{SW}\dot{\alpha}_g + A_g^2 h_{EW} I_{SW}\dot{\alpha}_g + A_g^2 h_{EW} I_{SW}\dot{\alpha}_g,$$

$$B_2 = A_b A_g h_{1wW} h_{1g} + A_b^2 h_{1wE} h_{1wW} + A_b A_g h_{1wE} h_{EW} + A_b A_g h_{1wW} h_{EW},$$

$$C_1 = A_g^2 h_{1g}^2 T_a + A_b A_g h_{1g} h_{1wN} T_a - 2A_g{}^2 h_{1g} h_{NS} T_a + A_g^2 h_{1g} I_{SS}\dot{\alpha}_g + A_b A_g h_{1wN} I_{SS}\dot{\alpha}_g - A_g^2 h_{NS} I_{SN}\dot{\alpha}_g - A_g^2 h_{NS} I_{SS}\dot{\alpha}_g,$$

$$C_2 = A_b A_g h_{1wS} h_{1g} + A_b^2 h_{1wN} h_{1wS} - A_b A_g h_{1wN} h_{NS} - A_b A_g h_{1wS} h_{NS},$$

$$D_1 = A_g^2 h_{1g}^2 T_a + A_b A_g h_{1g} h_{1wS} T_a - 2A_g{}^2 h_{1g} h_{NS} T_a + A_g^2 h_{1g} I_{SN}\dot{\alpha}_g + A_b A_g h_{1wS} I_{SN}\dot{\alpha}_g - A_g^2 h_{NS} I_{SN}\dot{\alpha}_g - A_g^2 h_{NS} I_{SS}\dot{\alpha}_g,$$

$$D_2 = A_b A_g h_{1wN} h_{1g} + A_b^2 h_{1wN} h_{1wS} - A_b A_g h_{1wN} h_{NS} - A_b A_g h_{1wS} h_{NS},$$

$$P_1 = A_b A_g h_{1wE} h_{1g} + A_g^2 h_{1g}^2 + A_b^2 h_{1wE} h_{1wW} + A_b A_g h_{1wW} h_{1g} + A_b A_g h_{1wE} h_{EW} + 2A_g^2 h_{1g} h_{EW} + A_b A_g h_{1wW} h_{EW},$$

$$P_2 = A_b A_g h_{1wN} h_{1g} + A_g^2 h_{1g}^2 + A_b^2 h_{1wN} h_{1wS} + A_b A_g h_{1wS} h_{1g} - A_b A_g h_{1wS} h_{NS} - 2A_g^2 h_{1g} h_{NS} - A_b A_g h_{1wN} h_{NS}$$

$$I_e = -\left(A_g^2 h_{1g} I_{SE}\dot{\alpha}_g + A_b A_g h_{1wW} I_{SE}\dot{\alpha}_g + A_g^2 h_{EW} I_{SE}\dot{\alpha}_g + A_g^2 h_{EW} I_{SW}\dot{\alpha}_g\right),$$

$$I_w = -\left(A_g^2 h_{1g} I_{SW}\dot{\alpha}_g + A_b A_g h_{1wE} I_{SW}\dot{\alpha}_g + A_g^2 h_{EW} I_{SE}\dot{\alpha}_g + A_g^2 h_{EW} I_{SW}\dot{\alpha}_g\right),$$

$$I_s = -\left(A_g^2 h_{1g} I_{SS}\dot{\alpha}_g + A_b A_g h_{1wN} I_{SS}\dot{\alpha}_g - A_g^2 h_{NS} I_{SS}\dot{\alpha}_g - A_g^2 h_{NS} I_{SN}\dot{\alpha}_g\right),$$

$$I_n = -\left(A_g^2 h_{1g} I_{SN}\dot{\alpha}_g + A_b A_g h_{1wS} I_{SN}\dot{\alpha}_g - A_g^2 h_{NS} I_{SN}\dot{\alpha}_g - A_g^2 h_{NS} I_{SS}\dot{\alpha}_g\right),$$

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$U_e = \left(A_g^2 h_{1g}^2 + A_b A_g h_{1g} h_{1wW} + 2A_g^2 h_{1g} h_{EW}\right),$$

$$U_w = \left(A_g^2 h_{1g}^2 + A_b A_g h_{1g} h_{1wE} + 2A_g^2 h_{1g} h_{EW}\right),$$

$$U_s = \left(A_g^2 h_{1g}^2 + A_b A_g h_{1g} h_{1wN} - 2A_g^2 h_{1g} h_{NS}\right),$$

$$U_n = \left(A_g^2 h_{1g}^2 + A_b A_g h_{1g} h_{1wS} - 2A_g^2 h_{1g} h_{NS}\right),$$

$$E = \left(\frac{h_{ewgE} I_e + h_{ewgW} I_w}{P_1} + \frac{h_{ewgS} I_s + h_{ewgN} I_n}{P_2}\right)\left(\frac{A_b}{4}\right),$$

$$F_1 = \left(\frac{h_{1wE} U_e + h_{1wW} U_w}{P_1} + \frac{h_{1wS} U_s + h_{1wN} U_n}{P_2}\right)\left(\frac{A_b}{4}\right),$$

$$h_{ewE} = 0.016273\, h_{cwgE}\left[\frac{P_w - P_{gE}}{T_w - T_{gE}}\right], (\text{Cooper}, 1973).$$

$$h_{ewgW} = 0.016273\, h_{cwgW}\left[\frac{P_w - P_{gW}}{T_w - T_{gW}}\right], (\text{Cooper}, 1973).$$

$$h_{ewS} = 0.016273\, h_{cwgS}\left[\frac{P_w - P_{gS}}{T_w - T_{gS}}\right], (\text{Cooper}, 1973).$$

$$h_{ewgN} = 0.016273\, h_{cwgN}\left[\frac{P_w - P_{gN}}{T_w - T_{gN}}\right], (\text{Cooper}, 1973).$$

$$h_{cwgE} = 0.884\left[(T_w - T_{gE}) + \frac{(P_w - P_{gE})(T_w + 273)}{268.9 \times 10^3 - P_w}\right]^{\left(\frac{1}{3}\right)}, (\text{Dunkle}, 1961).$$

$$h_{cwgW} = 0.884\left[(T_w - T_{gW}) + \frac{(P_w - P_{gW})(T_w + 273)}{268.9 \times 10^3 - P_w}\right]^{\left(\frac{1}{3}\right)}, (\text{Dunkle}, 1961).$$

$$h_{cwgS} = 0.884\left[(T_w - T_{gS}) + \frac{(P_w - P_{gS})(T_w + 273)}{268.9 \times 10^3 - P_w}\right]^{\left(\frac{1}{3}\right)}, (\text{Dunkle}, 1961).$$

$$h_{cwgN} = 0.884\left[(T_w - T_{gN}) + \frac{(P_w - P_{gN})(T_w + 273)}{268.9 \times 10^3 - P_w}\right]^{\left(\frac{1}{3}\right)}, (\text{Dunkle}, 1961).$$

$$P_w = \exp\left[25.317 - \frac{5144}{(T_w + 273)}\right]; P_{gE} = exp\left[25.317 - \frac{5144}{(T_{gE} + 273)}\right],$$

$$P_{gW} = \exp\left[25.317 - \frac{5144}{(T_{gW} + 273)}\right]; P_{gS} = \exp\left[25.317 - \frac{5144}{(T_{gS} + 273)}\right],$$

$$P_{gN} = \exp\left[25.317 - \frac{5144}{(T_{gN} + 273)}\right],$$

$$h_{EW} = 0.034 \times 5.67 \times 10^{-8}\left[(T_{gE} + 273)^2 + (T_{gW} + 273)^2\right][T_{gE} + T_{gW} + 546],$$

$$h_{NS} = 0.034 \times 5.67 \times 10^{-8}\left[(T_{gN} + 273)^2 + (T_{gS} + 273)^2\right]\left[T_{gN} + T_{gS} + 546\right].$$

19447450, 0, Downloaded from https://aiche.onlinelibrary.wiley.com/doi/10.1002/ep.14527 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

AIP Conference Proceedings

RESEARCH ARTICLE | DECEMBER 20 2024

Artificial intelligence based broken instrument from root canal and pulp retrieval of tooth using 33kHz ultrasonics signal FREE

Rahul Kumar ✉; Rajiv Kapoor

AIP Conf. Proc. 3217, 020004 (2024)
https://doi.org/10.1063/5.0234443

View Online Export Citation

Articles You May Be Interested In

Ultrasonics for process monitoring and control

J Acoust Soc Am (October 1999)

Robust shallow water reverberation reduction methods based on low-rank and sparsity decomposition

J. Acoust. Soc. Am. (April 2022)

Ultrasonic inspection of fiber suspensions

J Acoust Soc Am (November 1982)

26 December 2024 04:45:45

AIP Publishing

Artificial Intelligence Based Broken Instrument from Root Canal and Pulp Retrieval of tooth using 33Khz Ultrasonics Signal

Rahul Kumar[a] and Rajiv Kapoor [b]

Electronics and Communication Engineering, Delhi Technological University, Delhi, India

[a] *Corresponding author: rahul007ks@gmail.com*
[b]*rajivkapoor@dce.ac.in*

Abstract: Removal of broken pieces of instrument most difficult aspect of root canal treatment (RCT) for endodontists and currently no technique available for real-time localization of broken pieces of instrument inside the root canal and tooth pulp. This paper proposes a novel way of recording the signals from vitro teeth where the 2D-visualization of these signals helps in detecting the broken pieces of instrument inside the canal. The authors used ResNet50 2D CNN architecture as a feature extractor; MLP has been used for broken instrument localization within, top, bottom, and middle positions inside the canal. An acoustic signal of 33 kHz has been used. The recorded datasets were limited to vitro teeth only. For 2D visualization, the current work use LRSD (Low Rank Sparse Matrix Decomposition) of maximum peak change of received signal with respect to transmitted signal along with spectrogram information of received signal. Using the proposed methodology, the broken pieces of instruments can be located with an accuracy of 85%.

26 December 2024 04:45:45

INTRODUCTION

In endodontics, root canal therapy (RCT) is one of the major and important procedures [1] used by dental surgeons to diagnose the teeth related problems like tooth pain, cavities. The common diagnosis techniques used by the endodontists include: x-ray, cold test, bioimpedance meter, apex locator, etc. But the major issues associated with all these techniques are: false detection, inappropriate testing procedures, and these techniques tend to harm the human tissue. To overcome these issues, the authors of [2], [3], [4], and [5] suggested the use of ultrasonic equipment in dentistry. Ultrasound is one of the safest technologies for medical diagnosis as it is a non-invasion and non-ionized technique. Ultrasound is basically a form of mechanical energy which vibrates and travels through the mediums.

In healthcare sector, the multiple modes of ultrasound [6] have been used to diagnose health issues. In medical imaging, a frequency of 2 MHz or above is used for generating the ultrasound image [6]. In [7], the authors compared ultrasonography with computer tomography (CT) for the examination of the preoperative evaluation of a solid abdominal tumor. These results were then used to determine the relationship with their relevant physical parameters in children. It was found that the results with ultrasound were better as compared to the ones produced by CT. In [8], authors suggested ultrasound techniques to diagnose the diseases of farm animals where they used colour doppler techniques of ultrasonography for evaluation of testicular blood perfusion (TBP). The authors of [9] also focused on using ultrasound techniques for diagnosing health issues in a veterinary patient kept under ICU. In [10], oral contrast enhanced transducer abdominal ultrasonography (OCTU) techniques had been used to diagnose gastric cancer in its initial stages using Wisonic Clover 60 colour doppler ultrasonic diagnostic system. This system was a portable ultrasonic machine with its transducer frequency set between 3 and 5 MHz Lung ultrasonography (LUS) [11] was used to diagnose pulmonary complications after cardiac surgery which is similar to postoperative heart failures. Later, [12] presented a systematic review about the evaluation of the placenta in normal as well as abnormal conditions using ultrasound techniques in which they investigate the placenta thickness and placenta thickness (calcifications). From the available literature, it can be inferred that the ultrasound has been used to diagnose soft tissue. However, the differences in tooth structure and properties limited the scope of ultrasound in dentistry.

Proceedings of the 1st International Conference on Artificial Intelligence, Advanced Materials, and Mechatronics Systems
AIP Conf. Proc. 3217, 020004-1–020004-9; https://doi.org/10.1063/5.0234443
Published under an exclusive license by AIP Publishing. 978-0-7354-5040-0/$30.00

Some researchers presented the use of ultrasonography in dentistry as well. One of the most common ultrasonic equipment used by dentists is a teeth scaler [13]. In 1963, Baum et al. [14] observed the ultrasound imaging in dentistry for the first-time where they used a 15 MHz ultrasonic transducer. However, the key limitation of this work was that the obtained image of internal tooth structure was not clear. In [15], authors used ultrasonography for root canal visualization where the authors used the 25 MHz ultrasound frequency on three different age groups, i.e., 22, 33, and 59-year-old patients with no periodontal disease. However, the authors did not get any evidence of a root canal in the tooth. In [16], authors used the 15 MHz ultrasound frequency for observing the teeth's internal images but the resulting image was not very clear. The authors concluded that, if the resolution of the system is increased, then the detection of caries will also increase. In [17], authors proposed a technique for the measurement of tooth layer thickness using Fractional Fourier transformation (FFT). For recording the signal, the research group used an Olympus NDT inc. Waltham MA based probe with a 15 MHz supportive ultrasonic frequency. The probe is in contact with the sample and the authors used glycerin as the matching layer. Later, in [18], Kapoor R. et al. proposed Modified Sliding Singular Spectrum Analysis (MSSSA), for detecting the foreign objects in the root canal but authors did not target the root canal and pulp of tooth simultaneously.

From the conducted literature survey, it can be concluded that there is a big potential of using ultrasound in dentistry and the localization of broken pieces of instrument. But none of the papers claim or prove that the ultrasound can penetrate through the teeth and perform a real-time localization of a broken piece of instrument inside the canal. The authors of [19]–[21] applied AI-based techniques for tooth treatments. This paper proposed a technique based on AI and MLP in conjunction with novel recording method for retrieving broken pieces of instrument inside the root canal.
The major contributions of this paper are:

- Novel recording methodology in conjunction of 1D Rx signal to obtain 2D visualization of 1D recoded signal.
- Features extraction using AI based method.
- Localization of broken piece instrument inside the root and pulp using MLP model.

The reminder of the paper has been split into following sections: the methodology used in this work is presented in section 2; where data acquisition, data preprocessing, features extraction and localization of broken pieces of instrument have been discussed. Section 3 shows the outcomes of the proposed methodologies. Section 4 presents the future aspect of the proposed scheme followed by the conclusion.

26 December 2024 04:45:45

PROPOSED METHODOLOGY

This paper proposed an AI-based novel localization methodology for broken pieces of instrument during the RCT. The proposed method has been presented briefly in figure 1. This paper uses ResNet50 CNN [22][23] architectures as the feature extractor in conjunction with LRSD (low rank sparse matrix decomposition) of maximum peak change and spectrogram of received signal. MLP has been used as a classifier. The acoustic data is recorded by transducers designed by authors where a novel recording system has been used. During designing, properties like acoustics impedances, dimensions of tooth, number of piezoelectrical sensors are to be considered where Tx and Rx is the transmitted signal and received signal respectively. More than 100 vitro teeth are used for signals recording. The x-ray's samples of few teeth are used for verification as shown in figure 2. Frequency of 33 kHz is used for transmission. The received waveform shows change in pattern of waveform. This change occurs due to the presence of the broken pieces of instrument inside the canal.

$$Rx = Tx * [.] \tag{1}$$

In the equation (1), Rx and Tx represent the received and transmitted signals. While the teeth properties that cause the change in the Rx with respect to the Tx. Since Rx and Tx are 1D signals, LSRD in conjunction with maximum peak change and a spectrogram are used for the visual representation of 1D signals.

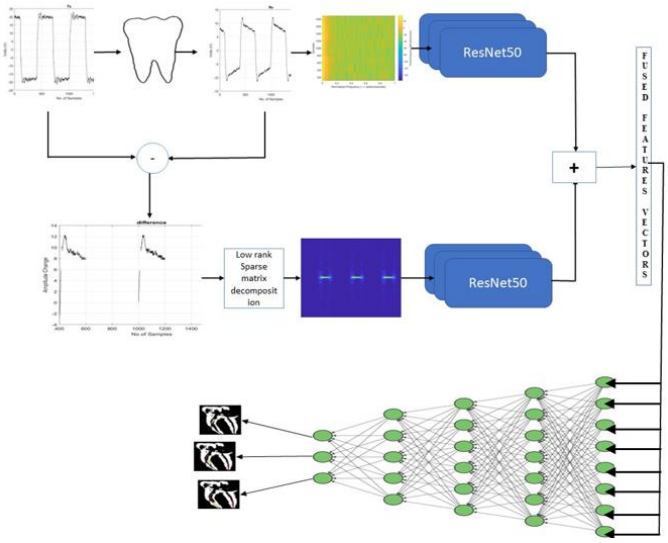


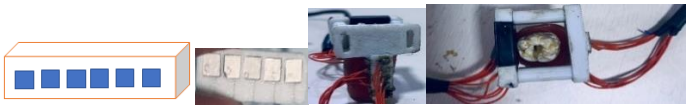
FIGURE 1. Proposed methods for the localization of the broken instrument pieces

(a) (b) (c)

FIGURE 2. X-ray imaging of broken instrument pieces

Data acquisition

This paper proposes a modified version of an ultrasonic transducer [18]. In previous work, authors used hard epoxy to prevent cross-talk between adjacent transducers and dental gel as a coupling agent. But authors did not mention the front matching layer for the transducer and used ADC ADS1115, which has a sampling rate of 860 samples/sec, which does not satisfy Nyquist criteria. In a modified version of the transducer, a dental gel pad is used as a coupling agent and front layer. In the modified version, an MCP3208 ADC with a sampling rate of 100 kHz at 5 volts of supply voltage is used. In this paper, an array of five transducers is used, as shown in figure 3. In which each transducer of size 3x3 mm is used. The depth of 1 mm is designed for dental gel pads. In previous work, a frequency range of 20–80 kHz was used, but in this version, a single frequency of 33 kHz is used because a range of frequencies is more susceptible to additive noise.



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

FIGURE 3 (a) array design (b) fabricated array (c) size view of transducer (d) top view of transducer

Data preprocessing

Data acquisition, explained in the previous section, needs to be preprocessed for feature extraction and localization of broken pieces of instrument. In this paper, preprocessing is used to convert the 1D signal to a 2D visualization. A spectrogram and LSRD of maximum peak change are used for 2D visualization of a 1D recorded signal.

26 December 2024 04:45:45

Spectrogram

It is the visual representation of a time-series signal. In visual representations, a spectrogram plots frequencies v/s time interval, where the power of frequencies is denoted by the intensity of the frequency-time plot. Higher power in dB (decibels) means brighter intensity, and vice versa. In the proposed method, a spectrogram is used to visualize the 1D Rx signal in 2D for features extracted from the CNN architecture, i.e., ResNet50.

Maximum peak changes

The deviation in received signal is calculated with respect to the transmitted signal. The change in peak values throughout the signal is estimated using (2) and (3).

$$Rx = \sum_{i=1}^{n} a_i \tag{2}$$

$$Tx = \sum_{i=1}^{n} b_i \tag{3}$$

Where, a_i, b_i and $[*]$ represents the instantaneous amplitude of received and transmitted and difference between them.

$$error = |(polv(c_i) - Rx)| \tag{4}$$

$$Dx = f_e * Rx_i \tag{5}$$

$$f_e = error_i \tag{6}$$

The above equations derive the maximum peak change between the transmitted and received signals. While algorithm 1 represents pseudo-code for maximum peak change.

ALGORITHM 1. Maximum Peak Change

```
Rx : received signal
Tx : transmitted signal
for j = n
  for i = n/path_length
     coef[i,:] = curvefitting(Tx,Rx,th,order)
     pol[i,:] = val(coef,Tx)
     f_error = abs((pol-Rx))^2
  end
  fe =  f_error[j,:]
  if fe<error_thress
     fe = NaN
  else
     fe = 1
   Dx[j] = fe*Rx[j]
end
```

Low Rank Sparse Matrix Decomposition (LSRD)

The received signal is a real-time, nonstationary signals. Therefore, the Short-Time Fourier transform abbreviated as STFT has been used for decomposition of non-stationary signal using a fix-sized sliding window. The sliding window moves through the input signal. It is assumed that the given non-stationary signal behaves like a stationary signal. Therefore, the STFT of a received and transmitted signals are described by (7).

$$X_{Dx}(j,k) = \sum_i w(t-i) Dx * exp((j * 3.14ik/N) \tag{7}$$

26 December 2024 04:45:45

where, k, i, j,and N denote frequency components, index of signals, time samples, and total frequency of samples respectively.

However, STFT has a major drawback the temporal and spectral resolution of the transmitted (Tx) and received (Rx) signal. These signals depend on the length of the window function. A larger window function has a good spectral resolution quantity, whereas a shorter window shows better temporal resolution. Therefore, the adaptive window function is generally or in this work used for the STFT decomposition of the Tx and Rx high-frequency acoustic signal. In this algorithm, optimizing the low rank matrix G and sparing matrix S as described as follows

$$X_{Dx} = B + S + noise \tag{8}$$

where, $\mathrm{S}, \mathrm{X_{Dx}}, \mathrm{B}$ and noise denote sparse matrix, the time-frequency distribution of differenced received and transmitted signals, low-rank matrix, and noise respectively. In LSRD [24] various algorithms like SVD, RPCA and GoDec is used in this paper. The GoDec is used as the matrix decomposition problem.

$$\min|X_{Dx} - B - S|^2 \ \ s.t. \ \ \mathrm{rank(B)} \leq \mathrm{r}, \mathrm{card(S)} \leq \mathrm{kM} \tag{9}$$

$$B_k = \arg\min |X_{Dx} - B_{k-1} - S_{k-1}|^2 \tag{10}$$

$$S_k = \arg\min |X_{Dx} - B_{k-1} - S_{k-1}|^2 \tag{11}$$

The iteration of $\mathrm{B_k}$, need two random projections i.e., $Z_1 = X_{Dx}P_1$ and $\mathrm{Y_2} = X_{Dx}^T P_2$, here P_1 and P_2 are random matrix. The approximation of L can be obtained by

$$B = Z_1(P_2^T Z_1)^{-1} P_2^T \tag{12}$$

The pseudo-code for the algorithm and visual representation are mention in algorithm 2 and figure 4.

ALGORITHM 2. LSRD

X_{Dx}: stft of D_x
r: low-ranked matrix
t: cardinality of sparse *matrix*
e: error range
I: maximum permissible iterations
While $\leq$ *I:*
  k = k+1
  $Z_1 = (X_{Dx} - B_{k-1})\, P_1, P_2 = Z_1$
  $Z_2 = (X_{Dx} - S_{k-1})^T P_2$
  If rank $(rank(P_2^T Z_1) \leq r$
    $r = rank(P_2^T Z_1)$
    break
  $B_k = Z_1(P_2^2 Z_1)^{-1} Z_2$
  $S_k = G_o(X_{Dx} - B_{k-1})$

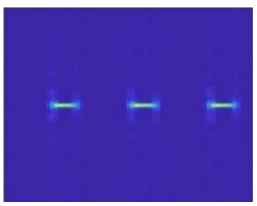


FIGURE 4. Visual representation by LSRD of maximum peak changed signal

26 December 2024 04:45:45

Features Extractor

In the proposed methodology, the 2D CNN architecture of ResNet50 is used as a feature extractor as the received signals are 1D. Therefore, the LSRD of the maximum peak change of Rx with respect to Tx and the spectrogram of Rx are used for 1D to 2D visualization. Those flattened vectors of 2D visualized data have prominent features of the given input signals the mathematical representation of flattened vectors is shown below.

$$f_x = [\sim] * S(Rx) \tag{13}$$
$$f_y = [\sim] * LRSD(Dx) \tag{14}$$

In the above equations 13 & 14, [~], S(.) and LRSD(.) and Dx denoted as the ResNet50 convolutions network, spectrogram, LRSD and maximum peak changes of Rx with respect to Tx.
There is a fusion of extracted features LSDR and spectrogram visualization, vector concatenation is used as the fuser.

Localization of broken pieces of instrument

In the proposed method, the MLP (multi-layer perceptron model) is used as shown in figure 1 for retrieval of position of broken pieces of instrument. In this paper, retrieval of broken pieces of instrument refers to classification between three positions, i.e., top, bottom, and middle. Hence, in the proposed method, three classes are taken for the given samples. The classification is made on the basis of input feature vectors extracted through the CNN architecture of ResNet50.

RESULTS AND DISCUSSION

This proposed method shows effective results for localization of a broken pieces of instrument inside the canal and obtained result in real-time with an approximate time delay of 1 second. figure 5 shows the confusion matrix of the proposed method. An accuracy level of 85% is achieved. The predicted results are verified with the labelled x-rays, as shown in fig. 2. in [18], the authors only focused on the location of the broken instrument within three positions, i.e., the top, middle, and bottom. However, in the present research work we took teeth which have broken pieces of instrument in multiple locations, broken pieces of instrument localize between the middle and bottom as shown in figure 6.

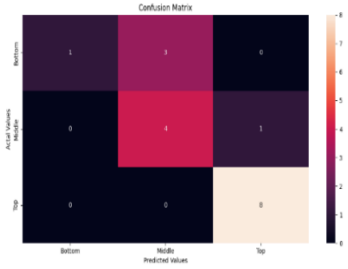


FIGURE 5. Confusion matrix

26 December 2024 04:45:45

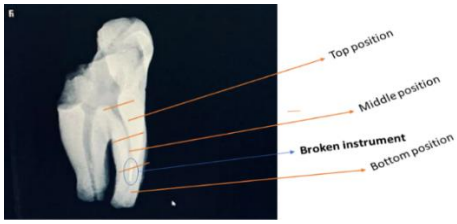


FIGURE 6. Segregation of relative position of root canal

In the proposed method, the signal is non-hazardous and offers several advantages with respect to conventional methods, as mentioned in Table 1. In [25]–[29], authors suggested the methodology for the extraction of broken pieces of instrument inside the canal. But all of these suggested methods have a low success rate because they suffer from a major drawback, i.e., no real-time localization can be made. But with the proposed methods, it is possible to localize the broken instrument. The method proposed in the present research overcomes this drawback and can localize the broken piece of instrument in the root canal.

TABLE 1. Comparison of proposed methodology

S.No.	Methods name	Advantages/	Disadvantages
1.	[25]–[29]	1. No use of electronic circuit.	1. No real time tracking of broken pieces of instrument. 2. Less success rates. 3. Chance to permanently damage of tooth during procedure. 4. Required highly skill endodontics surgeon. 5. Require x-ray's for analyzing the position of broken instrument inside the root canal.
2	Proposed method	1. Real-time tracking during retrieving of broken pieces of instrument. 2. Non-invasion or non-hazardous Ultrasonics signal is used. 3. Handy and easy to use. 4. More success rate as compare to conventional methods.	1. Misclassification between adjacent position 2. Delay of 1 sec.
3.	X-rays [30]	1. More accurate and precise.	1. No real time tracking 2. Ionized ray's hazardous to human tissue 3. Need separate room with skilled doctors

CONCLUSION & FUTURE SCOPE

Localization of broken pieces of instrument inside the root canal and pulp is complex tasks for dentist. This paper provides insight that it is feasible to detect the broken pieces of instrument using ultrasound signals at a low frequency of 33 kHz. The accuracy of the detection with the help of initial prototype is about 85%. However, the accuracy level can be increased by making design modifications and by collecting bigger datasets. The work will still be in progress for the exact and accurate localization of the broken instrument. This work will continue to detect broken pieces of instrument inside the vivo as well.

26 December 2024 04:45:45

In future, this work will be continued where the authors will extend the work to vivo teeth as well. This will increase the complexity of the task as it will present the option of fetching various other hyperparameters which are not possible to be fetched with the vitro version but will present the possibility of locating broken pieces of instrument in vivo tooth as well.

ACKNOWLEDGEMENTS

Authors pay gratitude to AICTE for funding this project. (Ref No. 8-122/FDC/RPS(Policy-1)/2019-20).

REFERENCES

1. M. Zehnder and G. N. Belibasakis, "On the dynamics of root canal infections—what we understand and what we don't," Virulence, vol. 6, no. 3. Landes Bioscience, pp. 216–222, Feb. 05, 2015. doi: 10.4161/21505594.2014.984567.
2. M. C. Ziskin, "Specification of Acoustic Output Level and Measurement Uncertainty in Ultrasonic Exposimetry," 2003.
3. G. R. Harris, "Progress in Medical Ultrasound Exposimetry," 2005.
4. M. Moderiano, M. McEvoy, J. Childs, and A. Esterman, "Safety of Ultrasound Exposure: Knowledge, Attitudes, and Practices of Australasian Sonographers," Journal of Diagnostic Medical Sonography, vol. 34, no. 5, pp. 357–367, Sep. 2018, doi: 10.1177/8756479318791512.
5. K. Maeda and A. Kurjak, "Diagnostic ultrasound safety," Donald School Journal of Ultrasound in Obstetrics and Gynecology, vol. 8, no. 2. Jaypee Brothers Medical Publishers (P) Ltd, pp. 178–183, 2014. doi: 10.5005/jp-journals-10009-1353.
6. A. Carovac, F. Smajlovic, and D. Junuzovic, "Application of Ultrasound in Medicine," Acta Informatica Medica, vol. 19, no. 3, p. 168, 2011, doi: 10.5455/aim.2011.19.168-171.
7. I. R. S. Lucena et al., "A comparison between ultrasonography and single-phase computed tomography for preoperative assessment of solid abdominal tumors in children," J Pediatr (Rio J), 2022, doi: 10.1016/j.jped.2022.05.003.
8. H. Samir, F. Radwan, and G. Watanabe, "Advances in applications of color Doppler ultrasonography in the andrological assessment of domestic animals: A review," Theriogenology, vol. 161. Elsevier Inc., pp. 252–261, Feb. 01, 2021. doi:10.1016/j.theriogenology.2020.12.002.
9. G. R. Lisciandro, "Cageside Ultrasonography in the Emergency Room and Intensive Care Unit," Veterinary Clinics of North America – Small Animal Practice, vol. 50, no. 6. W.B. Saunders, pp. 1445–1467, Nov. 01, 2020. doi: 10.1016/j.cvsm.2020.07.013.
10. L. Shen et al., "Initial screening of gastric cancer using oral contrast-enhanced trans-abdominal ultrasonography in rural asymptomatic individuals," Cancer Epidemiol, vol. 80, Oct. 2022, doi: 10.1016/j.canep.2022.102236.
11. S. M. Efremov et al., "Lung Ultrasonography and Cardiac Surgery: A Narrative Review," Journal of Cardiothoracic and Vascular Anesthesia, vol. 34, no. 11. W.B. Saunders, pp. 3113–3124, Nov. 01, 2020. doi: 10.1053/j.jvca.2020.01.032.
12. V. Schiffer et al., "Ultrasound evaluation of the placenta in healthy and placental syndrome pregnancies: A systematic review," European Journal of Obstetrics and Gynecology and Reproductive Biology, vol. 262. Elsevier Ireland Ltd, pp. 45–56, Jul. 01, 2021. doi:10.1016/j.ejogrb.2021.04.042.
13. A. A. R. Hashem, "Ultrasonic Vibration: Temperature Rise on External Root Surface during Broken Instrument Removal," J Endod, vol. 33, no. 9, pp. 1070–1073, Sep. 2007, doi: 10.1016/j.joen.2007.06.005.
14. G. Baum, I. Greenwood, S. Slawski, and R. Smirnow, "Observation of internal structures of teeth by ultrasonography," Science (1979), vol. 139, no. 3554, pp. 495–496, 1963, doi: 10.1126/science.139.3554.495.
15. B. Salmon and D. le Denmat, "Intraoral ultrasonography: Development of a specific high-frequency probe and clinical pilot study," Clin Oral Investig, vol. 16, no. 2, pp. 643–649, 2012, doi: 10.1007/s00784-011-0533-z.
16. K. T. Szopinski and P. Regulski, "Visibility of dental pulp spaces in dental ultrasound," Dentomaxillofacial Radiology, vol. 43, no. 1, 2014, doi: 10.1259/dmfr.20130289.
17. P. Kunche and N. Manikanthababu, "Fractional Fourier Transform," SpringerBriefs in Speech Technology, vol. 58, no. 10, pp. 25–49, 2020, doi: 10.1007/978-3-030-42746-7_2.

26 December 2024 04:45:45

18. R. Kapoor, D. Sharma, and A. Kapoor, “Modified Sliding Singular Spectrum Analysis-Based Noncontact-Type Detection of Separated Instrument in Root Canal Therapy Using Low-Frequency Ultrasonic Pattern Transceiver Design,” IEEE Trans Instrum Meas, vol. 70, 2021, doi: 10.1109/TIM.2021.3094635.
19. A. Aminoshariae, J. Kulild, and V. Nagendrababu, “Artificial Intelligence in Endodontics: Current Applications and Future Directions, Journal of Endodontics, vol. 47, no. 9. Elsevier Inc., pp. 1352–1357, Sep. 01, 2021. doi: 10.1016/j.joen.2021.06.003.
20. M. A. Saghiri, F. Garcia-Godoy, J. L. Gutmann, M. Lotfi, and K. Asgar, “The Reliability of artificial neural network in locating minor apical foramen: A cadaver study,” J Endod, vol. 38, no. 8, pp. 1130–1134, Aug. 2012, doi: 10.1016/j.joen.2012.05.004.
21. P. Lahoud et al., “Artificial Intelligence for Fast and Accurate 3-Dimensional Tooth Segmentation on Cone-beam Computed Tomography,” J Endod, vol. 47, no. 5, pp. 827–835, May 2021, doi: 10.1016/j.joen.2020.12.020.
22. S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, “1D convolutional neural networks and applications: A survey,” Mech Syst Signal Process, vol. 151, Apr. 2021, doi: 10.1016/j.ymssp.2020.107398.
23. B. Xiao, X. Yin, and S. C. Kang, “Vision-based method of automatically detecting construction video highlights by integrating machine tracking and CNN feature extraction,” Autom Constr, vol. 129, Sep. 2021, doi: 10.1016/j.autcon.2021.103817.
24. Q. Lyu et al., "SAR Interference Suppression Algorithm Based on Low-Rank and Sparse Matrix Decomposition in Time–Frequency Domain," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 4008305, doi: 10.1109/LGRS.2020.3048161.
25. M. Hulsmann and med dent, “Removal of Fractured Instruments Using a Combined Automated/Ultrasonic Technique,” 1994.
26. W. Nehme, “CLINICAL AID A New Approach for the Retrieval of Broken pieces of instrument,” 1999.
27. M. Fu, Z. Zhang, and B. Hou, “Removal of broken files from root canals by using ultrasonic techniques combined with dental microscope: A retrospective analysis of treatment outcome,” J Endod, vol. 37, no. 5, pp. 619–622, May 2011, doi: 10.1016/j.joen.2011.02.016.
28. U. G. H. Fors and J.-O. Berg, “CLINICAL AID A Method for the Removal of Broken Endodontic Instruments from Root Canals,” 1983.
29. C. D’arcangelo, G. Varvara, and P. de Fazio, “CASE REPORTS Broken Instrument Removal-Two Cases,” 2000.
30. N. Gencoglu and D. Helvacioglu, “Comparison of the Different Techniques to Remove Fractured Endodontic Instruments from Root Canal Systems”, European Journal of Dentistry, vol. 3, pp. 90-95, April-2009, DOI: 10.1055/s-0039-1697413.

26 December 2024 04:45:45

Artificial Neural Network based Modelling for Variational Effect on Double Metal Double Gate Negative Capacitance FET

Yash Pathak, Laxman Prasad Goswami *Student Member, IEEE,* Bansi Dhar Malhotra, *Member, IEEE,* and Rishu Chaujar, *Senior Member, IEEE*

arXiv:2412.14216v1 [cond-mat.mtrl-sci] 18 Dec 2024

***Abstract*—In this work, we have implemented an accurate machine-learning approach for predicting various key analog and RF parameters of Negative Capacitance Field-Effect Transistors (NCFETs). Visual TCAD simulator and the Python high-level language were employed for the entire simulation process. However, the computational cost was found to be excessively high. The machine learning approach represents a novel method for predicting the effects of different sources on NCFETs while also reducing computational costs. The algorithm of an artificial neural network can effectively predict multi-input to single-output relationships and enhance existing techniques. The analog parameters of Double Metal Double Gate Negative Capacitance FETs (D2GNCFETs) are demonstrated across various temperatures (T), oxide thicknesses (T_{ox}), substrate thicknesses (T_{sub}), and ferroelectric thicknesses (T_{Fe}). Notably, at $T = 300K$, the switching ratio is higher and the leakage current is 84 times lower compared to $T = 500K$. Similarly, at ferroelectric thicknesses $T_{Fe} = 4nm$, the switching ratio improves by 5.4 times compared to $T_{Fe} = 8nm$. Furthermore, at substrate thicknesses $T_{sub} = 3nm$, switching ratio increases by 81% from $T_{sub} = 7nm$. For oxide thicknesses at $T_{ox} = 0.8nm$, the ratio increases by 41% compared to $T_{ox} = 0.4nm$. The analysis reveals that $T_{Fe} = 4nm$, $T = 300K$, $T_{ox} = 0.8nm$, and $T_{sub} = 3nm$ represent the optimal settings for D2GNCFETs, resulting in significantly improved performance. These findings can inform various applications in nanoelectronic devices and integrated circuit (IC) design.**

***Index Terms*—Artificial Neural Network (ANN), NCFET, Machine learning, TCAD.**

I. Introduction

SINCE the 4 decades in the semiconductor industry, the size of FET devices has been decreased to attain better efficiency and performance, in possession with Moore's law [1]–[5]. The new trend of VLSI is governing the variability issue, such as global variability (GV) and local variability (LV) sources for both planar and gate-all-around (GAA) devices. In the case of LV source, a high-k/metal gate device with various work functions, which is vital for statistical variability, would be at extreme risk for device performance because of the random occupied grain orientation of gate material [6], [7]. Additionally, in the case of the GV source, the impact of critical dimension is variability in varying the device performance, and as device size is reduced, this effect cannot be ignored.

The power and thermal management of integrated circuits (ICs) are vital aspects of the transistor industry. Although with engineering technology and material science advancement transistor generates a low amount of heat. However, the ICs have millions of transistors on the chip, which generate huge amounts of heat. Due to its lesser supply voltage, tunnel FETs have also been cited by many research organizations in the semiconductor world as the ideal FET over traditional MOSFETs for future low-power applications. But dealing with TFET can lead to low Ion, so to get the amalgamation of improved Ion and subthreshold swing, the NC technology is an apt candidate to meet the demand of future electronics. [4], [8]–[10].

However, the variability focuses on the accurate prediction of correlation among different sources. Cogenda Visual TCAD and Python are proven tools for doing complicated analysis [3,4]. However, a large number of samples is required for better performance. Hence, a large computational budget is inevitable for precise prediction. In this study, a method of variability analysis is proposed for modeling for device level as well as circuit level with minimum error [11]–[15]. Consequently, we propose a new method of machine learning (ML) algorithm with a precise match of the TCAD simulator for improved efficiency. This article consists of an analysis of analog parameters of D2GNCFET for different oxide thicknesses, ferroelectric thickness, temperature, and substrate thickness, which results in reduced subthreshold swing, low input power consumption, reduced leakage current, reduced drain-induced barrier lowering, and better controllability of the device.

II. Device Architecture and Methodology

Fig.1 shows the schematic diagram of the Double metal double gate NCFET (D2GNCFET) [16], [17]. The param-

This work was funded in part by the Council of Scientific and Industrial Research under Grant 08/133(0050)/2020-EMR-I for Yash, and BDM thanks the Science & amp; Engineering Research Board (SERB), Govt. of India for the honor of a Renowned Fellowship (SB/DF/011/2019).

Y. Pathak and R. Chaujar are with the Department of Applied Physics, Delhi Technological University, Delhi, 110042 India (e-mail: yash_2k19phdap505@dtu.ac.in; chaujar.rishu@dtu.ac.in).

L. P. Goswami is with the Department of Physics, Indian Institute of Technology, Delhi, 110016 India (e-mail: goswami.laxman@gmail.com).

B. D. Malhotra is with the Department of Biotechnology, Delhi Technological University, Delhi, 110042 India (e-mail: bansi.malhotra@dce.ac.in).

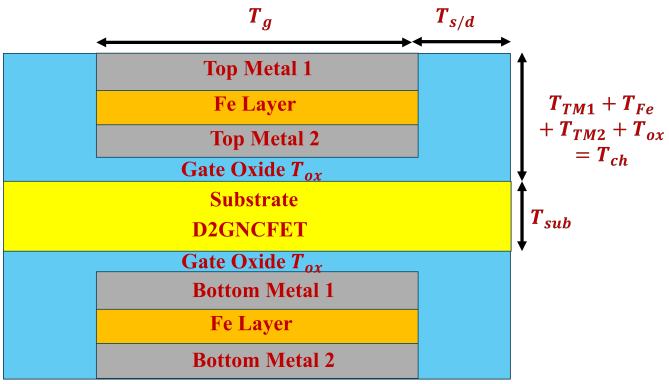


Fig. 1. The 2D schematic figure of double metal double gate NCFET.

TABLE I
THE DEVICE PARAMETERS FOR D2GNCFET ARE USED FOR THE SIMULATION

Parameter	Symbol	D2GNCFET
Substrate thickness	T_{sub}	$5nm$
Ferroelectric thickness	T_{fe}	$7nm$
Oxide Thickness	T_{ox}	$0.6nm$
Gate length	L_g	$50nm$
Concentration of Source/Drain	$N_{s/d}$	$10^{20}cm^{-3}$
Source/Drain length	$L_{s/d}$	$5nm$
Top/Bottom Metal thickness	T_{m1}/T_{m1b}	$10nm$

eters of D2GNCFET are summarized in table-I. The channel is made with silicon material and has a length of $L_g = 60nm$. The drain and source are made with aluminum which has length $L_{d/s} = 50nm$. The thickness of the substrate (T_{sub}), oxide (T_{ox}), top metal1/bottom metal1 (T_{TM1}/T_{TB1}), top metal2/bottom metal2 (T_{TM2}/T_{TB2}) are $5nm$, $0.6nm$, $10nm$, $5nm$ respectively. The total thickness of channel (T_{ch}) is $10nm$ as illustrated in table-I. The material of top metal1/bottom metal1, top metal2/bottom metal2 is Aluminium. The material taken for ferroelectric material, and gate oxide are HfO_2FE, SiO_2. The work function (ϕ) of aluminum material is fixed at $4.2eV$ for the double gate. The concentration of the doping profile for Source/Drain ($N_{s/d}$) is $1.0 \times 10^{20}cm^{-3}$.

The entire simulation is done by Cogenda Visual TCAD and Python high-level language. The gate voltage (V_{gs}) is varied from $0 - 0.5V$, with drain voltage fixed at $0.05V$. The various physical models are used in this work, such as Shockley Read Hall (SRH) which includes generation and recombination effects, the Arora model which incorporates the effects on carrier mobility due to temperature and impurities concentration, Crowell size for consideration shows the impact ionization, Fermi Dirac statistics is used for improved efficiency and accuracy.

III. Artificial Neural Network (ANN) Implementation

We constructed a tensor flow-based ANN model with a Rectified Linear Unit (ReLU) activation function [18]–[20] for D2GNCFET. The architecture of the ANN model is shown in Fig.2. The input to the input layer is an array of size 5 for each variable Temperature T, Substrate thickness T_{sub}, Oxide thickness T_{ox}, Ferroelectric thickness T_{fe}, and the voltage V_{ds}. The output layer provides the ANN prediction for the drain current I_d. Further, we consider three hidden layers in our ANN model. The first and third layers have 256 neurons while there are 512 neurons in the middle layer. The output of each neuron in the Neural Network (NN) is given as:

$$a_p^{[q]} = g\left(\vec{W}_p^{[q]} \cdot \vec{a}^{[q-1]} + b_p^{[q]}\right) \quad (1)$$

where, q represents the layers of the neural network (NN), and p is the node number such that $a_p^{[q]}$ is the output of p^{th} node from q^{th} layer. g is the activation function. $\vec{W}$ and b are the weight and biasing parameters for each node.

This function is repeated and runs up to the Taylor series to get the precise result. The input value put in ANN does not exclude work function variation (WFV) due to WFV impact and varies the T_{ox}, T, T_{sub}, T_{Fe} variability. In addition, it contains fixed parameters to study various dimensions of D2GNCFET. The algorithm of the implemented D2GNCFET ANN is illustrated in fig. 3.

IV. Result and discussion

Fig.4 represents the drain current (I_d) vs gate voltage (V_{gs}) of the TCAD simulation study (in red dash line) and predicted data (blue dash line) by ANN model for D2GNCFET at $V_{ds} = 0.05V$. The variation with oxide thickness (T_{ox}) are displayed in subplots (a1)-(a4) at $T_{Fe} = 7nm$, $T = 300K$, $T_{sub} = 5nm$. The predicted and simulated data shows a precise match at $T_{ox} = 0.6nm$, which reflects the lower error at $0.3V$ of gate voltage. The variation with ferroelectric thickness (T_{Fe}) is shown in subplots (b1)-(b4) at $T_{ox} = 0.6nm$, $T = 300K$, $T_{sub} = 5nm$. The precise data of predicted and simulated is matched at $T_{Fe} = 8nm$, reflecting the lower error after $0.3V$ of gate voltage. The variation with substrate thickness (T_{sub}) is illustrated in subplots (c1)-(c4) at $T_{ox} = 0.6nm$, $T = 300K$, $T_{Fe} = 7nm$. The precise data of predicted and simulated is matched at $T_{sub} = 4nm$, which reflects the lower error at $0.3V$ of gate voltage. The variations with temperature is shown in subplots (d1)-(d4) at $T_{ox} = 0.6nm$, $T_{Fe} = 7nm$, and $T_{sub} = 5nm$. The predicted and simulated data are matched at $T = 300k$, which reflects the minimum error after $0.3V$ of gate voltage. The different variations of all data variables turned out to coincide at gate voltage at 0.3V. The trainer of predicted data works successfully as a transfer characteristics start from off current to linear current.

Fig.5 shows the curve of output characteristics of D2GNCFET in linear and log scales for various temperatures and thicknesses. The device characteristics with variation in oxide thickness(T_{ox}) are shown in Fig.5(a) at fixed $V_{ds} = 0.05V$, $T = 300K$, $T_{Fe} = 7nm$, $T_{sub} = 5nm$. The better performance of the device is represented at $T_{ox} = 0.8nm$, which has a lower leakage current,

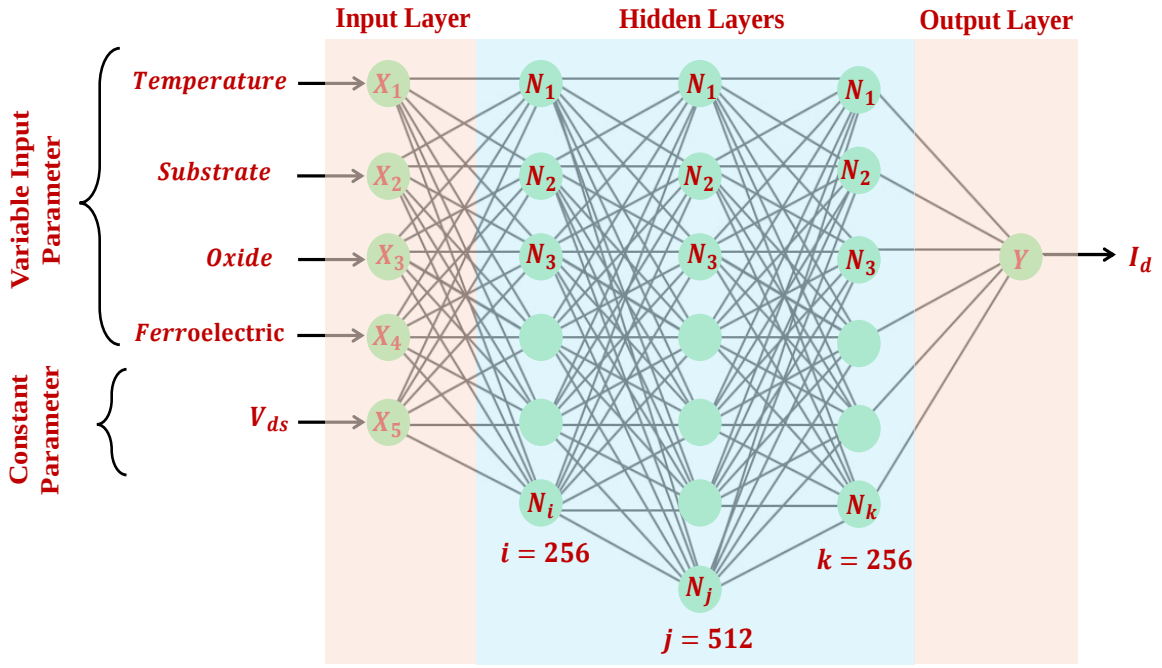


Fig. 2. The figure shows the D2GNCFET Artificial Neural Network (ANN) with one input, one output, and three hidden layers. There are 256 neurons in the first and third hidden layers, and 512 neurons in the middle layer. All I/O values include the source and the device parameter. The model predicts the drain current based on the input parameters Temperature, Substrate thickness, Oxide thickness, Ferroelectric thickness, and the drain-source voltage V_{ds}.

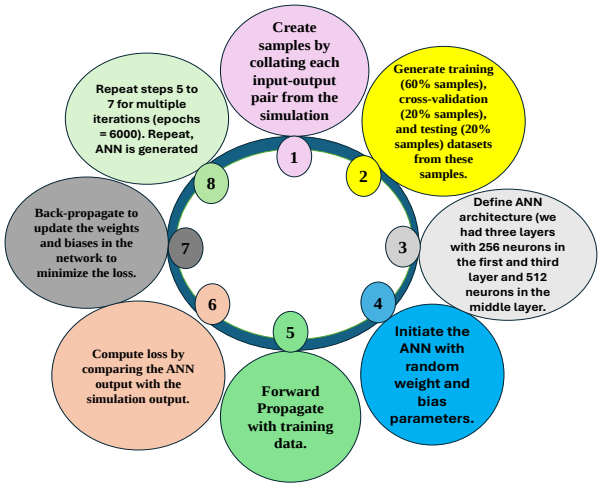


Fig. 3. Algorithm for D2GNCFET Artificial Neural Network (ANN)

higher drain current, and better switching ratio, confirming the improved gate controllability. The drain current of D2GNCFET is improved with higher oxide thickness. The impact of variation of ferroelectric thickness(T_{Fe}) is shown in Fig.5(b) at fixed $V_{ds} = 0.05V$, $T = 300K$, $T_{ox} = 0.6nm$, $T_{sub} = 5nm$. The better performance of the device is represented at $T_{Fe} = 4nm$ with lower leakage current and higher drain current, resulting in a better switching ratio with low power consumption. The drain current of D2GNCFET is improved with lesser ferroelectric thickness. The effect of variation of substrate thickness (T_{sub}) is shown in Fig.5(c) at fixed $V_{ds} = 0.05V$, $T = 300K$, $T_{ox} = 0.6nm$, $T_{Fe} = 7nm$. The better performance of the device is exhibited at $T_{sub} = 3nm$, which has a higher drain current, better switching ratio due to lower leakage current, and improved gate controllability. The drain current of D2GNCFET is improved with lesser substrate thickness. The influence of variation in temperature (T) is shown in Fig.5(d) at fixed $V_{ds} = 0.05V$, $T_{Fe} = 7nm$, $T_{ox} = 0.6nm$, and $T_{sub} = 5nm$. The better performance of the device is represented at $T = 300K$, which has a lower leakage current and a higher drain current, thus ensuring a better switching ratio, improved gate controllability, and low power consumption. The drain current of D2GNCFET is improved at low temperatures.

The simulation and ANN prediction results are compared in Fig.6. The linear relation between the prediction and the TCAD simulation results signifies the accuracy of our ANN model as depicted in subplot Fig.6(a). We further computed the losses during each epoch of training the ANN model. The loss reduces by increasing the number of epochs as shown in subplot Fig.6(b).

Fig.7 shows the plot of transconductance and TGF as a function of gate voltage for various temperatures, and different thicknesses at $V_{ds} = 0.05V$, $T_{Fe} = 7nm$,

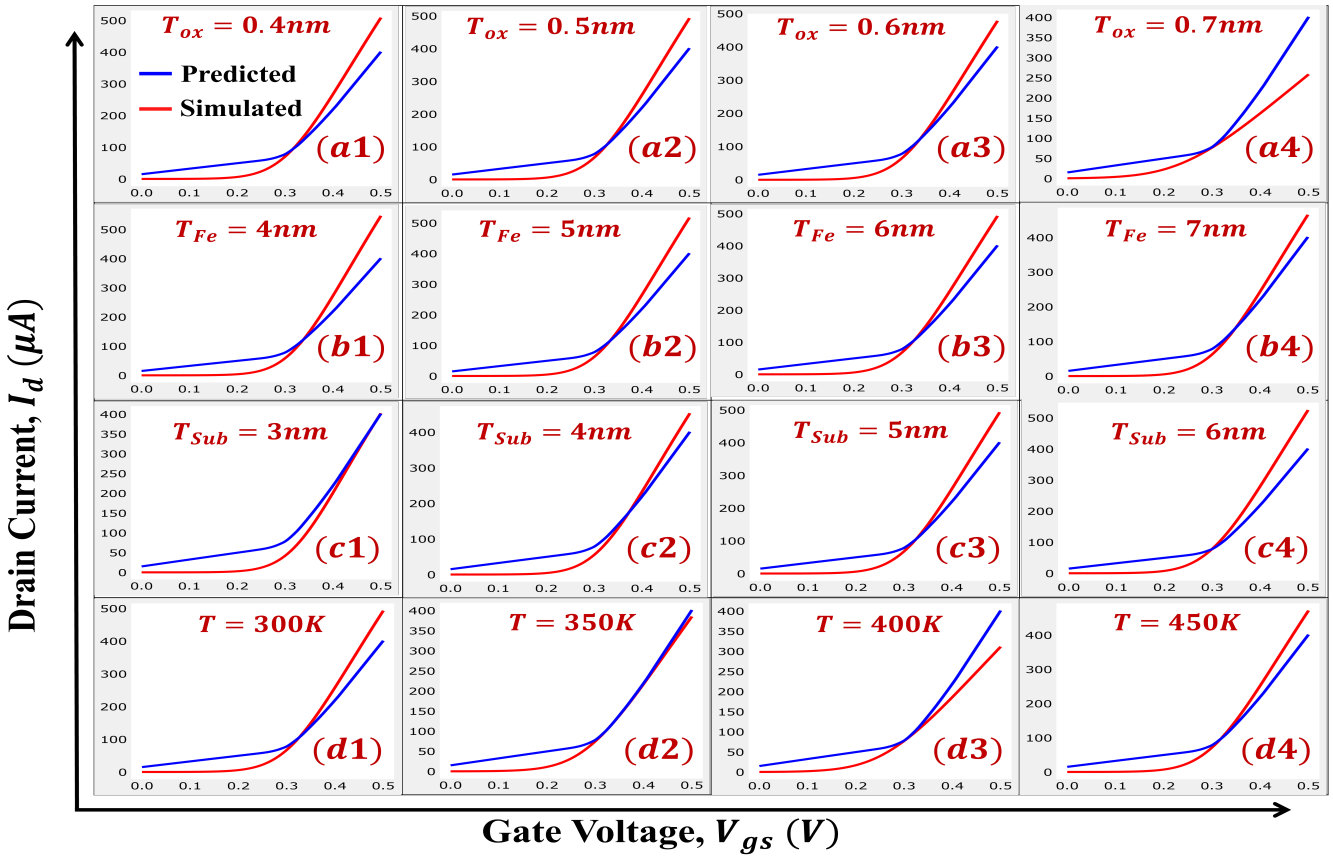


Fig. 4. The graph of drain current vs gate voltage for predicted and simulated data of D2GNCFET for different variations of temperature, oxide, thickness, ferroelectric thickness, substrate thickness at V_d=0.05V.

TABLE II
D2GNCFET COMPARISON FOR TEMPERATURE, SUBSTRATE THICKNESS, FERROELECTRIC, OXIDE THICKNESS.

No.	Parameter	$I_{off}(10^{-8}A)$	$I_{on}/I_{off}(10^5)$	$V_{th}(mV)$	$SS(mV/dec.)$
$(a1)$	$T = 300k,\ T_{ox} = 0.4nm,\ T_{sub} = 5nm,\ T_{Fe} = 7nm$	1.92	0.273	77.23	76.121
$(a2)$	$T = 300k,\ T_{ox} = 0.5nm,\ T_{sub} = 5nm,\ T_{Fe} = 7nm$	1.62	0.314	82.278	75.464
$(a3)$	$T = 300k,\ T_{ox} = 0.7nm,\ T_{sub} = 5nm,\ T_{Fe} = 7nm$	1.25	0.383	89.52	74.33
$(a4)$	$T = 300k,\ T_{ox} = 0.8nm,\ T_{sub} = 5nm,\ T_{Fe} = 7nm$	1.13	0.412	92.18	73.86
$(b1)$	$T = 300k,\ T_{ox} = 0.6nm,\ T_{sub} = 5nm,\ T_{Fe} = 4nm$	0.382	1.52	118.2	68.4
$(b2)$	$T = 300k,\ T_{ox} = 0.6nm,\ T_{sub} = 5nm,\ T_{Fe} = 5nm$	0.605	0.89	107.77	70.53
$(b3)$	$T = 300k,\ T_{ox} = 0.6nm,\ T_{sub} = 5nm,\ T_{Fe} = 6nm$	0.935	0.55	97.2	72.69
$(b4)$	$T = 300k,\ T_{ox} = 0.6nm,\ T_{sub} = 5nm,\ T_{Fe} = 8nm$	2.08	0.227	76.065	77.061
$(c1)$	$T = 300k,\ T_{ox} = 0.5nm,\ T_{sub} = 3nm,\ T_{Fe} = 7nm$	0.523	0.768	115.107	72.22
$(c2)$	$T = 300k,\ T_{ox} = 0.5nm,\ T_{sub} = 4nm,\ T_{Fe} = 7nm$	0.906	0.5	99.462	73.719
$(c3)$	$T = 300k,\ T_{ox} = 0.5nm,\ T_{sub} = 6nm,\ T_{Fe} = 7nm$	2.06	0.2548	74.58	75.89
$(c4)$	$T = 300k,\ T_{ox} = 0.5nm,\ T_{sub} = 7nm,\ T_{Fe} = 7nm$	2.88	0.192	63.84	76.86
$(d1)$	$T = 300k,\ T_{sub} = 5nm,\ T_{Fe} = 7nm,\ T_{ox} = 0.5nm$	1.41	0.347	86.68	74.95
$(d2)$	$T = 350k,\ T_{sub} = 5nm,\ T_{Fe} = 7nm,\ T_{ox} = 0.5nm$	6.96	0.0553	40.23	87.55
$(d3)$	$T = 400k,\ T_{sub} = 5nm,\ T_{Fe} = 7nm,\ T_{ox} = 0.5nm$	0.225	0.0138	−5.21	100.498
$(d4)$	$T = 450k,\ T_{sub} = 5nm,\ T_{Fe} = 7nm,\ T_{ox} = 0.5nm$	0.55	0.0047	−50.04	113.927
$(d5)$	$T = 500k,\ T_{sub} = 5nm,\ T_{Fe} = 7nm,\ T_{ox} = 0.5nm$	0.011	0.0012	−94.86	128.28

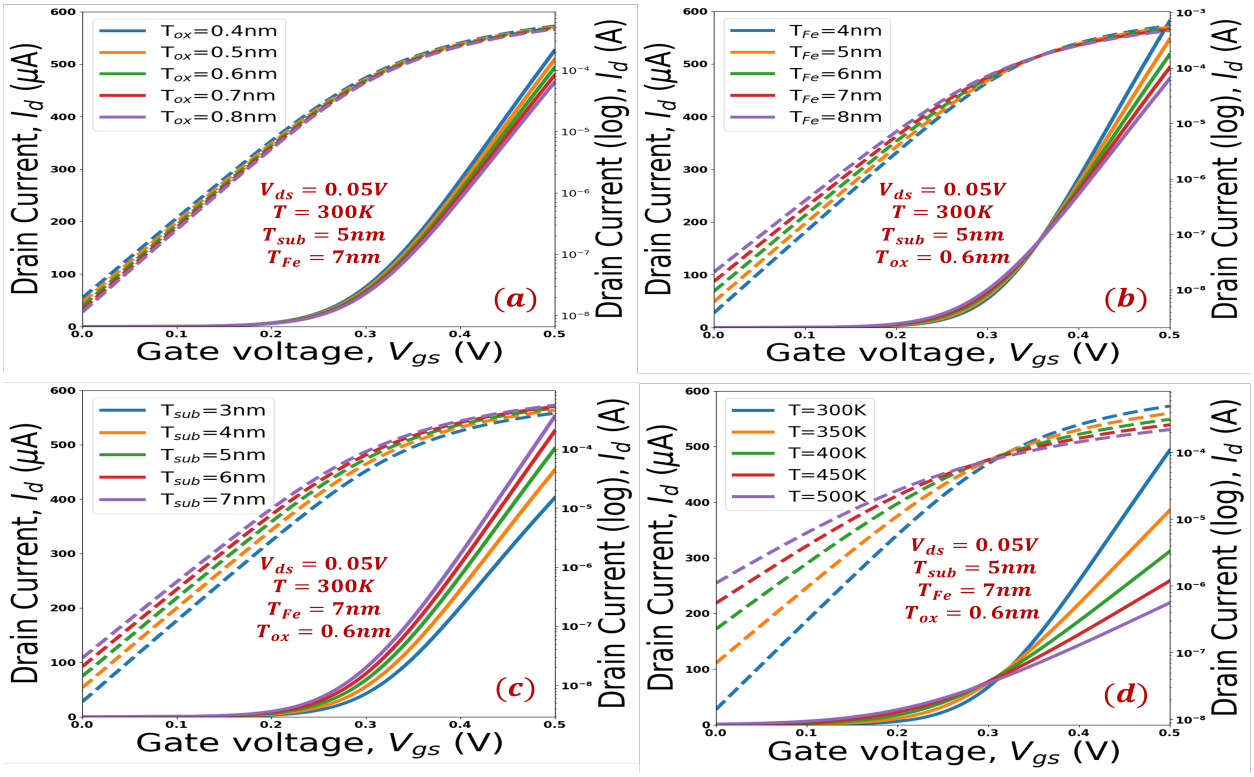


Fig. 5. The graph of drain current vs gate voltage D2GNCFET for different variations of temperature, oxide, thickness, ferroelectric thickness, substrate thickness at V_d=0.05V.

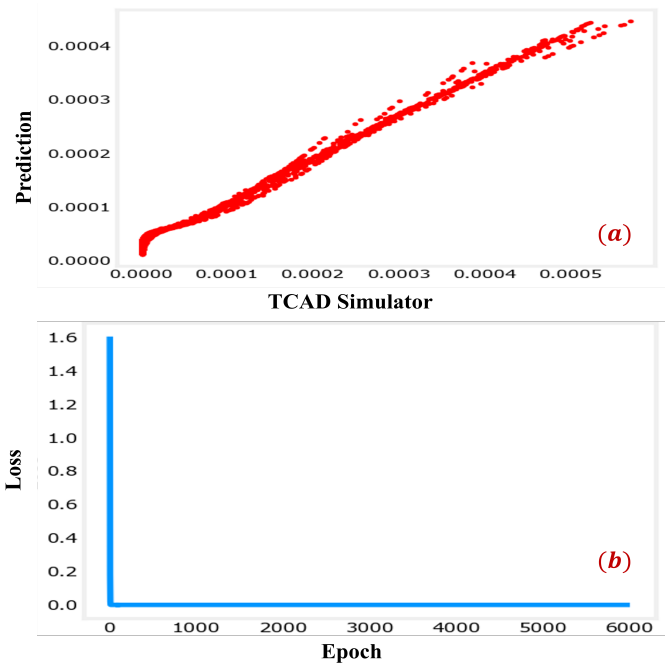


Fig. 6. The figure shows the comparison of TCAD simulated output and the output predicted by our ANN model. (a) The graph of predicted versus simulated data and (b) model loss at each epoch during the training of ANN model.

$T_{ox} = 0.6nm$, $T = 300K$, $T_{sub} = 5nm$. The subplots (a) show the transconductance (g_m) and transconductance gain factor (TGF) of D2GNCFET at $T_{ox} = 0.8nm$ is higher than other oxide thickness, which reflects the high On current, better gain generated per unit cell, reduces the leakage current $V_{ds} = 0.05V$, $T_{Fe} = 7nm$, $T = 300K$, $T_{sub} = 5nm$. The subplot (b) shows the transconductance (g_m) and transconductance gain factor (TGF) of D2GNCFET at T_{Fe}=4nm is higher than other thickness of ferroelectric layer, which reveals the high On current, better gain generated per unit cell, reduces the leakage current V_{ds}=0.05V, T=300K, T_{ox}=0.6nm, T_{sub}=5nm. The subplot (c) The transcoductance (g_m) and transconductance gain factor (TGF) of D2GNCFET at T_{sub}=3nm is higher than other thickness of substrate, which reflects the high On current, better gain generated per unit cell, reduces the leakage current V_{ds}=0.05V, T_{Fe}=7nm, T_{ox}=0.6nm, T=300K. The subplot (d) The transcoductance (g_m) and transconductance gain factor (TGF) of D2GNCFET at T=300K is higher than the other temperatures, which reflects the high On current, better gain generated per unit cell, reduced the leakage current at V_{ds}=0.05V, T_{Fe}=7nm, T_{ox}=0.6nm, T_{sub}=5nm.

Fig.8 spider chart of leakage current, on current, threshold voltage, SS for different temperatures, and thicknesses of different variable sets. (a) The oxide thickness of 0.8nm shows high On current, lower leakage current, higher threshold voltage, reduced subthreshold swing at V_{ds}=0.05V, $T_{Fe} = 7nm$, $T_{sub} = 5nm$, $T = 300K$. (b)

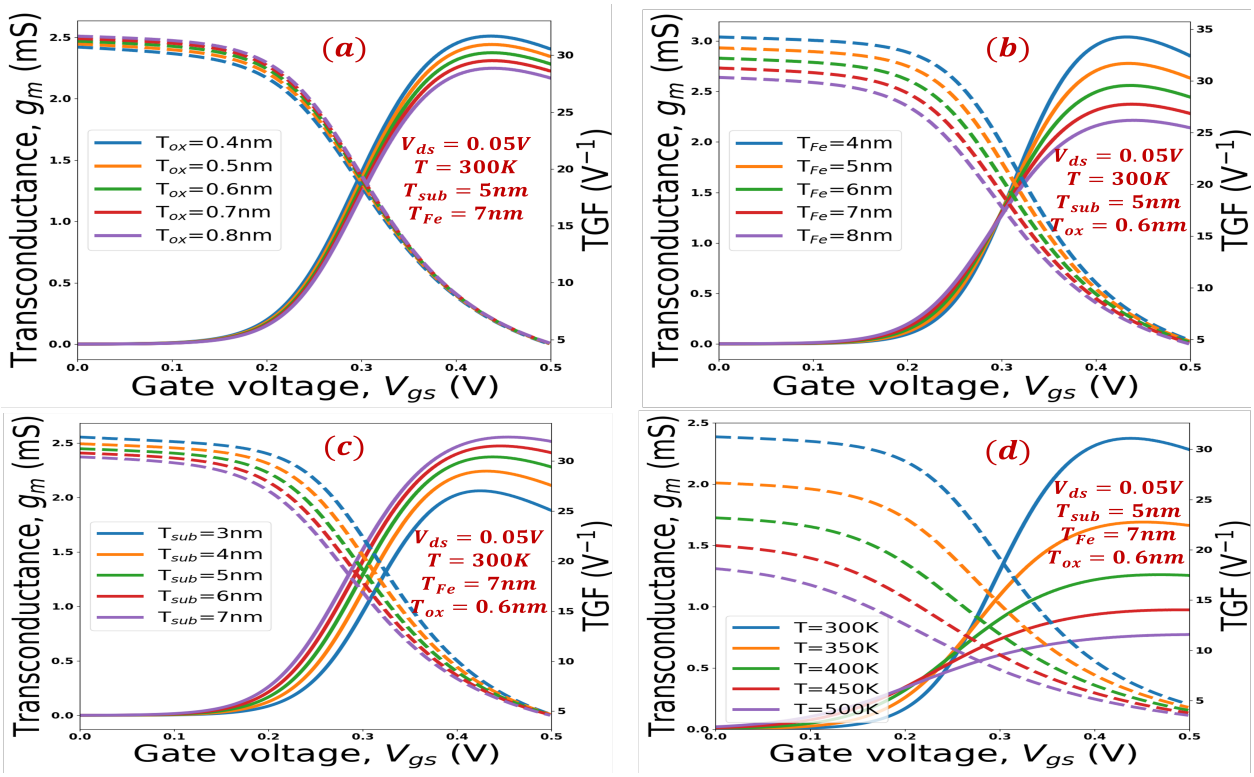


Fig. 7. The plot of transconductance and TGF vs V_{gs} for various temperature, thicknesses at V_d=0.05V.

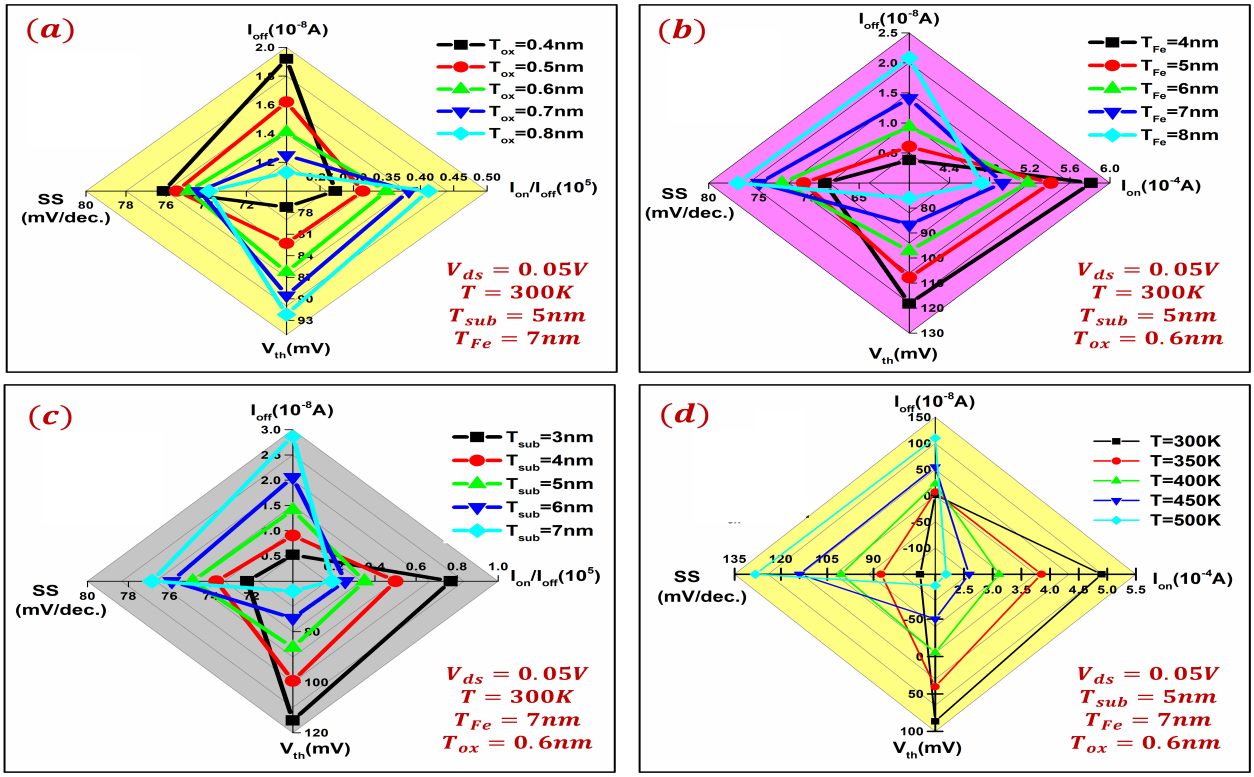


Fig. 8. The spider chart of leakage current, On current, Threshold Voltage, SS for different temperature, thicknesses at V_d=0.05V.

The ferroelectric thickness of $4nm$ results in high On current, lower leakage current, higher threshold voltage, reduced subthreshold swing at $V_{ds} = 0.05V$, $T_{ox} = 0.6nm$, $T_{sub} = 5nm$, $T = 300K$. (c) The substrate thickness of $3nm$ showcases high On current, lower leakage current, higher threshold voltage, reduced subthreshold swing at $V_{ds} = 0.05V$, $T_{Fe} = 7nm$, $T_{ox} = 0.6nm$, $T = 300K$. (d) The temperature of 300K reflects high On current, lower leakage current, higher threshold voltage, reduced subthreshold swing at $V_{ds} = 0.05V$, $T_{Fe} = 7nm$, $T_{sub} = 5nm$, $T_{ox} = 0.6nm$.

V. Conclusion

In this study, we have analyzed ANN-based machine learning to access the freedom of variability in the ULSI domain. The Cogenda Visual TCAD simulator and Python high-level language are used throughout the process. The cost of computational work is substantially reduced by ANN-based ML with precise calculation and efficiency. The method investigating variability processes can help in better optimization and designing the useful for ULSI application. The analog parameter of D2GNCFET is studied for various oxide thickness, substrate thickness, temperature, and ferroelectric thickness, such as lower leakage current, higher switching ratio at $T = 300K$ by 3000 times as compared to $T = 500K$, at $T_{Fe} = 4nm$ by 588% in contrast at $T_{Fe} = 8nm$, at $T_{sub} = 3nm$ by 300% in comparison at $T_{sub} = 7nm$, $T_{ox} = 0.8nm$ by 50% from $T_{ox} = 0.4nm$. All the analyzed results reveal that $T_{Fe} = 4nm$, $T = 300K$, $T_{ox} = 0.8nm$, $T_{sub} = 3nm$ as the most optimum set of variable sources for the D2GNCFET with the most improved performance and can further be used for various applications in nanoelectronic devices and IC designing.

Acknowledgment

The authors thanks to Vinod Dham Centre of Excellence for Semiconductors and Microelectronics (VDCE4SM), Nano Bioelectronics and Microelectronics Lab, Delhi Technological University for giving essential facilities.

References

[1] K. Ko, J. K. Lee, M. Kang, J. Jeon, and H. Shin, "Prediction of process variation effect for ultrascaled gaa vertical fet devices using a machine learning approach," *IEEE Transactions on Electron Devices*, vol. 66, no. 10, pp. 4474–4477, 2019.

[2] X. Wang, B. Cheng, D. Reid, A. Pender, P. Asenov, C. Millar, and A. Asenov, "Finfet centric variability-aware compact model extraction and generation technology supporting dtco," *IEEE Transactions on Electron Devices*, vol. 62, no. 10, pp. 3139–3146, 2015.

[3] X. Wang, B. Cheng, A. R. Brown, C. Millar, J. B. Kuang, S. Nassif, and A. Asenov, "Interplay between process-induced and statistical variability in 14-nm cmos technology double-gate soi finfets," *IEEE Transactions on Electron Devices*, vol. 60, no. 8, pp. 2485–2492, 2013.

[4] S. B. Rahi, S. Tayal, and A. K. Upadhyay, "A review on emerging negative capacitance field effect transistor for low power electronics," *Microelectronics Journal*, vol. 116, p. 105242, 2021.

[5] L. Gerrer, S. M. Amoroso, P. Asenov, J. Ding, B. Cheng, F. Adamu-Lema, S. Markov, A. Asenov, D. Reid, and C. Millar, "Interplay between statistical reliability and variability: A comprehensive transistor-to-circuit simulation technology," in *2013 IEEE International Reliability Physics Symposium (IRPS)*. IEEE, 2013, pp. 3A–2.

[6] X. Wang, A. R. Brown, B. Cheng, and A. Asenov, "Statistical variability and reliability in nanoscale finfets," in *2011 International electron devices meeting*. IEEE, 2011, pp. 5–4.

[7] H. F. Dadgour, K. Endo, V. K. De, and K. Banerjee, "Grain-orientation induced work function variation in nanoscale metal-gate transistors—part i: Modeling, analysis, and experimental validation," *IEEE Transactions on Electron Devices*, vol. 57, no. 10, pp. 2504–2514, 2010.

[8] C.-I. Lin, A. I. Khan, S. Salahuddin, and C. Hu, "Effects of the variation of ferroelectric properties on negative capacitance fet characteristics," *IEEE transactions on electron devices*, vol. 63, no. 5, pp. 2197–2199, 2016.

[9] Y. Pathak, B. D. Malhotra, and R. Chaujar, "Dft based atomic modeling and analog/rf analysis of ferroelectric hfo2 based improved fet device," *Physica Scripta*, vol. 98, no. 8, p. 085933, 2023.

[10] K. Verma and R. Chaujar, "Analysis of vertically stacked ferroelectric based finfet for switching applications," in *2023 2nd International Conference on Futuristic Technologies (INCOFT)*. IEEE, 2023, pp. 1–4.

[11] T. Hayasaka, A. Lin, V. C. Copa, L. P. Lopez Jr, R. A. Loberternos, L. I. M. Ballesteros, Y. Kubota, Y. Liu, A. A. Salvador, and L. Lin, "An electronic nose using a single graphene fet and machine learning for water, methanol, and ethanol," *Microsystems & nanoengineering*, vol. 6, no. 1, p. 50, 2020.

[12] K. Ko, J. K. Lee, M. Kang, J. Jeon, and H. Shin, "Prediction of process variation effect for ultrascaled gaa vertical fet devices using a machine learning approach," *IEEE Transactions on Electron Devices*, vol. 66, no. 10, pp. 4474–4477, 2019.

[13] T. Wu and J. Guo, "Multiobjective design of 2-d-material-based field-effect transistors with machine learning methods," *IEEE Transactions on Electron Devices*, vol. 68, no. 11, pp. 5476–5482, 2021.

[14] Y. Pathak, P. Mishra, M. Sharma, S. Solanki, V. V. Agarwal, R. Chaujar, and B. D. Malhotra, "Experimental circuit design and tcad analysis of ion sensitive field effect transistor (isfet) for ph sensing," *Materials Science and Engineering: B*, vol. 299, p. 116951, 2024.

[15] Y. Pathak, B. D. Malhotra, and R. Chaujar, "Detection of biomolecules in dielectric modulated double metal below ferroelectric layer fet with improved sensitivity," *Journal of Materials Science: Materials in Electronics*, vol. 33, no. 17, pp. 13 558–13 567, 2022.

[16] G. Pahwa, T. Dutta, A. Agarwal, and Y. S. Chauhan, "Physical insights on negative capacitance transistors in nonhysteresis and hysteresis regimes: Mfmis versus mfis structures," *IEEE Transactions on Electron Devices*, vol. 65, no. 3, pp. 867–873, 2018.

[17] G. Pahwa, A. Agarwal, and Y. S. Chauhan, "Numerical investigation of short-channel effects in negative capacitance mfis and mfmis transistors: Subthreshold behavior," *IEEE Transactions on Electron Devices*, vol. 65, no. 11, pp. 5130–5136, 2018.

[18] K. Hara, D. Saito, and H. Shouno, "Analysis of function of rectified linear unit used in deep learning," in *2015 international joint conference on neural networks (IJCNN)*. IEEE, 2015, pp. 1–8.

[19] S. Das, K. Saxena, J.-C. Tinguely, A. Pal, N. L. Wickramasinghe, A. Khezri, V. Dubey, A. Ahmad, V. Perumal, R. Ahmad, *et al.*, "Sers nanowire chip and machine learning-enabled classification of wild-type and antibiotic-resistant bacteria at species and strain levels," *ACS Applied Materials & Interfaces*, vol. 15, no. 20, pp. 24 047–24 058, 2023.

[20] T. Kessler, G. Dorian, and J. H. Mack, "Application of a rectified linear unit (relu) based artificial neural network to cetane number predictions," in *Internal Combustion Engine Division Fall Technical Conference*, vol. 58318. American Society of Mechanical Engineers, 2017, p. V001T02A006.

ANNALI DELL'UNIVERSITA' DI FERRARA (2025) 71:18
https://doi.org/10.1007/s11565-024-00562-0

Bivariate generalization for operators involving Appostol-Genocchi polynomial

Km. Lipi[1] · **Naokant Deo**[1]

Received: 21 June 2021 / Accepted: 6 August 2024
© The Author(s) under exclusive license to Università degli Studi di Ferrara 2024

Abstract
This article is primarily concerned with the bivariate generalization of operators involving a class of orthogonal polynomials called Apostol-Genocchi polynomials. The rate of convergence can be determined in terms of partial and total modulus of continuity as well as the order of approximation can be achieved by means of a Lipschitz-type function and Peetre's K-functional. In addition, we put forth a conceptual extension known as the "generalized boolean sum (GBS)" for these bivariate operators, which aims to establish the degree of approximation for Bögel continuous functions. In this study, we utilize the Mathematica Software to present a series of graphical illustrations that effectively showcase the rate of convergence for the bivariate operators. The graphs indicate that, in the case of certain functions, the bivariate operators exhibits superior convergence when α is less than β. Based on our analysis and comparison of the error of approximation between the bivariate operators and the corresponding GBS operators, it can be deduced that the GBS operators exhibit a faster convergence towards the function.

Keywords Apostol-Genocchi polynomial · GBS operators

Mathematics Subject Classification 41A25 · 41A36

✉ Naokant Deo
naokantdeo@dce.ac.in

Km. Lipi
chaudhary.lipi123@gmail.com

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

Published online: 04 December 2024

1 Introduction

The Genocchi polynomials $g_k(x)$ are generally proposed by means of the following generating functions

$$\frac{2te^{xt}}{e^t+1}=\sum_{k=0}^{\infty} g_k\,(x)\,\frac{t^k}{k!}.$$

In particular, $g_k := g_k(0)$ for $n \geq 0$ are called Genocchi numbers with $g_{2k+1} = 0$ for $k \geq 1$.

Luo [20], extended the Genocchi polynomial in x. These extended polynomials are called Apostol-Genocchi polynomials $g_k(x;\lambda)$ and given by the means of the following generating functions:

$$\frac{2te^{xt}}{\lambda e^t+1}=\sum_{k=0}^{\infty} g_k\,(x;\lambda)\,\frac{t^k}{k!}\,(|t|<|\log(-\lambda)|)\,.$$

Also (see [18]) Apostol-Genocchi polynomials $g_k^{(a)}(x;\lambda)$ of order α in variable x are proposed by means of the generating function:

$$\left(\frac{2t}{1+\beta e^t}\right)^{\alpha} e^{xt}=\sum_{k=0}^{\infty} g_k^{\alpha}\,(x;\beta)\,\frac{t^k}{k!},\qquad (|t|<\pi). \tag{1.1}$$

For $f \in C[0,\infty)$, Prakash et al. [25] considered the following operators:

$$G_n^{\alpha,\beta}(f;x)=e^{-nx}\left(\frac{1+e\beta}{2}\right)^{\alpha}\sum_{k=0}^{\infty}\frac{g_k^{\alpha}\,(nx;\beta)}{k!}f\left(\frac{k}{n}\right), \tag{1.2}$$

where $g_k^{\alpha}\,(x;\beta)$ is generalized Apostol-Genocchi polynomials with generating function given in (1.1). The Apostol-Genocchi polynomials and their properties are studied by many researchers. For details, readers can refer (cf. [20, 21, 23, 26]).

In [21], the following explicit series representation for the Apostol-Genocchi polynomials is given as:

$$g_k^{\alpha}\,(x;\beta)=2^{\alpha}\alpha!\binom{k}{\alpha}\sum_{i=0}^{k-\alpha}\frac{\beta^i}{(1+\beta)^{\alpha+i}}\binom{k-\alpha}{i}\binom{\alpha+i-1}{i}$$
$$\times\sum_{j=0}^{i}(-1)^j\binom{i}{j}j^i(x+j)^{k-i-\alpha}\,{}_2F_1[\alpha+i-k,i;i+1;j/(x+j)],$$

where $k,\alpha \in \mathbb{N}\cup\{0\}$, $\beta \in \mathbb{R}\setminus\{-1\}$ and ${}_2F_1[a,b;c;x]$ indicates the Gaussian hypergeometric function given by

$$ {}_2F_1[a, b; c; x] = {}_2F_1[b, a; c; x] = \sum_{j=0}^{\infty} \frac{(a)_j (b)_j}{(c)_j} \frac{x^j}{j!}, $$

where $(a)_0 = 1$, $(a)_j = a(a+1)\cdots(a+j-1) = \frac{\Gamma(j+a)}{\Gamma(a)}$

Deo and kumar [12] introduced Durrmeyer variant of operators (1.2) and gave approximation results. Here we propose to study the bivariate generalization of the Apostol-Genocchi operators defined in (1.2). In last couple of years, researchers have proposed bivariate of various operators and demonstrated their convergence behaviour. Readers can go through these papers ([2, 11, 14, 17, 19, 22, 24]) to get some knowledge about this topic.

We have arranged the whole paper in the following order. We first define the bivariate generalization of operators (1.2) followed by results on order of approximation. Next, we construct the associated GBS and estimate the rate of convergence for these operators. Lastly, we show a few graphical examples to demonstrate the rate of convergence for these operators.

2 Construction of bivariate operators

Let $C(I)$ be the class of all real valued continuous functions on $I = [0, \infty] \times [0.\infty]$. For a function $f \in C(I)$ the bivariate generalization of the operators (1.2) is defined as:

$$ G^{\alpha,\beta}_{n_1,n_2}(f; x_1, x_2) = e^{-(n_1x_1+n_2x_2)} \left(\frac{1+e\beta}{2}\right)^{2\alpha} \times \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \frac{g^{\alpha}_{k_1}(n_1x_1; \beta)}{k_1!} \frac{g^{\alpha}_{k_2}(n_2x_2; \beta)}{k_2!} f\left(\frac{k_1}{n_1}, \frac{k_2}{n_2}\right). \tag{2.1} $$

Some preliminaries for the above-mentioned operators which are useful throughout the paper are as follows:

Lemma 2.1 *Let $e_{rs}(t_1, t_2) = t_1^r t_2^s$, $0 \le r + s \le 2$. For $(x_1, x_2) \in I$, we have*

$$ G^{\alpha,\beta}_{n_1,n_2}(1; x_1, x_2) = 1; $$

$$ G^{\alpha,\beta}_{n_1,n_2}(t_1; x_1, x_2) = x_1 + \frac{\alpha}{n_1(1+e\beta)}; $$

$$ G^{\alpha,\beta}_{n_1,n_2}(t_2; x_1, x_2) = x_2 + \frac{\alpha}{n_2(1+e\beta)}; $$

$$ G^{\alpha,\beta}_{n_1,n_2}\left(t_1^2; x_1, x_2\right) = x_1^2 + \frac{(1+2\alpha+e\beta)}{n_1(1+e\beta)} x_1 + \frac{\alpha^2 - 2\alpha e\beta - \alpha e^2\beta^2}{n_1{}^2(1+e\beta)^2}; $$

$$ G^{\alpha,\beta}_{n_1,n_2}\left(t_2^2; x_1, x_2\right) = x_2^2 + \frac{(1+2\alpha+e\beta)}{n_2(1+e\beta)} x_2 + \frac{\alpha^2 - 2\alpha e\beta - \alpha e^2\beta^2}{n_2{}^2(1+e\beta)^2}. $$

Remark 2.1 Using lemma 2.1, we have

$$G^{\alpha,\beta}_{n_1,n_2}\left((t_1-x_1);x_1,x_2\right)=\frac{\alpha}{n_1\left(1+e\beta\right)};$$
$$G^{\alpha,\beta}_{n_1,n_2}\left((t_2-x_2);x_1,x_2\right)=\frac{\alpha}{n_2\left(1+e\beta\right)};$$
$$G^{\alpha,\beta}_{n_1,n_2}\left((t_1-x_1)^2;x_1,x_2\right)=\frac{x_1}{n_1}+\frac{\alpha^2-2\alpha e\beta-\alpha e^2\beta^2}{{n_1}^2(1+e\beta)^2};$$
$$G^{\alpha,\beta}_{n_1,n_2}\left((t_2-x_2)^2;x_1,x_2\right)=\frac{x_2}{n_2}+\frac{\alpha^2-2\alpha e\beta-\alpha e^2\beta^2}{{n_2}^2(1+e\beta)^2}.$$

Furthermore,

$$\lim_{n_1\to\infty} n_1 G^{\alpha,\beta}_{n_1,n_2}((t_1-x_1);x_1)=\frac{\alpha}{(1+e\beta)} \text{ and}$$
$$\lim_{n_2\to\infty} n_2 G^{\alpha,\beta}_{n_1,n_2}(t_2-x_2);x_2)=\frac{\alpha}{(1+e\beta)},$$
$$\lim_{n_1\to\infty} n_1(G^{\alpha,\beta}_{n_1,n_2}(t_1-x_1)^2;x_1)=x_1 \text{ and } \lim_{n_2\to\infty} n_2(G^{\alpha,\beta}_{n_1,n_2}(t_2-x_2)^2;x_2)=x_2.$$

3 Main results

We determine the order of approximation of operators (2.1) in the space of continuous functions on the set $I_{\mu\kappa}=[0,\mu]\times[0,\kappa]$. For $f\in C(I_{\mu\kappa})$ the total modulus of continuity in case of functions of two variable is given by:

$$\omega_{total}(f;\delta_{n_1},\delta_{n_2})=\sup\{|f(t_1,t_2)-f(x_1,x_2)|:(t_1,t_2),(x_1,x_2)\in I_{\mu\kappa},|t_1-x_1|\le\delta_{n_1},|t_2-x_2|\le\delta_{n_2}\}.$$

Furthermore, the partial modulus of continuity is given as:

$$\omega^1(f;\delta)=\sup\{|f(x',x_2)-f(x'',x_2)|:|x'-x''|\le\delta,0\le x_2\le\kappa\}$$

and

$$\omega^2(f;\delta)=\sup\{|f(x_1,y')-f(x_1,y'')|:|y'-y''|\le\delta,0\le x_1\le\mu\}.$$

Both partial and total modulus of continuity and usual modulus of continuity holds the same properties and can be explored in [3].

Theorem 3.1 *If $f\in C(I)$, then*

$$\lim_{(n_1,n_2)\to\infty} G^{\alpha,\beta}_{n_1,n_2}\left(f;x_1,x_2\right)=f\left(x_1,x_2\right).$$

The above convergence is uniform in each compact $I_{\mu\kappa}$.

Proof The proof follows from the Volkov Theorem [27]. □

Theorem 3.2 *Let $f \in C(I)$, then for all $(x_1, x_2) \in I$, the inequality*

$$(i)\ \left|G^{\alpha,\beta}_{n_1,n_2}(f; x_1, x_2) - f(x_1, x_2)\right| \leq 2\left(\omega^1(f; \delta_{n_1}) + \omega^2\left(f; \delta_{n_2}\right)\right),$$
$$(ii)\ \left|G^{\alpha,\beta}_{n_1,n_2}(f; x_1, x_2) - f(x_1, x_2)\right| \leq 4\omega\left(f; \delta_{n_1}, \delta_{n_2}\right),$$

holds true. Where $\delta_{n_1}(x_1) = \sqrt{G^{\alpha,\beta}_{n_1}\left((t_1 - x_1)^2; x_1\right)}$ *and* $\delta_{n_2}(x_2) = \sqrt{G^{\alpha,\beta}_{n_2}\left((t_2 - x_2)^2; x_2\right)}$.

Proof (i) Taking into consideration the partial modulus of continuity of $f(x_1, x_2)$ and applying Cauchy-Schwarz inequality, we get

$$\begin{aligned}
\left|G^{\alpha,\beta}_{n_1,n_2}(f; x_1, x_2) - f(x_1, x_2)\right| &\leq G^{\alpha,\beta}_{n_1,n_2}\left(|f(t_1, t_2) - f(x_1, x_2)|; x_1, x_2\right) \\
&\leq G^{\alpha,\beta}_{n_1,n_2}\left(|f(t_1, t_2) - f(x_1, t_2)|; x_1, x_2\right) \\
&\quad + G^{\alpha,\beta}_{n_1,n_2}\left(|f(x_1, t_2) - f(x_1, x_2)|; x_1, x_2\right) \\
&\leq G^{\alpha,\beta}_{n_1,n_2}\left(\omega^1(f; |t_1 - x_1|); x_1, x_2\right) \\
&\quad + G^{\alpha,\beta}_{n_1,n_2}\left(\omega^2(f; |t_2 - x_2|); x_1, x_2\right) \\
&\leq \omega^1\left(f; \delta_{n_1}\right)\left(1 + \frac{1}{\delta_{n_1}} G^{\alpha,\beta}_{n_1}\left(|t_1 - x_1|; x_1\right)\right) \\
&\quad + \omega^2\left(f; \delta_{n_2}\right)\left(1 + \frac{1}{\delta_{n_2}} G^{\alpha,\beta}_{n_2}\left(|t_2 - x_2|; x_2\right)\right) \\
&\leq \omega^1\left(f; \delta_{n_1}\right)\left(1 + \frac{1}{\delta_{n_1}} G^{\alpha,\beta}_{n_1}\left((t_1 - x_1)^2; x_1\right)^{\frac{1}{2}}\right) \\
&\quad + \omega^2\left(f; \delta_{n_2}\right)\left(1 + \frac{1}{\delta_{n_2}} G^{\alpha,\beta}_{n_2}\left((t_2 - x_2)^2; x_2\right)^{\frac{1}{2}}\right) \\
&\leq 2\left(\omega^1(f; \delta_{n_1}) + \omega^2\left(f; \delta_{n_2}\right)\right),
\end{aligned}$$

(ii) Taking into consideration linearity of $G^{\alpha,\beta}_{n_1}$, $G^{\alpha,\beta}_{n_2}$ and monotonicity of ω, we can write

$$\begin{aligned}
\left|G^{\alpha,\beta}_{n_1,n_2}(f; x_1, x_2) - f(x_1, x_2)\right| &\leq G^{\alpha,\beta}_{n_1,n_2}\left(|f(t_1, t_2) - f(x_1, x_2)|; x_1, x_2\right) \\
&\leq \omega\left(f; \delta_{n_1}, \delta_{n_2}\right)\left(G^{\alpha,\beta}_{n_1}(1; x_1) + \frac{1}{\delta_{n_1}} G^{\alpha,\beta}_{n_1}\left(|t_1 - x_1|; x_1\right)\right) \\
&\quad \times \left(G^{\alpha,\beta}_{n_2}(1; x_2) + \frac{1}{\delta_{n_2}} G^{\alpha,\beta}_{n_2}\left(|t_2 - x_2|; x_2\right)\right).
\end{aligned}$$

Applying Cauchy Schwarz inequality, we have

$$\begin{aligned}
\left|G^{\alpha,\beta}_{n_1,n_2}(f; x_1, x_2) - f(x_1, x_2)\right| &\leq \omega\left(f; \delta_{n_1}, \delta_{n_2}\right) \\
&\left(G^{\alpha,\beta}_{n_1}(1; x_1) + \frac{1}{\delta_{n_1}} G^{\alpha,\beta}_{n_1}\left((t_1 - x_1)^2; x_1\right)^{\frac{1}{2}}\right)
\end{aligned}$$

$$\times \left(G_{n_2}^{\alpha,\beta}(1; x_2) + \frac{1}{\delta_{n_2}} G_{n_2}^{\alpha,\beta}\left((t_2 - x_2)^2; x_2\right)^{\frac{1}{2}} \right).$$

Therefore, the desired result is achieved by selecting $\delta_{n_1}(x_1)$ and $\delta_{n_2}(x_2)$ as mentioned in the statement.

□

Now we determine the order of approximation of bivariate operators (2.1) with the help of Lipschitz class functions. We define Lipschitz class $Lip_{\mathcal{M}}(f; \zeta_1, \zeta_2)$ for bivariate functions for $0 < \zeta_1 \leq 1$ and $0 < \zeta_2 \leq 1$ as follows:

$$Lip_{\mathcal{M}}(f; \zeta_1, \zeta_2) = \left\{ f : |f(t_1, t_2) - f(x_1, x_2)| \leq \mathcal{M}|t_1 - x_1|^{\zeta_1}|t_2 - x_2|^{\zeta_2} \right\}.$$

Theorem 3.3 *If $f \in Lip_{\mathcal{M}}(f; \zeta_1, \zeta_2)$, then for $\zeta_1, \zeta_2 \in (0, 1]$*

$$\left| G_{n_1,n_2}^{\alpha,\beta}(f; x_1, x_2) - f(x_1, x_2) \right| \leq \mathcal{M}\delta_{n_1}^{\zeta_1}\delta_{n_2}^{\zeta_2}.$$

Proof For $f \in Lip_{\mathcal{M}}(f; \zeta_1, \zeta_2)$ and making use of Hölder's inequality, we have

$$\begin{aligned}
&\left| G_{n_1,n_2}^{\alpha,\beta}(f; x_1, x_2) - f(x_1, x_2) \right| \\
&\leq G_{n_1,n_2}^{\alpha,\beta}(|f(t_1, t_2) - f(x_1, x_2)|; x_1, x_2) \\
&\leq \mathcal{M} G_{n_1}^{\alpha,\beta}\left(|t_1 - x_1|^{\zeta_1}; x_1, x_2\right) G_{n_2}^{\alpha,\beta}\left(|t_2 - x_2|^{\zeta_2}; x_1, x_2\right) \\
&\leq \mathcal{M}\left(G_{n_1}^{\alpha,\beta}\left((t_1 - x_1)^2; x_1, x_2\right)\right)^{\frac{\zeta_1}{2}} \left(G_{n_1}^{\alpha,\beta}(1; x_1, x_2)\right)^{\frac{2-\zeta_1}{2}} \\
&\times \left(G_{n_2}^{\alpha,\beta}\left((t_2 - x_2)^2; x_1, x_2\right)\right)^{\frac{\zeta_2}{2}} \left(G_{n_2}^{\alpha,\beta}(1; x_1, x_2)\right)^{\frac{2-\zeta_2}{2}} \\
&\leq \mathcal{M}\delta_{n_1}^{\zeta_1}\delta_{n_2}^{\zeta_2}.
\end{aligned}$$

□

Theorem 3.4 *Let $f \in C^1(I_{\mu\kappa})$. Then, for any $(x_1, x_2) \in I_{\mu\kappa}$, we have the inequality*

$$\left| G_{n_1,n_2}^{\alpha,\beta}(f; x_1, x_2) - f(x_1, x_2) \right| \leq \delta_{n_1}(x_1) \left\| f'_{x_1} \right\|_{C(I_{\mu\kappa})} + \delta_{n_2}(x_2) \left\| f'_{x_2} \right\|_{C(I_{\mu\kappa})}.$$

Proof For $f \in C^1(I_{\mu\kappa})$, and $(x_1, x_2) \in I_{\mu\kappa}$, we have

$$f(t_1, t_2) - f(x_1, x_2) = \int_{x_1}^{t_1} f'_u(u, t_2)\, du + \int_{x_2}^{t_2} f'_v(x_1, v)\, dv \quad \text{for } (t_1, t_2) \in I_{\mu\kappa}.$$

Applying $G_{n_1,n_2}^{\alpha,\beta}$ to both sides, we obtain

$$G_{n_1,n_2}^{\alpha,\beta}(f; x_1, x_2) - f(x_1, x_2)$$

$$\leq G_{n_1,n_2}^{\alpha,\beta}\left(\int_{x_1}^{t_1} f_u^{'}(u,t_2)\,du;x_1,x_2\right)+G_{n_1,n_2}^{\alpha,\beta}\left(\int_{x_2}^{t_2} f_v^{'}(x_1,v)\,dv\right).$$

Now, using the sup norm on $I_{\mu\kappa}$, we get

$$\left|\int_{x_1}^{t_1} f_u^{'}(u,t_2)\,du\right|\leq |t_1-x_1|\left\|f_{x_1}^{'}\right\|_{C(I_{\mu\kappa})}$$

and

$$\left|\int_{x_2}^{t_2} f_v^{'}(x_1,v)\,dv\right|\leq |t_2-x_2|\left\|f_{x_2}^{'}\right\|_{C(I_{\mu\kappa})}.$$

By using these inequalities, we have

$$\begin{aligned}\left|G_{n_1,n_2}^{\alpha,\beta}(f;x_1,x_2)-f(x_1,x_2)\right|&\leq G_{n_1,n_2}^{\alpha,\beta}\left(\left|\int_{x_1}^{t_1} f_u^{'}(u,t_2)du\right|;x_1,x_2\right)\\&+G_{n_1,n_2}^{\alpha,\beta}\left(\left|\int_{x_2}^{t_2} f_v^{'}(x_1,v)dv\right|;x_1,x_2\right)\\&\leq G_{n_1,n_2}^{\alpha,\beta}(|t_1-x_1|;x_1,x_2)\left\|f_{x_1}^{'}\right\|_{C(I_{\mu\kappa})}+G_{n_1,n_2}^{\alpha,\beta}(|t_2-x_2|;x_1,x_2)\left\|f_{x_2}^{'}\right\|_{C(I_{\mu\kappa})}.\end{aligned}$$

Using Cauchy Schwarz inequality, we get required result. □

Next result is a Voronovoskaya type estimate for the operators (2.1).

Theorem 3.5 *For a function f, which is differentiable two times on the interval I, we have*

$$\begin{aligned}\lim_{n\to\infty} n\left(G_{n,n}^{\alpha,\beta}(f;x_1,x_2)-f(x_1,x_2)\right)=&\frac{\alpha}{(1+e\beta)}\left(f_{x_1}(x_1,x_2)+f_{x_2}(x_1,x_2)\right)\\&+x_1 f_{x_1x_1}(x_1,x_2)+x_2 f_{x_2x_2}(x_1,x_2),\end{aligned}$$

uniformly on $I_{\mu\kappa}$.

Proof Let $(x_1,x_2)\in I$ be arbitrary. Then by Taylor's theorem we have

$$\begin{aligned}f(t_1,t_2)=&f(x_1,x_2)+f_{x_1}(x_1,x_2)(t_1-x_1)+f_{x_2}(x_1,x_2)(t_2-x_2)\\&+\frac{1}{2}\{f_{x_1x_1}(x_1,x_2)(t_1-x_1)^2+2f_{x_1x_2}(x_1,x_2)(t_1-x_1)(t_2-x_2)\\&+f_{x_2x_2}(x_1,x_2)(t_2-x_2)^2\}+\psi(t_1,t_2;x_1,x_2)\sqrt{(t_1-x_1)^4+(t_2-x_2)^4},\end{aligned}\tag{3.1}$$

where $\psi(t_1,t_2;x_1,x_2)\in C(I_{\mu\kappa})$ and $\psi(t_1,t_2;x_1,x_2)\to 0$ as $(t_1,t_2)\to(x_1,x_2)$.

Applying $G_{n,n}^{\alpha,\beta}(f;x_1,x_2)$ on both sides of (3.1), we get

$$\begin{aligned} G_{n,n}^{\alpha,\beta}(f;x_1,x_2) =& f(x_1,x_2) + f_{x_1}(x_1,x_2)G_n^{\alpha,\beta}((t_1-x_1);x_1) \\ &+ f_{x_2}(x_1,x_2)G_n^{\alpha,\beta}((t_2-x_2);x_1) \\ &+ \frac{1}{2}\{f_{x_1x_1}(x_1,x_2)G_n^{\alpha,\beta}((t_1-x_1)^2;x_1) \\ &+ f_{x_2x_2}(x_1,x_2)G_n^{\alpha,\beta}((t_2-x_2)^2;x_2) \\ &+ 2f_{x_1x_2}(x_1,x_2)G_{n,n}^{\alpha,\beta}((t_1-x_1)(t_2-x_2);x_1,x_2)\} \\ &+ G_{n,n}^{\alpha,\beta}\left(\psi(t_1,t_2;x_1,x_2)\sqrt{(t_1-x_1)^4+(t_2-x_2)^4};x_1,x_2\right). \end{aligned} \tag{3.2}$$

Now, using Hölder's Inequality, we have

$$\begin{aligned} &\left|G_{n,n}^{\alpha,\beta}\left(\psi(t_1,t_2;x_1,x_2)\sqrt{(t_1-x_1)^4+(t_2-x_2)^4};x_1,x_2\right)\right| \\ &\le \left\{G_{n,n}^{\alpha,\beta}(\psi^2(t_1,t_2;x_1,x_2);x_1,x_2)\right\}^{1/2}\left\{G_{n,n}^{\alpha,\beta}((t_1-x_1)^4+(t_2-x_2)^4);x_1,x_2)\right\}^{1/2} \\ &\le \left\{G_{n,n}^{\alpha,\beta}(\psi^2(t_1,t_2;x_1,x_2);x_1,x_2)\right\}^{1/2}\left\{G_n^{\alpha,\beta}((t_1-x_1)^4;x_1)+G_n^{\alpha,\beta}((t_2-x_2)^4;x_2)\right\}^{1/2}. \end{aligned}$$

In view of Theorem (3.1), $G_{n,n}^{\alpha,\beta}(\psi^2(t_1,t_2;x_1,x_2);x_1,x_2)\to 0$ as $n\to\infty$ uniformly on I and since $G_n^{\alpha,\beta}((t_1-x_1)^4;x_1) = O\left(\frac{1}{n^2}\right)$, $G_n^{\alpha,\beta}((t_2-x_2)^4;x_1) = O\left(\frac{1}{n^2}\right)$, therefore we have

$$\lim_{n\to\infty} nG_{n_1,n_2}^{\alpha,\beta}\left(\psi(t_1,t_2;x_1,x_2)\sqrt{(t_1-x_1)^4+(t_2-x_2)^4};x_1,x_2\right)=0,$$

uniformly on I. Also using Remark (2.1)

$$\lim_{n\to\infty} nG_n^{\alpha,\beta}((t_1-x_1);x_1) = \frac{\alpha}{(1+e\beta)},\ \lim_{n_2\to\infty} n(G_n^{\alpha,\beta}(t_2-x_2);x_2) = \frac{\alpha}{(1+e\beta)},$$
$$\lim_{n\to\infty} n(G_n^{\alpha,\beta}(t_1-x_1)^2;x_1) = x_1,\ \lim_{n\to\infty} n(G_n^{\alpha,\beta}(t_2-x_2)^2;x_2) = x_2$$

and

$$\lim_{n\to\infty} nG_n^{\alpha,\beta}((t_1-x_1);x_1)G_n^{\alpha,\beta}((t_2-x_2);x_2)=0.$$

Above estimate together with (3.2) gives the desired result. □

In the next result, we use Peetre's K- functional to determine the rate of convergence of bivariate operators (2.1). Let $C^2(I_{\mu\kappa})$ denote the set of all continuous functions on $I_{\mu\kappa} = [0,\mu]\times[0,\kappa]$, which is continuous and have first and second derivatives on

$I_{\mu\kappa}$. The norm is defined as

$$\|\hbar\|_{C^2(I_{\mu\kappa})} = \|\hbar\|_{C(I_{\mu\kappa})} + \sum_{i=1}^{2}\left(\left\|\frac{\partial^i \hbar}{\partial x_1^i}\right\|_{C(I_{\mu\kappa})} + \left\|\frac{\partial^i \hbar}{\partial x_2^i}\right\|_{C(I_{\mu\kappa})}\right).$$

In [10], the second order modulus of continuity and Peetre's K-functional for $\hbar \in C^2(I_{\mu\kappa})$ is defined as:

$$\varpi_2\left(\hbar, \sqrt{\sigma}\right) = \sup\left\{\left|\sum_{i=0}^{2}(-1)^{2-i}\hbar\left(x_1 + ix_1', x_2 + ix_2'\right)\right| : (x_1, x_2), \left(x_1 + ix_1{}', x_2 + ix_2{}'\right) \in I_{\mu\kappa}, \left|x_1'\right| \le \sigma, \left|x_2'\right| \le \sigma\right\}$$

and

$$K(\hbar; \sigma) = \inf_{t \in C^2(I_{\mu\kappa})}\left\{\|\hbar - t\|_{C^2(I_{\mu\kappa})} + \sigma\|t\|_{C^2(I_{\mu\kappa})}, \sigma > 0\right\}$$

respectively.

Theorem 3.6 *For $f \in C(I_{\mu\kappa})$, the following inequality*

$$\left|\tilde{G}_{n_1,n_2}^{\alpha,\beta}(g; x_1, x_2) - g(x_1, x_2)\right| \le 4K(f; \sigma_{n_1,n_2}(x_1, x_2)) + \varpi_2\left(f; \sqrt{\rho_{n_1,n_2}(x_1, x_2)}\right),$$

where $\rho_{n_1,n_2}(x_1, x_2) = \left(\frac{\alpha}{n_1(1+e\beta)}\right)^2 + \left(\frac{\alpha}{n_2(1+e\beta)}\right)^2$

and $\sigma_{n_1,n_2}(x_1, x_2) = \delta_{n_1}^2(x_1) + \delta_{n_2}^2(x_2) + \rho_{n_1,n_2}(x_1, x_2)$ *holds true.*

Proof Before initiating the proof, we define the following modified operators:

$$\tilde{G}_{n_1,n_2}^{\alpha,\beta}(g; x_1, x_2) = G_{n_1,n_2}^{\alpha,\beta}(g; x_1, x_2) - g\left(x_1 + \frac{\alpha}{n_1(1+e\beta)}, x_2 + \frac{\alpha}{n_2(1+e\beta)}\right) + g(x_1, x_2). \tag{3.3}$$

Then using Remark (2.1), we have

$$\tilde{G}_{n_1,n_2}^{\alpha,\beta}((t_1 - x_1); x_1, x_2) = 0, \ \tilde{G}_{n_1,n_2}^{\alpha,\beta}((t_2 - x_2); x_1, x_2) = 0. \tag{3.4}$$

Let $h \in C^2(I_{\mu\kappa})$, using Taylor's theorem, we can write

$$\begin{aligned} h(t_1, t_2) - h(x_1, x_2) &= h(t_1, x_2) - h(x_1, x_2) + h(t_1, t_2) - h(t_1, x_2) \\ &= \frac{\partial h(x_1, x_2)}{\partial x_1}(t_1 - x_1) + \int_{x_1}^{t_1}(t_1 - u)\frac{\partial^2 h(u, x_2)}{\partial u^2}du \end{aligned}$$

$$+\frac{\partial h(x_1,x_2)}{\partial x_2}(t_2-x_2)+\int_{x_2}^{t_2}(t_2-v)\frac{\partial^2 h(x_1,v)}{\partial v^2}dv. \quad (3.5)$$

Applying $\tilde{G}^{\alpha,\beta}_{n_1,n_2}$ on both sides of (3.5) and using (3.4), we get

$$\tilde{G}^{\alpha,\beta}_{n_1,n_2}(h;x_1,x_2)-h(x_1,x_2)=\tilde{G}^{\alpha,\beta}_{n_1,n_2}\left(\int_{x_1}^{t_1}(t_1-u)\frac{\partial^2 h(u,x_2)}{\partial u^2}du;x_1,x_2\right)$$
$$+\tilde{G}^{\alpha,\beta}_{n_1,n_2}(\int_{x_2}^{t_2}(t_2-v)\frac{\partial^2 h(x_1,v)}{\partial v^2}dv;x_1,x_2).$$

Hence

$$\left|\tilde{G}^{\alpha,\beta}_{n_1,n_2}(h;x_1,x_2)-h(x_1,x_2)\right|\le G^{\alpha,\beta}_{n_1,n_2}\left(\int_{x_1}^{t_1}|t_1-u|\left|\frac{\partial^2 h(u,x_2)}{\partial u^2}\right|du;x_1,x_2\right)$$
$$+\int_{x_1}^{x_1+\frac{\alpha}{n_1(1+e\beta)}}\left|x_1+\frac{\alpha}{n_1(1+e\beta)}-u\right|\left|\frac{\partial^2 h(u,x_2)}{\partial u^2}\right|du$$
$$+G^{\alpha,\beta}_{n_1,n_2}(\int_{x_2}^{t_2}|t_2-v|\left|\frac{\partial^2 h(x_1,v)}{\partial v^2}\right|dv;x_1,x_2)$$
$$+\int_{x_2}^{x_2+\frac{\alpha}{n_2(1+e\beta)}}\left|x_2+\frac{\alpha}{n_2(1+e\beta)}-v\right|\left|\frac{\partial^2 h(x_1,v)}{\partial v^2}\right|dv$$
$$\le\left\{G^{\alpha,\beta}_{n_1,n_2}((t_1-x_1)^2;x_1,x_2)+\left(\frac{\alpha}{n_1(1+e\beta)}\right)^2\right\}||h||_{C^2(I_{\mu\kappa})}$$
$$+\left\{G^{\alpha,\beta}_{n_1,n_2}((t_2-x_2)^2;x_1,x_2)+\left(\frac{\alpha}{n_2(1+e\beta)}\right)^2\right\}||h||_{C^2(I_{\mu\kappa})}$$
$$\le\left\{\delta^2_{n_1}(x_1)+\delta^2_{n_2}(x_2)+\left(\frac{\alpha}{n_1(1+e\beta)}\right)^2+\left(\frac{\alpha}{n_2(1+e\beta)}\right)^2\right\}\|h\|_{C^2(I_{\mu\kappa})}.$$

Also

$$\left|\tilde{G}^{\alpha,\beta}_{n_1,n_2}(g;x_1,x_2)\right|\le 3||g||_{C(I_{\mu\kappa})}. \quad (3.6)$$

In view of (3.6), we can write

$$\left|\tilde{G}^{\alpha,\beta}_{n_1,n_2}(g;x_1,x_2)-g(x_1,x_2)\right|\le\left|\tilde{G}^{\alpha,\beta}_{n_1,n_2}(g-h;x_1,x_2)\right|$$
$$+\left|\tilde{G}^{\alpha,\beta}_{n_1,n_2}(h;x_1,x_2)-h(x_1,x_2)\right|+|h(x_1,x_2)-g(x_1,x_2)|$$
$$+\left|g\left(x_1+\frac{\alpha}{n_1(1+e\beta)},x_2+\frac{\alpha}{n_2(1+e\beta)}\right)-g(x_1,x_2)\right|$$
$$\le\left(4||g-h||_{C(I_{\mu\kappa})}+\sigma_{n_1,n_2}(x_1,x_2)||h||_{C^2(I_{\mu\kappa})}\right)$$
$$+\varpi_2\left(f;\sqrt{\rho_{n_1,n_2}(x_1,x_2)}\right).$$

Taking the infimum on the right hand side over all $h \in C^2(I_{\mu\kappa})$, it follows

$$\left|\tilde{G}^{\alpha,\beta}_{n_1,n_2}(g;x_1,x_2) - g(x_1,x_2)\right| \leq 4K(f;\sigma_{n_1,n_2}(x_1,x_2)) + \varpi_2\left(f;\sqrt{\rho_{n_1,n_2}(x_1,x_2)}\right),$$

which completes the proof. □

4 Associated GBS operators

In the past few decades, GBS operators have gained substantial importance amongst researchers to study functions in Bögel space. In 1934 Bögel ([8, 9]) developed the study of B-continuous and B-differentiable functions. Later, Dobrescu and Matei [13] illustrated that the convergence of GBS operators corresponding to the bivariate Bernstein polynomial is uniform. Badea and Cottin [6], instituted Korovkin-type theorem for GBS operators. Thereafter Badea et al. [4] proved the well known "Test function theorem" to approximate these kinds of functions. Badea and Badea [5] for these functions established a quantitative variant of the Korovkin-type theorem. In recent years many researchers (cf. [1, 7, 15, 16]) contributed in the area of approximation theory.

Let $A = I \times J$ where I and J be compact intervals. For any $f : A \to R$ and any$(t_1, t_2)(x_1, x_2) \in A$, let $\triangle_{(t_1,t_2)} f(x_1, x_2)$ be the bivariate mixed difference operators defined as:

$$\triangle_{(t_1,t_2)} f(x_1,x_2) = f(t_1,t_2) - f(t_1,x_2) - f(x_1,t_2) + f(x_1,x_2).$$

The function $f : A \to R$ is $B-$bounded on D if there exists $K > 0$ such that $\left|\triangle_{(t_1,t_2)} f(x_1,x_2)\right| \leq K$ for any $(t_1,t_2),(x_1,x_2) \in A$. We denote by $B_b(A)$, the space of all $B-$bounded functions on A equipped with the norm:

$$\|f\|_B = \sup_{(x_1,x_2),(t_1,t_2)\in A} \left|\triangle_{(t_1,t_2)} f(x_1,x_2)\right|.$$

Further let $C_b(A)$ be the subspace consisting of the functions from $B_b(A)$ and is defined by

$$C_b(A) = \left\{ f | \lim_{(t_1,t_2)\to(x_1,x_2)} \triangle_{(t_1,t_2)} f(x_1,x_2) = 0 \forall (x_1,x_2) \in A \right\}.$$

Here, for a compact subset A, every $B-$continuous function is a B-bounded function.

We denote the space of all $B-$differentiable functions on A by

$$D_b(A) = \left\{ f | f : A \to R, and \lim_{(t_1,t_2)\to(x_1,x_2)} \frac{\triangle_{(t_1,t_2)} f(x_1,x_2)}{(t_1-x_1)(t_2-x_2)} = D_B f(x_1,x_2) < \infty \forall (x_1,x_2) \in A \right\}.$$

The mixed modulus of continuity of $f \in B_b\left(I_{\mu\kappa}\right)$, where $I_{\mu\kappa} = [0,\mu] \times [0,\kappa]$, is the compact subset of A, is the function $\omega_B : [0,\infty) \times [0,\infty) \to \mathbb{R}$, defined by

$$\omega_B\left(f;\delta_1,\delta_2\right) := \sup\left\{\left|\Delta f\left[(t,s);(x_1,x_2)\right]\right| : |x_1 - t_1| < \delta_1, |x_2 - t_2| < \delta_2\right\},$$

for any $(t_1,t_2),(x_1,x_2) \in A$.

We define the GBS of the operators $G^{\alpha,\beta}_{n_1,n_2}$ given by (2.1), for any $f \in C_b\left(I_{\mu\kappa}\right)$ and $n,m \in \mathbb{N}$, by

$$\widehat{G}^{\alpha,\beta}_{n_1,n_2}\left(f(t_1,t_2);x_1,x_2\right) = G^{\alpha,\beta}_{n_1,n_2}\left(f(t_1,x_2) + f(x_1,t_2) - f(t_1,t_2);x_1,x_2\right), \tag{4.1}$$

that is

$$\begin{aligned}
&\widehat{G}^{\alpha,\beta}_{n_1,n_2}\left(f;x_1,x_2\right)\\
&= e^{-(n_1x_1+n_2x_2)}\left(\frac{1+e\beta}{2}\right)^{2\alpha}\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\frac{g^{\alpha}_{k_1}\left(n_1x_1;\beta\right)}{k_1!}\frac{g^{\alpha}_{k_2}\left(n_2x_2;\beta\right)}{k_2!}\left(f\left(\frac{k_1}{n_1},x_2\right)\right.\\
&\left.+f\left(x_1,\frac{k_2}{n_2}\right) - f\left(\frac{k_1}{n_1},\frac{k_2}{n_2}\right)\right).
\end{aligned}$$

Theorem 4.1 *For every $f \in C_b(I_{\mu\kappa})$, at each point $(x_1,x_2) \in I_{\mu\kappa}$, the operators (4.1) verifies the following inequality*

$$\left|\widehat{G}^{\alpha,\beta}_{n_1,n_2}\left(f;x_1,x_2\right) - f(x_1,x_2)\right| \le 4\omega_B\left(f;\delta_{n_1},\delta_{n_2}\right).$$

Proof By using the property of $\omega_B\left(f;\delta_{n_1},\delta_{n_2}\right)$, we have

$$\begin{aligned}
\left|\Delta_{(t_1,t_2)}f\left(x_1,x_2\right)\right| &\le \omega_B\left(f;|t_1-x_1|,|t_2-x_2|\right)\\
&\le \left(1+\frac{|t_1-x_1|}{\delta_{n_1}}\right)\left(1+\frac{|t_2-x_2|}{\delta_{n_2}}\right)\omega_B\left(f;\delta_{n_1},\delta_{n_2}\right).
\end{aligned}$$

Now

$$\begin{aligned}
\left|\widehat{G}^{\alpha,\beta}_{n_1,n_2}\left(f;x_1,x_2\right) - f(x_1,x_2)\right| &\le G^{\alpha,\beta}_{n_1,n_2}\left(\left|\Delta_{(t_1,t_2)}f\left(x_1,x_2\right)\right|;x_1,x_2\right)\\
&\le \left(G^{\alpha,\beta}_{n_1,n_2}\left(1;x_1,x_2\right) + \frac{1}{\delta_{n_1}}\left(G^{\alpha,\beta}_{n_1,n_2}\left((t_1-x_1)^2;x_1,x_2\right)\right)^{1/2}\right.\\
&+ \frac{1}{\delta_{n_2}}\left(G^{\alpha,\beta}_{n_1,n_2}\left((t_2-x_2)^2;x_1,x_2\right)\right)^{1/2}\\
&+ \frac{1}{\delta_{n_1}}\left(G^{\alpha,\beta}_{n_1,n_2}\left((t_1-x_1)^2;x_1,x_2\right)\right)^{1/2}\\
&\left.\frac{1}{\delta_{n_2}}\left(G^{\alpha,\beta}_{n_1,n_2}\left((t_2-x_2)^2;x_1,x_2\right)\right)^{1/2}\right)\omega_B\left(f;\delta_{n_1},\delta_{n_2}\right)\\
&\le 4\omega_B\left(f;\delta_{n_1},\delta_{n_2}\right).
\end{aligned}$$

□

Next, let us define the Lipschitz class for $B-$continuous functions. For $f \in C_b\left(I_{\mu\kappa}\right)$ the Lipschitz class $Lip_{\mathcal{M}}\left(\zeta_1, \zeta_2\right)$ with $\zeta_1, \zeta_2 \in (0, 1]$ is defined by

$$\begin{aligned} &Lip_M\left(\zeta_1, \zeta_2\right) \\ &\quad = \left\{f \in C_b\left(I_{\mu\kappa}\right) : \left|\Delta f_{(t_1,t_2)}(x_1, x_2)\right|\right. \\ &\left.\leq \mathcal{M}\left|t_1 - x_1\right|^{\zeta_1}\left|t_2 - x_2\right|^{\zeta_2}, \text{for}\left(t_1, t_2\right), \left(x_1, x_2\right) \in I_{\mu\kappa}\right\}. \end{aligned}$$

In our next result, we determine the order of approximation for the operators $\widehat{G}^{\alpha,\beta}_{n_1,n_2}$ by means of the class $Lip_{\mathcal{M}}(\zeta_1, \zeta_2)$.

Theorem 4.2 *For $f \in Lip_M\left(\zeta_1, \zeta_2\right)$, we have*

$$|\widehat{G}^{\alpha,\beta}_{n_1,n_2}(f; x_1, x_2) - f(x_1, x_2)| \leq \mathcal{M}(\delta_{n_1}(x_1))^{\zeta}_1(\delta_{n_2}(x_2))^{\zeta}_2,$$

for $\mathcal{M} > 0$, $\zeta_1, \zeta_2 \in (0, 1]$.

Theorem 4.3 *If $f \in D_b\left(I_{\mu\kappa}\right)$ and $D_B f \in B\left(I_{\mu\kappa}\right)$, then for each $(x_1, x_2) \in I_{\mu\kappa}$, we get*

$$\left|\widehat{G}^{\alpha,\beta}_{n_1,n_2}\left(f; x_1, x_2\right) - f(x_1, x_2)\right| \leq \frac{M}{n_1^{1/2} n_2^{1/2}}\left(\|D_B f\|_{\infty} + \omega_B\left(D_B f; n_1^{-1/2}, n_2^{-1/2}\right)\right).$$

Proof Since $f \in D_b\left(I_{\mu\kappa}\right)$, we have the identity

$$\triangle f\left[(t_1, t_2); (x_1, x_2)\right] = (t_1 - x_1)(t_2 - x_2) D_B(\varepsilon, n), \text{ with } x_1 < \varepsilon < t_1; x_2 < n < t_2.$$

Eventually,

$$D_B f(\varepsilon, n) = \triangle D_B f(\varepsilon, n) + D_B f(\varepsilon, x_2) + D_B f(x_1, n) - D_B f(x_1, x_2).$$

Since $D_B f \in B\left(I_{\mu\kappa}\right)$, by above relation, we can write

$$\begin{aligned} &\left|G^{\alpha,\beta}_{n_1,n_2}\left(\triangle f\left[(t_1, t_2); (x_1, x_2)\right]; x_1, x_2\right)\right| \\ &= \left|G^{\alpha,\beta}_{n_1,n_2}\left((t_1 - x_1)(t_2 - x_2) D_B(\varepsilon, n); x_1, x_2\right)\right| \\ &\quad \leq G^{\alpha,\beta}_{n_1,n_2}\left(\left|t_1 - x_1\right|\left|t_2 - x_2\right| \triangle D_B f(\varepsilon, n); x_1, x_2\right) \\ &\quad + G^{\alpha,\beta}_{n_1,n_2}\left(\left|t_1 - x_1\right|\left|t_2 - x_2\right|\left(\left|D_B f(\varepsilon, x_2)\right|\right.\right. \\ &\quad + \left.\left.\left|D_B f(x_1, n)\right| + \left|D_B f(x_1, x_2)\right|\right); x_1, x_2\right) \\ &\quad \leq G^{\alpha,\beta}_{n_1,n_2}\left(\left|t_1 - x_1\right|\left|t_2 - x_2\right| \omega_B\left(D_B f; \left|\varepsilon - x_1\right|, \left|n - x_2\right|\right); x_1, x_2\right) \\ &\quad + 3\|D_B f\|_{\infty} G^{\alpha,\beta}_{n_1,n_2}\left(\left|t_1 - x_1\right|\left|t_2 - x_2\right|; x_1, x_2\right). \end{aligned}$$

Using above inequality along with linearity of $G^{\alpha,\beta}_{n_1,n_2}$ and Cauchy-Schwarz inequality, we get

$$
\begin{aligned}
&\left|\widehat{G}^{\alpha,\beta}_{n_1,n_2}(f;x_1,x_2)-f(x_1,x_2)\right| = \left|G^{\alpha,\beta}_{n_1,n_2}(f\,[(t_1,t_2);(x_1,x_2)];x_1,x_2)\right| \\
&\quad \le 3\|D_B f\|_\infty \sqrt{G^{\alpha,\beta}_{n_1,n_2}\left((t_1-x_1)^2(t_2-x_2)^2;x_1,x_2\right)} \\
&\quad + \left(G^{\alpha,\beta}_{n_1,n_2}(|t_1-x_1|\,|t_2-x_2|;x_1,x_2)\right. \\
&\quad + \delta_{n_1}^{-1} G^{\alpha,\beta}_{n_1,n_2}\left((t_1-x_1)^2|t_2-x_2|;x_1,x_2\right) \\
&\quad + \delta_{n_2}^{-1} G^{\alpha,\beta}_{n_1,n_2}\left(|t_1-x_1|(t_2-x_2)^2;x_1,x_2\right) \\
&\quad + \delta_{n_1}^{-1}\delta_{n_2}^{-1} G^{\alpha,\beta}_{n_1,n_2}\left((t_1-x_1)^2(t_2-x_2)^2;x_1,x_2\right)\omega_B\left(D_B f;\delta_{n_1},\delta_{n_2}\right) \\
&\quad \le 3\|D_B f\|_\infty \sqrt{G^{\alpha,\beta}_{n_1,n_2}\left((t_1-x_1)^2(t_2-x_2)^2;x_1,x_2\right)} \\
&\quad + \left(\sqrt{G^{\alpha,\beta}_{n_1,n_2}\left((t_1-x_1)^2(t_2-x_2)^2;x_1,x_2\right)}\right. \\
&\quad + \delta_{n_1}^{-1}\sqrt{G^{\alpha,\beta}_{n_1,n_2}\left((t_1-x_1)^4(t_2-x_2)^2;x_1,x_2\right)} \\
&\quad + \delta_{n_2}^{-1}\sqrt{G^{\alpha,\beta}_{n_1,n_2}\left((t_1-x_1)^2(t_2-x_2)^4;x_1,x_2\right)} \\
&\quad \left.\left. + \delta_{n_1}^{-1}\delta_{n_2}^{-1} G^{\alpha,\beta}_{n_1,n_2}\left((t_1-x_1)^2(t_2-x_2)^2;x_1,x_2\right)\right)\omega_B\left(D_B f;\delta_{n_1},\delta_{n_2}\right). \qquad (4.2)
\end{aligned}
$$

In view of Remark 2.1, for $(t_1,t_2)\in I_{\mu\kappa}$, $(x_1,x_2)\in I_{\mu\kappa}$ and $i,j=1,2$

$$
\begin{aligned}
&G^{\alpha,\beta}_{n_1,n_2}\left((t_1-x_1)^{2i}(t_2-x_2)^{2j};x_1,x_2\right) \\
&\quad = G^{\alpha,\beta}_{n_1,n_2}\left((t_1-x_1)^{2i};x_1,x_2\right) G^{\alpha,\beta}_{n_1,n_2}\left((t_2-x_2)^{2j};x_1,x_2\right) \\
&\quad \le \frac{M_1}{n_1^i}\frac{M_2}{n_2^j}, \qquad (4.3)
\end{aligned}
$$

for some constant $M_1, M_2 > 0$.

Choosing $\delta_{n_1} = n_1^{-1/2}$, and $\delta_{n_2} = n_2^{-1/2}$ and combining (4.2) and (4.3), we get the desired result. □

Example 4.1 Fig. 1 illustrates the convergence of operators $G^{\alpha,\beta}_{n_1,n_2}$ to $f(x_1,x_2) = (x_1-1)^2-(x_2-1)^2$ for $\alpha=\beta=2$.

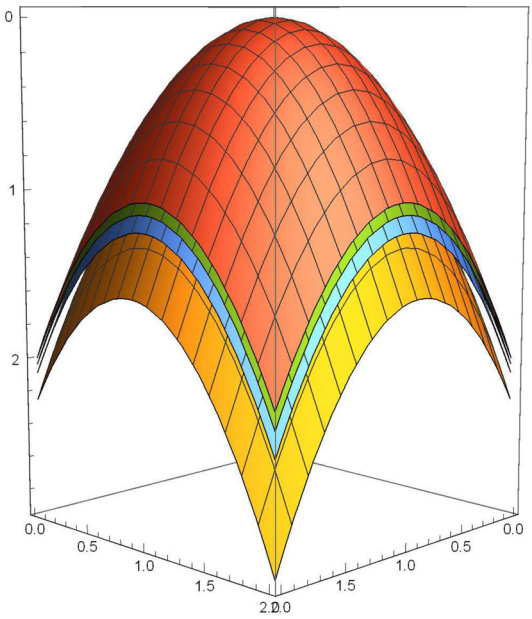


Fig. 1 The Convergence of operators $G^{\alpha,\beta}_{n_1,n_2}$ to the function $f(x_1, x_2)$ (Red f, Green $G^{2,2}_{50,50}$, Blue $G^{2,2}_{20,20}$, Yellow $G^{2,2}_{5,5}$)

Example 4.2 In Table 1, we estimate the absolute error $E^{\alpha,\beta}_{n_1,n_2}(f; x_1, x_2) = \left|G^{\alpha,\beta}_{n_1,n_2}(f; x_1, x_2) - f(x_1, x_2)\right|$ for functions $f_1(x_1, x_2) = (x_2 - 2)^2 - (x_1 - 2)^2$ and $f_2(x_1, x_2) = x_1 - 2x_1x_2 + 1$. Further, the convergence of the operators $G^{\alpha,\beta}_{n_1,n_2}$ to the function $f_1(x_1, x_2)$ is illustrated in Fig. 2. Both table and figure show a very interesting fact that the convergence of operators $G^{\alpha,\beta}_{n_1,n_2}$ is faster when $\alpha < \beta$ and slower when $\beta < \alpha$.

Table 1 Error estimation of operators $G^{\alpha,\beta}_{n_1,n_2}$ to the functions $f_1(x_1, x_2)$ and $f_2(x_1, x_2)$ (for cases $\alpha > \beta$, $\beta > \alpha$)

(x_1, x_2)	$E^{5,4}_{15,15}$	$E^{4,5}_{15,15}$
(1,1)	0.0858002	0.0554948
(2,2)	0.198099	0.128597
(3, 3)	0.310397	0.2017
(4,4)	0.422695	0.274802
(5,5)	0.534994	0.347904
(x_1, x_2)	$E^{6,2}_{5,5}$	$E^{2,6}_{5,5}$
(0,4)	2.29148	0.984868
(0.5, 3.5)	1.71861	0.738651
(1, 3)	1.14574	0.492434
(1.5, 2.5)	0.57287	0.246217
(2, 2)	0	0

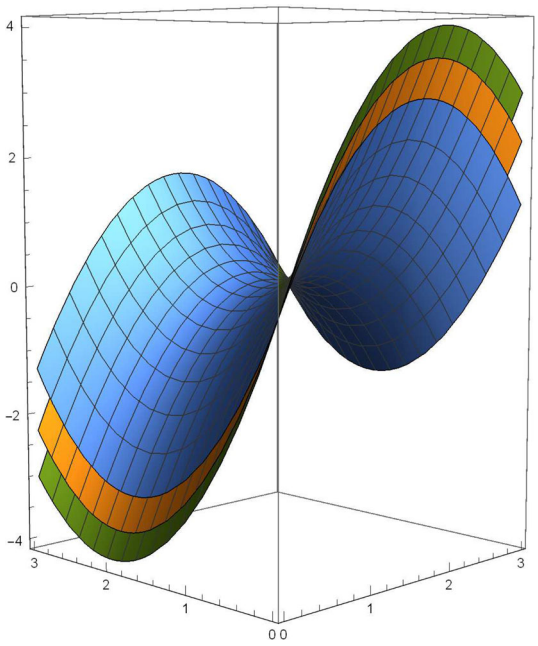


Fig. 2 Convergence of operators $G^{\alpha,\beta}_{n_1,n_2}$ to the function $f_1(x_1, x_2)$ (Green f_1, yellow $G^{2,6}_{5,5}$, $(\beta > \alpha)$, Blue $G^{6,2}_{5,5}$, $(\alpha > \beta)$)

Example 4.3 Define $\widehat{E}^{\alpha,\beta}_{n_1,n_2} = \left|\widehat{G}^{\alpha,\beta}_{n_1,n_2}(f; x_1, x_2) - f(x_1, x_2)\right|$. In Table 2, we estimate the error between the operators $G^{\alpha,\beta}_{n_1,n_2}$ and its corresponding GBS $\widehat{G}^{\alpha,\beta}_{n_1,n_2}$. We observe that GBS operators converge to the function parallely and the convergence is faster as well.

Table 2 Comparison between bivariate and its corresponding GBS for the function $f_2(x_1, x_2)$.

Considering equal values of n_1 and n_2		
(x_1, x_2)	$E^{1,10}_{1,1}$	$\widehat{E}^{1,10}_{1,1}$
(0.4,0.4)	0.0238076	0.00251803
(0.8,0.8)	0.0805798	0.00251803
(1.2,1.2)	0.137352	0.00251803
(1.6,1.6)	0.194124	0.00251803
(2,2)	0.250896	0.00251803
Considering different values of n_1 and n_2		
(x_1, x_2)	$E^{1,10}_{8,15}$	$\widehat{E}^{1,10}_{8,15}$
(2,2)	0.022789	0.0000209836
(4,4)	0.0499923	0.0000209836
(6,6)	0.0771957	0.0000209836
(8,8)	0.104399	0.0000209836
(10,10)	0.131602	0.0000209836

Acknowledgements I am indebted to an anonymous referee who made excellent suggestions to improve the manuscript.

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.

Data availability Any sources of data are indicated in the text.

Declarations

Conflict of interest There are no Conflict of interest.

References

1. Acu, A.M., Acar, T., Muraru, C.V., Radu, V.A.: Some approximation properties by a class of bivariate operators. Math. Methods Appl. Sci. **42**, 1–15 (2019)
2. Agratini, O.: On a class of linear positive bivariate operators of King type. Stud. Univ. Babeş-Bolyai Math. **51**(4), 13–22 (2006)
3. Anastassiou, G.A., Gal, S.: Approximation Theory: Moduli of Continuity and Global Smoothness Preservation. Birkhauser, Boston (2000)
4. Badea, C., Badea, I., Gonska, H.H.: A test function theorem and approximation by pseudo polynomials. Bull. Austral. Math. Soc. **34**, 95–108 (1986)
5. Badea, C., Badea, I., Gonska, H.H.: Notes on the degree of approximation of B-continuous and B-differentiable functions. J. Approx. Theory Appl. **4**, 95–108 (1988)
6. Badea C., Cottin, C.: Korovkin-type theorems for generalised Boolean sum operators. In: Approximation Theory (Kecskemét, Hungary). Colloquia Mathematica Societatis János Bolyai, vol. 58, pp. 51–67 (1990)
7. Baxhaku, B., Kajla, A.: Blending type approximation by bivariate generalized Bernstein type operators. Quaestiones Mathematicae. (2019). https://doi.org/10.2989/16073606.2019.1639843
8. Bögel, K.: Mehrdimensionale differentiation von funtionen mehrerer veränderlicher. J. Reine Angew. Math. **170**, 197–217 (1934)
9. Bögel, K.: Über die mehrdimensionale differentiation, integration und beschränkte variation. J. Reine Angew. Math. **173**, 5–29 (1935)
10. Butzer, P.L., Berens, H.: Semi-Groups of Operators and Approximation. Springer, Berlin-Heidelberg-New York (1967)
11. Deo, N., Bhardwaj, N.: Some approximation theorems for multivariate Bernstein operators. Southeast Asian Bull. Math. **34**, 1023–1034 (2010)
12. Deo, N., Kumar, S.: Durrmeyer variant of Apostol-Genocchi-Baskakov operators. Quaestiones Mathematicae **44**, 1817–1834 (2020)
13. Dobrescu, E., Matei, I.: The approximation by Bernstein type polynomials of bidimensionally continuous functions. An. Univ. Timisoara Ser. Sti. Mat-Fiz. **4**, 85–90 (1996)
14. Gupta, V., Deo, N.: On the rate of convergence for bivariate beta operators. Gen. Math. **13**(3), 107–114 (2005)
15. Gupta, V., Rassias, T.M., Agrawal, P.N., Acu, A.M.: Recent Advances in Constructive Approximation Theory. Springer, New York (2018)
16. Heilmann, M., Raşa, I.: Note on a proof for the representation of the *k*th order Kantorovich Modification of Linking Baskakov Type Operators. In: Deo, N., Gupta, V., Acu, A., Agrawal, P. (eds) Mathematical Analysis I: Approximation Theory. ICRAPAM 2018. Springer Proceedings in Mathematics & Statistics, pp. 111–120. Springer, Singapore (2020)
17. Ispir, N., B*ü*y*ü*kyazici, I.: Quantitative estimates for a certain bivariate Chlodowsky-Szász-Kantorovich type operators. Math. Commun. 21, 31–44 (2016)
18. Jolany, H., Sharifi, H., Alikelaye, R.E.: Some results for the Apostol-Genocchi polynomials of higher order. Bull. Malays. Math. Sci. Soc. **36**(2), 465–479 (2013)
19. Lipi, K., Deo, N.: General family of exponential operators. Filomat **34**(12), 4043–4061 (2020)

20. Luo, Q.M.: Extensions of the Genocchi polynomials and their Fourier expansions and integral representations. Osaka J. Math. **48**, 291–309 (2011)
21. Luo, Q.M., Srivastava, H.M.: Some generalizations of the-Genocchi polynomials and the Stirling numbers of the second kind. Appl. Math. Comput. **217**, 5702–5728 (2011)
22. Mishra, N.S., Deo, N.: Kantorovich variant of Ismail-May operators. Iran. J. Sci. Technol. Trans. A Sci. **44**, 739–748 (2020)
23. Ozden, H., Simsek, Y.: Modification and unification of the Apostol-type numbers and polynomials and their applications. Appl. Math. Comput. **235**, 338–351 (2014)
24. Özergin, E., Özarslan, M.A., Srivastava, H.M.: Some families of generating functions for a class of bivariate polynomials. Math. Comput. Model. **50**, 1113–1120 (2009)
25. Prakash, C., Verma, D.K., Deo, N.: Approximation by a new sequence of operators involving Apostol-Genocchi polynomials. (In press)
26. Srivastava, H.M., Özarslan, M.A., Kaanoğlu, C.: Some generalized Lagrange-based Apostol-Bernoulli. Apostol-Euler and Apostol-Genocchi polynomials. Russ. J. Math. Phys. 20(1), 110–120 (2013)
27. Volkov, V.I.: On the convergence of sequences of linear positive operators in the space of continuous functions of two variables, Dokl Akad Nauk SSSR (N.S.) 115, 17–19 (1957)

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

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

Plasmonics
https://doi.org/10.1007/s11468-024-02715-5

RESEARCH

Black Phosphorus-Based Slotted D-Shaped PCF SPR Sensor for Cancer Detection

Arhee Bhuyan[1] · Akash Khamaru[1] · Ajeet Kumar[1]

Received: 19 November 2024 / Accepted: 13 December 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

This work proposes a slotted D-shaped photonic crystal fiber (PCF) surface plasmon resonance (SPR) sensor for the detection of cancerous cells. Cancerous analytes are used in different concentrations having different refractive indices (RI) in the sensing medium. To confine the light in PCF, two types of air holes are incorporated, i.e., circular and oblong-shaped air holes. The designed PCF is coated with a gold layer (50 nm) along with TiO_2 (12 nm) that works as a plasmonic enhancer as well as an adhesive to stick the PCF with gold. Apart from gold + TiO_2, an additional seven layers of emerging plasmonic material—black phosphorus are coated on the frame to enhance the sensitivity having a monolayer thickness (0.53 nm). Since healthy and infected cells have different RI, therefore the coupling conditions also differ, which is reflected in peak confinement loss wavelength shift. With the help of the spectral shift, respective wavelength sensitivity (WS) is determined. The WS obtained for cervical, blood, skin, adrenal gland, and breast (types 1 and 2) cancerous cells are 5417 nm/RIU, 6071 nm/RIU, 4750 nm/RIU, 9286 nm/RIU, 11,429 nm/RIU, and 10,714.29 nm/RIU, respectively. The recorded results show that our proposed sensor is a better counterpart among previously proposed sensors and suitable for cancer cell detection.

Keywords Slotted PCF · Black phosphorus · Oblong air holes · Surface plasmon resonance

Introduction

The body is made up of millions of microscopic cells, each of which functions as a completely different living organism. Every single cell in our body works collectively with every other cell to form our tissues and organs. An illustration of this cooperation can be seen in the process by which cell division takes place. They begin to regenerate when it becomes necessary to fix damaged or dead cells. When the process of cellular replication becomes unregulated, cancer develops [1]. Each cell has a nucleus that contains a huge bunch of chromosomes, and these chromosomes are made of genes in the form of DNA strands containing coded messages that guide the cell about the work they will perform. This is how the information about the presence of dead cells reaches the brain, and the cells undergo the process of cell division. During cell division, a mutation occurs which may be the outcome of smoking or any other external harmful factor. The discussed process of cell division gets misled, and without any instructions, cells start to grow abruptly, resulting in the formation of a lump known as cancer [2]. As soon as it is discovered, it is already in a stage where no treatment can help with the recovery. Cancer is actually the growth of cells in an abnormal way which changes the usual gene replication and eventually, it becomes a multiple-cell system that disbalances the cell replication in the body [3]. It ultimately transforms into a huge cell group that can invade any tissues in any part of the body. Through invading tissues, they infect the entire organ system, causing a painful death of the host.

Therefore, early detection of cancerous cells is one of the most talked topic. Generally, people avail detection techniques like MRI (magnetic resonance imaging), PET-CT (positron emission tomography-computed tomography), biopsy, X-ray, and CT scans to monitor cancerous presence

✉ Ajeet Kumar
ajeetdph@dtu.ac.in

Arhee Bhuyan
arheebhuyan_23mscphy72@dtu.ac.in

Akash Khamaru
akashkhamaru_23phdap02@dtu.ac.in

[1] Advanced Photonics Simulation Research Laboratory, Department of Applied Physics, Delhi Technological University, Shahbad Daulatpur, Bawana Road, Delhi 110042, India

Published online: 19 December 2024

in the body regularly but they have numerous disadvantages like slow [4], complicated [5], and inefficient procedures, with a lot of radiation exposure [6], and need of a lot of trained people to operate equipment making it more and more expensive.

Over the years, various efficient sensing techniques for the detection of cancer have evolved a lot. Some of them are colorimetric biosensors; fluorescence-based biosensors; plasmon resonance-based sensing; photonic crystal fiber (PCF) based sensors, interferometer, and grating-based biosensors and resonator-based biosensors [7]. Among all these sensors, surface plasmonic resonance (SPR) based sensors are considered very efficient due to its label-free sensing mechanism and deliver higher sensitivity toward its analytes [8].

When incident light strikes on thin metal film, like gold or silver, some portion of light gets refracted and converted into electric fields known as the evanescent wave. The evanescent wave penetrates through the metal layer, interacts with the free electrons at the metal surface, and geared them to oscillate collectively in response to the interacted waves. This oscillation of electrons is called surface plasmons (SP). The SPs oscillate or resonate as they interact with refracted photons, resulting in the generation of a surface plasmon polariton (SPP). SPPs are defined as types of waves that combine the oscillation of SPs in the metal with electromagnetic waves (polaritons) in the dielectric or air. When the frequency of the incident light matches the natural frequency of these collective oscillations, resonance occurs, and the same is referred to as SPR. As a portion of the light gets utilized for the phenomena of SPR, the incident light no longer be completely reflected. Therefore, SPR is examined by the dip in the intensity of the reflected light. The resonance conditions can be varied by associating any medium with the plasmonic material.

PCFs are a class of optical fiber that uses a periodic structure of air holes within the cladding to confine light. This structure creates unique properties not found in traditional optical fibers, such as tailored dispersion characteristics, high birefringence, and enhanced nonlinearity [9], PCFs can be utilized for both light guiding and fluid flow channels. This results in a significant interaction of light with samples. They are also functionalized with bio-sensing layers that can bind and collect bimolecular targets. Due to the small size of the PCF core and cladding air holes, the intense interaction between the sample and light may be achieved, which results in the detection of that particular sample. Sensing needs very small samples of fluid volume. Small volume consumption is of great interest in chemical and biological applications.

Various different works have been proposed in recent times to enhance sensitivity in the field of early cancer detection. Sardar et al. have suggested a dual-core sensor with two polished sides with two types of air holes, i.e., circular and rectangular air holes for various cancer cells HeLa cell, basal cell, PC-12, Jurkat, MDA-MB-231, and MCF-7. Gold is used as a plasmonic material and a layer of PML and achieved a maximum wavelength sensitivity of 7143 nm/RIU for MDA-MB-231 and MCF-7, with an amplitude sensitivity of -270 RIU^{-1} and a sensor resolution of 2.9×10^{-5} RIU [10]. Yasli et al. have designed a PCF-SPR biosensor with the PCF core curved in a crescent shape and three types of circular air holes of three different diameters and gold as a plasmonic material. They have used FEM to analyze the sensitivity of all the cancer cells (HeLa cell, basal cell, PC-12, Jurkat, MDA-MB-231, MCF-7). The proposed biosensor has exhibited the highest wavelength sensitivity as 7142.86 nm/RIU for MCF-7 with -757 RIU^{-1} amplitude sensitivity for HeLa cancer cells [11]. In another study by Majeed et al., a dual-core PCF sensor is proposed with oblong air holes and two types of circular air holes of different diameters of gold as a plasmonic material to obtain the sensitivity of all the cancerous cells. They have demonstrated a maximum wavelength sensitivity of 11,429 nm/RIU for MCF-7 and an amplitude sensitivity of -1251.18 RIU^{-1} for MDA-MB-231 with a maximum resolution of 8.75×10^{-6} RIU [12]. Chaudhary et al. have presented a PCF-SPR sensor with two circular air holes positioned in a hexagonal pattern and gold as a plasmonic material. They have successfully achieved a maximum wavelength sensitivity of 10,714.28 nm/RIU for MCF-7 and maximum amplitude sensitivity as -1599 RIU^{-1} for Jurkat and a sensor resolution of 1.82×10^{-5} RIU for basal cell [13]. Ibrahim et al. have designed a sensor with an analyte channel to sense the target, a metal plasmonic coating, i.e., gold and a PCF section for guiding light. They have managed to obtain a maximum wavelength sensitivity of 7142.86 nm/RIU for MCF-7, amplitude sensitivity of -4220.99 RIU^{-1} for HeLa, and a sensor resolution of 3.33×10^{-5} RIU for basal cell [14]. In another research, Kumar et al. have proposed a sensor with Mxene and gold as plasmonic materials with five layers of air holes arranged in a vertical line. They have managed to achieve a maximum wavelength sensitivity of 10,714 nm/RIU for MCF-7 in X-polarization and 13,071 nm/RIU in Y-polarization with a sensor resolution of 10^{-6} RIU [15].

The proposed cancer cell PCF SPR sensor consists of gold and TiO_2 as plasmonic materials connecting the fiber to the analyte. In view of suitable plasmonic material, gold is chosen since it is environment-friendly, corrosion-free, and bio-compatible [16]. At the same time, TiO_2 works as an adhesive between the fiber surface and gold. Another plasmonic material, black phosphorus (BP) is used in the proposed biosensor. BP is an emerging 2-D plasmonic material that has a very significant effect on increasing the sensitivity of a sensor. BP has a layer-dependent bandgap covering 0.3 to 2 eV. The bandgap tuning range is larger than other 2D materials, which gives flexibility in electronic or photonic

applications. Black phosphorus includes a direct band gap, smaller work function, high charge carrier mobility, and better binding of molecules on the surface of sensors. Direct band gap allows BP to interact with light at specific wavelengths which hence, enhances the sensitivity of the sensor [17]. Smaller work function allows the transfer of electrons between the surrounding environment and BP layers. High charge carrier mobility makes the sensor transmit signals fast and efficient which results in rapid response and higher sensitivity. The ability to bind molecules tightly on the surface allows BP to be used in biological and chemical sensing [18]. BP monolayers inherit a very small thickness of 0.53 nm. So, the said plasmonic material is used as a stack of layers one over the other to analyze the effect of each layer on the SPR phenomenon. The proposed sensor is designed to detect infected cancer cells with respect to their normal and uninfected ones. The infected counterpart of the healthy cells shows an increase in refractive index (RI) as the protein content in the nucleus of the infected cell increases. Cancer cells have many types: skin cancer (basal cell), blood cancer (Jurkat), adrenal gland cancer (PC-12), cervical cancer (HeLa cell), breast cancer (types 1 and 2), etc. The RI of healthy basal cells, Jurkat, PC-12, and HeLa cells lies between 1.36 and 1.39. Whereas, the infected counterpart of these cells shows an increase in RI and lies within 1.38 to 1.41 [19]. These healthy and infected cell forms the sample analyte for our analysis and based on their RI difference we will calculate the sensitivity of the infected cell with respect to the healthy cell.

Design of Proposed PCF-Based SPR Biosensor

Figure 1 is the cross-sectional demonstration of the proposed biosensor. Different geometric parameters were chosen and placed in a way to create a passage for the interaction between the core mode and the SPP mode of the sensor. In this designed sensor, two types of air holes are used: the first is oblong air holes and the second one is circular air holes. The PCF consists of two sets of oblong air holes: one has a height and width defined as H and W, whereas the other slanted oblongs have the same width but the height attain the magnitude of $H' = 1.35 \times H$. The slanted oblongs are at an angle of 60° with each other. Therefore, to direct the evanescent field to the plasmonic layers (gold-TiO_2), the first type of oblong air holes are placed in a horizontal manner. Smaller air holes having diameter (d_c) are placed near the vicinity of the core to prevent any extent of leakage of light. The thickness of TiO_2 (t_{Ti}) used is 12 nm for the

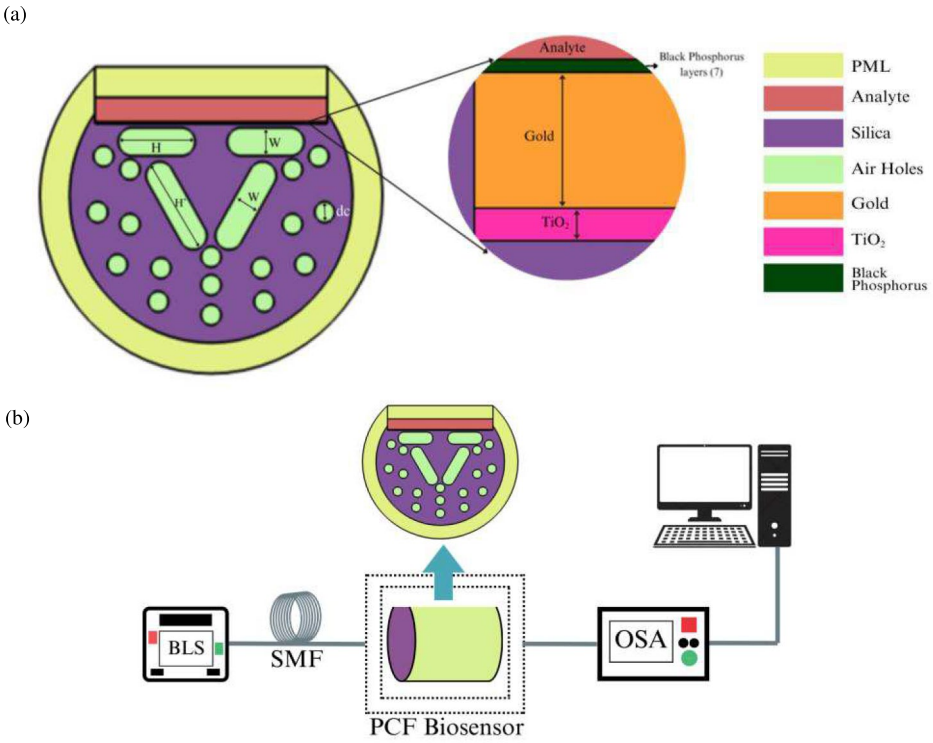


Fig. 1 **a** Proposed PCF SPR bio-sensor, **b** proposed experimental setup of the proposed PCF SPR sensor

initial analysis and similarly gold layer has a thickness (t_g) of 50 nm. Here, seven layers of BP are used each of thickness 0.53 nm united giving a total thickness of 3.71 nm to initiate the investigation.

The series of plasmonic layers over the PCF are as TiO_2, then gold (Au) over which thin films of BP are layered. TiO_2 is known to enhance the sensitivity by behaving as an adhesive between the fiber and gold layer. The bio-sample analyte is injected over the gold layer using a "selective filling approach" and then a Perfectly Matched Layer (PML) layer is added. The main purpose of adding PML is to prevent any reflected radiation from being scattered.

A concentration of 30–70% is for healthy cells and 80% of the concentration is used for infected cancerous cells in the analyte for detection. Table 1 shows the refractive indices of the healthy and the infected cells.

For the fundamental material, the proposed sensor is composed of fused silica. It is crucial to determine the refractive index of fused silica which is done by using the known wavelength-dependent Sellmeier Equation, which states a relationship between wavelength and the refractive index and helps to find the refractive index of fused silica at different wavelengths. The equation of Sellmeier equation is given as [22]

$$n(\lambda) = \sqrt{1 + \frac{0.696163\lambda^2}{\lambda^2 - 0.0046791486} + \frac{0.4079426\lambda^2}{\lambda^2 - 0.0135120631} + \frac{0.8974794\lambda^2}{\lambda^2 - 97.934003}} \quad (1)$$

λ here represents the optical wavelength (in μm) and n(λ) represents the refractive index of fused silica.

Drude–Lorentz model is used here to compute the dielectric constant of gold. The Drude-Lorentz model expression is as follows [23]

$$\varepsilon_{Au} = \varepsilon_\infty - \frac{w_D^2}{w\left(w + j\gamma_D\right)} - \frac{\Delta\varepsilon.\Omega_L^2}{\left(w^2 - \Omega_L^2\right) + j\Gamma_L w} \quad (2)$$

Here,ε_{Au} represents the permittivity of gold, $w = 2\pi c/\lambda$ is the angular frequency, where c is the speed of light. $\varepsilon_\infty = 5.9673$, $w_D/2\pi = 2113.6$ THz, $\gamma_D/2\pi = 15.92$ THz, $\Omega_L/2\pi = 650.07$ THz, $\Gamma_L/2\pi = 104.86$ THz, $\Delta\varepsilon = 1.09$. The refractive index of TiO_2 is determined by the following equation [24]

$$n_{TiO_2} = \sqrt{5.913 + \frac{2.441 \times 10^7}{\lambda^2 - 0.803 \times 10^7}} \quad (3)$$

Here, n_{TiO_2} is the refractive index of TiO_2 and λ represents optical wavelength.

The refractive index of Black Phosphorous used in the sensor is measured with the help of the polynomial fitted equation [25]

$$n(\lambda) = A + B\frac{10^4}{\lambda^2} + C\frac{10^9}{\lambda^4}, \quad (4)$$

$$k(\lambda) = D \times 10^{-5} + E\frac{10^4}{\lambda^2} + F\frac{10^9}{\lambda^4} \quad (5)$$

Here, n(λ) is the real part, k(λ) is the imaginary part of the refractive index, and A = 3.57, B = 6.79, C = 39.99, D = 3206, E = 0.521, F = 10.26 are constants.

Fabrication Possibilities for the Presented Sensor

The proposed PCF SPR sensor can be fabricated through procedures such as stack-and-draw, injection molding, as well as through sol–gel casting. Stack and Drawing is the most used method for the fabrication because of its versatility and as it has the most inexpensive casting [26]. Stack and draw, sol–gel methods, and capillary stacking are used for making fibers with circular air holes. A preform is generated by stacking a small length capillary with a known diameter into a larger diameter glass tube. Then cane is formed with the help of the preform, and a fiber having a definite

Table 1 Refractive index of various cancerous cells with their cancer types and references

Cancer type	Name of the cell	Refractive index of healthy cell	Refractive index of infected cell	References
Skin	Basal cell	1.360	1.380	[20, 21]
Blood	Jurkat	1.376	1.390	
Adrenal gland	PC12	1.381	1.395	
Cervical	HeLa	1.368	1.392	
Breast cancer-type 1	MDA-MB231	1.385	1.399	
Breast cancer-type 2	MCF7	1.387	1.401	

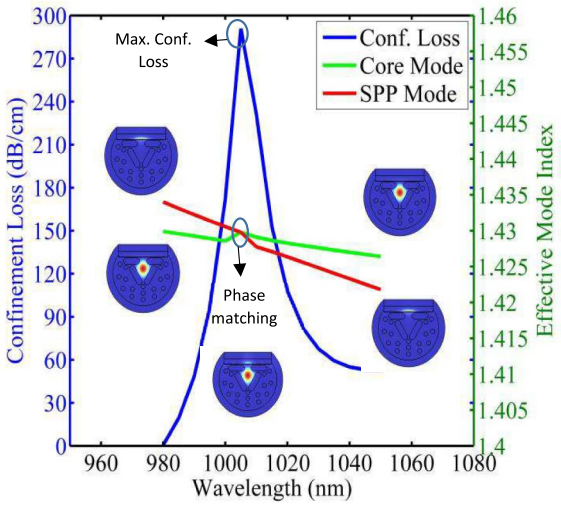


Fig. 2 Dispersion curve for the core and SPP mode for PC12 cancerous cells

diameter draws the cane into it. By shutting one end of any air-holes of the cane, air hole size can be changed. On the other hand, 3D printing, and extrusion methods have the potential to fabricate complex airholes (rectangular, oblong) within PCF structures. The previous research works had already proposed rectangular and oblong air holes with their fabrication techniques and feasibility [27, 28]. Thermal evaporation, radio frequency sputtering, chemical vapor deposition (CVD), and wet chemical deposition are the different methods that may be used to deposit thin layers of metals like TiO_2 and gold [29]. CVD technique may be preferred for generating a homogeneous metal layer with nm-range thickness [30]. BP is fabricated using various techniques like mechanical exfoliation, liquid-phase exfoliation, and electrochemical exfoliation [31]. Using the selective filling approach, the analyte or the cell samples can be filled over the plasmonic layer into the sample channel [32]. The discussion covers all the processes and ways by which the proposed biosensor can be fabricated.

Dispersion Relation of the Biosensor

The COMSOL Multiphysics software is employed for stimulating the SPR-PCF sensor using FEM (finite element method). FEM works with Maxwell's equations to get the real and imaginary parts of the effective modal index (n_{eff}). The real part of n_{eff} describes the guided light modes (core and SPP modes). Whereas the imaginary part does the analysis over the losses incurred in the fiber. The confinement loss (CL) can be obtained by the following equation [33]

$$Loss\left(\frac{db}{cm}\right) = 8.686 \times \frac{2\pi}{\lambda} \times Im\left(n_{eff}\right) \times 10^4 \tag{6}$$

The dispersion relationship is taken in the y-polarization as all the plasmonic materials lie on the top of the sensor. Figure 2 shows the respective core mode, SPP mode, and SPR mode. The dispersion relationship between core mode and SPP mode is also shown in Fig. 2. In the

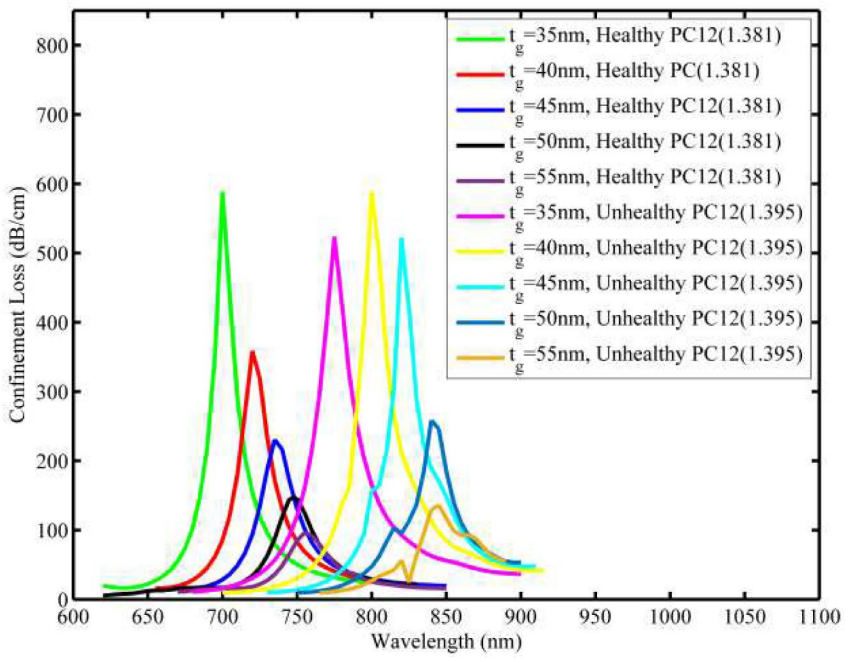


Fig. 3 CL spectra for various gold thicknesses (t_g) at RI of 1.385 and 1.391

Table 2 Optimization of thickness of gold layer (t_g)

t_g (nm)	Max. CL (Normal cell) (dB/cm)	Max. CL (Infected cell) (dB/cm)	ΔWL (nm)	Sensitivity (nm/RIU)
35	588.52	523.63	75	5357.14
40	358.76	589.04	80	5714.29
45	230.62	522.44	85	6071.43
50	145.99	258.39	90	6428.50
55	96.16	136.08	85	6071.53

figure, n_{eff} is marked on the right side, and confinement loss is marked on the left side. The blue curve is taken in the range from 980 to 1050 nm that corresponds to the confinement loss. Whereas, the n_{eff} for the core and the SPP modes are denoted by the light green and the red line respectively over a range of wavelengths. The figure provides information that the peak for the confinement loss of the core mode arrives as 290 dB/cm at 1050 nm, and both the curves for the real part of the n_{eff} for core and SPP mode meet at the exact same wavelength. The point at which both these curves meet each other is known as the phase-matching position and simultaneously maximum loss observed that defines the resonance (SPR) condition. At that particular point, maximum energy is transferred from the core mode to the SPP mode and indicates coupling between both modes.

To evaluate the wavelength sensitivity of any cell, the following equation is used [34]

$$S_\lambda\left(\frac{nm}{RIU}\right) = \frac{\Delta\lambda_{Peak}}{\Delta n_a} \tag{7}$$

Here, $\Delta\lambda_{Peak}$ represents the difference between the peak wavelengths of the healthy and the infected cells, Δn_a represents the difference in the refractive index between the healthy and the infected cells.

Result and Discussion of the Proposed PCF SPR Sensor

Optimization of the Plasmonic Materials and Geometrical Parameters

The optimization of the structural parameters of different constituents involved in the proposed slotted PCF is to be done in this section. Therefore, it is ideal to investigate different values of the parameters that can provide maximum sensitivity. Here, PC-12 (Cervical Cancer) is taken as the base cancer type for optimizing the structural parameters, where 1.381 and 1.395 are the refractive index of normal and the infected cell for PC-12. The respective change in the refractive index is 0.014.

Gold (Au) Thickness Optimization

The given study was conducted about the effects of the thickness of gold (t_g). The gold thickness hugely affects the sensing performance of the sensor. The optimal thickness

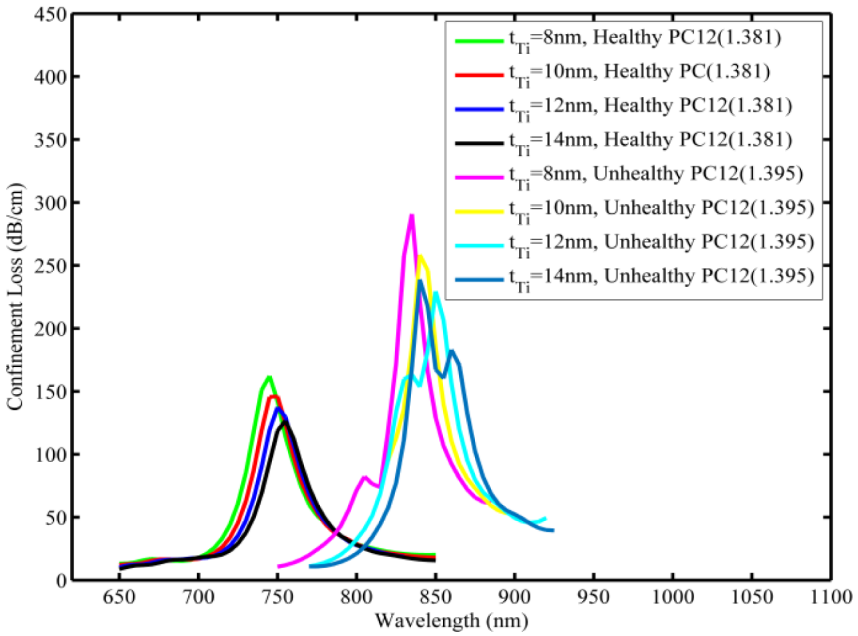


Fig. 4 CL spectra for various TiO_2 thicknesses at RI of 1.385 and 1.391

Table 3 Optimization of the thickness of TiO_2 layer (t_{Ti})

t_{Ti} (nm)	Max. CL (Normal cell) (dB/cm)	Max. CL (Infected cell) (dB/cm)	ΔWL (nm)	Sensitivity (nm/RIU)
8	162.03	290.82	90	6428.57
10	145.99	258.39	90	6428.57
12	137.25	229.42	100	7142.86
14	126.56	238.82	85	6071.53

Table 4 Optimization of width of the oblong air hole (W)

W (μm)	Max. CL (Normal cell) (dB/cm)	Max. CL (Infected cell) (dB/cm)	ΔWL (nm)	Sensitivity (nm/RIU)
1.05	142.66	341.43	90	6428.57
1.1	137.25	229.43	100	7142.86
1.15	126.97	182.92	95	6785.71

of the gold is determined by varying it in the range of 35–55 nm at a difference of 5 nm. Figure 3 depicts the graphical representation of confinement loss versus the wavelength with varying thicknesses. Table 2 is a systematic representation of the variations of peak wavelengths (WL) and confinement loss (CL) in the healthy and unhealthy refractive indices with respect to the thickness of gold, as observed in the figure. From the figure, it is interpreted that with the increment of the thickness of the gold layer, the interaction of the incident light with the analyte gets minimized, which results in decrement in the peak loss. Whereas, a thin layer of gold disrupts the sensor performance due to high confinement loss. From the table, it is clear that out of all the thicknesses, 50 nm gold has the maximum change in the wavelength (ΔWL) which delivers maximum sensitivity. With the increase in the t_g, the peak loss value is decreased. Therefore, the optimum thickness for gold will be taken as 50 nm, which shows a moderate loss magnitude for convenient detection.

TiO_2 Thickness Optimization

TiO_2 is known as a remarkable plasmonic material that would definitely alter the sensor's performance. The optimal thickness of the TiO_2 (t_{Ti}) is determined by varying it in the range of 8–14 nm at a difference of 2 nm keeping the gold thickness optimal at 50 nm. Figure 4 depicts the graph of confinement loss versus the wavelength with varying thicknesses. Table 3 is a systematic representation of the variations of wavelengths and confinement loss (CL) in the healthy and unhealthy refractive indices with respect to the thickness of TiO_2 (t_{Ti}). Since, TiO_2 is applied as a sensitivity enhancer and as an adhesive, therefore it has a minor influence on the confinement loss. The confinement loss suffers minor decrement with the increase of thickness of TiO_2. In addition, a thicker layer tends to generate some extra loss peaks which should be avoided for appropriate sensing performance. From the table, it is clear that, out of

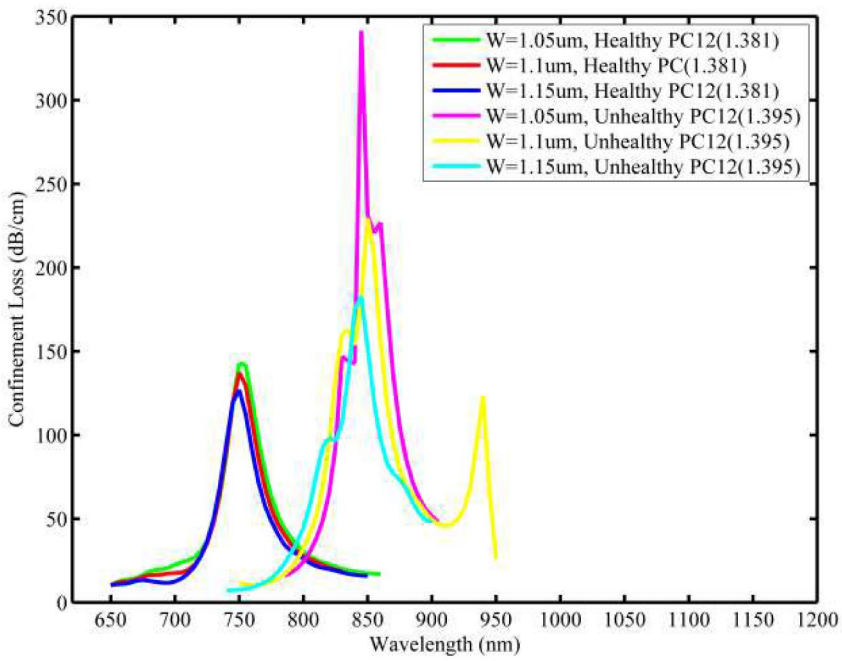


Fig. 5 CL spectra for various width of the oblong air hole (W) at RI of 1.385 and 1.391

all the thicknesses 12 nm has the maximum change in the wavelength (ΔWL). Optimizing the thickness of Titanium Dioxide increased the wavelength sensitivity from 6428.571 to 7142.857 nm/RIU. Therefore, the optimum thickness for further analysis of TiO_2 is 12 nm.

In the proposed biosensor, two types of oblong air holes are present where one is dependent on the other. The optimal width and height of the oblong air hole will be determined in this section, as these air holes are highly responsible for altering the sensitivity. Here the height and width of the air hole will be varied keeping the optimal thickness of gold and TiO_2 as 50 nm and 12 nm, respectively.

Width (W) Optimization

The optimal width of the oblong air hole is determined by varying it in the range of 1.05–1.15 μm at a difference of 0.05 μm. Figure 5 depicts the graphical representation of confinement loss versus the wavelength with the varying width.

Table 4 displays the systematic representation of the variations of wavelengths and confinement loss (CL) in the healthy and unhealthy refractive indices with respect to the width of the oblong air hole (WL). As the width increases from 1.05 to 1.15 μm, the light confinement gets stronger in the core and made it difficult to reach toward the plasmonic material, which eventually reduces the interaction between the core and the plasmon mode, ultimately minimize the confinement loss and affects the sensitivity.

Table 5 Optimization of height of the oblong air hole (H)

H (μm)	Max. CL (Normal cell) (dB/cm)	Max. CL (Infected cell) (dB/cm)	ΔWL (nm)	Sensitivity (nm/RIU)
1.85	201.50	372.33	100	7142.86
1.90	179.14	318.27	100	7142.86
1.95	157.42	270.85	100	7142.86
2.0	137.24	229.42	100	7142.86
2.05	118.3138	178.46	80	5714.29

From the table, it is clear that out of all the width values, 1.1 μm has the maximum change in the wavelength (ΔWL). Therefore 1.1 μm width is taken as the optimal magnitude for further analysis.

Height (H) Optimization

The optimal height of the oblong air hole is determined by varying it in the range of 1.9–2.1 μm at a difference of 0.05 μm. Figure 6 and Table 5 provide the systematic representation of the variations of confinement loss with wavelength in the healthy and unhealthy refractive indices with respect to the height of the oblong air hole. From 1.85 to 2.05 μm, no major obstruction is faced by core mode to make a proper interaction with the plasmon mode ultimately shows no change in the observed sensitivity with a minor decrement in the confinement loss since, increasing height

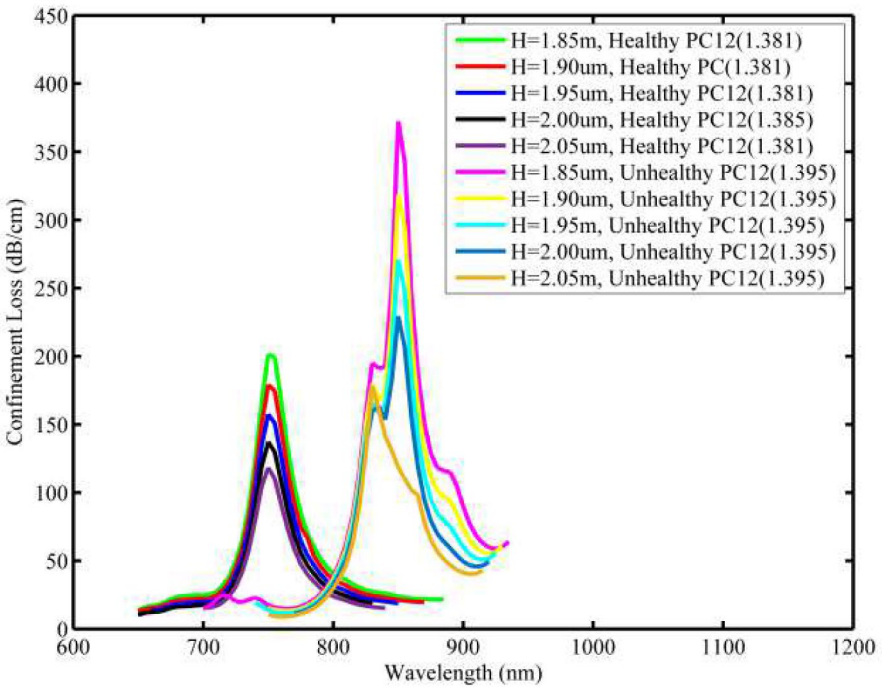


Fig. 6 CL spectra for various height of the oblong air hole H at RI of 1.385 and 1.391

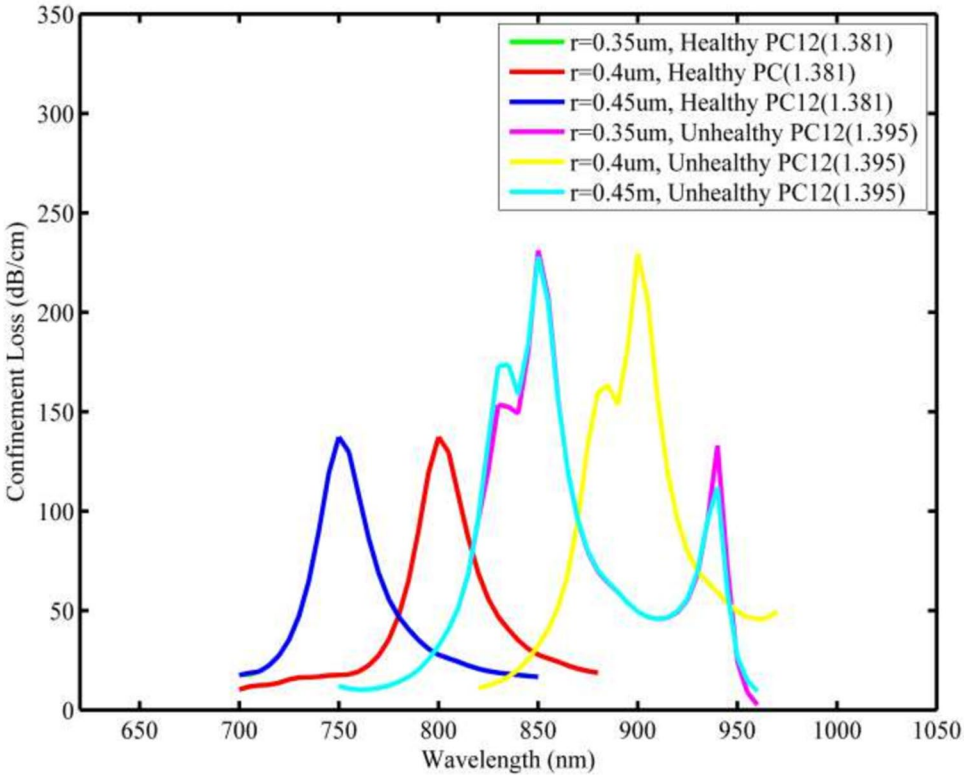


Fig. 7 CL spectra for various radius of the circular air hole (r) at RI of 1.385 and 1.391

especially upper oblong air holes blocks the core mode to interact with SPP mode. Therefore, to ensure a nominal confinement loss, 2 μm height is taken as the optimized value.

Circular Hole Radius (r) Optimization

The optimal radius of the circular air hole is determined by varying it in the range of 0.35–0.45 μm at a gap of 0.05 μm. Figure 7 along with Table 6 shows the systematic representation of the variations in the confinement loss in the healthy and unhealthy refractive indices with respect to the width of the circular air hole. As it is noticeable from the table that all the radius corresponds to the same change in the peak wavelength with negligible variation in loss, that is because the circular air holes are not playing major role in interaction of the light with the plasmonic materials and also are not connected in any way with the core They are just placed to obstruct the core field from leaking outside core region. Therefore, by maintaining a nominal loss and fabrication possibilities, 0.4 μm is considered as the optimum radius.

Table 6 Optimization of radius of circular holes (r)

r (μm)	Max. CL (Normal cell) (dB/cm)	Max. CL (Infected cell) (dB/cm)	ΔWL (nm)	Sensitivity (nm/RIU)
0.35	137.26	231.10	100	7142.86
0.4	137.24	229.42	100	7142.86
0.45	137.23	228.01	100	7142.86

Black Phosphorus (BP) Optimization

Black phosphorus are thin layers of 2-D plasmonic material that have a fixed monolayer thickness of 0.53 μm. Here, in this section, a number of BP layers used will be optimized. As the thickness is very small, the range of the thickness used is 5–9 layers at a difference of 2. Figure 8 shows the change in confinement loss versus wavelength with the varying BP layers.

When black phosphorus layers are increased, the sensitivity of the sensor increases rapidly but the FWHM (full width half maximum) also increases. From Table 7, it is observed that sensitivity increases in the beginning and then decreases with the increase in the number of layers of BP. Therefore, the optimum number of layers for black phosphorus for the proposed biosensor will be taken as 7 (seven).

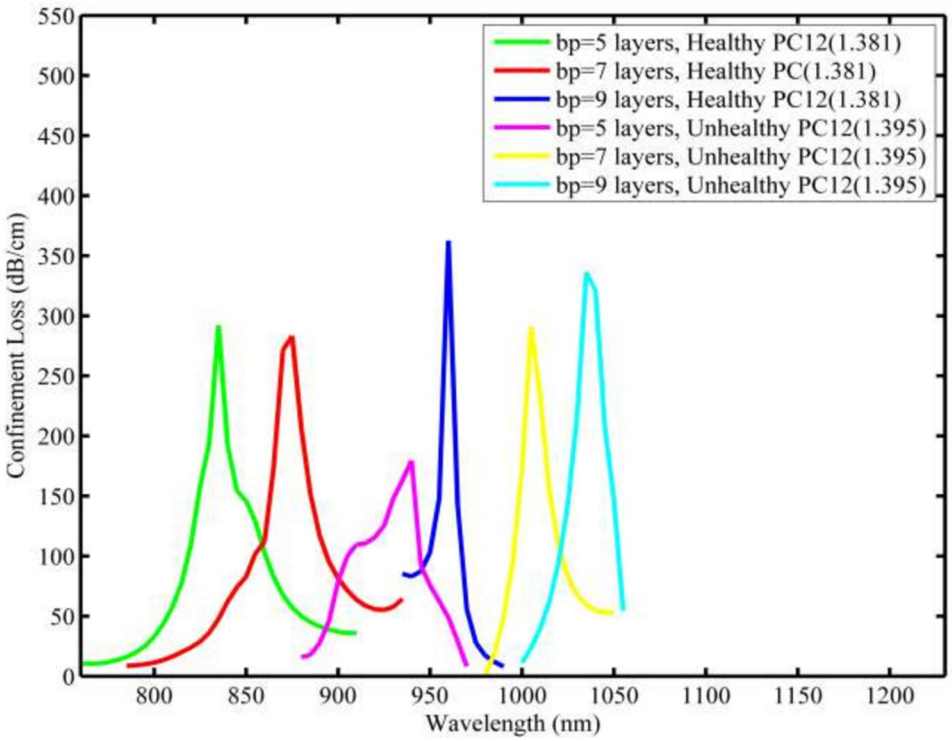


Fig. 8 CL spectra for various number of layers of BP at RI of 1.385 and 1.391

Sensor Performance

While studying the sensitivity analysis at different cancer cell through the presented PCF SPR sensor, different sets of loss spectrum were found which are listed in Fig. 9a to f. In the figure, we have marked the spectra for healthy cells in blue and infected cells in red. It can be seen evidently that all the cells have gone through a wavelength shift, when infected counterparts are associated. The wavelength shift for PC12 is observed as 130 nm, for Jurkat (~ 85 nm), for basal cell (~ 95 nm), for HeLa (~ 130 nm), for MCF-7, we calculated a shift of 160 nm, and for MDA-MB-231 cell, the respective shift is 150 nm. The former data provides a sensitivity of 9285.71 nm/RIU, 6071.43 nm/RIU, 4750 nm/RIU, 5416.67 nm/RIU, 11,429 nm/RIU, and 10,714 nm/RIU, respectively. The BP layers increased the sensitivity by 30%. This shows that the proposed biosensor is suitable for cancer detection.

Table 8 represents the comparison of this work with the recently published work to give a clear view on the reliability of the sensor. This proves that the proposed sensor can detect any minute change in the refractive index in the cell of a human body and gives a required sign of any cancerous presence.

Table 7 Optimization of number of BP layers (bp)

No. of BP layers (bp)	Max. CL (Normal cell) (dB/cm)	Max. CL (Infected cell) (dB/cm)	ΔWL (nm)	Sensitivity (nm/RIU)
5	108.37	192.39	100	7142.86
7	283.12	290.81	130	9285.71
9	362.55	336.34	75	5357.14

Conclusion

The proposed slotted PCF SPR sensor is designed with two types of oblong air holes and circular air holes that show a high sensitivity toward the detection of cancerous cells. In the proposed sensor a 50 nm thick layer of gold and 7 layers of black phosphorus with a monolayer thickness of 0.53 nm is used as plasmonic materials. Titanium Dioxide (TiO_2) is used as an adhesive between the fiber and gold having thickness (~ 12 nm), but it also has remarkable plasmonic properties leading to its contribution in the enhancement of sensitivity. COMSOL Multipurpose software is used to determine the sensor's performance using finite element method (FEM). All the different parameters

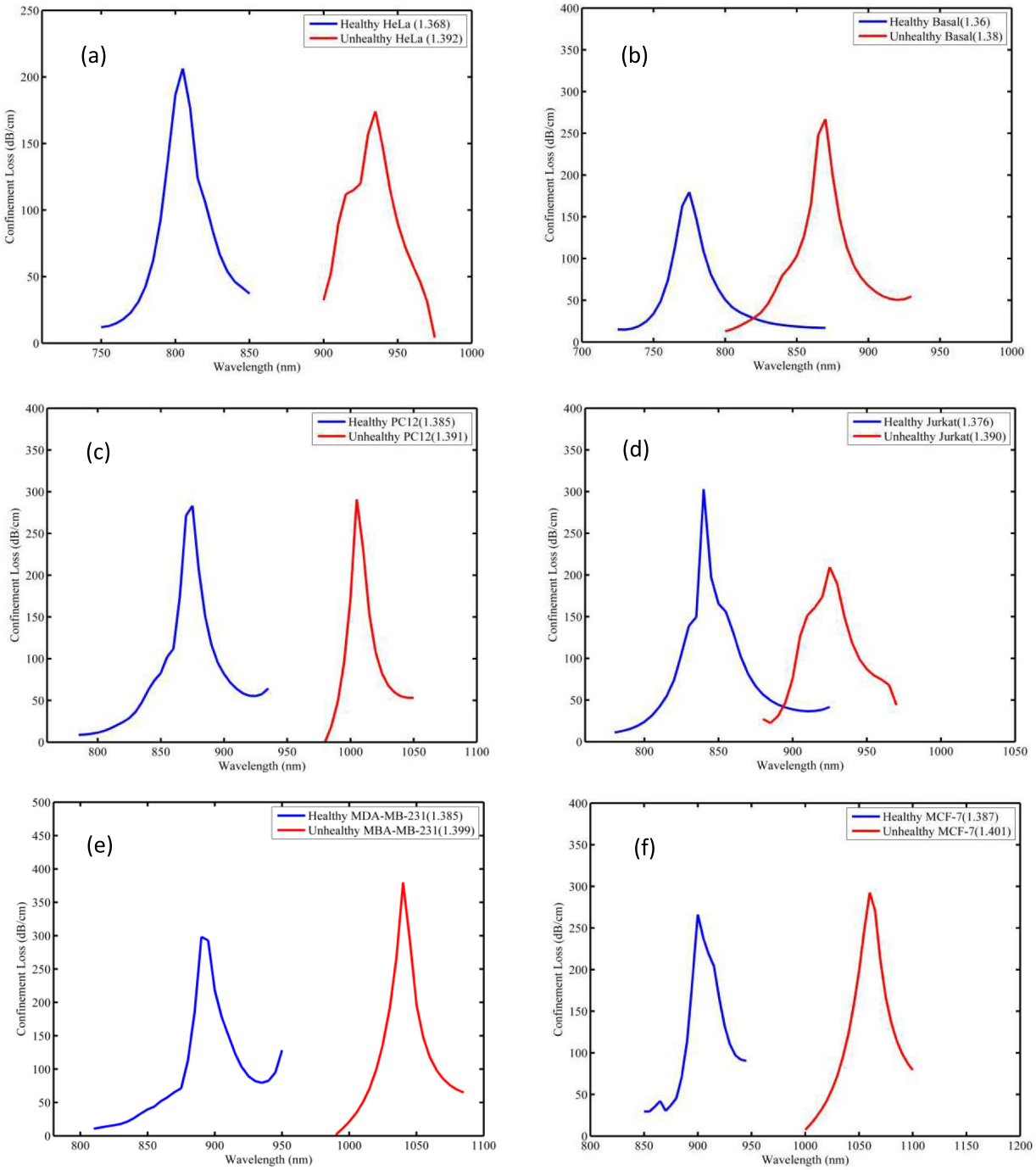


Fig. 9 Loss spectra of **a** HeLa (cervical cancer); **b** basal cell (skin cancer); **c** PC12 (adrenal gland); **d** Jurkat (blood cancer); **e** MDA-MB-231(breast cancer type-1); **f** MCF-7 (breast cancer type-1)

Table 8 Comparison of proposed work with past research works

Reference	Cancer type	Wavelength sensitivity (nm/RIU)
[10]	Skin	3500
	Cervical	5000
	Blood	5714
	Adrenal gland	6429
	Breast type-1	7143
	Breast type-2	7143
[11]	Skin	3150
	Cervical	4333.33
	Blood	4642.86
	Adrenal gland	5500
	Breast type-1	6428.57
	Breast type-2	7142.86
[12]	Skin	3750
	Cervical	5417
	Blood	6071
	Adrenal gland	7500
	Breast type-1	9643
	Breast type-2	11,429
[13]	Skin	–
	Cervical	5500
	Blood	6000
	Adrenal gland	7571.43
	Breast type-1	9428.57
	Breast type-2	10,714.28
[34]	Skin	3800
	Cervical	5008.33
	Blood	–
	Adrenal gland	–
	Breast type-1	6214.28
	Breast type-2	–
Proposed PCF SPR sensor	**Skin**	**4750**
	Cervical	**5417**
	Blood	**6071**
	Adrenal gland	**9286**
	Breast type-1	**10,714**
	Breast type-2	**11,429**

involved in the calculations are optimized to achieve the atmost accuracy. The proposed sensor can detect six kinds of cancer cells (PC12, HeLa, MDA-MB-231, Jurkat, basal, MCF-7). The sensor achieves a very high sensitivity of MCF-7 for the cell 11,429 nm/RIU.

Acknowledgements Arhee Bhuyan, a student of Masters of Physics in the Department of Applied Physics at Delhi Technological University, acknowledges for providing excellent research facilities for conducting research.

Author Contribution A. Bhuyan did the simulation and presented results and wrote the original manuscript, A. Khamaru helped in the explanation of results and helped in paper writing, A. Kumar. reviewed paper, concept building, Supervision, Software, and Formal analysis.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing Interests The authors declare no competing interests.

References

1. Kaushansky K, et al. 2016 Myeloma. In: Williams Hematology. 9th ed. McGraw Hill
2. Labi V, Erlacher M (2015) How cell death shapes cancer. Cell Death Dis 6(3):e1675–e1675
3. Saini A, Kumar M, Bhatt S, Saini V, Malik A (2020) Cancer causes and treatments. Int J Pharm Sci Res 11(7):3121–3134
4. Venkatram P (2024) Other Diagnostic modalities, advantages and disadvantages in different pathologic conditions. In Heart Diseases and Echocardiogram: Principles in Practice Cham. Springer Nature, Switzerland, pp 783–787
5. Mittal S, Kaur H, Gautam N, Mantha AK (2017) Biosensors for breast cancer diagnosis: a review of bioreceptors, biotransducers and signal amplification strategies. Biosens Bioelectron 88:217–231
6. Rankin SC (2005) CT and MRI. Surgery (Oxford) 23(5):162–165
7. Kaur B, Kumar S, Kaushik BK (2022) Recent advancements in optical biosensors for cancer detection. Biosens Bioelectron 197:113805
8. Das S, Devireddy R, Gartia MR (2023) Surface plasmon resonance (SPR) sensor for cancer biomarker detection. Biosensors 13(3):396
9. Khamaru A, Kumar A (2024) As38 Se62 based segmented cladgraded index photonic crystal fiber for supercontinuum generation covering 3–9.5 μm with moderate peak power. Opt Quant Electr 56(7):1246
10. Sardar MR, Faisal M (2024) Dual-core dual-polished PCF-SPR sensor for cancer cell detection. IEEE Sensors Journal 24:9843
11. Yasli A (2021) Cancer detection with surface plasmon resonance-based photonic crystal fiber biosensor. Plasmonics 16:1605–1612
12. Majeed MF, Ahmad AK (2024) Design and analysis of a dual-core PCF biosensor based on SPR for cancerous cells detection. Opt Quant Electron 56(6):1–19
13. Chaudhary VS, Kumar D, Kumar S (2022) Au-TiO 2 coated photonic crystal fiber based SPR refractometric sensor for detection of cancerous cells. IEEE Trans Nanobiosci 22(3):562–569
14. Ibrahimi KM, Kumar R, Pakhira W (2024) Early detection of cancer cells using high-sensitivity dual-side polished photonic crystal fiber biosensors based on surface plasmon resonance. Opt Quant Electron 56(5):888
15. Kumar A, Verma P, Jindal P (2023) Surface plasmon resonance sensor based on MXene coated PCF for detecting the cancer cells with machine learning approach. Microelectron Eng 267:111897
16. Zuhayer A, Abd-Elnaby M, Ahammad SH, Eid MM, Sorathiya V, Rashed ANZ (2022) A gold-plated twin core d-formed photonic crystal fiber (PCF) for ultrahigh sensitive applications based on surface plasmon resonance (SPR) approach. Plasmonics 17(5):2089–2101

17. Tong K, Wang J, Cai Z, Liu Y (2021) Study on increasing measurement range and enhancing sensitivity of PCF surface-plasmon-resonance biosensor using black phosphorus. J Russ Laser Res 42:283–291
18. Zhang G, Huang S, Wang F, Yan H (2021) Layer-dependent electronic and optical properties of 2D black phosphorus: fundamentals and engineering. Laser Photonics Rev 15(6):2000399
19. Karki B, Uniyal A, Pal A, Srivastava V (2022) Advances in surface plasmon resonance-based biosensor technologies for cancer cell detection. Int J Optics 2022(1):1476254
20. Ayyanar N, Raja GT, Sharma M, Kumar DS (2018) Photonic crystal fiber-based refractive index sensor for early detection of cancer. IEEE Sens J 18(17):7093–7099
21. Jabin MA, Ahmed K, Rana MJ, Paul BK, Islam M, Vigneswaran D, Uddin MS (2019) Surface plasmon resonance based titanium coated biosensor for cancer cell detection. IEEE Photonics J 11(4):1–10
22. Islam N, Arif MFH, Yousuf MA, Asaduzzaman S (2023) Highly sensitive open channel based PCF-SPR sensor for analyte refractive index sensing. Results in Physics 46:106266
23. Opoku G, Danlard I, Dede A, Akowuah EK (2023) Design and numerical analysis of a circular SPR based PCF biosensor for aqueous environments. Results in Optics 12:100432
24. Gangwar RK, Amorim VA, Marques PVS (2019) High performance titanium oxide coated D-shaped optical fiber plasmonic sensor. IEEE Sens J 19(20):9244–9248
25. Lee SY, Yee KJ (2021) Black phosphorus phase retarder based on anisotropic refractive index dispersion. 2D Mater 9:015020
26. Pysz D et al (2014) Stack and draw fabrication of soft glass microstructured fiber optics. Bulletin of the Polish Academy of Sciences. Techn Sci 62(4):667–682
27. Atakaramians S (2011) Terahertz waveguides: a study of microwires and porous fibres (Doctoral dissertation)
28. Rifat et al (2017) Photonic crystal fiber based plasmonic sensors. Sens Actuators, B Chem 243:311–325
29. Boehm J, François A, Ebendorff-Heidepriem H, Monro TM (2011) Chemical deposition of silver for the fabrication of surface plasmon microstructured optical fibre sensors. Plasmonics 6:133–136
30. Wu Y et al (2021) Two-dimensional black phosphorus: properties, fabrication and application for flexible supercapacitors. Chem Eng J 412:128744
31. Nielsen K, Noordegraaf D, Sørensen T, Bjarklev A, Hansen TP (2005) Selective filling of photonic crystal fibres. J Opt A: Pure Appl Opt 7(8):L13
32. Majeed MF, Ahmad AK (2024) Design and analysis of a high sensitivity open microchannel PCF-based surface plasmon resonance refractometric sensor. Opt Mater 147:114617
33. Ibrahimi KM, Kumar R, Pakhira W (2023) Enhance the design and performance analysis of a highly sensitive twin-core PCF SPR biosensor with gold plating for the early detection of cancer cells. Plasmonics 18(3):995–1006
34. Abdelghaffar et al (2023) Highly sensitive V-shaped SPR PCF biosensor for cancer detection. Opt Quant Electron 55(5):472

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

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

Wireless Personal Communications
https://doi.org/10.1007/s11277-024-11657-4

Circular RFID Antenna for White Space TV Band

Richa Sharma[1,2] · **N. S. Raghava**[1] · **Asok De**[1]

Accepted: 11 November 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
The proposed antenna design is intended to be used in the 470 MHz to 806 MHz frequency band, which is now available for commercial use following the transfer of television signal transmission from the TV to satellite band. The antenna is more compact than existing RFID antennas in India, with four concentric rings sliced from a circular patch, and patch is shorted to ground by shorting pin. The design was created using FR4, a commonly used and durable substrate. The antenna's gain and directivity are 8.11 dBi and 9.9 dB, respectively. It can operate in two bands within the desired frequency range: 479–492 MHz and 643–653 MHz. These bands have individual bandwidths of 13 MHz and 10 MHz, respectively, which are higher than commercially available UHF RFID antennas. The return loss in these bands is − 29 dB and -14 dB, respectively. The impact of changes in the ground plane is also discussed, and the return loss and bandwidth have been increased to − 29 dB and to 13 MHz, respectively.

Keywords Television White Space (TVWS) band · Radio Frequency Identification (RFID) · Ultra-high frequency (UHF) band

1 Introduction

Radio Frequency Identification (RFID) is a wireless technology that enables the identification of objects using radio waves. RFID antennas are an essential component of RFID systems, as they allow for the transmission and reception of radio signals between RFID readers and tags. The design of RFID antennas has become a subject of great interest in past few years, largely due to growing demand for RFID technology in various applications such as inventory management, logistics, and security. Various compact RFID antenna

✉ Richa Sharma
richa.sharma70@gmail.com

N. S. Raghava
nsraghava@dtu.ac.in

Asok De
asok.de@gmail.com

[1] ECE Department, Delhi Technological University, New Delhi, India

[2] ECE Department, Galgotias College of Engineering & Technology, Greater Noida, U.P., India

design techniques are presented by many researchers, which are working on different frequencies varying from low frequency to high frequency [1].

RFID antennas are designed for two frequency ranges in the ultra-high frequency (UHF) band: 433 MHz and 860–960 MHz. Commercial use of the 433–860 MHz frequency range is avoided by manufacturers, as it was previously reserved for television services. However, with the relocation of television services to the C-band and Ku-band, this frequency range is now considered "white space" in the TV band, and available for commercial use. This presents an opportunity for RFID technology to operate in this range without interfering with television transmissions. The white space TV band also provides the potential for long-range and high-bandwidth RFID communication, enabling new and innovative applications such as improved inventory management and tracking in supply chain logistics [2]. RFID technology operates similarly to barcodes and magnetic strips on credit cards, but with the advantage of not requiring a specific orientation in relation to the reader [3]. The growth of RFID technology provides opportunities for antenna designers to create various types of RFID antennas [4]. When an RFID tag comes into range of an RFID scanning antenna, electromagnetic waves trigger it, which the antenna picks up and sends to an RFID reader for decoding [5]. RFID tags operate in different frequency ranges depending on their intended use.

Governments around the world allocate bandwidth for RFID applications, with each country dividing its allocated bandwidth into different channels. In India, the 865–867 MHz band with 2 MHz bandwidth is used for RFID applications [6]. Different shape antennas are available for RFID applications in the literature, but microstrip techniques have some disadvantages such as small bandwidth, low gain, and low efficiency [7, 8]. The authors suggest that by printing different slots on a microstrip antenna, the benefits of using FR-4 substrate for RFID technology can be realized [9]. Researchers have proposed solutions to these issues, including designing microstrip patch antennas [10], changing ground [11, 12], cutting various shaped slots [13–16], using different substrates with varying heights [17, 18], and using meander-lines [19]. Planar antenna technology can also be used in compact wireless systems, such as RFID applications [20].

The main purpose of antenna design is to create an RFID antenna with a high gain that can function effectively in the TVWS band. Here a circular patch antenna on an FR4 substrate that includes four circular rings etched onto the patch is developed. The effect of changes in the ground plane is also studied. To enhance the antenna's gain, a finite ground plane is used in the final design. Engraving circular rings on a patch shifts its center frequency of the multiple resonance. The novelty of designing RFID (Radio Frequency Identification) antenna in white space TV band includes: i) to utilize unused TV spectrum (also known as "white space") for RFID communication because white space TV band is the portion of the TV spectrum that is not being used by any TV broadcast or wireless communication system in a given geographic area ii) This approach allows for longer range communication and better penetration of obstacles, making it suitable for different applications. iii) using the white space TV band for RFID communication avoids interference with existing wireless systems, which is a critical concern in densely populated areas. iv) It also provides a cost-effective solution as the infrastructure for TV broadcasting is already in place, reducing the need for additional infrastructure for RFID communication.

The paper is divided into six parts. The antenna design equations are defined in the second half of the paper. The parametric analysis of antenna design is explored in third part of paper. The proposed design results are compared in the paper's fourth part. The results of proposed designed is compared with existing work in fifth part of paper. The proposed idea's findings and observations are summarized in the paper's sixth part.

2 Proposed Antenna Design

The structure of the proposed antenna is depicted in Fig. 1 below which is working at 480 MHz. The radius of the microstrip patch is denoted by the letter 'r'. The patch antenna's radius is computed from the Eq. (1).

$$r = \frac{F}{\left\{1 + \frac{2h}{\pi\varepsilon_r F\left[\ln\left(\frac{0.5\pi F}{h}\right)+1.7726\right]}\right\}^{1/2}} \tag{1}$$

The resonant frequency of the antenna is achieved by the formula $F = \frac{K_{nm}c}{2\pi a_e\sqrt{\varepsilon_e}}$ [8], where K_{nm} is the m.th root of the derivative of the Bessel function of order n and c is the velocity of light. The field variations of a circular patch are elucidated in terms of the Bessel function. The ε_e and a_e are the effective dielectric constant and the effective radius of the patch antenna, respectively. The fringing fields along the boundaries of the CMSA are contemplated by substituting the radius r by the effective radius a_e given by the following equation: [16]

$$a_e = a\sqrt{\left[1 + \frac{2h}{\pi a\varepsilon_r}\left\{In\left(\frac{0.5a}{h}\right) + \frac{1}{\sqrt{2}}\varepsilon_r + \frac{1}{0.564} + \frac{h}{a}\left(\frac{1}{3.73}\varepsilon_r + 1.65\right)\right\}\right]} \tag{2}$$

The value of ε_e is obtained by using the ratio of the total capacitance of the dominant mode of CSMA with and without a dielectric substrate [17, 18].

In the proposed design FR4 substrate of 1.6 mm height with permittivity 4.4 and loss tangent, 0.02 is used. The antenna is designed in three steps- First basic circular patch is designed and parameters studied done, in the second step, rings are etched on the circular patch and finally center shorted circular patch in which four rings are etched with a finite ground plane is designed and various parameters of the final design are studied. High

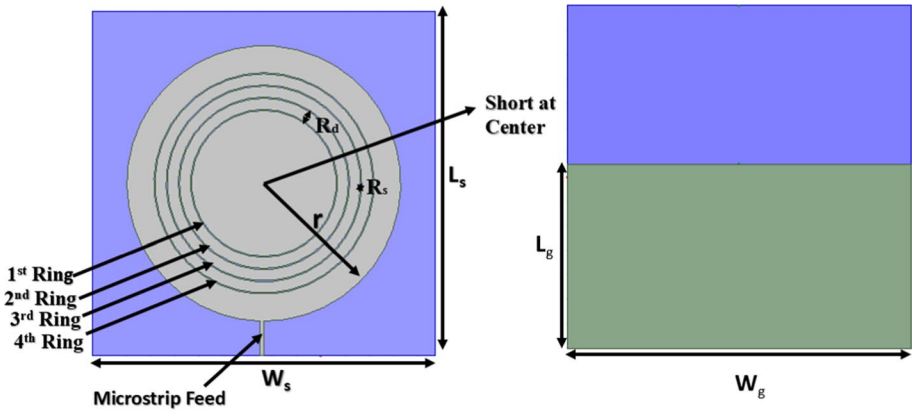


Fig. 1 Proposed antenna **a** Front View and **b** Back View

Frequency Simulation Software (HFSS) is used for simulation and microstrip line feed is chose as excitation of antenna.

Here 'L_g' and 'W_g' denotes the length and width of ground plane. 'W_s' and 'L_s' is the width and length of substrate. The distance between ring is denoted by 'R_d' and size of ring slot is denoted by 'R_s'. In this work, a patch antenna is designed on which four rings are etched from the copper patch. The size of each ring is 1 mm and each ring is separated from other ring by a 10 mm distance. The antenna is energized by a microstrip feed line at the center of the width. The addition of rings and a finite ground gives desired results of bandwidth and gain. After optimization the proposed design parameters are: r = 112 mm, W_s = 280 mm, L_s = 280 mm, W_g = 280 mm, L_g = 150 mm, R_d = 1 mm, R_s = 10 mm.

3 Parametric Analysis

This section explores the impact of L_g (ground size), R_s (ring size), and R_d (distance between ring) on the performance of the proposed RFID antenna.

3.1 Effects of ground size L_g

Parametric analysis has been done to find out the optimum size of ground plane. A larger ground plane tends to act as a more effective reflector, confining the radiated energy. Conversely, reducing the ground plane size changes the current distribution and the effective radiation patterns of the antenna. The size of L_g is varied from 280 to 150 mm. The resonant frequency and S_{11} are dependent on the ground dimensions, as depicted in Fig. 2. It has been found that there is an enhancement of S_{11} from − 10.4 dB to − 29 dB, and the resonance shifted from of 480 MHz to 487 MHz when L_g decreased. When ground size is reduced up to 210 mm, S_{11} improved but resonance shifted to higher side. As we reduce the ground plane size up to 210 mm, the effective electrical length of the antenna is shortened, which leads to a higher resonant frequency. This is because the smaller ground plane

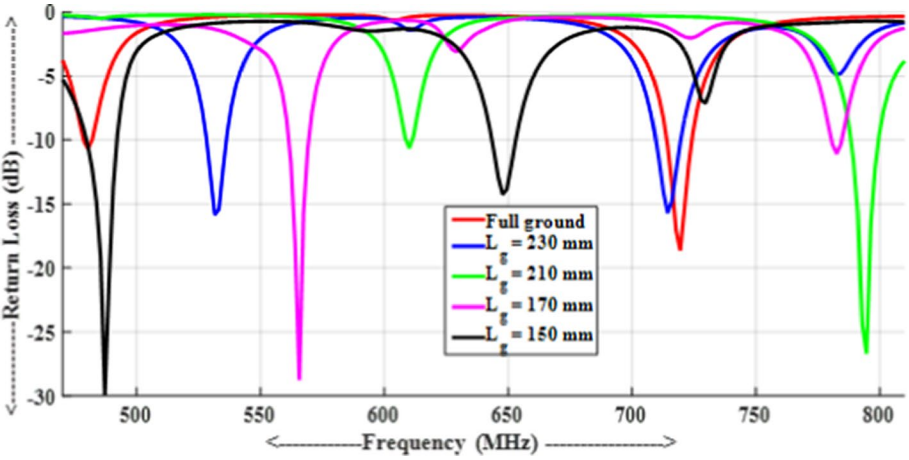


Fig. 2 Simulated S-parameter of proposed antenna as a function of L_g

allows for less confinement of the electromagnetic fields, effectively reducing the antenna's electrical length and pushing the resonance upwards. Up to 210 mm, the ground plane size reduction improves the impedance matching, because the smaller ground plane better matches the free-space radiation impedance of the antenna.

After 210 mm resonance starts shifting towards lower side with better return loss. This happens because, at smaller ground plane sizes, the ground plane no longer support efficient reflection and radiation. The fields become less confined, and the overall effective electrical length of the antenna increases. This increase in effective electrical length causes a shift to a lower frequency. While the resonance frequency shifts lower, the return loss S_{11} continues to improve due to better matching. This is likely due to the optimized coupling between the radiating element and the ground plane at this size, where the energy radiated is efficiently transferred with minimal reflection. The optimum value of L_g obtained is 150 mm. Due to finite ground, bandwidth has increased from 3 to 13 MHz. The increase in bandwidth is due to the broader impedance matching at lower Lg values. A smaller ground plane tends to result in a more "leaky" structure, which enhances the radiation efficiency across a wider range of frequencies, thus increasing the bandwidth.

3.2 Effects of Gap between Ring R_d

To find out the optimum gap between rings, R_d is varied for desired results. The effects of gap between ring are depicted in Fig. 3. The gap between rings is varied from 5 to 15 mm. As the gap between rings are increased multiple resonance is received because when the rings are close together, they are strongly coupled, and the electromagnetic fields of one ring influence the behavior of the other. This strong coupling generally leads to a single dominant resonance, as the rings work together as a composite resonant structure. When the gap between the rings increases, the coupling weakens. Each ring begins to resonate more independently, with its own natural resonant frequency. This decoupling allows each ring to contribute its own resonance to the system, leading to the appearance of multiple resonant frequencies. The optimum gap of 10 mm achieves the desired resonance because it strikes a balance between strong and weak coupling, supporting the resonance mode at the target frequency while still allowing some impedance between the rings.

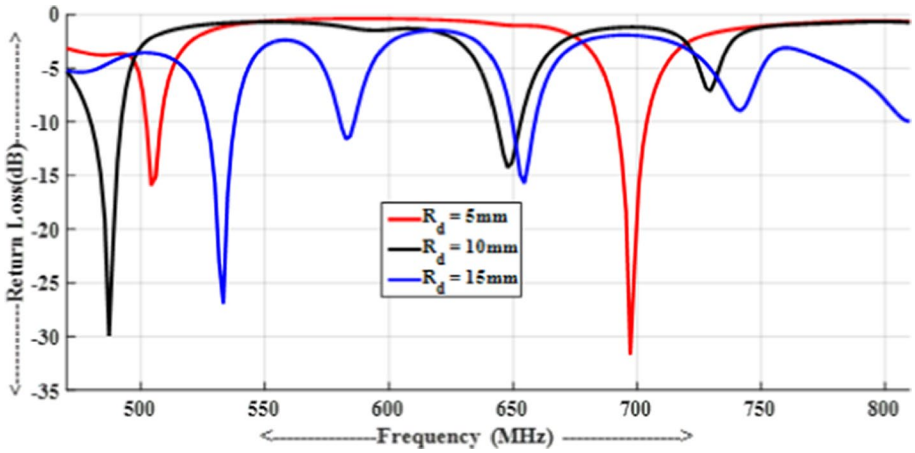


Fig. 3 Simulated S-parameters of proposed antenna as a function of R_d

3.3 Effects of Ring Slot Size R_s

Effects of size of ring slot are shown in Fig. 4. The size of ring slot (R_s) is varied from 0.5 to 1.5 mm. It is observed that the optimum value for R_s is 1 mm for best matching of $S_{11.}$ The change in the size of ring slots is less sensitive in this design because after changing the ring size only return loss is going to change. There is no change in the resonance of antenna. The S_{11} is changed to −30 dB with a change of 2 MHz in resonant frequency.

4 Simulated and Measured Results

The suggested work demonstrates that by reducing the ground plane dimensions, the antenna bandwidth can be increased. Due to a shorting pin connecting the patch to the ground, two frequency bands are received. From Fig. 5, it is depicted that the S_{11} is improved from − 10.4 dB to − 29 dB and bandwidth improved from 3 to 13 MHz due to a change in ground plane size.

The proposed antenna has been manufactured and hardware results has been compared with simulated results. The fabricated design is shown in Fig. 6. The standard antenna used for gain measurement is the Amitech DRH 20 Horn, which has a working range of MHz. The measurements are performed in an anechoic chamber as shown in Fig. 7 to avoid reflections and external interference. The standard antenna and the proposed antenna under test are placed in the far-field region, ensuring that the antennas are aligned properly and are at the same height. In the measurement process, the signal generator (Anritsu MG3692B) is connected to the reference antenna, and it is pointed towards the proposed antenna. Then, the power level is measured using the spectrum analyzer (HP 8563E) connected to the proposed antenna under test (PAUT). Subsequently, the positions of the reference antenna and the PAUT are swapped, and connect the signal generator to the PAUT. Point the PAUT towards the reference antenna and measure the received power level using the spectrum analyzer connected to the reference antenna. It is cleared from Fig. 8 that bandwidth is more than 2 MHz which is used for RFID operation in the Indian region that is why the proposed antenna is suitable for

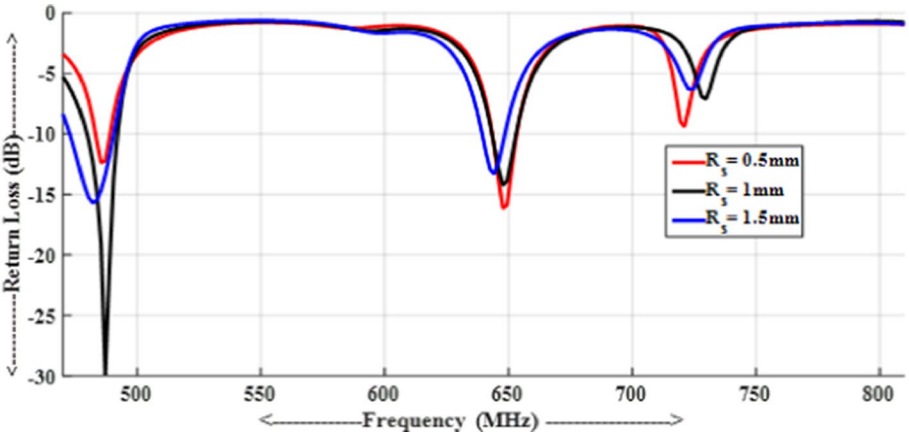


Fig. 4 Simulated S-Parameter as a function of R_s

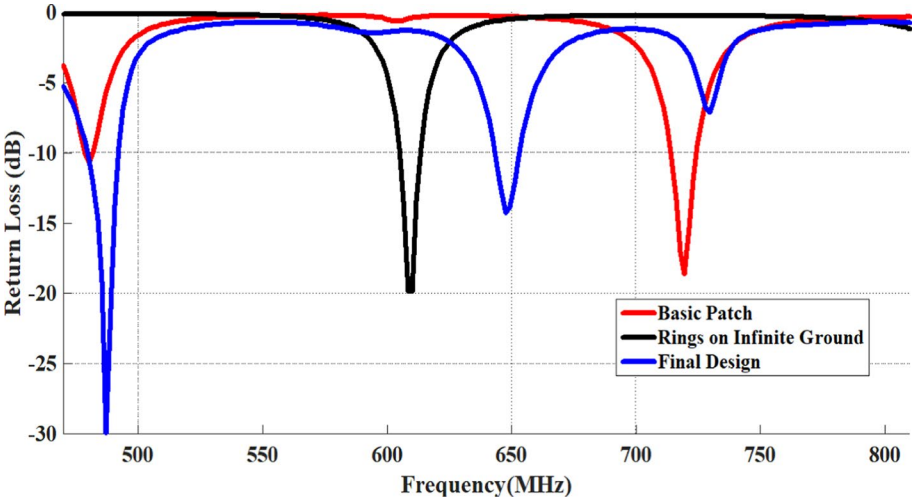


Fig. 5 Simulated return loss of proposed antenna for three stages

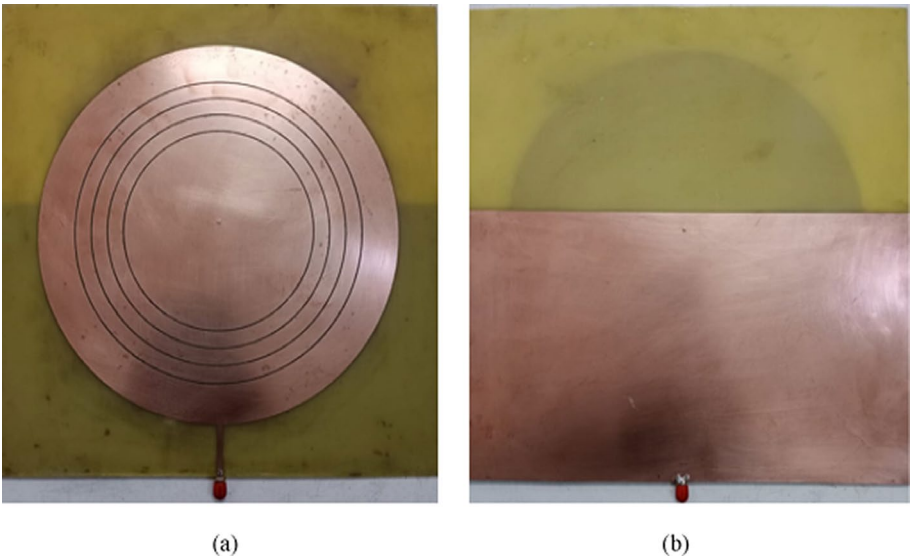


Fig. 6 Fabricated Design **a** Front View and **b** Back View

RFID application with a return loss of − 29 dB. Due to ring structure on main patch, the operational frequency is moved from 480 to 487 MHz in simulation. In terms of the resonant frequency, there was an error of 1.4% in simulated results. The comparison between simulated and experimental results reveals that the experimental results exhibit a higher resonant frequency. The S_{11} of fabricated design is − 23 dB at 482 MHz frequency. The identical antenna may be used at 647 MHz frequency for bandwidth of 10 MHz and return loss of − 14 dB.

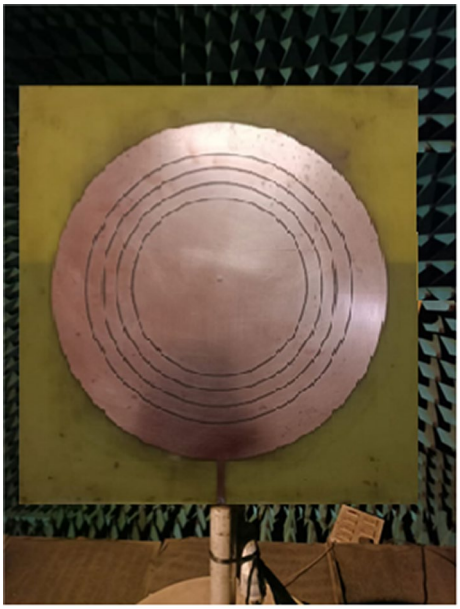


Fig. 7 Proposed antenna under test in anechoic chamber for gain measurement

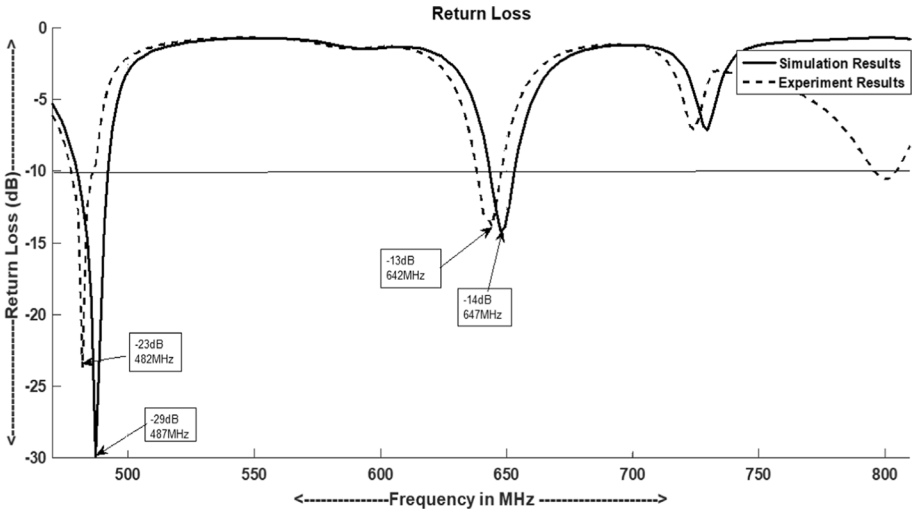


Fig. 8 Simulated and Experimental S-parameter

The efficiency of an antenna design can be determined by its voltage standing wave ratio (VSWR). Furthermore, VSWR is a reliable indicator of how much power an antenna reflects. From Fig. 9 it is indicated that the antenna is adequate for practical use as the simulated VSWR value is 1.07 while the experimental VSWR is 1.13.

Figure 10 depicts the radiation plot of the presented antenna. Most of the signals are only transmitted in one direction. The directivity of the antenna is 9.9 dB at 480 MHz frequency. Figure 11 displays the antenna's radiation pattern at phi = 0^0 and 90^0, at a

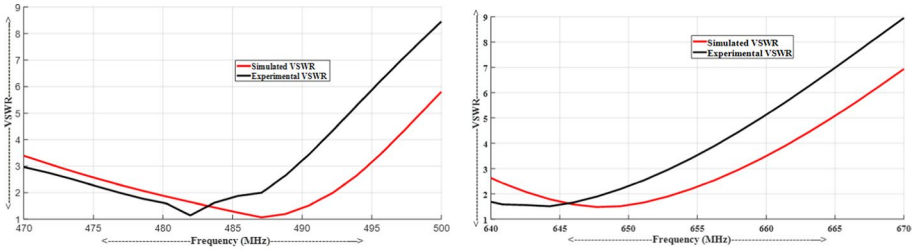


Fig. 9 VSWR plot for **a** 1st band **b** 2nd band of proposed design

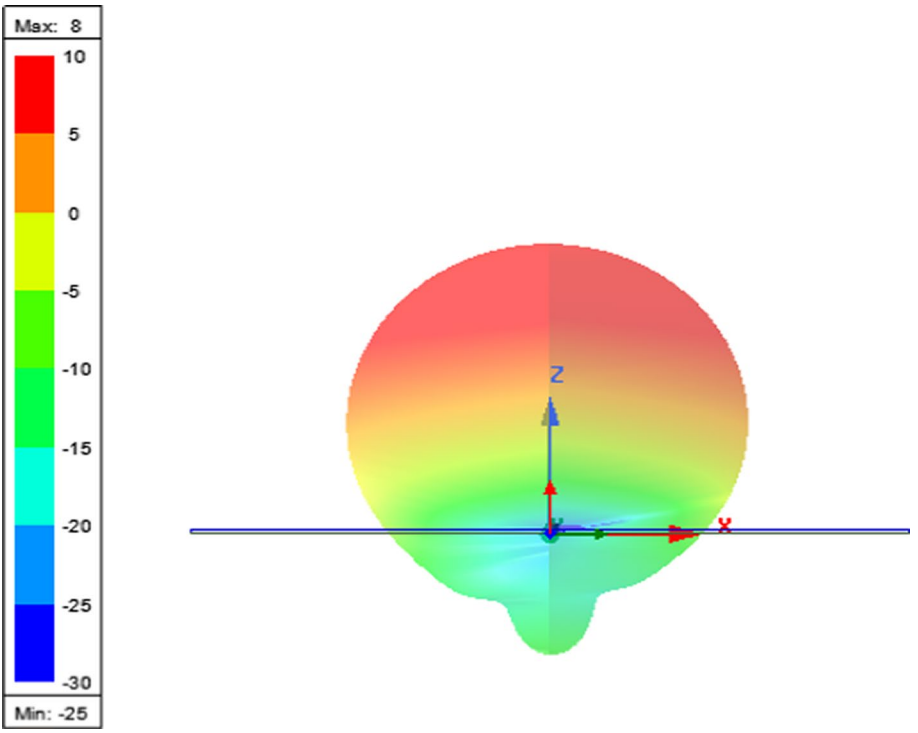


Fig. 10 Radiation plot of proposed antenna

resonant frequency of 487 MHz. The gain magnitude in the main lobe is 8.11 dBi. As indicated in the graph, maximum gain is directive. Figure 12 shows the total gain for the antenna at phi = 0^0 and 90^0 at 647 MHz in 2nd band. The normalized value of cross polarization is less than 20 dB for both the bands. Figure 13 shows the comparison of measured and simulated gain versus frequency. There is some variation between measured and simulated gain. The measured gain achieved is 7.5 dBi while the simulated gain is 8.11 dBi over the desired frequency range. In a given direction, the phase and amplitude of the co- polarization and cross-polarization components are entirely

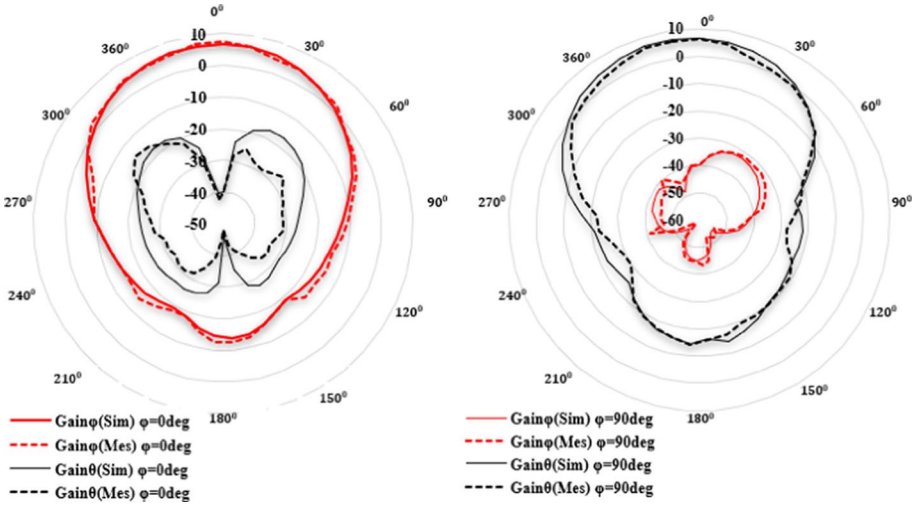


Fig. 11 Simulated and experimental radiation patterns of the proposed antenna at: 487 MHz

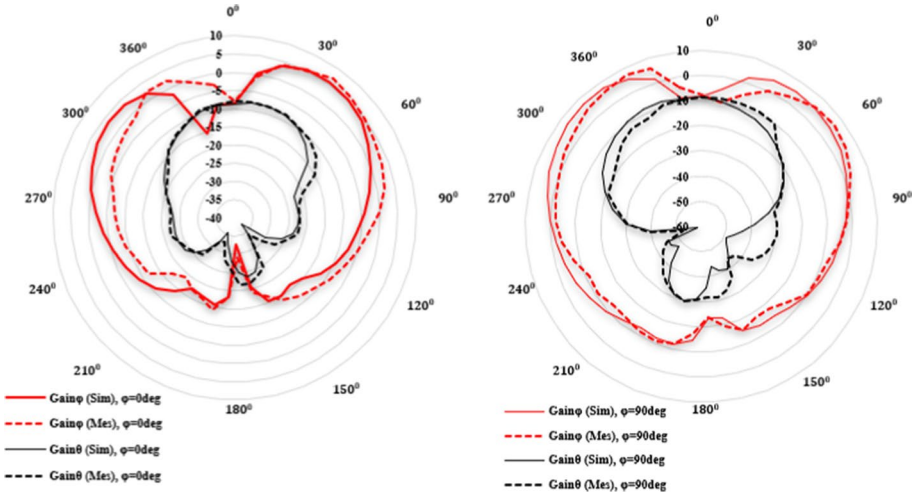


Fig. 12 Simulated and experimental radiation patterns of the proposed antenna at: 647 MHz

coupled. The real and imaginary impedance is plotted with frequency and its impedance is matching at a resonant frequency. The impedance plot is depicted in Fig. 14.

Surface current density of proposed design for 487 MHz frequency is depicted in Fig. 15. Surface current density is maximum at phase angle of 90^0. Electric field of proposed antenna design at 487 MHz is shown in Fig. 16. From electric field it is cleared that all the four rings are radiating. The outer ring radiates less as compared to inner ring corresponding to 480 MHz.

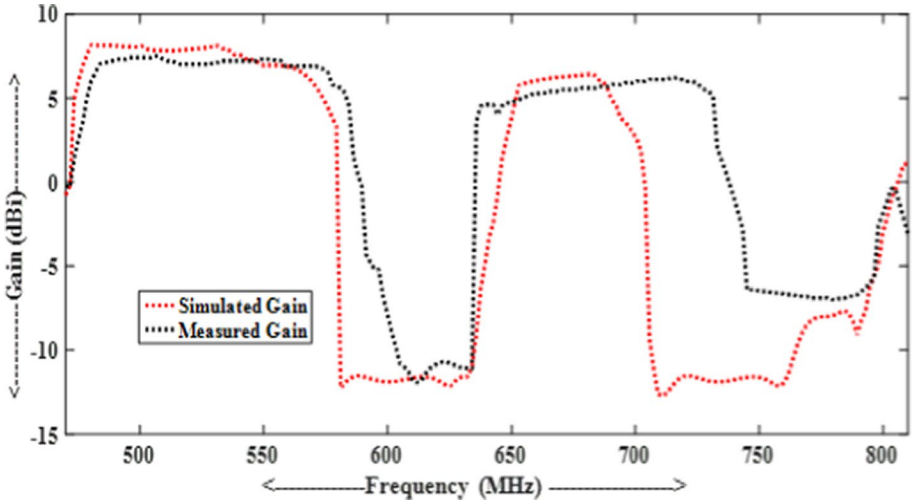


Fig. 13 Simulated and experimental gain vs frequency

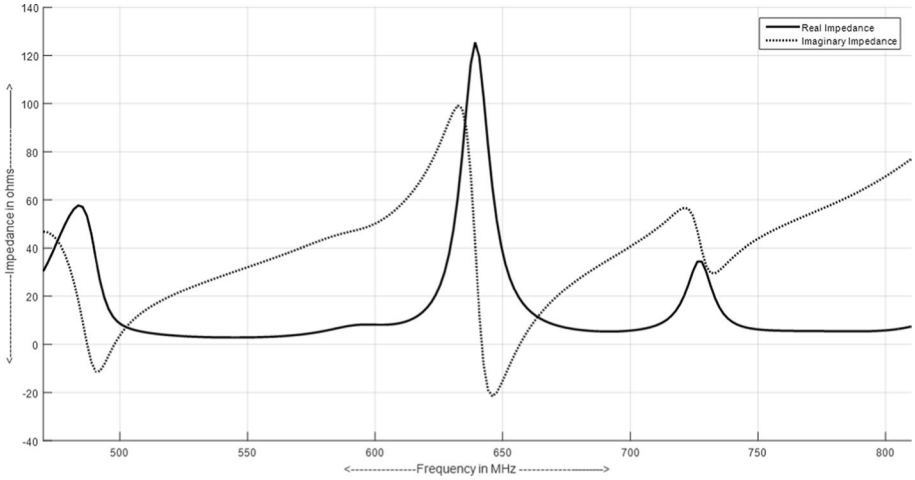


Fig. 14 Real and imaginary impedance of proposed design

5 Antenna Observations and Comparisons with Existing Designs

Parameters of designed antenna have been compared with the commercially available RFID antenna in Table 1. The proposed design is having the lowest VSWR of 1.07 as compared to 1.4 for all compared antennas. Bandwidth is also increased from 3 to 13 MHz. It was observed that the proposed design has a compact size with a good agreement of gain, bandwidth, VSWR, and front-to-back ratio and limited ground size.

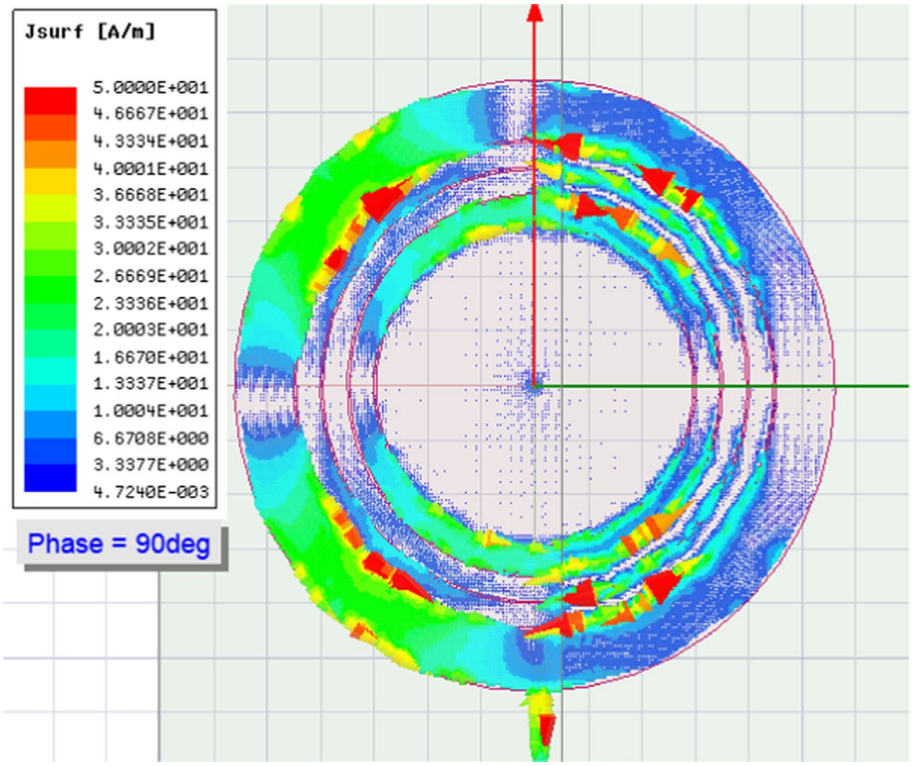


Fig. 15 Surface current density of proposed design

6 Conclusion

In this paper a RFID antenna is proposed to use 470 MHz to 806 MHz (white space T V frequency band) frequency band which is now available for commercial use as the transmission of television signals has transferred from the TV to satellite band. Antenna is fabricated on FR4 substrate which is fairly common and durable substrate. The design consists of four rings sliced from a circular patch, connected to the ground via a shorting pin. In this design finite ground is used. The proposed antenna experimental gain is 7.5 dBi. The antenna may operate on two bands in the intended frequency spectrum, namely 479–492 MHz and 643–653 MHz. The bandwidth obtained in these bands are 13 MHz and 10 MHz, individually, that is greater than commercially available RFID antenna. Return loss of two bands is − 29 dB and − 14 dB, respectively. The impact of ground plane changes is also discussed, and return loss and bandwidth has been increased from − 10.4 to − 29 dB and 3 to 13 MHz respectively. In comparison to the currently existing RFID antenna in India, the proposed antenna is compact in dimensions. An effective low-profile compact design with restricted ground size and circular ring slots has been modelled for the bandwidth of 13 MHz and S_{11} of − 29 dB. Overall, the novelty of designing RFID antenna in white space TV band lies in the potential to leverage an underutilized portion of the radio spectrum for reliable and efficient RFID communication.

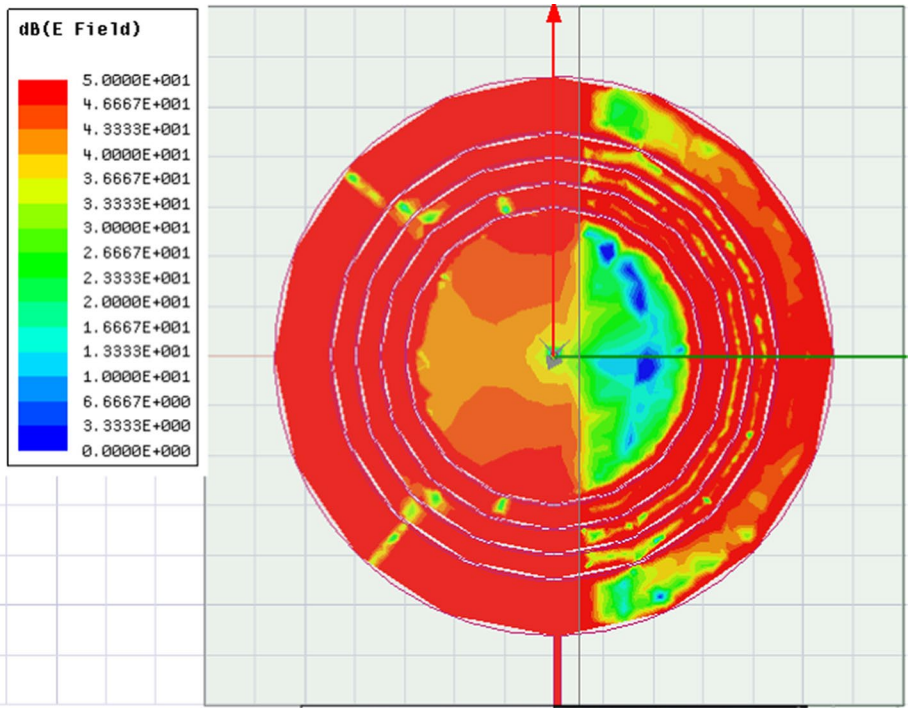


Fig. 16 Electric Field of proposed design

Table 1 Comparison of proposed design with commercially available RFID antenna

Model No	Size (mm)	Freq. (MHz)	Gain	FBR	weight	HPBW
ETS-RA 02	260×260×45	865–868	8dBi	≥20	1.2 kg	$65^0/65^0$
ETS-RA 03	140×120x 20	865–868	5dBi	≧23	0.4 kg	$65^0/55^0$
ETS-UFMAS 15	900×500×25	865–868	5dBi	≧24	1.5 kg	$95^0/84^0$
CP11 RFID UHF	180×180×17.5	865–868	0.2 dBi	-	485 g	120°/120°
p11 RFID UHF	230×230×25	865–868	3.2–3.4 dBiC	<−15	545 g	100°/100°
p12 RFID UHF	277×137×3.3	865–868	5.4–5.5 dBiC	-	230 kg	60°/90°
p13 RFID UHF	417×137×3.3	865–868	1.1–3.2 dB	<−20	350 g	40°/90°
p14 RFID UHF	557×137×3.3	865–868	8.0–8.1 dBiC	<−18	460 g	30°/90°
p16 RFID UHF	837×137×3.3	865–868	9.4–9.5 dBiC	<−20	725 g	20°/90°
p22 RFID UHF	277×277×3.3	865–868	0.5–2.5 dB	<−20	470 g	60°/60°
p33 RFID UHF	417×417×3.3	865–868	10.7–10.8 dBiC	<−20	1050 g	40°/40°
SP11 RFID UHF	207×207×11.7	865–868	7.7–8.0 dBiC	<−15	1055 g	70°/70°
SP12 RFID UHF	417×207×11.7	865–868	9.3–9.5 dBiC	<−20	2520 g	40°/70°
Proposed Design	228×228×1.6	479–492	7.5 dBi	>21	280 g	40°/40°

Author Contributions The novelty of designing RFID (Radio Frequency Identification) antenna in white space TV band is to utilize unused TV spectrum (also known as "white space") for RFID communication. White space TV band is the portion of the TV spectrum that is not being used by any TV broadcast or wireless communication system in a given geographic area. This approach allows for longer range communication and better penetration of obstacles, making it suitable for different applications. Using the white space TV band for RFID communication avoids interference with existing wireless systems, which is a critical concern in densely populated areas. It also provides a cost-effective solution as the infrastructure for TV broadcasting is already in place, reducing the need for additional infrastructure for RFID communication.

Funding Not Applicable.

Data Availability Yes.

Code Availability Not Applicable.

Declarations

Conflicts of interest There is no conflict of interest.

References

1. Abdulghafor, R., et al. (2021). Recent advances in passive UHF-RFID tag antenna design for improved read range in product packaging applications: A comprehensive review. *IEEE Access, 9*, 63611–63635. https://doi.org/10.1109/ACCESS.2021.3074339
2. Sardroud, J. M. (2012). Influence of RFID technology on automated management of construction materials and components. *Scientia Iranica, 19*(3), 381–392.
3. Ariff, M. H., Hisyam, M. Y., Ibrahim, M. Z., Ismarani, I., & Shamsuddin, N. (2018). Circular microstrip patch antenna for UHF RFID reader. *Journal of Telecommunication, Electronic and Computer Engineering, 10*(1–2), 61–65.
4. Ariff, M. H., Ismarani I., and Shamsuddin, N., (2015) Microstrip antenna based on rectangular patch with arms and partial ground plane for UHF RFID readers, *in: IEEE 6th Control and System Graduate Research Colloquium (ICSGRC), Shah Alam*. pp. 61–65.
5. Zhu, F., Xiao, B., Liu, J., Wang, B., Pan, Q., & Chen, L. J. (2016). Exploring tag distribution in multi-reader RFID systems. *IEEE Transactions on Mobile Computing, 16*(5), 1300–1314.
6. The Global language of business (2021) Regulatory status for using RFID in the EPC Gen2 (860 to 960 MHz) band of the UHF spectrum 13 March.
7. Shen, L., Huang, C., Wang, C., Tang, W., Zhuang, W., Xu, J., & Ding, Q. (2016). A Yagi-Uda antenna with load and additional reflector for near-field UHF RFID. *IEEE Antennas and Wireless Propagation Letters, 16*, 728–731.
8. Shen, J., Fang, Y., Kong, L., and Liu, X., (2021) A Slot-based RFID Reader Antenna Design with Broad Left to Right Space Detections, *in: 2021 IEEE 4th International Conference on Electronic Information and Communication Technology (ICEICT)*, pp. 620–622, https://doi.org/10.1109/ICEICT53123.2021.9531039.
9. Kulkarni, S., & Kasabegoudar, V. (2017). Bandwidth enhancement of compact circular slot antenna for UWB applications. *Global Journal of Researches in Engineering, 17*(1), 9.
10. Tashi, M., Hasan. S., and Yu, H., (2011) A complete planner design of microstrip patch antenna for a passive UHF RFID tag, *in: The 17th International conference on automation and computing*, pp. 12–17.
11. Salman, K. N., Ismail, A., Raja Abdullah, R. S., & Saeedi, T. (2017). Coplanar UHF RFID tag antenna with U-shaped inductively coupled feed for metallic applications. *PLoS ONE, 12*(6), e0178388.
12. Aznabet, I., Ennasar, M. A., El Mrabet, O., Andia-Vera, G., Khalladi, M., & Tedjni, S. (2017). A broadband modified T-shaped planar dipole antenna for UHF RFID tag applications. *Progress Electromagnetics Research C, 73*, 137–144.
13. Faudzi, N. M., Ibrahim, A., Mozi, A. M., & Hanaffi, W. N. W. (2021). Wideband slotted microstrip patch antenna for UHF-RFID reader. *Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/1755/1/012028

14. Nakano, H., & Vichien, K. (1989). Dual-frequency square patch antenna with rectangular notch. *Electronics Letters, 25*, 1067–1068.
15. Singh, R., De, A., & Yadava, R. S. (1990). Resonant frequency of elliptical microstrip patch radiator. *International Journal of Electronics Theoretical and Experimental, 69*(3), 385–388.
16. Wang, P., Wen, G., Li, J., Huang, Y., Yang, L., & Zhang, Q. (2012). Wideband circularly polarized UHF RFID reader antenna with high gain and wide axial ratio beamwidths. *Progress In Electromagnetics Research, 129*, 365–385.
17. Helmi, M., & Ragheb, H. (2021). Novel monopole antenna design for UHF circularly polarized RFID tag antenna. *International Applied Computational Electromagnetics Society Symposium (ACES)*. https://doi.org/10.1109/ACES53325.2021.00009
18. Balanis, C. (2014). *Antenna theory: Analysis and design* (3rd ed., pp. 811–865). Wiley.
19. Rushingabigwi, G., Sun, L., He, Y., Zhu, M., Li, Y., & de Dieu Ntawangaheza, J. (2016). The impact of substrate materials to the design of UWB modern antennas. *Journal of Computer and Communications, 4*(03), 20.
20. Babar, A., Ukkonen, L., & Sydanheimo, L. (2010). Dual UHF RFID band miniaturized multipurpose planar antenna for compact wireless systems. *International Workshop on Antenna Technology (iWAT), 2010*, 1–4. https://doi.org/10.1109/IWAT.2010.5464676

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

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

Richa is presently working as Associate Professor in Galgotias College of Engineering and Technology, Greater Noida, U.P., India. She has completed her PhD from Delhi Technological University, New Delhi in 2023. She has received B.Tech and M.Tech in Electronics and Communication Engineering from Sant Longowal Institute of Engineering and Technology (SLIET) in 2009 and 2011 respectively. She has published more than 20 research papers in International Journals/Conferences. She has filled two patent, one of them is granted. She has attended/presented more than 10 National/International Conferences. Her area of interest is Antenna, RF & microwave and Data communication.

Nallanthighal Srinivasa Raghava is working as Professor & Head of the Department in Electronics and Communication Engineering Department in Delhi Technological University (Formerly Delhi College of Engineering), New Delhi. His area of specialization is Antenna and Propagation, Microwave Engineering, Digital Communication, Wireless Communication, Cloud Computing, Information Security.

Asok De obtained his Bachelor and Master of Engineering with specialization of Electronics and Communication from Jadavpur University Kolkata in 1978 and 1980, respectively. He was awarded Ph.D. degree from Indian Institute of Technology Kharagpur in the year 1986. He joined Delhi University, Delhi, India as Lecturer in the year 1984 and promoted as Reader in the year 1987. In 1991 he joined Kolkata University, Kolkata, India as Reader and subsequently joined Delhi College of Engineering (presently Delhi Technological University), Delhi, India as Professor in the year 1997 and worked as Professor in Electronics and Communication Engineering Department, Head of Computer Engineering and Head of Information Technology. In 2005, he joined Ambedkar Institute of Advanced Communication Technology and Research Delhi as Principal and continued with this position till July 2012. Then he worked as Director of National Institute of Technology Patna from 2012–2017. At that period of time, he also had an additional charge of Director of National Institute of Technology Durgapur, India. His area of interest includes Antennas, Transmission Lines, Microwave Circuits, and Computational Electromagnetics. He has published over more than 120 research articles in international journals and more than 130 research articles in international/national conference proceedings. He is a senior member IEEE (USA), fellow IEI (India), fellow IETE (India), and life member CSI (India).

JMEPEG
https://doi.org/10.1007/s11665-024-10551-1

©ASM International
1059-9495/$19.00

ORIGINAL RESEARCH ARTICLE

Comparison of Tin and Indium Thin Film's Structural, Morphological, Adhesion, and Electrical Properties for Ohmic Contact in HgCdTe-Based Infrared Detectors

Shailendra Kumar Gaur, Qasim Murtaza, and R.S. Mishra

Submitted: 3 July 2024 / Revised: 19 September 2024 / Accepted: 20 October 2024

As the requirement for high-performance, low weight, and efficient Infrared devices continues to rise, the formation of good ohmic is a critical step in fabrication of IR detector arrays. Structural and adhesion properties of tin (Sn) and indium (In) films and their contact formation n-HgCdTe substrate are studied using the thermal evaporation technique. The x-ray diffractogram indicated polycrystalline bcc phases with (200) orientation plane and (101) of β-Sn and indium films, respectively. The microstrain and dislocation densities were 4.10 and 9.10 × 10^{10} lines/cm^2 and 3.83 and 9.03 × 10^{10} lines/cm^2 of Sn and In, respectively. SEM examination indicated lumpy deposition of particle sizes from a few nanometers to microns in Sn film and uniform continuous films over the entire HgCdTe substrate with very small numbers of randomly oriented particles of 1-2 μm sizes in In film. EDS compositional analysis confirmed the presence of pure Sn and In films, respectively. Using AFM force microscopy in contact mode, adhesion force and thermodynamic adhesion work were 26.95 nN, 0.42 J/m^2(DMT), and 0.57 J/m^2(JKR) and 50.73 nN, 0.80 J/m^2(DMT), and 1.07 J/m^2 of In and Sn, respectively. Based on the transfer length method (TLM), the Sn/n-HgCdTe interface showed a higher specific contact resistance with nearly ohmic behavior, but the In/n-HgCdTe specific contact resistance of ρ_c = 3 × 10^{-4} Ωcm^2 indicated that the In/n-HgCdTe interface exhibited ohmic features. Based on the experimental results, it was concluded that indium was the most suitable material to use for forming the ohmic n-contact in HgCdTe-based infrared detectors.

Keywords adhesion, AFM, indium, thermal evaporation, thin film, tin

1. Introduction

Infrared detector arrays, particularly those utilized for imaging infrared systems for vital military purposes, are made using tiny-gap $Hg_{1-x}Cd_xTe$. Mercury cadmium telluride (HgCdTe) is an excellent material to use in the development of infrared detectors because of its optical and electrical characteristics. Technology for forming contacts with low and homogeneous contact resistance in HgCdTe detector elements is required for avalanche photodiode detector arrays, high operating temperature performance, reproducibility, and long-term stability. The metal/HgCdTe ohmic contact and low specific contact resistance are two of the key elements influencing the efficiency and dependability of a HgCdTe infrared detector (Ref 1-5). In order to ensure the linearity of current–voltage (I–V) characteristics, a contact is considered "ohmic" if the voltage drop across it is significantly less than the voltage drop across the device. The repeatability and dependability of ohmic connections are additional crucial characteristics (Ref 6, 7). During device fabrication, $Hg_{1-x}Cd_xTe$ alloys are more prone to damage than elemental semiconductors. On HgCdTe, the formation of metallic contacts at room temperature breaks the lattice and results in mixing of the alloy's constituent elements with the contact metal (Ref 8, 9) Contact degradation brought on by this reaction may eventually lead to surface dissociation and heat instability during packing and field use.

When the interfacial barrier's height and width are at their minimum, an effective ohmic contact forms, allowing the current transport to be dominated by a massive tunneling component. On elemental semiconductors, forming an ohmic metallic contact is simple. However, in compound semiconductors, it has become more difficult because pinning the Fermi level in the mid-gap region causes the creation of a high interfacial barrier width (Ref 10-12).

To create an ohmic electrical metal contact on HgCdTe-based IR photodetectors or devices, a variety of metal schemes, including single, double, and triple metal layers, have been presented. These schemes include Cr, In, Al, Au, Ti, Ni, etc (Ref 13-16). A.R. Momen-Baghdadabad et al. (Ref 17) examined the iron thin film characteristics such as magnetization, hardness, friction coefficient, charge transfer resistance, and elemental composition. Authors have studied the charge transfer phenomena in film with XPS and anticipated the improvement in corrosion resistance. They have also evaluated the material hardness and scratch resistance using nanoindentation and nanoscratch tests. The vibrating sample magnetome-

Shailendra Kumar Gaur, Department of Mechanical Engineering, Delhi Technological University, Bawana road, Delhi 110042, India; and Solid State Physics Laboratory, Delhi 110054, India; and **Qasim Murtaza** and **R.S. Mishra**, Department of Mechanical Engineering, Delhi Technological University, Bawana road, Delhi 110042, India. Contact e-mail: gaur.shailendra10@gmail.com.

Published online: 26 December 2024

ter (VSM) study indicates that multiphase steel films exhibit paramagnetic behavior, indicating the possibility of a retained austenite percentage dominating the films.

The TLM method can be used to evaluate the electrical characteristic of the prepare Sn/n-$Hg_{o.4}Cd_{o.6}Te$ junctions. The creation of a surface region that is more strongly doped with n-type than the bulk and the inward diffusion of Sn. At temperatures of up to 300 °C, Sn was identified as an electrically active donor atom. Hg is ejected and replaced by Sn in the uppermost monolayer of the $Hg_{0.78}Cd_{0.22}Te$ lattice at the interface Sn/HgCdTe. An epitaxial layer of SnTe with inclusions of CdTe was shown to increase with additional Sn deposition (and Hg loss) (Ref 15, 18-20).

According to a report (Ref 21), indium (In100 nm) can be used to achieve good ohmic contact for the In/n-ZnSe interface. In/n-HgCdTe, HgTe/In (30 nm)/Au (200 nm)/HgCdTe, Ti/Pt/Au/HgCdTe, and organic (3,4-polyethylene-dioxythiophene) poly (styrene sulfonate) (PEDOT: PSS) /HgCdTe are the other documented ohmic contact schemes [10, 15, 20, 22, 23). With the exception of the PEDOT: PSS /n-type HgCdTe interface, which has demonstrated a modest rectifying behavior, all interfaces had ohmic I–V characteristics (Ref 23). The n-type-In (1.4 μm) metal layer was used in the fabrication of the HgCdTe-based injection laser, which is a light-emitting device with well-behaved electrical properties. On InGaN light-emitting diodes (LEDs), similar In and In (20 nm) /indium tin oxide (ITO-200 nm)-based metal schemes were put into practice to achieve high-quality ohmic connections. The interface based on In and In (200 nm)/ITO (200 nm) was reported to have specific contact resistances of 2.2×10^{-5} Ωcm^2 and 1.8×10^{-3} Ωcm^2, respectively (Ref 24-26). A flat surface topography for the metallic indium/indium oxide/ZnO interface has been achieved by Palimar et al. (Ref 27). According to the literature (Ref 29), two-layered metal contacts (Mo/Au and Mo/In contacts) were used to fabricate various infrared detectors (Mo/Au contact: ohmic contact in discrete detectors, Mo/In contact: indium bumps for multi element FPA). According to a recent report (Ref 28), the thermally grown In (100 nm) film was used to fabricate the ultraclean ohmic van der Waals contacts on molybdenum disulfide (MoS_2). The specific contact resistance (ρ_c) of the In/MoS_2 interface was 5.0×10^{-6}-5.0×10^{-7} Ω cm^2 with a uniform surface morphology of the In film. Consequently, it is widely known that indium (In) can create an ohmic contact on n-type HgCdTe. The substrate/metal interface, surface treatments (during processing), deposition circumstances, and other factors are related to the electrical characteristics of metallic contacts. Thus, it is possible to generate ohmic electric contacts to a HgCdTe-based photodetector while preserving elements such as low-temperature deposition, a highly adherent metallic layer, and low noise levels (Ref 28, 29). Thus, by creating an ohmic metal contact on the HgCdTe substrate and high-density connectivity indium film (HgCdTe detector with Si-ROIC), it is possible to create IR detectors with high reliability and performance. The indium/HgCdTe structure, encompassing all of its structural, compositional, morphological, and electrical features, has not been studied thoroughly earlier.

The physical–chemical phenomenon accountable for the joining of two surfaces when they come into contact is called adhesion. Adhesion force (pull-off force) in nano- or microsystem is controlled by topography of thin film and tip and chemical composition. Corina Birleanu et al. (Ref 30) determined the tribo-mechanical features of 100-500 nm thickness Cr-Au films. They determined the frictional force, work of adhesion, and adhesion force (pull-off force). The Cr-Au pull-off force measured 87-190 nN, the work of adhesion was 1-2.5 J/m^2, and the coefficient of friction was 0.015. Due to their low specific contact resistance and low surface stickiness (adhesion), thin layer gold and silver were utilized in MEMS appliances, particularly in microgrippers (Ref 31, 32).

The aim of the present work is to study the structural, morphological, and electrical properties both In and Sn films on n-type $Hg_{.71}Cd0_{.29}Te$ material to understand their role for making good ohmic contact on n region of HgCdTe detector. In and Sn materials are chosen for the present study as these two materials form a good ohmic contact on n-HgCdTe among all other regularly used metals as reported by P. Leach et al. (Ref 15). This work is specifically done for MWIR HgCdTe detector array for which x value is 0.29. Specific contact resistance is also estimated experimentally for both films to choose the best material between them for MWIR HgCdTe detector array. In addition to that surface adhesion force and adhesion work for both film on HgCdTe are estimated which is a novel work as it is not reported in the literature for HgCdTe material. This is useful for determining the reliability of film.

The characteristics of thermally produced tin and indium films were assessed using a variety of characterization methods, including x-ray diffraction (XRD), energy-dispersive x-ray (EDS), atomic force microscopy (AFM), and scanning electron microscopy (SEM). The results show that the grown indium films, which were used to fabricate indium film (in the configuration of In/n-HgCdTe) for transfer length method (TLM) structures (in the configuration of In/n/p-HgCdTe), are optimal. SEM analysis was used to assess the microstructural characteristics of the (Sn/HgCdTe and In/HgCdTe) interface. I–V/resistance measurements have been used to analyze the formed TLM structures. The resulting In/HgCdTe interface electrical properties exhibit minimal contact resistance and ohmic behavior. In light of this, we may conclude that the indium is best suited material for forming n-type ohmic contact on HgCdTe-based IR photodetector using thermal evaporation method.

2. Materials and Methods

Using the TLM, test structures were created to gauge the Sn-HgCdTe and In-HgCdTe system's contact resistance. The investigation of thin films was conducted using the n-type $Hg_{1-x}Cd_xTe$ epilayers, and a homegrown technique known as vertical dipping liquid phase epitaxy (VD-LPE) was used to create the epilayers on CdZnTe (111) substrate. To carry out the proposed investigation work, epilayers with the following parameters were used: dimensions: 30×30 mm^2, thickness: 10 μm, composition: $x = 0.29$, concentration: 1×10^{16} cm^3, mobility: 416 cm^2 V^{-1} s^{-1}. These were refined utilizing bromine in methanol etching solution in chemo-mechanical and chemical etching procedures. After polishing, ellipsometric parameters were tested to guarantee a defect-free and contamination-free surface. ZnS-capped thermally evaporated CdTe was an easy way to passivate recently polished and etched surfaces. ZnS/CdTe films were etched subsequent to the PLG contact pattern generation. Using a regulated junction depth (1.0-1.4 μm), boron ions are gradually diffused into the HgCdTe substrate to form native defects responsible for type conversion,

resulting in a n/p area on the sample. At room temperature, the implantation process is usually carried out with the following parameters: angle of incidence: 7°, implantation dose: 5×10^{14} atom/cm^2, and ion energy: 50-140 KeV. The epilayer was then descummed using oxygen plasma for 30 seconds (pressure: $10^{-1} - 10^{-2}$ Torr, power: 100 W, O_2 flow rate: 200 mL/min), and then transferred into the coating unit's vacuum chamber. Following the achievement of the maximum chamber vacuum, which was 4.0×10^{-7} Torr, the tin and indium thin film (5 μm) were deposited on the HgCdTe by vacuum evaporation process. A quartz crystal was employed to determine the film thickness and film deposition rate. The 99.999% pure tin and indium small shots were placed in molybdenum boat for actual evaporation. After completing a photolithography lift-off procedure (in acetone) of the deposited sample, the final TLM structure was obtained.

The structural properties of thermally evaporated films were evaluated using a high-resolution XRD system (Model: Analytical B.V.-X'Pert PRO MRD; Material Research Diffractometer), with a CuKα radiation source (λ = 0.1542 nm) and Xe-proportional counter. Using the AFM (Model-5600LS, Agilent Technologies) instrument, the morphological statistics of thin films were obtained at the parameters (grazing angle = 3° and 2θ = 20-85°). The microstructural and compositional characteristics of tin and indium films grown on HgCdTe were determined using a FESEM (Model: Carl Zeiss SUPRA55VP) assisted by an Oxford Instrument X-MAX EDX Spectrometer (operating voltage: 5-20 kV). The surface quality of the film was measured by LSE MS STOKES Ellipsometer Model-7109-C370 M from GAERTNER. Monochromatic He-Ne laser beam was used at measuring angle of 70 degree and wavelength used 6328 Å. The thicknesses of films were determined by KLA alpha step D500 surface profiler. The Sn/HgCdTe and In/HgCdTe-based TLM structure's current–voltage (I–V) and resistance measurements were carried out using a cryoprober (Model: Keithley Parametric analyzer, Model-4200, with integrated 4200-CUV card). At 80 K, the I–V properties of the TLM structure based on HgCdTe were observed.

3. Results and Discussion

3.1 Crystallinity Measurement

Tin (Sn) as well as indium (In) diffraction pattern of both Sn and In indicated good agreement with typical Bragg's peak of Sn and In with a tetragonal bcc phase as shown in Fig. 1(a) and (b), respectively. The good crystalline properties of the Sn and In films were interpreted by the fine intensity and close line-broadness of the analyzed XRD peaks. The XRD intensive peaks obtained related to tetragonal bcc phase (200) preferential plane at 2θ = 30.66° and (101) preferential plane at 32.94° of β-Sn and In films, respectively. The β-Sn film's other peaks obtained were at 2θ = 43.85°, 44.95°, 55.39°, 63.84°, 64.62°, 72.45°, 73.17°, and 79.55° of (220), (211), (301), (400), (321), (420), (411), and (312) planes, respectively (JCPDS no. 89-7898) (Ref 33, 34). The indium film's low intense peaks were at 2θ = 36.30°, 39.16°, 54.47°, 56.58°, 63.20°, 67.05°, 69.12°, and 77.15° of (002), (110),(112), (200), (103), (211), (202), and (004) planes, respectively (JCPDS no. 85-1409) (Ref 35-40, 42). The average crystallite sizes (D) of Sn and In films were calculated using the Debye Scherrer's equation, ($D = (0.94\lambda)/(\beta Cos\theta)$) (Ref 41, 42, 48), where θ is the angle of diffraction and β is full-width half maximum (FWHM). The lattice misfit due to film deposition conditions generates the strain in the film. The strain (ε) in the Sn and In was calculated by using the equation (ε = β/4Tanθ) (Ref 48). The dislocation density (δ) of the Sn and In was determined by employing the equation ($\delta = 1/D^2$)

The β-Sn and In film's crystalline structural parameters for (200) and (101) plane calculated from x-ray diffraction results are given in Table 1. From Table 1, the obtained values of crystallite size, microstrain and dislocation density were comparable with the earlier reported results (Ref 35, 37, 39, 43-47). The x-ray results revealed that the grain sizes 33.14 nm and 33.27 nm of Sn and In films, respectively. The microstrain and dislocation density of In film were lower than the Sn film which indicated that In film was good crystalline structure.

3.2 Surface Morphology and Compositional Analysis

The SEM micrographs of the Sn as well as In are shown in Fig. 2(a) and (b) and 3(a) and (b), respectively. Visually both Sn and In films were having white shiny surface with no pin holes or cracks. The SEM images of Sn film indicated the rough film with cloudy surfaces. The Sn films showed nonuniform particle sizes ranging from few nanometers to micron size distributed all over the entire film (Ref 49). The SEM images of In film indicated the continuous and uniform film with some uneven particle sizes of 1-2 μm (Ref 40, 50).

The EDS analysis was done to verify the wholesomeness of tin and indium in the layers. The EDS spectrum of Sn and In films is shown in Fig. 2(c) and 3(c), respectively. The obtained experimental values indicated that the atomic % 51.50 and 94.80 of tin and indium elements, respectively (Ref 40, 51). The small amount of C element was present in Sn film may be because of HgCdTe film exposed to atmosphere during mounting on the sample holder or during the handling of HgCdTe sample. The small amount of K was present in In film may be because of contagionity of substrate at the time of handling of HgCdTe. Broad region around the dominated peaks in EDS spectrum are due to background, surface plasmon, and interband transition. Interband transition plasmon and electrons that are scattered are a result of interactions that cause the sample material's electrons to oscillate plasmonically or to transition between distinct energy levels. Due to its plasmon resonance, gold nanorods in particular and gold nanostructures in general have been used in numerous applications, including biosensors. There are two surface plasmon resonance bands in gold nanorods and one surface plasmon resonance band in gold nanoparticles. The plasmon oscillations on the (PVA / NiPc) sample surface are the reason why the crystallites are not spherical in shape (Ref 52, 53). Vamsi Borra et al. (Ref 51) grown rough Sn film at 0.05 nm/s− 1.6 nm/s deposition rates and challenging smooth Sn film 2.6 nm/s− 2.8 nm/s deposition rate on Si and soda lime glass by thermal evaporation process. They did the EDS analysis of the smooth Sn film and found that the film was pure Sn only or no other elements were present which very well agreed with our study .

3.3 Surface Topography Analysis

The Sn and In film's topographies were analyzed with silicon tip in tapping mode. Figure 4(a), (c), (e) and (b), (d), (f)

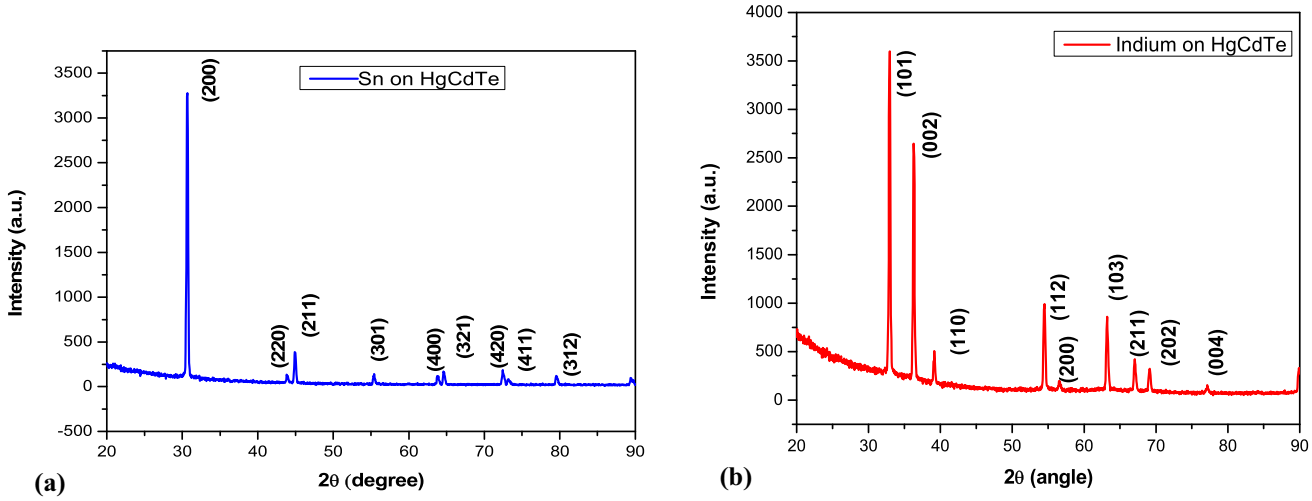


Fig. 1 XRD pattern (a) tin film (b) indium film

Table 1 Crystalline structural parameters of Sn and In for (200) and (101) preferred orientation planes

	Sn	In
Crystalline structural parameters	Calculated values	Calculated values
Lattice plane	(200)	(101)
2θ, degree	30.66	32.94
FWHM β, radian	0.004502	0.004537
Average crystallite size D, nm	33.14	33.27
Strain ε ($\times 10^{-3}$)	4.10	3.83
Dislocation density δ ($\times 10^{10}$ lines/cm^2)	9.10	9.03

shows the AFM images of Sn and In films, respectively. The scan size was 5 μm x 5 μm for Sn and In films. The root mean square (rms) roughness, average roughness, and other roughness values for both the films are given in Table 2. The AFM images indicated that rms roughness values were 44 nm and 20 nm of Sn and In, respectively. The Sn and In films roughness values were comparable with the previous study (Ref 42, 54, 55), but the Sn film surface roughness was double than the In film surface roughness.

The AFM line diagrams of Sn and In films scanned surface are shown in Fig. 4(a) and (b), respectively. The line profile of the films given information about the increase in the grain size and roughness of the film surfaces**.** R_z is the difference between the tallest peak and the deepest valley in the surface. R_{sk} is used to find the asymmetry of the roughness profile about the mean line. R_{ku} (roughness kurtosis) is used to find the peakedness of the profile about the mean line.

3.4 Calculation of Adhesion Properties

The AFM force microscopy in contact mode was utilized to determine the adhesion force (pull-off force) and adhesion work of tin as well as indium films. From force-distance curve (f–d curve), obtained from AFM force spectroscopy the adhesion force is determined from adhesion force among film and AFM tip. The f–d curves of Sn and In films on n-HgCdTe substrate are shown in Fig. 5 and 6, respectively. DMT and JKR theories are used to calculate the thermodynamic work of adhesion assuming that the forces of affinity at the adjoin region are similar to those outside region. Corina Birleanu et al. (Ref 30) determined the pull-off force and work of adhesion of Au thin film of different thickness and found that the adhesion properties vary with film thickness. For adhesion properties measurement in contact mode, the tip used was Si with Au reflective side, tip curvature radius 10 nm, tip height 14-16 μm, and resonant frequency 4-17 kHz. The maximum force, adhesion force, and work of adhesion of Sn and In films are given in Table 3.

From Table 3, we observed that the pull-off force and adhesion work of Sn film were almost double the In film though the density (7.31 g/cm ^{3}) of bulk tin and indium were same. Thus, this difference was due to different metal particles arrangement on HgCdTe surface during film growth, leading to different compact or denser films. The indium film's adhesion properties were studied and found comparable with our study (Ref 59). The adhesion force and adhesion work obtained by Corina Birleanu et al. (Ref 30) of Au films were different from our results due to different metal films physical properties as our metals (Sn and In) are soft with low density and low melting point compared to Au.

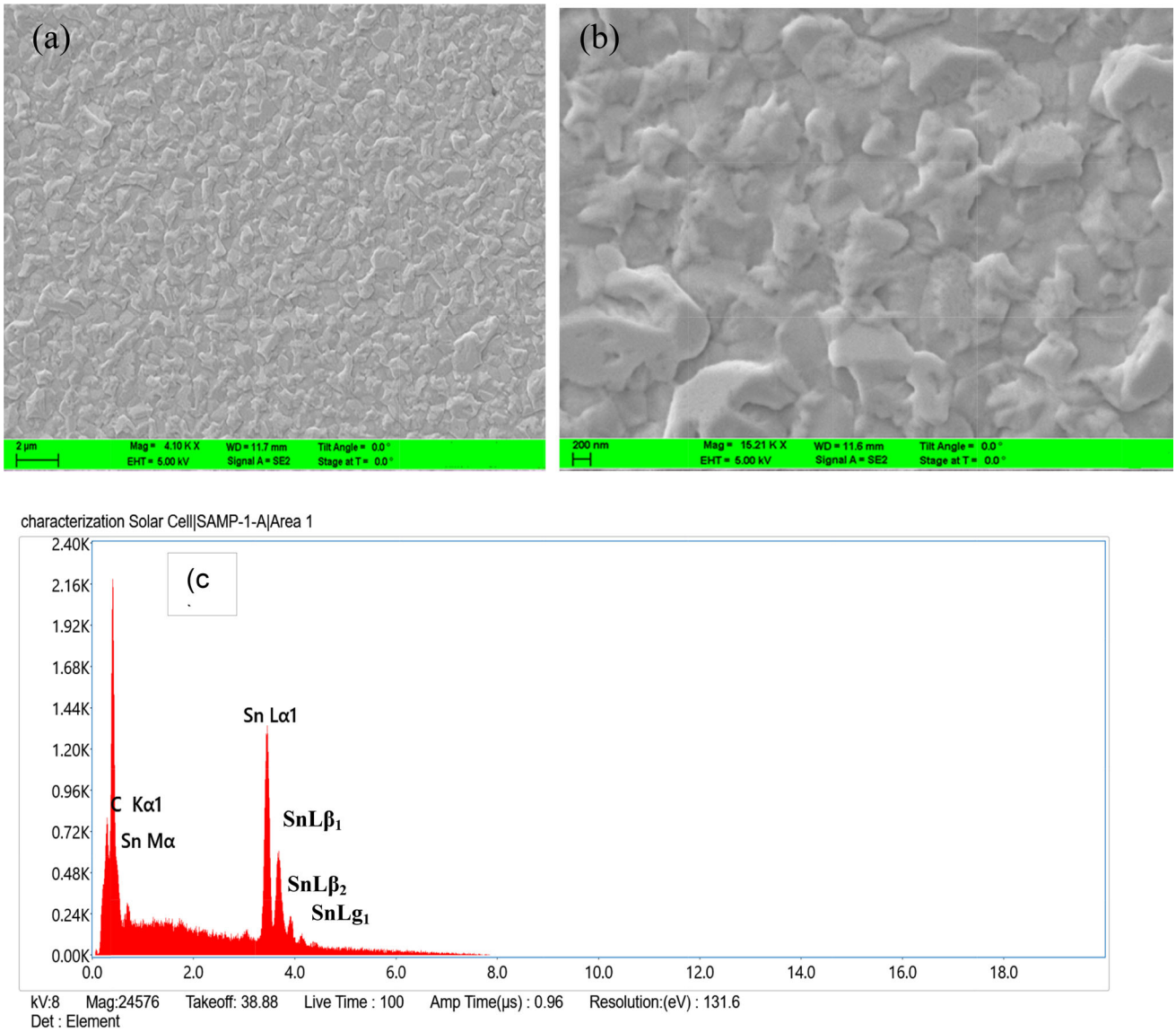


Fig. 2 SEM images of tin (a) SEM image at 4 KX magnification (b) Close view at 15KX magnification (c) EDS spectrum graph of tin

3.5 Ellipsometry Analysis

The ellipsometry analysis is identified by two factors psi (ψ) and delta (Δ). It measures the complex reflection ratio ρ:

$$\rho = \frac{r_p}{r_s} \mathrm{Tan}\psi \cdot e^{i\theta} \qquad \text{(Eq 1)}$$

where r_p and r_s are Fresnel reflection coefficients. Tanψ is the amplitude ratio after reflection and Δ is the phase difference. The ellipsometric (laser ellipsometry) measurement of Sn and In films was done to find out the quality of surface and examination of distribution by using eq. (1). Ellipsometric parameters psi (ψ) and delta (Δ) of both films are given in Table 4. Zang et al. (Ref 60) studied the β-Sn film on Si by ellipsometry at different temperatures. They studied the optical properties of melted and film and found that the dielectric parameter was different to two phases. Zang et al. (Ref 61) deposited In nanoparticle film on Si (100) and melting of nanoparticle was studied by ellipsometry. In particle of size larger than 44.9 nm, melting point was near to bulk In. For particle size less than 42.4 nm, no obvious melting point was found. Thus, our In film ellipsometry parameters were comparable with the above study.

3.6 Tin and Indium Film's Adhesion Test

The Scotch tape test of Sn and In films confirmed the good strength of adhesion as no part of the film removed with scotch tape. The Sn and In films were suitable with the metal contact requirement for infrared detectors.

3.7 Surface Profiler Thickness Measurement

A small portion on HgCdTe substrate covered such that no deposition takes place at that portion to make step height. Thus,

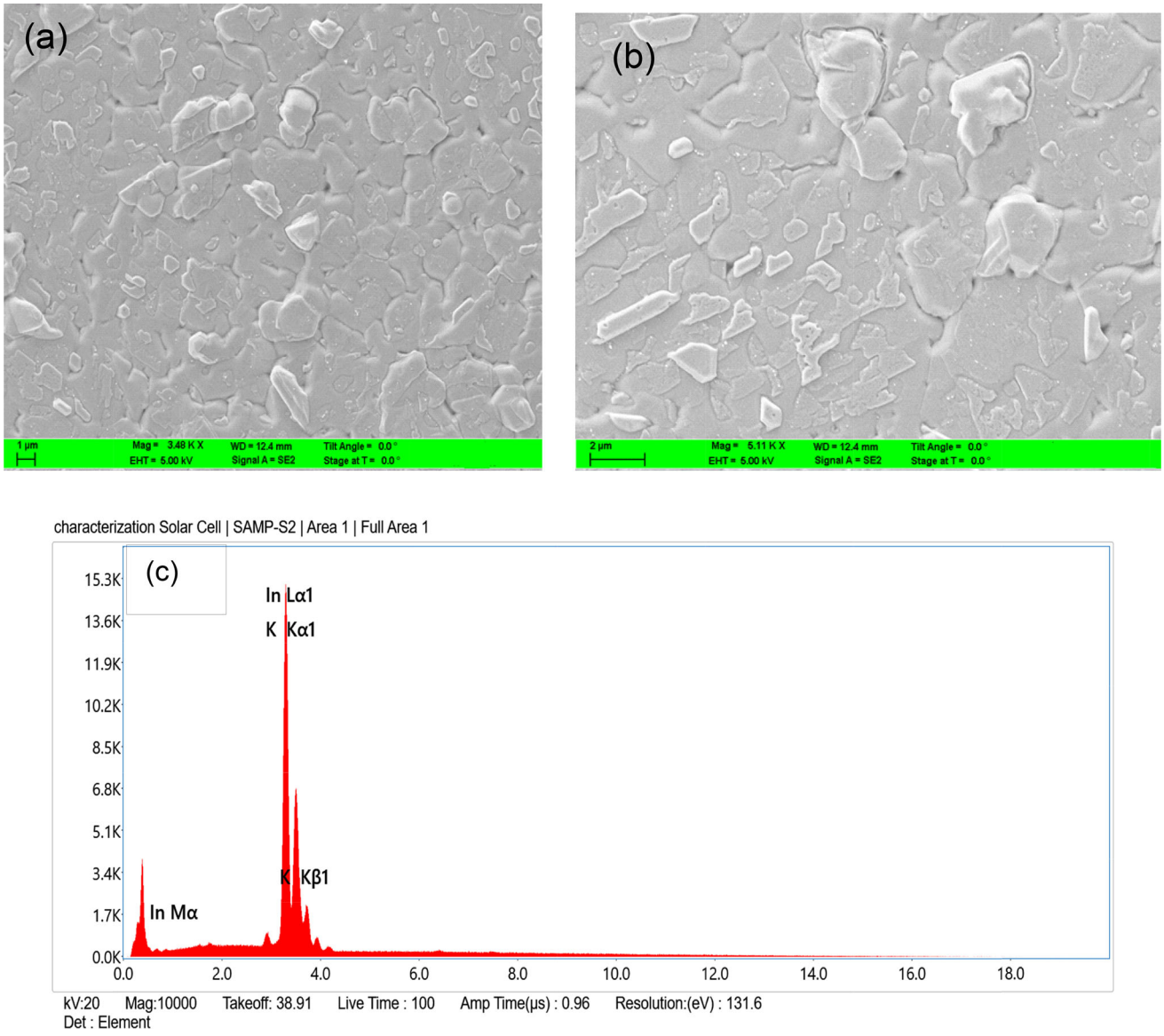


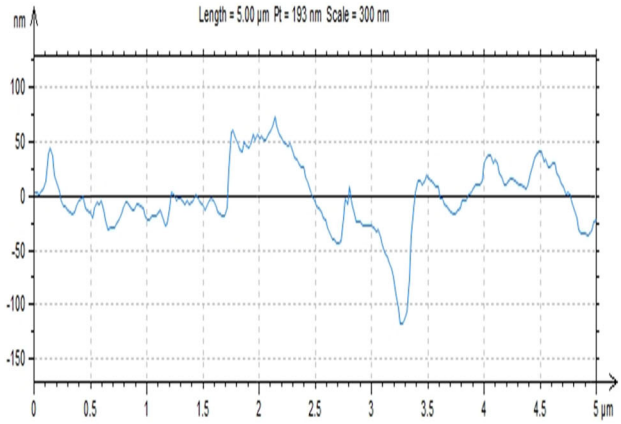
Fig. 3 SEM images of indium (a) SEM image at 3.48 KX magnification (b) SEM image at 5.11 KX magnification (c) EDS spectrum graph of indium

HgCdTe substrate has two regions one very small region with no deposition and large region with Sn and In films. Sn and In films thickness was confirmed by KLA Alpha step D500 surface profiler at several regions on the HgCdTe surface and found that average thickness was about 5 micrometer for both Sn and In as shown in Fig. 7.

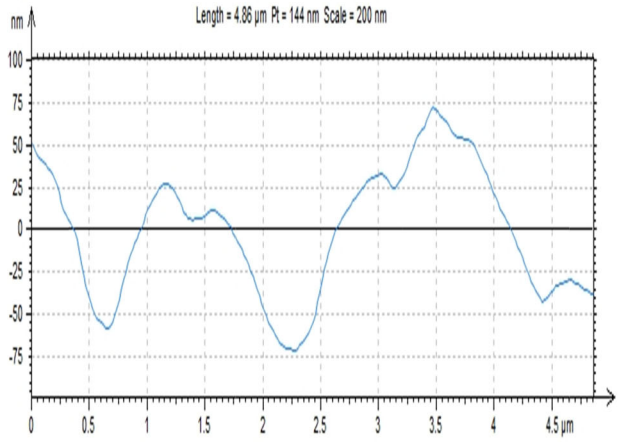
Thus, measurement of tin and indium film thickness by this method reconfirmed the thickness monitored by quartz crystal during the thermal evaporation process.

3.8 Electrical Properties of Tin and Indium Films

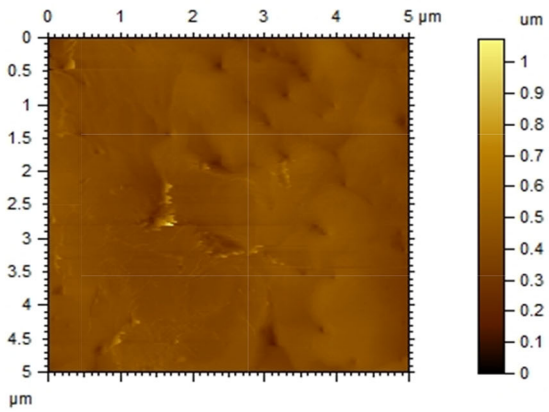
The HgCdTe samples D1 and D2 were used to grow the tin and indium thin films at an optimal deposition rate in order to create the TLM structure and determine the electrical characteristics of the grown tin and indium thin films. Measurements of the Sn/n/p-HgCdTe and In/n/p-HgCdTe-based TLM structure's I–V and resistances were then carried out. As illustrated in the schematic diagram (Fig. 8), the fabricated TLM structures have eight contact pads with the same area ($100 \times 200\ \mu m^2$; the contact pads' length, width, and length remain the same). The contact pads also have different interpad spacing, denoted as d1, d2, d3, d4, d5, d6, d7, and d8. d1—50 μm, d2—100 μm, d3—150 μm, d4—200 μm, d5—250 μm, d6—300 μm, d7—350 μm, and d8—400 μm are the values of the interpad spacing between consecutive contact pads. All adjacent contact pad pairs in the TLM structure had their resistances measured.



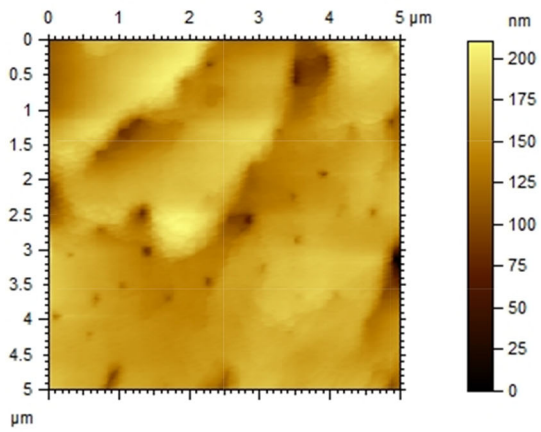
(a)Sn film line diagram



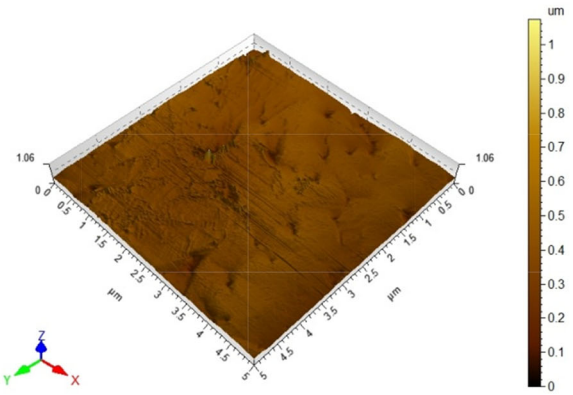
(b) In film line diagram



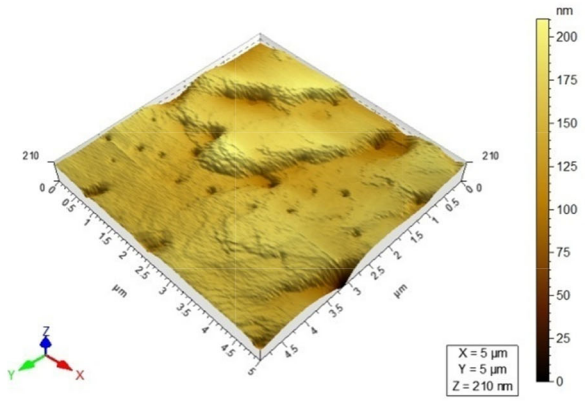
(c) Sn film 2-Dimensional image



(d) In film 2-Dimensional image



(e) Sn film 3-Dimensional image



(f) In film 3-Dimensional image

Fig. 4 AFM topography images of Sn and In films

Table 2 Roughness Table of Sn and In films

Film	R_a, nm	R_q, nm	R_z, nm	R_{sk}	R_{ku}
Sn	32	44	1060	0.19	10.62
In	15.2	20	210	0.94	5.78

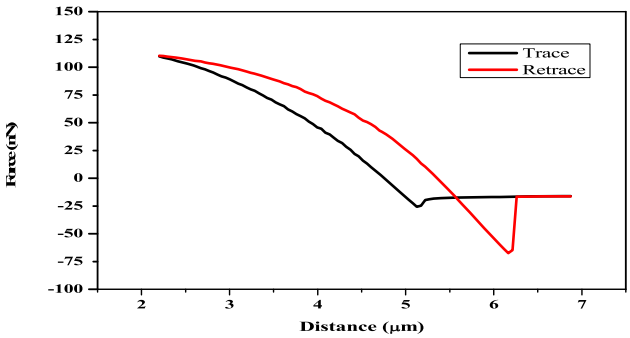


Fig. 5 AFM f–d curve of Sn on HgCdTe substrate

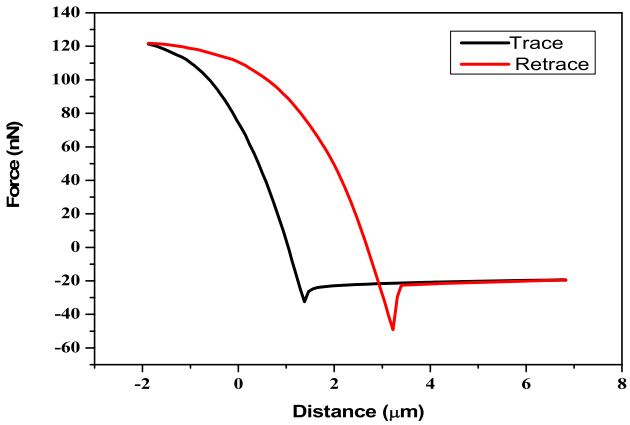


Fig. 6 AFM f–d curve of In on HgCdTe substrate

3.9 Transfer Length Method Calculation and Analysis

For every pair of adjacent contact pads, the total resistances R_T of the HgCdTe sample D1- and D2-based TLM structures (configuration: Sn/n/p-HgCdTe and In/n/p-HgCdTe) were measured at 80 K. Next, as seen in Fig. 9(a) and (b), a plot of total resistance (R_T) versus interspacing (d) was created of tin and indium, respectively. The electrical parameters (R_c, L_T, R_{sh}, and ρ_c) were estimated using Equations (7-11), and Table 5 lists the evaluated values of the electrical parameters of tin and indium. From Table 7, we found that the values of ρ_c for the tin-based

Table 3 Maximum force, pull-off force, and thermodynamic adhesion work based on DMT and JKR theories

Film	Maximum force, nN	Pull-off force, nN	Adhesion work, J/m^2	
…	…	…	γ_{DMT}	γ_{JKR}
In	121.73	26.95	0.42	0.57
Sn	110.25	50.73	0.80	1.07

Table 4 Ellipsometry parameters of Sn and In films

Film	Ψ, °	Δ, °
Sn	32.26	114.74
In	40.79	124.61

TLM structure (sample D1) and indium-based TLM structure (sample D2) were 2.0×10^{-2} Ω•cm^2 and 3.0×10^{-4} Ω•cm^2, respectively. Thus, the indium-based TLM structure consisted much lower specific contact resistance than the tin, which was favorable and comparable to the early reported values of low specific contact resistance of indium films ($10^{-3} - 10^{-5}$ Ω•cm^2), where these low contact resistances were achieved for Hg_{1-x} Cd_x Te (where $x = 0.28$ and 0.30) substrate using different combinations of metals, such as In/n-HgCdTe, In/HgTe/n-HgCdTe, Mo/In/n-HgCdTe, Ti/HgTe/n-HgCdTe, Ti/Au/PEDOT:PSS/n-HgCdTe, Au/HgTe/n-HgCdTe, and Ti/Pt/Au/n-HgCdTe (Ref 10, 11, 14-16, 18-20, 61). The specific contact resistance (ρ_c) of indium film has also demonstrated a strong correlation with the ρ_c of indium–metal contacts, which are employed in the progress of high-performance optoelectronic devices based on MoS_2 and CdTe-based IR detectors (Ref 65, 66).

The deposited film's surface roughness could be the cause of the low specific contact resistance observed at the In/n-HgCdTe interface as the roughness of indium film was lower than the tin film as investigated in AFM (Sn: R_q 44 nm and In: R_q 20 nm)

The more inward diffusion of In than Sn into the HgCdTe near-surface region, which increases the HgCdTe-doping concentration (because of the distribution of ionized donors, the doping density becomes relatively higher than the bare HgCdTe, which reduces the width of a potential barrier), may therefore be the second explanation for low specific contact resistance (Ref 10, 11, 64). This process is also responsible for the creation of metal contacts with low specific contact resistance. The presence of characteristics in the grown-in films, such as contaminants, nonuniform thickness, and impurities, which might make it difficult to generate the necessary

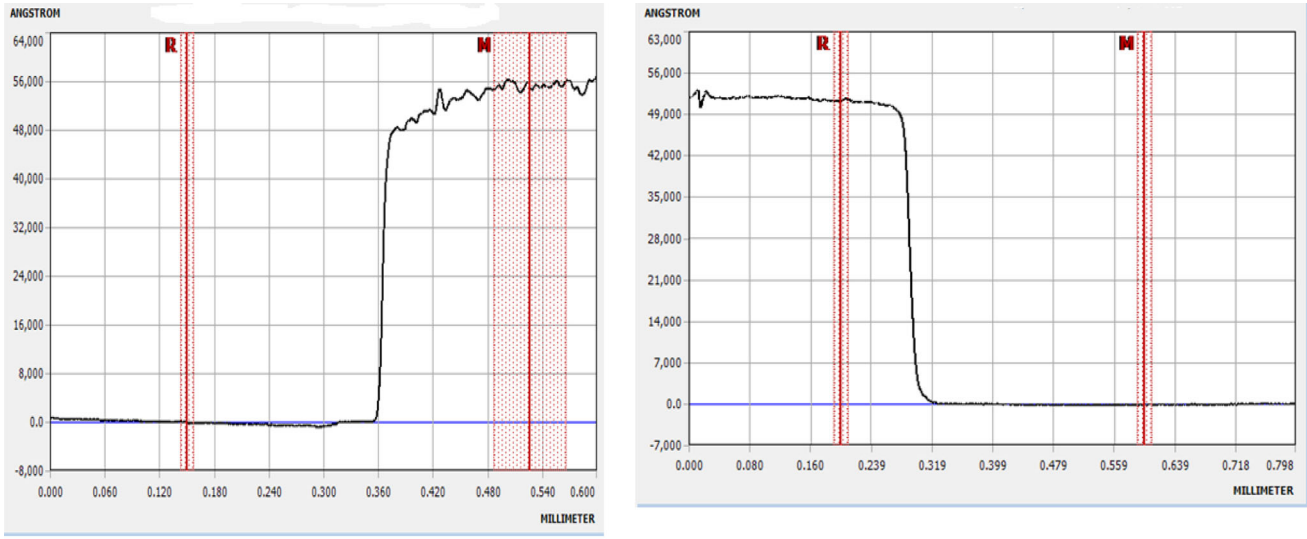


Fig. 7 Surface profiler thickness measurement graphs (a) Sn film thickness (step height) (b) In film thickness (step height)

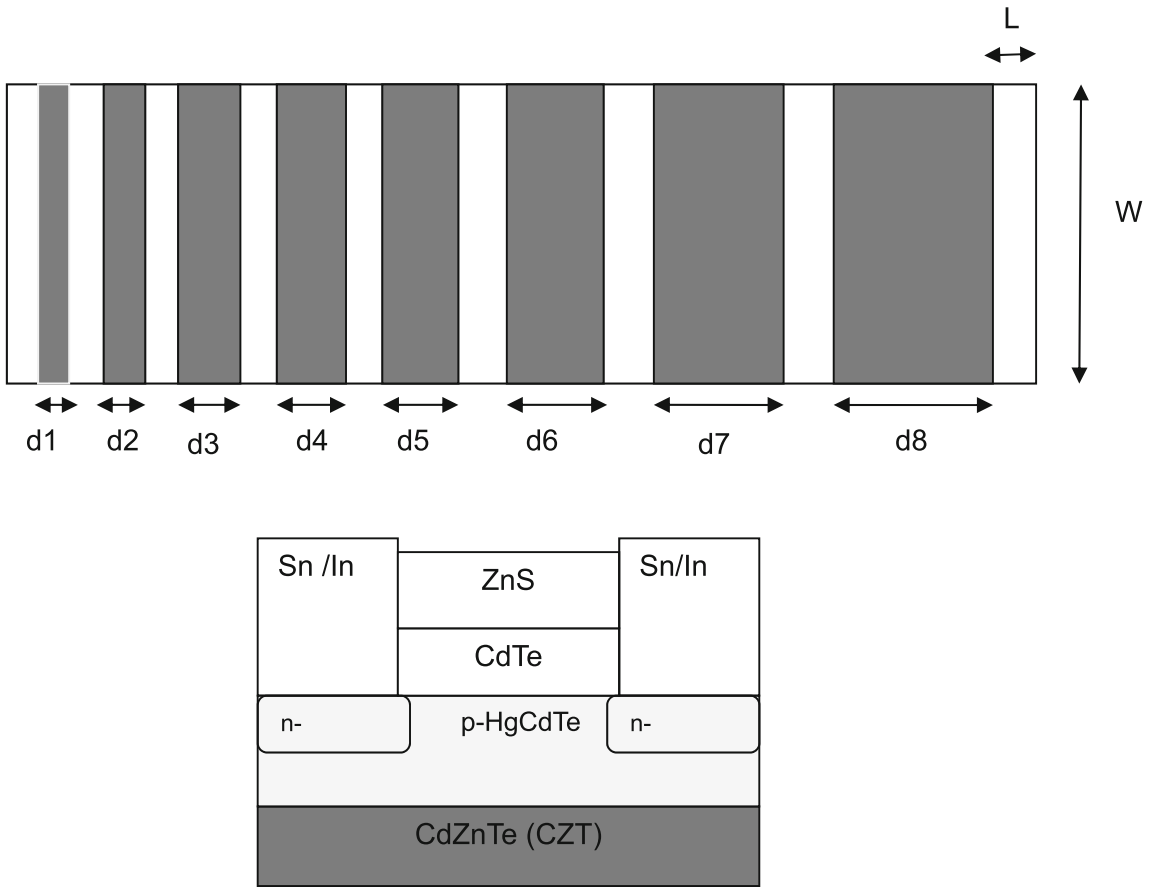


Fig. 8 Schematic representation of HgCdTe-based TLM test structure. (a) two TLM structures (configuration: Sn/n/p-HgCdTe and In/n/p-HgCdTe) of same dimensions are fabricated with eight pads each and each pad has same width and same length. The interpad spacings are 50, 100, 150, 200, 250, 300, 350, and 400 μm. (b) The cross section of a part of the test structure is shown in this figure

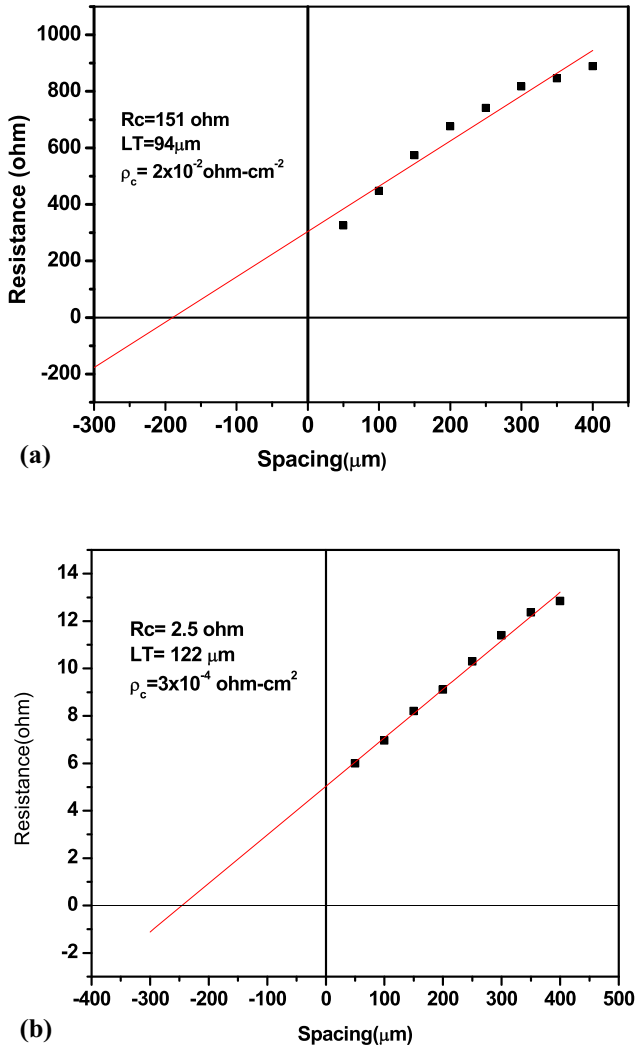


Fig. 9 (a) The total resistance vs. interpad spacing curve of tin/HgCdTe based on TLM structure. (b) The total resistance vs. interpad spacing curve of indium/HgCdTe based on TLM structure

Table 5 The estimated values of electrical variables for the HgCdTe-based TLM structure made-up (samples D1 and D2; configuration: Sn/n-HgCdTe and In/n-HgCdTe interface)

Sample No.	R_c, Ω	L_T, μm	ρ_c, Ωcm^2
D1	151	94	2×10^{-2}
D2	2.5	122	3×10^{-4}

electrical characteristics, may be the third element responsible for high contact resistance.

The characteristics of the films that were obtained from energy-dispersive x-ray, atomic force microscope, and scanning electron microscope examination reveal that the films are uniform, pure, and dense. The indium film also has little roughness and the appropriate small crystallite size as well as grain size. As a result, there would not be as many obstacles in the way of the charges as they pass over the In/n-HgCdTe

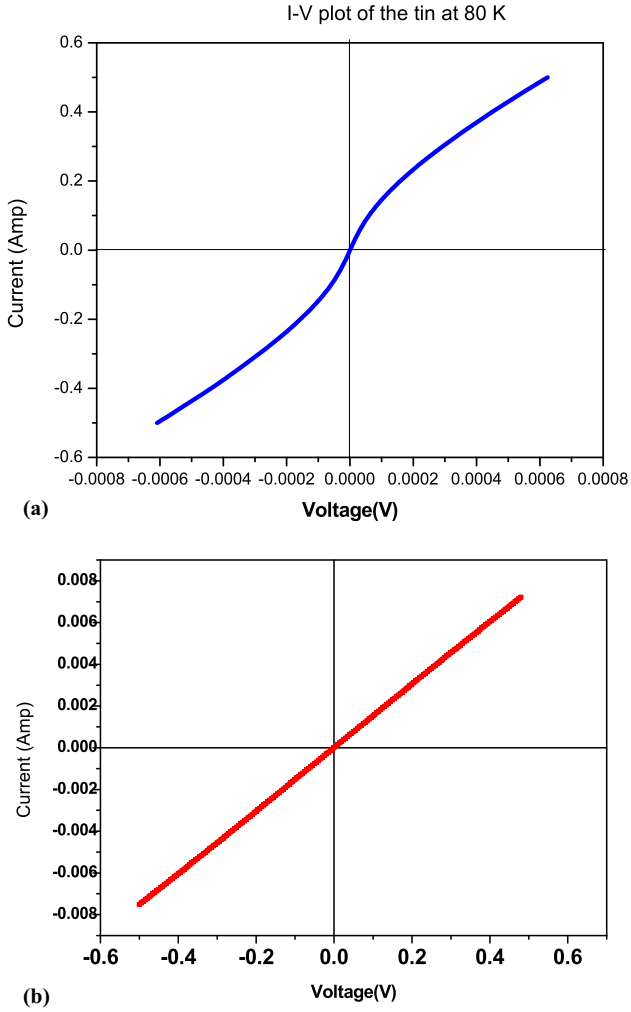


Fig. 10 (a) I–V characteristics of tin TLM structure. (b) I–V characteristics of indium TLM structure

interface, resulting in a low grown-in contact specific contact resistance (Ref 10, 11, 29).

To understand the behavior of the tin and indium metals contact as well as the current transport mechanism of the Sn/n-HgCdTe and In/n-HgCdTe interface, I–V measurements were carried out. The I–V characteristics of the constructed TLM structures (sample D1 and sample D2), measured at 80 K, are shown in Fig. 10(a) and (b) and indicated that the grown indium contacts on n-type HgCdTe exhibit ohmic behavior but the tin contacts on n-type HgCdTe exhibit nearly ohmic behavior. Hence, the linear I–V characteristics and ohmic behavior of IR photodetectors utilizing metals/n-HgCdTe interfaces are comparable to the dependable I–V features of thermally produced indium films (Ref 10, 11, 18-20).

The thermionic field emission of HgCdTe ($x = 0.30$) substrate at 77 K temperature realizes the current flow in the In/n-HgCdTe contacts due to two types of current transport mechanisms that operate simultaneously: quantum field emission tunneling and thermionic emission. This occurs for doping concentrations ranging from 3.5×10^{16} to 1.6×10^{18} cm^{-3} (Ref 10). Within the two aforementioned regimes, the specific contact resistance (R_c) has a clearly defined dependence. In the thermionic emission regime, the R_c is independent of the

doping density (concentration). It is determined by the quantum field tunneling's reciprocal exponential carrier concentration root. To enable the e^- to overcome the potential barrier with a high tunneling probability, a small depletion layer (thin potential barrier) is created by high doping concentration. Given that the HgCdTe-based TLM structure's parameters (composition $x = 0.29$, doping concentration: 1×10^{16} cm^{-3}, and I–V measurement temperature: 80 K) match those of the structures that have been reported (Ref 16, 18, 22, 67), it is likely that significant tunneling and negligible thermionic emission are the main causes of the current transport in the fabricated In/n-HgCdTe interface. Therefore, the properties of thermally generated indium films were suitable for generating an ohmic contact on the n-HgCdTe surface and may be utilized in the development of infrared photodetectors based on HgCdTe.

4. Conclusions

Metals (tin and indium) films were deposited on the n-HgCdTe substrate by the thermal evaporation method. A comparative study of tin(Sn) and indium(In) done based on their crystalline parameters, morphology, composition, topography, adhesion, surface quality, and electrical characteristics. The XRD examination confirmed that the films were polycrystalline tetragonal bcc phase along (200) and (101) planes orientation of tin and indium, respectively. The XRD peak intensity of indium film was very large than the tin film's XRD peak intensity for the corresponding preferential planes which indicated that the In film's crystallinity was much better than the Sn. From XRD results, the calculated 33.14 nm crystallite size of β-Sn film was of slight small size than the 33.27 nm crystallite size of In film but the microstrain and dislocation density values of indium were lower than β-Sn film. Thus, overall crystalline structure of indium film was better than the tin film. SEM images of β-Sn film shown crack free nonuniform particle size distribution, while indium film SEM images indicated a continuous and uniform film over entire surface with a few large micron size particles. So, the morphology of the indium film was better than the tin film. EDS analysis revealed that deposited films were pure of tin and indium elements. AFM scanned images indicated that tin film was rougher than the indium film. The 44 nm rms value and 32 nm average value were of β-Sn film while 20 nm rms value and 15.2 nm average value were of In film. This higher roughness of β-Sn film was due to nonuniform cluster growth and nonuniform distribution of Sn clusters. Adhesion force and thermodynamic adhesion work of tin film were nearly double than the indium film. The scotch tape test confirmed the good adhesion or bonding of Sn and In film molecules with HgCdTe surface. Surface profiler ensured the tin and indium 5 μm film thicknesses and validated the thickness monitored by the quartz crystal during evaporation process. Using TLM and I–V measurements for various contact geometries, the Sn/n-HgCdTe and In/n-HgCdTe interfaces were electrically examined. The In/n-HgCdTe interface indicated linear I–V characteristics (ohmic behavior), but the Sn/n-HgCdTe interface indicated nonlinear I–V characteristics (nearly ohmic behavior). In this instance, the thermionic field emission mechanism controls current transport. The In/n-HgCdTe interface consisted low specific contact resistance, $\rho_c = 3 \times 10^{-4}$ Ωcm^2 than the Sn/n-HgCdTe interface specific contact resistance, ρ_c 2×10^{-2} Ωcm^2.

Compared to tin film, the indium film had lower specific contact resistance and surface adhesion force—two vital characteristics needed for metal contact in infrared device fabrication. Thus, indium film's overall structural, morphological, compositional, topographical, adhesion, ellipsometric, and electrical properties were better than tin. Indium was found to be suitable as a good ohmic n-contact in optoelectronic, micro/nanoscale electronic devices and HgCdTe-based infrared detectors.

Acknowledgments

The author acknowledges the Director SSPL for aiding in doing this work. The author also acknowledges the CEO STARC for supporting this work.

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

Appendix

(a) Adhesion Characteristics

The potential or effort necessary to divide two superficial make contact with is called adhesion force (pull-off force). The interfacial forces and van der Waals force are the sources of adhesion force. Johnson, Kendall, and Roberts (JKR) revealed a novel theory explaining adhesion between two elastic bodies (Ref 56). They arrived at the following eq. (2) to define the contact radius between a sphere and a plane when adhesion was in effect:

$$a = \left(\frac{R}{K}\left(P + 3\gamma\pi R + \sqrt{6\gamma\pi RP + (3\gamma\pi R)^2}\right)\right)^{1/3} \quad \text{(Eq 2)}$$

where P denoted the applied load, γ denoted the thermodynamic adhesion work (or Dupré energy of adhesion), and K was a parameter that changed according on the Poisson's ratio and the elastic modulus that matched each facade. Among the characteristics of this theory were a nonzero contact region with no load, an infinite tensile stress at the perimeter of the contact zone, and the lowest stable normal load that still revealed a nonzero contact area. The lowest load that was established by the following was the pull-off force, also referred to as the critical load:

$$F_{JKR} = \frac{-3}{2}\pi\gamma R \quad \text{(Eq 3)}$$

A short time later, Derjaguin, Muller, and Toporov (DMT) added adhesion to an alternative expression (Ref 57). They postulated that, in accordance with the Hertz hypothesis, adhesion would result in an overall increase in load while the deformed contact profile would remain unchanged. This resulted in attractive interactions between the sphere and the plane that acted as a dead weight at all distances, as shown below:

$$a = \left(\frac{R}{K}(P + 2\pi\gamma R)\right)^{1/3} \quad \text{(Eq 4)}$$

and now the pull-off force or critical load was:

$$F_{\mathrm{DMT}} = -2\pi\gamma R \quad \text{(Eq 5)}$$

where R was radius of tip in nm (12 nm) and γ was thermodynamic adhesion work in J/m^2

In the Derjaguin–Muller–Toporov (DMT) theory, the contact area goes to zero at the pull-off force, and there was no singularity in the contact stresses. The theory assumed that the attraction forces at the interface region were equal to those outside of it in order to derive the thermodynamic adhesion work or surface energy. David S. Grierson et al. (Ref 58) explained the theories about the adhesion and friction properties between contact surfaces. They explained about the Derjaguin, Muller, and Toporov (DMT) and Johnson, Kendall, and Roberts (JKR) theories in continuum contact mechanics to determine the adhesion and frictional characteristics.

(b) Electrical Characteristics Using Transfer Length Method (TLM)

To determine the sheet resistance (R_{sh}) and transfer length (L_{T}), a resistance versus interpad separation plot was created. The other parameters, such as contact resistance (R_{c}) and specific contact resistance (ρ_{c}), were assessed using the resulting values of R_{sh} and L_{T}. I–V measurements of TLM structures were also carried out to look at the contacts' behavior. The measured resistance (R_{T}) for a pair of TLM structure contact pads (length L, width W, and interpad spacing d) may be found using eq. (6) (Ref 14, 20, 22, 62, 63).

$$R_{\mathrm{T}} = R_{\mathrm{semi}} + 2R_{\mathrm{c}} = R_{\mathrm{sh}}\left(\frac{d}{w}\right) + 2(R_{\mathrm{sh}}L_{\mathrm{T}})/W \quad \text{(Eq 6)}$$

Transfer length, sheet resistance, semiconductor resistivity, and contact resistance are represented, respectively, by the parameters L_{T}, R_{sh}, R_{semi}, and R_{c}. Equation (7) can be used to calculate the transfer length (Ref 14, 20, 22, 62, 63).

$$\mathrm{LT} = \sqrt{(\rho_c/R_{\mathrm{sh}})} \quad \text{(Eq 7)}$$

Transfer length, sheet resistance, semiconductor resistivity, and contact resistance are represented, respectively, by the parameters L_{T}, R_{sh}, R_{semi}, and R_{c}. Equation (8) can be used to calculate the transfer length (Ref 14, 20, 22, 62, 63).

$$\rho_{\mathrm{c}} = L_{\mathrm{T}}R_{\mathrm{sh}} \Rightarrow R_{\mathrm{c}} = \rho_{\mathrm{c}}/(L_{\mathrm{T}}W) = (R_{\mathrm{sh}}L_{\mathrm{T}})/W \quad \text{(Eq 8)}$$

Equations (6), (7), and (8) were used to the ohmic planar contact in order to evaluate R_{c} and ρ_{c} (Ref 14, 20, 22, 59, 60). Given that the definition of R_{c} for an ohmic planar contact is:

$$R_{\mathrm{c}} = ((R_{\mathrm{sh}}L_{\mathrm{T}})/W)\coth(L/L_{\mathrm{T}}) \approx (\rho_{\mathrm{c}}/(L_{\mathrm{T}}W))\coth(L/L_{\mathrm{T}}) \quad \text{(Eq 9)}$$

The value of R_{c}, which is expressed in the simplified form as follows, is altered by the limiting cases:

No considerable current penetration in the metal contact, or short contact limit (L«L_{T}) $\rightarrow$

$$\coth(L/L_{\mathrm{T}}) \approx L/L_{\mathrm{T}} \text{then } R_c = \rho_{\mathrm{c}}/(LW) \Rightarrow \rho_{\mathrm{c}} = R_{\mathrm{c}}LW \quad \text{(Eq 10)}$$

Long contact significant current transfer into the metal contact (limit L»L_{T}) $\rightarrow$

$$\coth(L/L_{\mathrm{T}}) \approx 1 \text{ then } R_{\mathrm{c}} = \rho_{\mathrm{c}}/(L_{\mathrm{T}}W) \Rightarrow \rho_{\mathrm{c}} = R_{\mathrm{c}}L_{\mathrm{T}}W \quad \text{(Eq 11)}$$

References

1. A. Rogalski, Recent Progress in Infrared Detector Technologies, *Infrared Phys. Technol.*, 2011, **54**, p 136–154
2. S.M. Sze, *Physics of Semiconductor Devices*, Wiley, New York, 1969
3. C.T. Tan and H. Mohsani, Emerging Technologies for High Performance Infrared Detectors, *Nanophotonics*, 2018, **7**, p 169–197
4. P. Capper and J. Garland, *Mercury Cadmium Telluride Growth, Properties and Application*, John Wiley and Sons Ltd., U.K., 2011, p 340–347
5. V. Srivastav, R. Pal, and V. Venkataraman, Modeling of Room Temperature Current-Voltage Measurements on Homojunction HgCdTe Diodes Exhibiting Nonequilibrium Effects, *J. Appl. Phys.*, 2012, **111**, p 033112
6. D.K. Schroder, *Semiconductor Material and Device Characterization*, John Wiley and Sons, New Jersey, 2006, p 127–150
7. G. K. Reeves, H. B. Harrison, Obtaining the Specific Contact Resistance from Transmission Line Model Measurements, IEEE Electron Device Lett., 3, No, 5, 1982
8. G.D. Davis, W.A. Beck, M.K. Kelly, N. Tache, and G. Margaritondo, Deposition of the Unreactive Metal Au onto Sputtered and Cleaved Hg1-xCdxTe Surfaces, *J. Appl. Phys.*, 1986, **60**(9), p 3157–3161
9. G.D. Davis, Overlayer Interactions with (HgCd)Te, *J. Vac. Sci. Technol.*, 1988, **A6**(3), p 1939–1945
10. P.W. Leech and G.K. Reeves, Specific Contact Resistance of Indium Ohmic Contacts to n-type Hg1-x CdxTe, *J. Vac. Sci. Technol. A*, 1992, **10**, p 105
11. W.E. Spicer, Metal Contacts on Hg1− xCdxTe, *J. Vac. Sci. Technol. A*, 1990, **8**(2), p 1174–1177
12. M. Arthur, Turner, HgTe Contacts to p-HgCdTe, *J. Vac. Sci. Technol. B*, 1992, **10**, p 1534
13. A. Wall, A. Raisanen, S. Chang, P. Philip, A. Franciosi, and D.J. Peterman, Local Stoichiometry and Atomic Interdiffusion During Reactive Metal/Mercury-Cadmium-Telluride Junction Formation, *J. Vac. Sci. Technol.*, 1987, **A5**, p 3193
14. D. Liu, C. Lin, S. Zhou, and Hu. Xiaoning, Ohmic Contact of Au/Mo on $Hg_{1-x}Cd_xTe$, *J. Electron. Mater.*, 2016, **45**(6), p 2802–2807
15. W. Patrick, The Specific Contact Resistance of Ohmic Contacts to $HgTe/Hg_{1-x}Cd_xTe$ Heterostructures, *J. Appl. Phys.*, 1990, **68**, p 907
16. P.W. Leech and M.H. Kibel, Properties of Schottky Diodes on n-Type Hg_{1-x} Cd_xTe, *J. Vac. Sci. Technol.*, 1991, **B9**, p 1770
17. A.R. Momen-Baghdadabad, A. Bahari, and F.M. Aghamir, Synthesis of Multi-Phase Steel Thin Films by a Low Energy Plasma Focus Device, *Mater. Chem. Phys.*, 2024, **319**, 129324
18. P.W. Leech, G.K. Reeves, and M.H. Kibel, Ohmic Contacts on N-Type $Hg_{0.4}Cd_{0.6}Te$, *MRS Online Proc. Lib.*, 1990, **216**, p 155–160
19. H. Zimmermann, R.C. Keller, P. Meisen, H.J. Richter, and M. Seelamnn-Eggebert, Interface Formation Between Deposited Sn and $Hg_{0.8}Cd_{0.2}Te$, *J. Electron. Mater.*, 1996, **25**(8), p 1293–1299
20. F. Sizov, Z. Tsybrii, M. Apatska, N. Dmytruk, V. Slipokurov, S. Bunchuk, Y. Bezsmolnyy, V. Popovych, M. Wiertel, and N. Mikhailov, Ohmic metal/Hg1-xCdxTe (x ≈ 03) Contacts, *Semicond. Sci. Technol.*, 2020, **35**, p 125030
21. Y.X. Wang and P.H. Holloway, Indium Ohmic Contacts to n-ZnSe, *Vacuum*, 1992, **43**(11), p 1149–1151
22. V. Srivastav, R. Pal, B.L. Sharma, V. Mittal, V. Gopal, and H.P. Vyas, Electrical Properties of Titanium-HgCdTe Contacts, *J. Electron. Mater.*, 2005, **34**(3), p 225–231
23. P.Y. Emelie, E. Cagin, J. Siddiqui, J.D. Phillips, C. Fulk, J. Garland, and S. Sivananthan, Electrical Characteristics of PEDOT: PSS Organic Contacts to HgCdTe, *J. Electron. Mater.*, 2007, **36**, p 841–845
24. M. Zandian, J.M. Arias, R. Zucca, R.V. Gil, and S.H. Shin, HgCdTe Double Heterostructure Injection Laser Grown by Molecular Beam Epitaxy, *Appl. Phys. Lett.*, 1991, **59**(9), p 1022–1024

25. J.M. KiMan Kang, J.K. Jo, H. Kim, Y.S. Kim, C. Sone, and Y. Park, In/ITO Ohmic Contacts to Ga-face and N-face n-GaN for InGaN-based Light-emitting Diodes, *J. Korean Phys. Soc.*, 2009, **55**(1), p 318–321
26. S.Y. Moon, J.H. Son, K.J. Choi, J.L. Lee, and H.W. Jang, Indium as an Efficient Ohmic Contact to N-face n-GaN of GaN-based Vertical Light Emitting Diodes, *Appl. Phys. Lett.*, 2011, **99**, 202106
27. S. Palimar, K.V. Bangera, and G.K. Shivakumar, Study of the Doping of Thermally Evaporated Zinc Oxide Thin Films with Indium and Indium Oxide, *Appl. Nanosci.*, 2013, **3**(6), p 549–553
28. B.K. Kim, T.H. Kim, D.H. Choi, H. Kim, K. Watanabe, T. Taniguchi, H. Rho, J.J. Kim, Y.H. Kim, and M.H. Bae, Origins of Genuine Ohmic van der Waals Contact Between Indium and MoS_2, NPJ 2D, *Mater Appl.*, 2021, **5**, p 9
29. Z. Tsybrii, Y. Bezsmolnyy, K. Svezhentsova, M. Vuichyk, I. Lysiuk, M. Apatska, M. Smolii, N. Dmytruk, S. Bunchuk, K. Andreeva, and F. Sizov, HgCdTe/CdZnTe LPE Epitaxial Layers: From Material Growth to Applications in Devices, *J. Cryst. Growth*, 2019, **529**, 125295
30. C. Birleanu, M. Pustan, V. Meire, R. Muller, R. Voicu, A. Baracu, and S. Craciun, Temperature Effect on the Mechanical Properties of Gold Nano Films with Different Thickness, *IOP Conf. Ser.: Mater. Sci. Eng.*, 2016, **147**, 012021
31. M. Casa, M. Sarno, R. Liguori et al., Conductive Adhesive Based on Mussel-Inspired Graphene Decoration with Silver Nanoparticles, *J. Nanosci. Nanotechnol.*, 2018, **18**(2), p 1176–1185
32. M. Kolahdoozan, M. Hamedi, and M. Nikkhah-Bahrami, A Novel Model for the Effect of Geometric Properties of Micro/Nanoscale Asperities on Surface Adhesion, *Int. J. Adhes. Adhes.*, 2014, **48**, p 280–287
33. J. Kaur, R.K. Jain, A. Khanna, and A.K. Chawla, Effect of Thickness on the Wettability and Electrical Properties of Sn Thin Films, *J. Vac. Sci. Technol. B*, 2021, **39**, 032205
34. S. Hishita, Z. Stryhal, I. Sakaguchi, N. Ohashi, N. Saito, and H. Hanoda, Sn Film Deposition on Silica Glass Substrates, *Thin Solid Films*, 2004, **464–465**, p 146–149
35. R.S. Parra, P.J. George, G.G. Sánchez, A.J. González, L. Baños, and P.K. Nair, Optical and Electrical Properties of Pbs + In Thin Films Subjected to Thermal Processing, *J. Phys. Chem. Solids*, 2000, **61**(5), p 659–668
36. P.N. Bartlett, D. Cook, C.H. Kees de Groot, A.L. Hector, R. Huang, A. Jolleys, G.P. Kissling, W. Levason, S.J. Pearce, and G. Reid, Non-aqueous Electrodeposition of p-block Metals and Metalloids from Halometallate Salts, *RSC Adv.*, 2013, **3**(36), p 15645
37. R. Das and R.K. Soni, Synthesis and surface-enhanced Raman scattering of Indium Nanotriangles and Nanowires, *RSC Adv.*, 2017, **7**(51), p 32255–32263
38. S.R. Kumar, B. Prajapati, S.K. Tiwari, and V.K. Tiwari, Growth and Characterization of Copper, Indium and Copper-Indium Alloy Films Non-Aqueous Method of Electrodeposition, *Indian J. Pure Appl. Phys.*, 2008, **46**, p 198–203
39. K. Ulutaş, D. Değer, N. Kalkan, S. Yildirim, Y.G. Çelebi, Y. Iskarlatos, M.L. Öveçoğlu, and A. Genç, Structural Properties of In-In_2O_3 Composite Films, *IOP Conf. Ser. Mater. Sci. Eng.*, 2010, **15**, p 012095
40. P. Lobaccaro, A. Raygani, A. Oriani, N. Miani, A. Piotto, R. Kapadia, M. Zheng, Z. Yu, L. Magagnin, D.C. Chrzan, R. Maboudian, and A. Javey, Electrodeposition of High-Purity Indium Thin Films and Its Application to Indium Phosphide Solar Cells, *J. Electrochem. Soc.*, 2014, **161**(14), p D794–D800
41. M. Prashant Kumar and A. Ghanashyam Krishna, A Comparative Study of laser- and Electric-Field-Induced Effects on the Crystallinity, Surface Morphology and Plasmon Resonance of indium and Gold Thin Films, *Phys. Status Solidi A*, 2010, **207**(4), p 947–954
42. J. Kaur, A. Khanna, and A.K. Chawla, Metallic to Semiconducting Transition and Hydrophobicity Properties of Indium Films, *Vacuum*, 2022, **203**, p 11281
43. C.Y. Tsay, H.C. Cheng, Y.T. Tung, W.H. Tuan, and C.K. Lin, Effect of Sn-Doped on Microstructural and Optical Properties of ZnO Thin Films Deposited by Sol-Gel Method, *Thin Solid Films*, 2008, **517**, p 1032–1036
44. R. Das and R. Kumar, Preparation of Nanocrystalline PbS Thin Films and Effect of Sn Doping and Annealing on Their Structural and Optical Properties, *Mater. Res. Bull.*, 2012, **47**, p 239–246
45. R.P. Reed, C.N. McCowan, R.P. Walsh, L.A. Delgado, and J.D. McColskey, Tensile Strength and Ductility of Indium, *Mater. Sci. Eng. A*, 1988, **102**, p 227–236
46. G. Lee, J.Y. Kim, A.S. Budiman, N. Tamura, M. Kunz, K. Chen, M.J. Burek, J.R. Greer, and T.Y. Tsui, Fabrication, Structure and Mechanical Properties of Indium Nanopillars, *Acta Mater.*, 2010, **58**(4), p 1361–1368
47. S.E. Al Garni and A.F. Qasrawi, Effect of Indium Nano-Sandwiching on the Structural and Optical Performance of ZnSe Films, *Results Phys.*, 2017, **7**, p 4168–4173
48. L. Soekhee, P. Juyun, O. Jin-W, and K. Yong-C, Effect of Compositional Ratio of Sn in SnZn Thin Films Morphological and Chemical Properties, *Phys. B: Phys. Condensed Matter*, 2020, **591**, 412257
49. F. Aleksandra, L. Yue, G. Iman, C.C. Peter, and C. Eric, Investigation of Tin (Sn) Film Using an Aerosol Jet Additive Manufacturing Deposition Process, *J. Electron. Mater.*, 2017 https://doi.org/10.1007/s11664-017-5524-7
50. J. Ivanco, T. Todar, A. Firsov, M. Brzhezinskaya, M. Sperling, W. Braun, and D.R.T. Zahn, Indium on a Copper Phthalocyanine Thin Film: Not a Reactive System, *Phys. Rev. B*, 2010, **81**, 115325
51. V. Borra, D.G. Georgiev, and C.R. Grice, Fabrication of Optically Smooth Sn Thin Films, *Thin Solid Films*, 2016, **616**, p 311–315
52. D. Shajari, A. Bahari, P. Gill, and M. Mohseni, Synthesis and Tuning of Gold Nanorods with Surface Plasmon Resonance, *Opt. Mater.*, 2017, **64**, p 376–383
53. Z.H. Ali, A. Bahari, and A.H. Alarajiy, Investigation the Effect of Concentration(x) in Polyvinyl Alcohol (PVA)x/ Nickel Phthalocyanine (NiPc)(1–x) nano Composites: Real & Imaginary Component of Permeability and Optical Conductivity, *Opt. Mater.*, 2024, **148**, p 114862
54. V. B. Neimash, P. Y. Shepelyavyi, A. S. Nekolenko, V.V. Strelchuk, and I. V. Chegel, 2022, Micro and Nanostructure of Layer Si/Sn/Si Films Formed by Vapor Deposition, J. Nanomater., 7910708
55. K. Santosh Kumar, C. Manoharan, S. Dhanapandian, A. Gowri Manohari, and T. Mahalingam, Effect of Indium Incorporation on the Properties of SnS Thin Films Prepared by Spary Pyrolysis, *Optik*, 2014, **25**(15), p 3996–4000
56. K.L. Johnson, K. Kendall, and A.D. Roberts, Surface Energy and the Contact of Elastic Solids, *Proc. R. Soc. A.*, 1971, **324**, p 301–313
57. K.L. Derijaguin, V.M. Muller, and Y.P. Toporov, Effect of Contact Deformations on the Adhesion of Particles, *J. Colloid Interface Sci.*, 1975, **53**, p 314–326
58. S.G. David, W.C. Robert, and E. Flater, Accounting for the JKR-DMT Transition in Adhesion and Friction Measurements with Atomic Force Microscopy, *Journal of Adh. Sci. Tech.*, 2005, **19**(3–5), p 291–311
59. C. Birleanu, M. Pustan, R. Müller, C. Dudescu, V. Merie, R. Voicu, and A. Baracu, Experimental Investigation by Atomic Force Microscopy on Mechanical and Tribological Properties of Thin Films, *Int. J. Mater. Res.*, 2016, **107**, p 1–10
60. D.X. Zang, B. Shen, Y.X. Zeng, S.Y. Wang, J.B. Jang, S.D. Yang, R.J. Zang, L.Y. Chen, C.Z. Wan, and K.M. Ho, Evolution of Optical Properties of Tin Film From Solid to Liquid Studied by Spectroscopic Ellipsometry and ab inito Calculation, *Appl. Phys. Lett.*, 2014, **104**, 121907
61. Z. Jingo, Z. Yuxiang, Z. Dongdong, Y. Shangdong, Y. Liao, L. Zhunhua, Z. Rongjun, W. Songyou, D. Zang, and C. Liangyao, Ellipsometric Study on Size-Dependent Melting Point of Nanometer-Sized Indium Particles, *J. Phys. Chem.*, 2016, **120**, p 10686–10690
62. P.N. Vinod, Specific Contact Resistance Measurements of the Screen-Printed Ag Thick Film Contacts in the Silicon Solar Cells by Three-Point Probe Methodology and TLM Method, *J. Mater. Sci. Mater. Electron.*, 2011, **22**, p 1248–1257
63. A.K. Anand Singh, S.J. Shukla, B.S. Yadav, and R. Pal, Electrical Characteristics of Electroless Gold Contacts on p-Type $Hg_{1-x}Cd_xTe$, *Mater. Sci. Semiconduct. Process.*, 2014, **26**, p 294–300
64. W.A. Beck, G.D. Davis, and A.C. Goldberg, Resistance and 1/f Noise of Au, Al, and Ge Contacts to (Hg, Cd)Te, *J. Appl. Phys.*, 1990, **67**, p 6340
65. S. Nozaki and A.G. Milnes, Specific Contact Resistivity of Indium Contacts to n - Type CdTe, *J. Electron. Mater.*, 1985, **14**(2), p 137–155

66. V. Krishnamurthy, A. Simmons, and C.R. Helms, Oxide Interfacial Layers in Au Ohmic Contacts to p-Type Hg_{1-x} Cd_xTe, *Appl. Phys. Lett.*, 1990, **56**, p 925
67. E.H. Rhoderick, *Metal Semiconductor Contacts*, Oxford University Press, Oxford, 1978, p 77–115

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

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

Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comprehensive study of Design attributes for sustainable packaging design: Analytical techniques and expert validation

Taruna Singh

tarunasingh@dtu.ac.in

Delhi Technological University

Ranganath M. Singari

Delhi Technological University

Sumer Singh

Indian Institute of Technology

Research Article

Keywords: Sustainable Packaging design (SPD), environment, sustainability, attributes and Kano model

Posted Date: December 23rd, 2024

DOI: https://doi.org/10.21203/rs.3.rs-5668960/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

A significant shift toward incorporating sustainable practices into packaging design with the increasing concerns about environmental sustainability. Sophisticated frameworks or tactics are necessary to lessen the increasing environmental impact of packaging waste. However, there are numerous opportunities to reduce waste and use more sustainable packaging design (SPD) techniques despite these initiatives. To address the need for a holistic strategy, this study introduces a comprehensive approach to sustainable packaging design. This technique covers packaging systems and packaged goods across the entire supply chain starting from the producer to the final consumer. This is presented as a strategic blueprint that aims to achieve not only more efficient but also more sustainable. It will provide a holistic perspective based on identified attributes on what constitutes a sustainable packaging solution. The approach incorporates several distinct analytical methods that ensure cost-effectiveness. The methodology is intended to be a practical tool for the packaging Industry and designers who seek to enhance sustainability within their packaging systems. Experts were selected from industry, academia and entrepreneurial background to enhance robustness of the proposed methodology. Moreover, attributes of this framework were validated on the Kano model that analysed whether its strength and relevance possessed significant impacts towards sustainable packaging design. The findings of this research will contribute to the development of a comprehensive framework for sustainable packaging design. These attributes will offer valuable guidance to manufacturers, designers, and companies seeking to prioritize sustainability in their operations. By providing a structured approach to decision-making, it aims to empower stakeholders to make well-informed choices that align with environmental goals and improve the overall sustainability of packaging systems.

1. Introduction

India generated 17 million tonnes of plastic in 2019, and throughout the past 30 years, the country's plastic consumption has increased by more than 20 times (India Plastics, 2019). It is anticipated that India's packaging market would expand by 18% a year, with rigid packaging rising at 15% and flexible packaging at 25%. Packaging is the economy of India's fifth-largest industry and one of its fastest-growing industries. The Packaging Industry Association of India (PIAI) estimates that the industry is expanding between 22%-25% annually. According to a research by the global consulting company EY and the trade and commerce trade association ASSOCHAM, India's packaging market is expected to rise to $73.6 billion by the 2020 fiscal year (FY2020) as a result of the country's rising income levels and population. It's regarded as one of the crucial Ps in the marketing mix. It is regarded as a "Five Second Commercial" and a "Silent Salesman." One of the key components of marketing is said to be packaging. While it was viewed as indifferent in the 1950s, it is now gaining a lot of relevance. It becomes much more significant in the modern day as it offers a variety of dimensions, including functional, technological, instructive, and visual. Customers are looking for better options/solutions to their issues. Therefore, packaging need to be viewed as one of the key components of the marketing mix rather than just a simple container. According to the data, shoppers often make a purchasing choice in a matter of seconds. Thus, it functions as a "Silent Salesman." Packaging often indicates the quality of the goods.

Packing needs to be good from a technical, functional, instructive, and aesthetic standpoint. Therefore, the need for recyclable, reusable and environmentally friendly packaging options is becoming a key driver of innovation in many sectors (Byggeth and Hochschorner, 2006; Beitzen-Heineke et al., 2015). Packaging affects customer pleasure and loyalty, it plays a crucial role in consumer buying behaviour. In this regard, a number of researchers stress the need to study the connection between the packaging's quality attribute and how it influences consumers' purchasing decisions (Oliver, 1997; Lofgren and Witell, 2005). While some studies try to look at every possible aspect of packaging and how it affects consumers' decisions to buy (Butkeviciene, Stravinskiene, & Rutelione, 2008), others concentrate on specific aspects of packaging and how they affect consumers' decisions to buy (Ampuero & Vila, 2006; Madden, Hewett, & Roth, M, 2000; Underwood, Klein, & Burke, 2001; Bloch, 1995). As sustainability becomes an increasingly important focus in packaging design, businesses are under significant pressure to move away from the traditional, linear models of production and consumption. There is a strong push for adopting more sustainable, circular practices that minimize waste, reduce environmental impact, and promote resource efficiency. Data suggests that consumers are not only aware of these issues but are also willing to invest in products with sustainable packaging. The perceived major hurdle when implementing SPD in product development and the design process is cost. The main reason why the majority of users did not utilize SPD tools is because they assumed that their customers would not be willing to pay the additional cost (Ma, X., & Moultrie, J. 2018). Many consumers now recognize that sustainable packaging solutions contribute to environmental conservation and broader sustainability efforts, further driving demand for eco-friendly options. In response to these changing consumer expectations and environmental challenges, packaging companies are reassessing their approaches to product packaging. Rather than merely focusing on the functional aspects of packaging, such as protection and convenience, industries are increasingly incorporating sustainability into the design and production processes. This shift is a reflection of the growing recognition that sustainable packaging is not just a trend, but a vital part of doing business responsibly in today’s environmentally conscious market (Singhal and Malik, 2018). A thorough grasp of consumer preferences is essential to creating packaging that works as well as possible. It is crucial to certify your understanding of how package features and traits affect overall sustainable growth. To promote sustainable development, the packages must meet the social, economic, and environmental standards of the community. (T. Singh et at., 2023).

The Kano Model was used when conducting an empirical study. The findings from Kano model demonstrate that, although boosting customer happiness is vital, preventing disappointment among consumers is much more crucial. However, how can buyers feel satisfied with a product or assistance? Customer satisfaction is correlated with customer satisfaction. Certain consumer requirements must be met by the product or service in order for those demands to be fulfilled. Different customer needs have varying effects on customer satisfaction over time because consumers become accustomed to them, because alternatives are offered, or because rivals seek to enhance their already-existing items. Then, in a market that is competitive, knowing the significance of the various product features and included them in the product is insufficient. It's also critical to monitor shifts in consumer demands and continuously assess how competitive the product is. Packaging industry should create premium packaging by

considering the needs of their target audience. Given the paucity of research in the field of packaging, it is imperative and necessary to investigate the quality attributes of packaging design. Traditionally, packaging has been regarded as a key element of the physical product (Brown 1950). However, with changing customer expectations and increasing demands, the role of packaging has grown in importance, as it now provides additional information and functionality. This evolution may lead to modifications or enhancements in packaging functions beyond merely protecting products and aiding storage and transportation. The inclusion of new features or services in packaging can be seen as a shift along the goods-to-services continuum. Industry considers packaging to be a crucial component of product strategy and consider it to be the fifth P in the marketing mix, along with the other four Ps of product, price, promotion, and place (Philip Kotler K. L., 2008). As customers perceive goods and services differently, this transition emphasizes the need for companies to reconsider their definitions of quality and how it ties to customer satisfaction. An important part of packaging industry is sustainability. However, there are numerous obstacles for adapting sustainable development in packaging because of its high visibility and significance as a crucial enabler for the cost, marketing, distribution, and safe use of consumer and other products. Examples of factors driving the development of new package designs and technologies that frequently run counter to the ideas of sustainable development include changes in distribution, market segmentation, and consumer behaviour and spending patterns (James et al. 2005). Therefore, organizations must focus on developing, designing, and delivering packaging that ensures high levels of perceived quality.

A pressing question arises: how do improved or enhanced packaging designs contribute to a better sustainable solution? In other words, how should packaging be designed to enhance product competitiveness while also promoting associations with sustainability? The authors set out to identify the most important design attributes of packaging design that contribute to sustainability. Using the Delphi approach, 30 experts were selected from various design fields like academia’s, entrepreneurs & industry. Drawing inspiration from Herzberg's Motivation-Hygiene Theory (M-H Theory), Professor Kano and his associates formulated the Theory of Attractive Quality (Kano et al., 1984). This theory is designed to enhance the understanding of how customers assess and perceive the key features or the attributes. According to the Theory of Attractive Quality, the relationship between the extent of fulfilment and customer satisfaction with a quality attribute can be categorized into five distinct types of perceived quality : I) attractive attributes; 2) must-be attributes; 3) reverse attributes; 4) one-dimensional attributes; and 5) indifferent attributes.

Experts were subsequently administered a structured questionnaire designed to assess their experiences with the functioning and dysfunction of 13 distinct attributes in context of SPD. The Kano model of excitement and basic quality (Kano et al., 1984; Berger et al., 1993; Mattler et al., 1996) provides a distinctive framework for analysing opportunities for product and service improvement. The Kano model considers the asymmetric and non-linear relationship between key features and customer satisfaction. It categorizes consumer needs into three types: basic attributes, performance attributes, and excitement attributes. (refer to Fig. 1). This categorization facilitates a deeper understanding of how

different attributes affects overall customer satisfaction and identifies areas where enhancements could lead to significant improvement.

2. Kano model

The Kano model provides insight on how the different attributes actually affects customer satisfaction and helps in prioritizing improvements based on the expectations of the customer. With this knowledge on how the attributes are dynamic—where, for example, attractive attributes become basic over time—companies are better able to adapt their strategy and improve and maintain customer satisfaction effectively. The Kano model can effectively address financial limitations by pinpointing features that yield the highest benefits with the least investment For instance, when multiple service aspects require improvement but resources and time are limited, the Kano model helps identify which attributes most significantly impact customer satisfaction, thus prioritizing those needing immediate attention (Matzler K et al., 1996). Over time, attribute classifications can change: desirable attributes can turn into one-dimensional attributes, and one-dimensional attributes can become must-be attributes (Zacarias D. 2018) By monitoring the changes in attribute classifications between different time periods, a retailer can be ahead of competition because this analysis shows the shifting preference of customers and how the importance of the attributes changes.

Retailers must comprehend the significance of different attributes in influencing customer satisfaction (Matzler K et al., 1998). It has become much easier to assess such attributes since the Kano model was introduced by Professor Noriaki in 1984. The Kano model establishes a connection between customer satisfaction and product quality and functionality (Ek Fİ, Çıkış Ş. 2015 and Avikal S et al.; 2014). It also provides a graphical representation based on survey data to depict customer satisfaction and dissatisfaction. In this graph, the horizontal axis represents the extent to which a feature's functionality is fulfilled, while the vertical axis reflects the resulting customer satisfaction.

The Kano model classifies features into three primary categories based on their effect on customer satisfaction: must-be quality, one-dimensional quality, and attractive quality (Sauerwein E et al., 1996 and (Matzler K et al., 1996). These other dimensions of customer perception include Avikal et al. notions of indifferent quality and reverse quality. Overall, the five categories have given an effective insight to how various characteristics will impact on customer satisfaction (Ek Fİ, Çıkış Ş. 2015, Shahin A, et al., 2015, Salehzadeh R et al., 2015).

In the Kano model, attributes are categorized into several distinct types:

a. **Attractive Attributes (Excitement Attributes)**: These are the attributes that exceed customer expectations and provide delight or surprise when present. Their absence does not lead to dissatisfaction, but their presence significantly enhances customer satisfaction and creates a sense of excitement.

b. **Must-Be Attributes (Basic Attributes)**: These are the essential features that customers expect to be included. The lack of these attributes causes significant dissatisfaction, while their presence is expected and does not necessarily boost satisfaction beyond a baseline level. They represent fundamental requirements for the product or service.
c. **Reverse Attributes**: These attributes elicit varying reactions from different customers; some may find them satisfying while others may find them dissatisfying. This duality makes them polarizing, as they can lead to opposite responses depending on individual preferences.
d. **One-Dimensional Attributes (Performance Attributes)**: These features have a direct and correlative relationship with customer satisfaction. The higher the level of performance of these attributes, the greater the customer satisfaction, and conversely, deficiencies in these areas lead to reduced satisfaction.
e. **Indifferent Attributes**: These attributes have minimal impact on customer satisfaction, regardless of their presence or absence. Customers generally do not have strong preferences or reactions to these features, and they do not significantly influence their overall perception of the product or service.

3. Methodology

An empirical study was conducted to evaluate the packaging's qualities. An expert panel from the field of design was chosen to conduct the research using the Delhi method. A survey was then conducted to determine the significance of various quality attributes and their ranking according to the Kano Model. The flow of the methodology is shown in Fig. 2.

3.1 Studying the dimensions of attributes for sustainable packaging design

The current study intends to classify and rank these traits using approaches like the Kano Model, offering insightful information for user-centred product creation (Lo 2021; Yang, 2005; Wenninger, Rau, and Röglinger, 2022). The Kano model takes user emotions and product quality into account while identifying basic, performance, and excitement aspects (Lo 2021).

Identifying Attributes of Sustainable Packaging Design (SPD)

In the initial phase, an extensive brainstorming session was conducted to compile a list of packaging design attributes, resulting in a total of 341 attributes. These attributes were subsequently validated for their relevance to sustainable packaging design through a comprehensive literature review, incorporating sources such as journals, books, online resources, catalogues, market surveys, and research papers. This validation process facilitated the refinement of the list, narrowing it down to 120 closely related attributes.

Refining the attributes for (Adjective, verbs etc) for the survey

A further refining procedure was carried out to get the qualities ready for efficient survey usage after an initial list of 341 features pertinent to sustainable package design was created and validated. This phase included examination of each property to determine its clarity, relevance, and conciseness so that it fits into the widely accepted descriptors that are often verbs, adjectives, and other simple phrases. The 17 attributes identified by this methodical refining were survey-ready, succinct, and comprehensive, allowing for more significant insights into the preferences and goals for sustainable package design as well as clearer replies.

Defining attributes and their definitions

In this phase, each sustainable packaging design (SPD) attribute was clearly defined to ensure a consistent understanding among survey respondents. To prevent misinterpretation, definitions were provided in plain, uncomplicated language. For instance, "biodegradable" was defined as "decomposes naturally without harming the environment, such as through composting or microbial breakdown," and "recyclable" as "capable of being reprocessed and reused in manufacturing." In order to emphasise the qualities of the attributes, they were also characterised using comparable adjectives and verbs, such as "durable," which is connected to Material, Strong, Sturdy, robust, Long-lasting, Robust, etc. This process ensured that every attribute is identified, which would create an appropriate foundation for accurate and trustworthy survey results. Attribute and their definition are as shown in Table 1.

Table 1
Definition of attributes.

S.No	Attribute	Definition	Other parameters
1	Accessibility	In the state of being always available	Accessible Availability Attainability Acquirable Within reach Easy to reach
2	Adaptable	Referring the ability of an identity to adapt it-self efficiently and fast to changed circumstances.	Adjustable Convertible Flexible Modifiable Scalability Compatible
3	Aesthetic	Being interested in how something looks, feels and seems pleasing to the senses and especially to the sight	Pleasing Beautiful Attractive Eye catchy High visibility Appreciable Desirable
4	Affordable	The state of being cost-effective enough for mass to be able to buy.	Cost-effective Inexpensive Low-cost Cost Accepted by mass
5	Contextual	A diversified solution that is convenient to use, generous and impactful in most situations in various demographic situations	Achievable Awareness Convenient Generous

S.No	Attribute	Definition	Other parameters
			Impactful Adequate Feasible Dependable Loyal
6	Durable	Made up of good quality material that can last for long	Material Strong Sturdy Resilient Long-lasting, Robust
7	Emotional	Consumers experience a feeling or attachment with a product.	Affectionate Caring Cultural Nostalgic Attachment Belongingness
8	Sustainable	Created to develop eco-friendly solutions that address immediate needs while integrating a long-term outlook that “Returns to its roots”. Not or less harmful to the environment.	Biodegradable Sustainable development Natural Eco-friendly Recyclable Minimal damage Eco-system Green Non-polluting Low carbon-footprint Energy-Efficient

S.No	Attribute	Definition	Other parameters
9	Ergonomic	Product to improve people's working/operating conditions and help them work more effortlessly.	Easy to use User friendly Flexibility in use Low pain points Used by mass
10	Functional	Relating to the way in which product can serve for longer time than intended, or relating to how useful it is.	Useful Easy to use Re-usability After-use Shelf-life
11	Inclusive	Should be easily used by as many people as possible of all backgrounds and abilities	Comprehensive For all
12	Informative	A component of design that uses visuals, text or content that strategically convey a message or express information.	Branding Readability Promotional Communicative Information
13	Intuitive	With an intuitively designed product, customers will understand how to use it without much effort.	Understandable Perceived value Perceived information Status
14	Materials	A substance that can be used for making a sustainable solution	Easily available Robust Long lasting Procurable Eco-friendly Production – friendly transportable

S.No	Attribute	Definition	Other parameters
			Durable Hygienic Re-growing
15	Portability	Should provide a ease factor in transportation / distribution of packaging or its material	Light weight Cost effective Manageable Convenient Compact
16	Protection	The state of being shielded from harm and being hygienic.	Shelf-life Hygienic Harm-less Low / Zero contamination
17	Social		Behavioural Attachment Awareness Diverse Responsible Empowering Collaborative Social Values & Responsibilities

Identification of experts by Delphi Method

The identification of experts for this study was conducted using the Delphi method, a structured approach that aims to gather insights from a panel of knowledgeable individuals through multiple rounds of questioning (Linstone & Turloff, 1975). The Delphi method (Dalkey & Helmer, 1963) is particularly useful for achieving consensus on complex issues where there may not be a single clear answer. Experts from this list were selected on their basis of experience in sustainable packaging design, environmental sustainability, and related fields. Based on the expertise, industry experience, and familiarity with the latest trends and challenges in packaging design, these experts were chosen for this study. For that, the

selection was carried out by identifying academics, professionals, and practitioners. After the elicitation of experts, questionnaires from the Delphi method were then sent out among them in a series of rounds. Responses were analysed, refine and improve the next set of questions, allowing the group to converge toward a consensus on key attributes and considerations in sustainable packaging design. This iterative process helped ensure the reliability and validity of the findings, capturing a broad range of expert opinions and leading to a more comprehensive understanding of the subject.

Conducting the survey: In this research study, the questionnaire was formulated for gathering opinions of experts concerning the attributes of sustainable packaging design. It comprised functional as well as dysfunctional questions on the Kano model, measuring the way the absence or presence of the concerned attributes was influencing experts' perceptions. In this study, sustainable packaging attributes are categorized into six Kano quality types: **A (Attractive Quality)**, **M (Must-be Quality)**, **O (One-dimensional Quality)**, **I (Indifferent Quality)**, **Q (Questionable/Doubtful Quality)**, and **R (Reverse Quality)**. Experts rated their responses on a five-point scale: "I like it," "I expect it," "I am neutral," "I can tolerate it," and "I dislike it." This allowed for a detailed understanding of each attribute's significance. The questionnaire was distributed to experts either in person or via Google Forms, enabling easy access and completion. Since no pilot survey was conducted with the experts, the initial responses were directly used to refine the findings and conclusions.

3.2 Analysis using the kano model

The Kano Model is utilized to evaluate the attributes that influence designers' decision-making and to categorize customer requirements. It helps aiming to improve the service's performance of a product or service. The following steps were implemented to apply the Kano Model in assessing and categorizing the attributes of packaging.

A questionnaire in which the experts indicates whether they are satisfied with a hypothetical scenario serves as the basis for the Kano model's identification of excitement and fundamental criteria (Fig. 3). Each product need is followed by two questions. The consumer has five options for responding to each inquiry (Kano et al., 1984). The first question, sometimes known as the functional question, asks the client how they would respond if the product met the demand or if it meets the requirement with appropriate performance. The response is indicated by the second question, often known as the dysfunctional question, if the product fails to meet those criteria or performs inadequately. The criterion may change based on how consumers respond to functional and dysfunctional inquiries.

The most important decision in the Kano approach is which of the alternatives to choose. The wording of the options, which were taken from (Berger et al. 1993), is comparable to the Japanese version proposed by (Kano et al. 1984). These choices include "I like it that way," "It must be that way," "I am neutral," "I can live with it that way," and "I dislike it that way." The wording employed by (Berger et al. 1993) was determined to be the most appropriate by the authors for their analysis of Swedish customers. However, they think that the phrasing should be adjusted based on the respondents that are being worked with.

The expert can respond to each section of the research by selecting one of the five options shown in Fig. 3. claims that the phrasing. The expert can respond to each section of the research by selecting one of the five options shown in Fig. 4. Berger et al. (1993) claims that the phrasing one should take into account the subtle differences in vocabulary between Japanese, English, and Swedish language.

Table 2
Kano evaluation table.

Functional	Dysfunctional					
		I LIKE IT	MUST BE	I AM NEUTRAL	I CAN LIVE WITH IT	I DISLIKE IT
	I LIKE IT	Questionable	Attractive	Attractive	Attractive	One-Dimensional
	MUST HAVE	Reverse	Indifferent	Indifferent	Indifferent	Must-have
	I AM NEUTRAL	Reverse	Indifferent	Indifferent	Indifferent	Must-have
	I CAN LIVE WITH IT	Reverse	Indifferent	Indifferent	Indifferent	Must-have
	I DISLIKE IT	Reverse	Reverse	Reverse	Reverse	Questionable

The classification of attributes based on the responses from participants was carried out using three distinct methods:

1. **Frequency-based attribute classification method:** This approach involves categorizing each attribute based on the response category with the highest frequency (M, O, A, I, R, Q). The attribute is classified according to the category that received the most responses.
2. **Comparison-based attribute classification method:** In this method, attributes are classified by comparing the total frequencies of two response groups. If the combined frequency of categories M, O, and A is higher than that of I, R, and Q, the attribute is classified under the category with the highest frequency within M, O, or A. If the total frequency of I, R, and Q is higher than M, O, and A, the attribute is assigned to the highest frequency category among I, R, or Q. If the sums are equal, the classification follows a priority order established by (Matzler et al. 1996), specifically M > O > A > I.
3. **Index-Based Attribute Classification Method:** This method introduces two indices: the Satisfaction Index (SI) and the Dissatisfaction Index (DI).

The Satisfaction Index is calculated as

$$SI = \frac{(A+O)}{(A+O+M+I)}$$

with a range from 0 to 1, while

The Dissatisfaction Index is determined as

$$DI = \frac{(M+O)}{(A+O+M+I)}$$

* (-1), which ranges from -1 to 0.

A diagram is used to plot the Satisfaction and Dissatisfaction Indices for the 12 attributes to give an overall perspective.
Various attributes were categorized using the Satisfaction and Dissatisfaction Indices, as outlined in Table 3.

Table 3
Index based attribute classification

Satisfaction Index (SI)	Dissatisfaction Index (DI)	Classification
< 0.5	≥ 0.5	Must-be (M)
≥ 0.5	≥ 0.5	One-Dimensional (O)
≥ 0.5	< 0.5	Attractive (A)
< 0.5	< 0.5	Indifferent (I)

Category Strength (CS) and Total Strength (TS), introduced by Lee and Newcomb in 1997, are two metrics used to assess attributes. Category Strength (CS) quantifies the difference in response percentages between the most frequently selected category and the second-most selected category. For example, if an attribute's highest response category is "O" with 40% of responses, and the second highest is "A" with 30%, then CS would be calculated as 40% − 30% = 10%. Total Strength (TS), on the other hand, represents the combined percentage of responses across the three categories: Must-be (M), One-dimensional (O), and Attractive (A). For example, if an attribute's responses are 20% for "M," 40% for "O," and 10% for "A," then TS would be the sum of these percentages, resulting in TS = 20% + 40% + 10% = 70%.

4. Results and discussions

The empirical study's results, detailed in Table 4, allowed for the definitive classification 12-quality attributes using the Kano model. It shows total number of ratings received from each category i.e. A = Attractive, O = One Dimension, M = Must have, I-Indifferent, R-Reverse and Q = Questionable. This classification was achieved through an analysis of category strength, total strength, and the frequency of questionable responses, complemented by statistical testing as shown in Table 4.

Table 4
Dimensions and attributes of packaging.

S.No	Attribute	A	O	M	I	R	Q	TOTAL
1	**Accessible**	4	5	10	8	0	0	27
2	**Adaptable**	3	5	10	9	0	0	27
3	**Aesthetics**	11	7	6	3	0	0	27
4	**Affordable**	3	10	12	1	1	0	27
5	**Contextual**	2	3	4	17	1	0	27
6	**Durable**	4	6	9	8	0	0	27
8	**Functional**	3	9	8	7	0	0	27
9	**Informative**	11	7	6	3	0	0	27
10	**Portable**	4	7	6	10	0	0	27
11	**Protection**	2	10	12	2	1	0	27
12	**Reusable**	3	10	12	1	1	0	27

The results of the study show high statistical significance, as 12 out of the measured attributes were shown to have a p-value of less than 0.01, which reflects very high confidence in the relationships found. The horizontal coordinate indicates the fulfilment of the quality attributes, while the vertical coordinate indicates consumer satisfaction; it also represents which product to be kept in a one-dimensional category of attributes, attractive attributes and must be attribute shown in Fig. 4. Attributes like Protection, Reusable, affordable, accessible and energy efficient came out to be the must have design attributes for SPD. Durable and Functional were one-dimensional.

Table 5
Represents the customer satisfaction and customer dissatisfaction coefficient for the attributes along with attribute strength.

S.NO	ATTRIBUTES	SI=(A + O)/(A + O + M + I)	DI=(O + M)/(A + O + M + I) x (-1)	Total Satisfaction Index	Attribute Strength	
1	**Protection**	0.44	-0.81	-0.37	**4.81**	**MUST HAVE**
2	**Reusable**	0.48	-0.81	-0.33	**4.59**	
3	**Affordable**	0.48	-0.81	-0.33	**4.66**	
4	**Adaptable**	0.29	-0.55	-0.26	**4.33**	
5	**Accessible**	0.33	-0.55	-0.22	**4.23**	
6	**Durable**	0.36	-0.55	-0.19	**4.26**	**ONE-DIMENSIONAL**
7	**Functional**	0.36	-0.55	-0.19	**3.74**	
8	**Aesthetics**	0.66	-0.48	0.18	**3.59**	**Attractive**
9	**Informative**	0.66	-0.48	0.18	**3.41**	

Furthermore, the horizontal coordinate in the diagram represents the degree to which the quality attributes of the product are fulfilled, providing a measure of how well each attribute is addressed in the design or features of the product. On the other hand, the vertical coordinate reflects consumer satisfaction, illustrating how much consumer's value and are satisfied with each of these attributes in the product. This two-dimensional framework allows us to understand not only how well the product meets specific quality criteria but also how those criteria translate into consumer perceptions and satisfaction.

Analysis categorizes the attributes into three different groups based on importance and their impact on consumer satisfaction, as shown in Fig. 4. These categories include:

Must-have attributes

Those are the attributes necessary to satisfy the consumer. Without these attributes, the satisfaction of a consumer and his/her perception toward the product may be negative. In SPD (Sustainable Product Design), Protection, Reusable, Affordable, Accessible, and Energy Efficient are the must-have attributes. These attributes are the basis for the basic functional and sustainability requirements that a product should meet. In addition, these attributes are what form the foundation of consumer expectations. If any of these attributes are lacking, the product may not be considered viable or satisfactory by consumers.

One-dimensional attributes

One-dimensional attributes are attributes that build satisfaction with a product in direct proportion. When these attributes fulfilled, satisfaction is correspondingly higher. Durability and Functionality are two such one-dimensional attributes in the research that have been reported. Both of them, though essential for a product, work straightforwardly; in other words, when a product scores well on the dimensions, say, in durability or functionality, it goes hand-in-glove that satisfaction with that product increases correspondingly. However, simply gratifying these characteristics is insufficient for high overall satisfaction because neglecting other more critical characteristic results in low overall satisfaction.

Attractive attributes

An attribute beyond the normal satisfaction with basic expectations may even give greater-than-normal pleasure to delight. While their absence doesn't create dissatisfaction, these may help consumers feel a form of pleasure or additional value from being present and done sufficiently. Aesthetics, for example, as well as the Informative features of this study, fall within this category of attractive characteristics. These elements enhance the appeal of the product by providing more layers of satisfaction, such as visual design or informative features that educate the consumer about the benefits and use of the product. The classification of attributes into 'must-have', 'one-dimensional', and 'attractive' helps to determine which attributes are most critical for meeting basic consumer expectations, which enhance overall satisfaction, and which are more likely to provide a delightful experience. The balance between these attributes is key to designing sustainable packaging that not only meets fundamental consumer needs but also offers extra value that can improve consumer satisfaction and loyalty. This expanded version offers more context to the categorization of attributes and their implications for product design, consumer expectations, and satisfaction.

Packaging has become a crucial component. Brand’s overall sustainability strategy as environmental consciousness grows in importance among both consumers and Industries. Nowadays, Industries see packaging as a powerful tool to demonstrate their sustainable philosophy rather than just a means of transporting their goods. Brands can respond to the growing number of environmentally conscious consumers by projecting a very eco-friendly image with the use of creative packaging designs and materials obtained responsibly. A long-lasting or sustainable business is a good business. Sustainability is a modern business strategy that considers social, environmental, and economic factors to ensure a company's long-term viability (K. Haanaes 2016) "What is the role of designers with their designs so that a business can be sustainable?" is the next question to be taken into consideration. Implementing the idea of sustainable development into the design is one duty that a competent designer might complete. A designer bears moral and ethical responsibility for the design and its effects (O. Diegel 2001; and D. Nemerson, J. Rosenbeck 2009). This study offers a thorough method for designing sustainable packaging that covers packaging systems and packaged commodities at every stage of the supply chain, from manufacturers to final customers. The suggested methodology, which serves as a "blueprint" for creating more environmentally friendly packaging solutions, pinpoints essential characteristics that support sustainability from all angles. According to this study, the five most crucial design attributes that consumers hunt for in packaging are Protection, reusable, affordable, adaptable and accessible. These

sustainable packaging attributes will provide a emerging perspective to creative approaches that suit consumer needs while striking a balance between ecological responsibility and affordability. These attributes can be used develop sustainable packaging design that can have more design qualities, which will lead to a better sustainable packaging design solution. Kano model have utilized to identify the most desirable attributes classifying customer needs into five categories—Basic must have, one-dimensional, attractive, Indifferent, and Reverse Needs—the Kano Model aids firms in better understanding and prioritizing customer needs, making it a useful research tool. Packaging Industry may concentrate on what are the needs for customers. This study will offer insights into how many aspects of a product or service impact customer happiness. Whether it's introducing unexpected attributes, boosting performance, or strengthening basic functions, it helps prioritize resources to increase consumer satisfaction. Furthermore, because the Kano Model is dynamic and acknowledges that customer expectations change over time, firms can benefit from continual research to stay abreast of market trends. Businesses can improve customer loyalty, lower unhappiness, and gain a competitive edge by employing this methodology to make well-informed, customer-centric decisions. Additionally, it promotes improved departmental communication by offering a common language for talking about consumer expectations and successfully directing product development.

5. Conclusions

To gain a competitive advantage, stakeholders in the packaging industry must understand the attributes that are most important to developing a better sustainable solution. This study not only identifies these attributes but also ranks them, enabling the industry to prioritize the most essential attributes. To meet the study’s objectives, an extensive literature review was conducted, 12 key attributes are critical for developing sustainable packaging, with must-have attributes like protection, reusability, affordability, adaptable and accessibility being central to customer satisfaction. Findings from this study reveal that respondents experience varying levels of customer satisfaction depending on the fulfilment of specific attributes. The findings also showed that a sustainable packaging must protect the content inside in order to maintain its basic purpose, since this is an essential quality, as it lack would drastically reduce consumer satisfaction. Other than Reusability and affordability have scored the same SI and should be considered while designing a SPD along with adaptability and Accessibility. Aesthetics and informative are the attributes from the appealing group that is proportionately related to customer satisfaction but whose absence does not result in discontent from customers. Attributes like offering Durability and Functional are proportionately related to consumer pleasure. Majority of the attributes with a high degree of attribute strength come from the must-be and one-dimensional categories, according to the self-stated importance questionnaire. The results of this study will assist stakeholders of the sustainable packaging design industry in learning more about the relative significance of design attributes and will allow them to assess how their present procedures affect the degree of client satisfaction. This study demonstrates the application of the Kano model in evaluating customer satisfaction within the sustainable packaging sector. However, the findings are derived from the insights and expertise of design professionals. It is important to note that the results predominantly reflect the perspectives of

this demographic and should not be generalized to other customer groups. The scope of this study will be expanded by include target participants who are clients from a number of different age groups. If there is a relationship between income level and customer satisfaction traits for the retail industry, it may be found by doing similar research in other regions of the world with varying economic circumstances. Likewise, this study might be practiced in industries other than retail, including the supply industry, to identify the traits that can be enhanced for increased output and contentment. The study reflects the perspectives of design professionals and should be expanded to include a broader range of customer demographics, including different age groups and income levels, to validate results. Further studies could explore regional and industry-specific variations in customer satisfaction and sustainable packaging preferences, enhancing the understanding of diverse consumer needs. The study's conclusions would be more comprehensive if these varied demographics were included, enabling packaging Industry better to match their packaging techniques with a larger clientele. Additionally, further studies could investigate regional and industry-specific variations in customer satisfaction and packaging preferences. This would enhance understanding of how cultural, geographic, and sectoral factors shape consumer expectations, providing more targeted insights. Understanding these regional and industry-specific differences could help businesses tailor their sustainable packaging approaches to meet the unique needs of diverse markets and industries, ultimately fostering greater consumer satisfaction and engagement.

Declarations

Conflict of interest:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contribution

Taruna Singh: Conceptualization, Methodology, Investigation. Ranganath M. Singari: Writing – review & editing. Sumer Singh: Supervision, Writing – review & editing.

Acknowledgement

AcknowledgementThe authors acknowledge the Department of Design, Delhi technical university, Delhi, India for providing the resources. The authors also acknowledge central facilities of Delhi technical university, Delhi, India, for providing research facilities.

Availability of data and materials

The raw processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

References

1. Ampuero O, Vila N (2006) Consumer perceptions of product packaging. J Consumer Mark 23(2):100–112
2. Avikal S, Jain R, Mishra PK A Kano model, AHP and M-TOPSIS method-based technique for disassembly line balancing under fuzzy environment. Appl Soft Comput 2014 Dec 1;25:519–529
3. Berger C, Blauth R, Boger D (1993) Kano's methods for understanding customer-defined quality
4. Beitzen-Heineke EF, Balta-Ozkan N, Reefke H (2015) The prospects of zero packaging grocery stores to improve the social and environmental impacts of the food supply chain. Clean Prod 140(3):1528–1541. https://doi.org/10.1016/j.jclepro.2016.09.227
5. Bloch PH (1995) Seeking the ideal form: product design and consumer response. J Mark 59(July):16–29
6. Byggeth S, Hochschorner E (2006) Handling trade-offs in Ecodesign tools for sustainable product development and procurement. J Clean Prod 14:15–16. https://doi.org/10.1016/j.jclepro.2005.03.024
7. Brown GH (1950) Measuring consumer attitude towards products. Journal of marketing 14, no. 5:691 to 98
8. Butkeviciene V, Stravinskiene J, Rutelione A (2008) 'Impact of consumer package communication on consumer decision making process'. Inzinerine Ekonomika-Engineering Econ (1), pp. 57–65
9. Dalkey NC, Helmer O (1963) An experimental application of the Delphi Method to the use of experts. Manage Sci 9(3):458–468
10. Ek Fİ, Çıkış Ş (2015. Apr) Integrating the Kano model into architectural design: quality measurement in mass-housing units. Total Qual Manage Bus Excellence 26(3):400–414
11. James K, Fitzpatrick L, Lewis H, Sonneveld K (2005) Sustainable Packaging System Development. In: Leal Filho W (ed) Handbook of Sustainability Research. Peter Lang Scientific Publishing, Frankfurt
12. Kano N, Seraku N, Takahashi F, Tsuji S (1984) Attractive Quality and Must-Be Quality. J Japanese Soc Qual Control 41:39–48
13. Lee MC, Newcomb JF (1997) Applying the Kano methodology to meet customer requirements: NASA's microgravity science program. Qual Manage J 4(3):95–110. https://doi.org/10.1080/10686967.1997.11918805
14. Linstone H, Turloff M (1975) The Delphi method: Techniques and applications. Addison-Wesley, London, UK
15. Löfgren M, Witell L (2005) Kano's Theory of Attractive Quality and Packaging. Qual Manage J 12(3):7–20. https://doi.org/10.1080/10686967.2005.11919257

16. Lo hiHung.2021.ApplicationofRefinedKano'sModeltoShoeProductionandConsumerSatisfac tion Assessment. Sustain (Switzerland) 13 (5): 1–22. https://doi.org/10.3390/su13052484
17. Ma X, Moultrie J (2018) Understand sustainable packaging design in practice. In *DS 92*: (pp. 2693–2704)
18. Madden TJ, Hewett K, Roth MS (2000) Managing images in different cultures: A cross-national study of color meanings and preferences. J Int Mark 8(4):90–107
19. Matzler K, Hinterhuber HH, Bailom F, Sauerwein E (1996) How to delight your customers. J Prod Brand Manage. Apr 1
20. Matzler K, Hinterhuber HH (1998. Jan) How to make product development projects more successful by integrating Kano's model of customer satisfaction into quality function deployment. Technovation 18(1):25–38
21. Oliver RL (1997) A Behavioral Perspective on the Consumer. McGraw Hill, New York
22. Salehzadeh R, Shahin A, Kazemi A, Barzoki AS (2015) Is organizational citizenship behaviour an attractive behavior for managers? A Kano model approach. *Journal of Management Development*. May 11
23. Sauerwein E, Bailom F, Matzler K, Hinterhuber HH (1996) The Kano model: How to delight your customers. *InInternational working seminar on production economics* Feb 19 (Vol. 1, No. 4, pp. 313–327)
24. Shahin A, Shahiverdi SM (2015) Estimating customer lifetime value for new product development based on the Kano model with a case study in automobile industry. Benchmarking: Int J. Jul 6
25. Tranfield D, Denyer D, Smart P (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br J Manag 14(3):207–222
26. Singh T, Singari RM, Singh S (2023) Exploring opportunities in Sustainable Packaging Design industry. ACED 2023, BRICSplus HFE 2023 and HWWE2023. Mumbai, Mharashtra, India, pp 14–16
27. Underwood RL, Klein NM, Burke RR (2001) Packaging communication: attentional effects of product imagery. J Prod Brand Manage 10(7):403–422
28. Wenninger A, Rau D, Röglinger M (2022) Improving customer satisfaction in proactive service design: A Kano model approach. Electron Markets 32(3):1399–1418
29. Zacarias D (2018) The complete guide to the Kano Model: Prioritizing customer satisfaction and delight. Fold Burritos
30. Diegel O Journal of http://156.62.60.45/handle/10292/1713
31. Nemerson D, Rosenbeck J A designer's role in sustainability the extent to which product and process designers can play a role in environmental sustainability and social responsibility. https://www.idsa.org/sites/default/files/Nemerson.pdf (n.d), p. 1–9

Figures

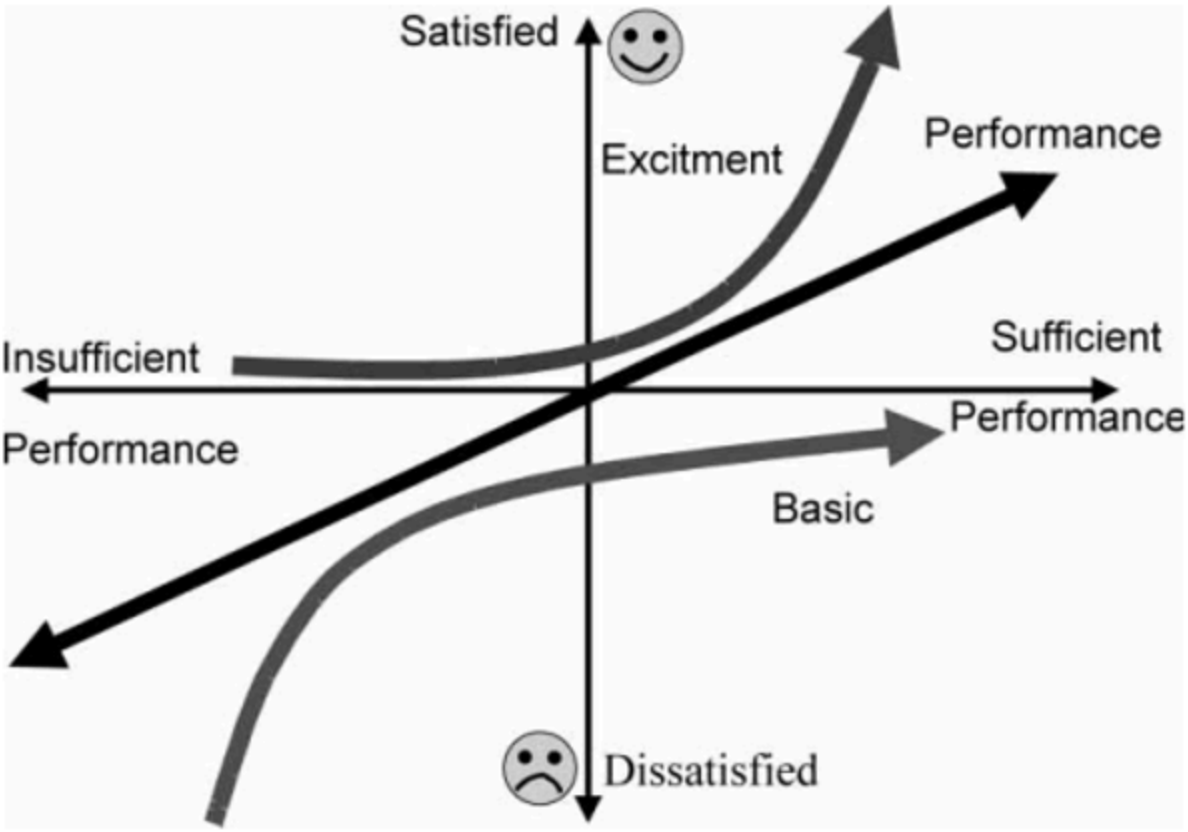


Figure 1

Kano model of excitement and basic quality (Matzler et al., 1996).

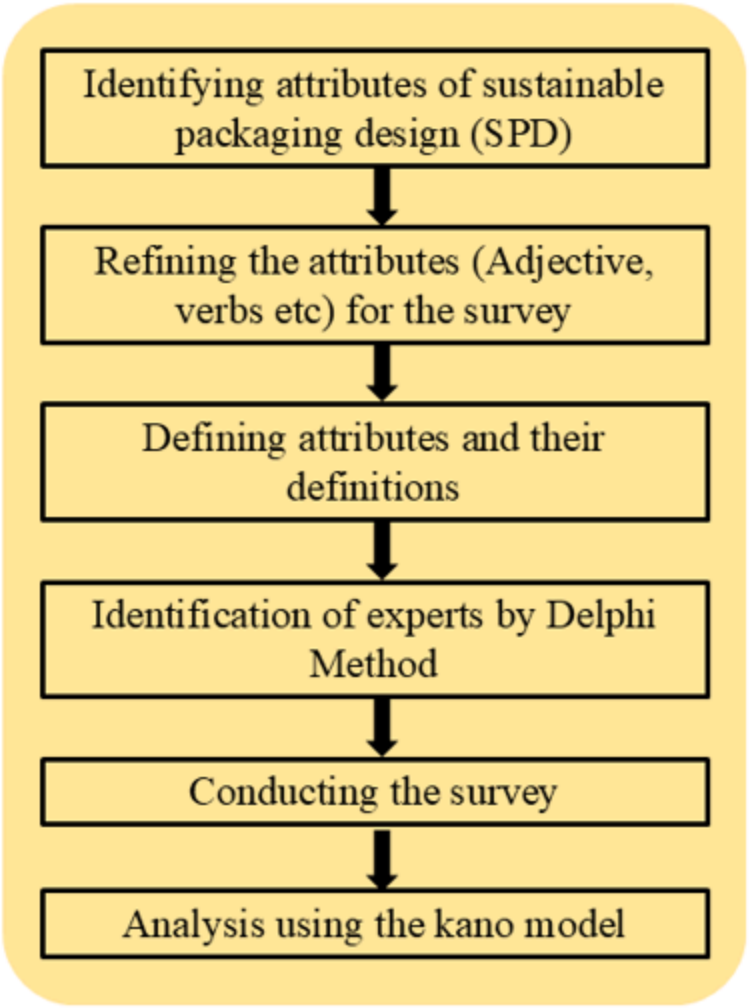


Figure 2

The flow of methodology

Functional form of the question →

Dysfunctional form of the question →

What if the Packaging is aesthetically pleasing?	1. I like it 2. Must be 3. I am neutral 4. I can live with it 5. I dislike it
What if the Packaging is NOT aesthetically pleasing?	1. I like it 2. Must be 3. I am neutral 4. I can live with it 5. I dislike it

Figure 3

Kano model questionnaires.

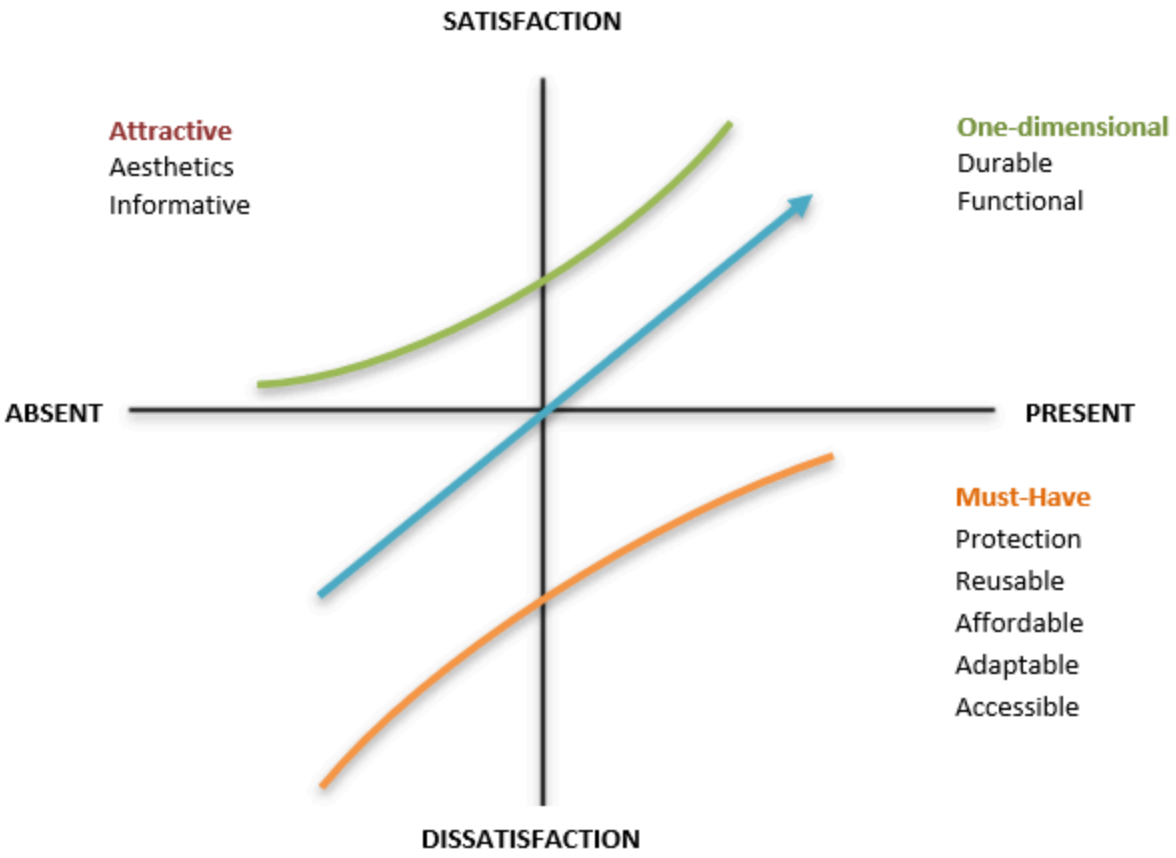


Figure 4

Analysis of the Kano model of SPD attributes

J Mater Sci: Mater Electron **(2024) 35:2106**

Comprehensive study on thermal, structural, and luminescent properties of $BiYWO_6$: Eu^{3+} phosphors synthesized by various methods

Indrajeet Maurya[1] and M. Jayasimhadri[1,*]

[1] *Department of Applied Physics, Luminescent Materials Research Lab (LMRL), Delhi Technological University, Delhi 110 042, India*

Received: 19 April 2024
Accepted: 3 November 2024
Published online: 19 November 2024

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

ABSTRACT

Europium-activated bismuth yttrium tungstate (BYW: Eu^{3+}) phosphors were synthesized by four various synthesis techniques such as solid-state reaction (SSR), sol–gel combustion (SGC), co-precipitation (CP), and hydrothermal (HT) methods. Relative investigations such as thermal, structural, morphological, and luminescence characterizations have been carried out to optimize the synthesis process of BYW:Eu^{3+}phosphor. The TGA–DSC curves signify the endothermic/exothermic peaks and corresponding weight loss during heating. X-ray diffraction analysis and Rietveld refinement have been used to identify the phase and crystal structure of the undoped and doped $BiYWO_6$ sample. The field emission scanning electron microscope with energy-dispersive X-ray was carried out to examine the morphological and compositional behavior of the synthesized BYW: Eu^{3+}phosphor. The luminescent spectral profiles indicate the strong absorption in the blue region (λ_{ex} = 465 nm) and intense emission in the red region (λ_{em} = 613 nm) ascribed to the $^5D_0 \rightarrow {}^7F_2$ transition. The compared photoluminescence (PL) results signify that the phosphor synthesized by the CP method at calcination temperature 900 °C exhibits the strongest emission than the phosphor synthesized via other methods (SGC, SSR, and HT) and is especially two times higher than the phosphor synthesized by the SSR method. Further, the PL intensity enhanced with increasing activator concentration of Eu^{3+} ions up to 20 mol%. The calculated CIE chromaticity coordinates (0.654, 0.345) of 20.0 mol% Eu^{3+}-doped BYW sample were situated in the red region, which is comparable with the commercially available red-emitting phosphors Y_2O_3:Eu^{3+} (0.645, 0.347) and Y_2O_2S:Eu^{3+} (0.647, 0.343). The average PL decay time of the synthesized phosphor was in the microseconds range. The obtained results suggest that the $BiYWO_6$ activated with Eu^{3+} ions phosphor synthesized by the CP method has distinctive PL characteristics with good morphology, which can be employed as an intense red-emitting component in photonics devices.

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

https://doi.org/10.1007/s10854-024-13852-2

1 Introduction

In the recent decade, phosphor-based white-light-emitting diodes (w-LEDs) have revolutionized the lighting industry by providing a more energy-efficient, small size, short response time, durable, and versatile lighting solution for various applications [1–3]. Also, w-LEDs have set the standard for energy-efficient lighting devices to minimize climate change and environmental costs due to carbon emissions during the generation of electric power based on the burning of fossil fuel [4]. Commercially available phosphor-converted w-LED is the combination of YAG: Ce^{3+} phosphor and blue-light-emitting InGaN chip [5]. This lighting device produces light in the blue to yellow portion of the visible spectrum and exhibits poor red-emitting content. As a result, it has a low color-rendering index (CRI), and high correlated color temperature (CCT > 6500 K) [6–8]. However, the readily accessible red phosphor (Y_2O_2S: Eu^{3+}) has poor color-rendering characteristics, is very sensitive to moisture, thermally unstable, and also releases sulfide gas into the environment [5, 9–11]. Therefore, it is indispensable to explore a thermally stable and efficiently blue pumped inorganic red-illuminating phosphor.

An efficient phosphor requires a good host lattice, which is one of the essential prerequisites of the potential phosphor. Among the various oxide phosphor hosts, the tungstate-based materials are attracted due to their good chemical, photochemical, and photothermal stability and have broad optical transparency from visible to the NIR region [12]. Tungstates have direct utility and applicability in the fields of photonics, optoelectronics, quantum electronics, scintillators, lasers, and humidity sensing [13, 14]. The Aurivillius family of layered perovskites includes the compound $BiYWO_6$, which recently received much attention for its microwave dielectric and photocatalytic capabilities [15].

After selecting the host lattice, the synthesis of phosphor using various synthesis routes may affect the structural, morphological and luminescence characteristics of phosphors. Hence, different synthesis routes have been used to optimize the synthesis process and morphology of the $BiYWO_6$ phosphor material for the enhancement of luminescence properties. Among various synthesis routes, the commercially feasible SSR method is one of the most generally accepted techniques to synthesize bulk phosphor materials due to its reproducibility, economically viable, and a large quantity of powder that can be easily produced [16]. Further, the sol–gel combustion approach is one of the attractive chemical methods for synthesizing powder materials in deep microns or nano-size [17]. This approach has a number of benefits, including low cost, short reaction time, simple preparation, and attaining products with high yield. Moreover, the hydrothermal synthesis takes place at a desired temperature and pressure during chemical reaction to produce the nanomaterials with the desired morphology [18]. The foremost feature of this method is that it generates nanomaterials with minimum loss of materials. In addition to the above-mentioned methods, the co-precipitation method has gained popularity due to its benefits, including its simplicity, homogenous mixing of the reactants, high yield, affordability, and short reaction time in the synthesis process [19]. Several researchers/scientists have analyzed and reported that numerous rare-earth-activated tungstates such as $NaY(WO_4)_2$: Eu^{3+}, Y_2WO_6, and $ZnWO_4$ have been synthesized by solid-state reaction, hydrothermal, and sol–gel process [20–24]. In recent years, R. Kamal et al. investigated the luminescent properties of $BaWO_4$:Tb^{3+} nanophosphor for latent print and anti-counterfeiting applications [25]. J. Huagng et al. synthesized $NaBi(WO_4)_2$:Dy^{3+}, Tm^{3+} phosphors and investigated its electronic and photoluminescence studies in detail. Also, Dy^{3+}/Tm^{3+}-co-doped $NaBi(WO_4)_2$ phosphor may be a promising candidate in UV-excited single host w-LEDs [26]. Further, it is confirmed that no study has been reported on the optimization of synthesis and luminescent features of Eu^{3+}-doped $BiYWO_6$ (BYW:Eu^{3+}) phosphor.

In the present work, Eu^{3+}-doped $BiYWO_6$ (BYW:Eu^{3+}) phosphors have been synthesized using four different methods: SSR, SGC, HT, and CP. The purpose of the current study is to reveal the most appropriate and efficient method to synthesize BYW:Eu^{3+}phosphor to act as a red component for w-LEDs. A comparative study of the structural, morphological, and luminescent characteristics was conducted for the phosphors synthesized via numerous methods to disclose the most suitable synthesis method. Optimized $BiYWO_6$ phosphor has been synthesized by varying dopant (Eu^{3+}) concentrations. Further, the suitability of the prospective BYW:Eu^{3+} phosphor for solid-state lighting applications was thoroughly investigated.

2 Experimental

2.1 Materials and synthesis

Eu^{3+}-doped $BiYWO_6$ phosphors have been synthesized via four different synthesis techniques as follows:

2.1.1 Synthesis of BYW: Eu^{3+} by SSR method

In SSR, the starting precursors Bi_2O_3, Y_2O_3, WO_3, and Eu_2O_3 were taken in the stoichiometric ratio as desired to prepare BYW:$x$$Eu^{3+}$ (x = 0.0 and 1.0 mol%) material. The initial precursors were ground in a pastel mortar for 2 h with acetone medium. The mixer was calcined in a high-temperature muffle furnace at 1100 °C for 4 h using heating rate of 2° per minute. The furnace spontaneously cooled to room temperature after calcination, and the resultant material is now available for further characterization.

2.1.2 Synthesis of BYW: $Eu^{3+}$by CP method

In co-precipitation synthesis, to prepare BYW:$x$$Eu^{3+}$ (where, x = 0.0, 1.0, 5.0, 10.0, 15.0, and 20.0 mol%) phosphor, the precursors $Y(NO_3)_3{\cdot}6H_2O$, $Na_2WO_4{\cdot}2H_2O$, $Bi(NO_3)_3{\cdot}5H_2O$, and Eu_2O_3 were taken in stoichiometric ratio. The initial precursors $Y(NO_3)_3{\cdot}6H_2O$ and $Na_2WO_4{\cdot}2H_2O$ were dissolved in an appropriate amount of deionized water and two other precursors $Bi(NO_3)_3{\cdot}5H_2O$, Eu_2O_3 dissolved nitric acid. In the second step, all the individual solutions of the precursors were mixed and stirred for 1 h. Mixed solution was found to be pH ~ 1, which was further neutralized with ammonia solution to obtain pH ~ 7. Then, the transparent solution was stirred continuously for 3 h at 60 °C to convert into precipitate. After a certain time, the precipitate material was centrifuged for 30 min and obtained clay type sample, which was dried in an oven at 120 °C for 24 h to get powder form. The dried sample was ground for 10–15 min and calcined at different temperatures (700 °C–1000 °C). After calcination, the obtained materials are ready for further investigation.

2.1.3 Synthesis of BYW: Eu^{3+} by HT method

In hydrothermal synthesis, the initial precursors $Y(NO_3)_3{\cdot}6H_2O$, $Na_2WO_4{\cdot}2H_2O$, $Bi(NO_3)_3{\cdot}5H_2O$, and Eu_2O_3 were taken in stoichiometric ratio to obtain the expected phosphor powder of BYW:$x$$Eu^{3+}$ (x = 0.0 and 1.0 mol%). The initial precursors $Y(NO_3)_3{\cdot}6H_2O$ and $Na_2WO_4{\cdot}2H_2O$ were dissolved in the appropriate amount of deionized water and two other precursors $Bi(NO_3)_3{\cdot}5H_2O$, Eu_2O_3 dissolved in appropriate amount of nitric acid. All the dissolved precursors were mixed together to have pH that is nearly equal to 1, which was neutralized using the ammonia solution at 60 °C. After continuous stirring for 2 h on magnetic stirrer, the obtained sample was kept in an autoclave at 180 °C for 24 h with auto-generated pressure in a compact reactor. After cooling down to room temperature, the obtained sample was centrifuged several times with deionized water for 30 min and obtained white precipitated clay, which was dried in the oven at 120 °C for 24 h. Finally, the dried sample was calcined at different temperatures (700 °C–1000 °C) for further characterization process.

2.1.4 Synthesis of BYW: Eu^{3+} by SGC method

In sol–gel citrate combustion synthesis, the precursors used in the preparation of BYW:$x$$Eu^{3+}$ (x = 0.0 and 1.0 mol%) phosphor, such as $Y(NO_3)_3{\cdot}6H_2O$, $H_{42}N_{10}O_{42}W_{12}$, $Bi(NO_3)_3{\cdot}5H_2O$, citric acid, and ethylene glycol were taken in stoichiometric ratio. The precursors, $Y(NO_3)_3{\cdot}6H_2O$ and $Bi(NO_3)_3{\cdot}5H_2O$, were dissolved in the desired amount of nitric acid, whereas $H_{42}N_{10}O_{42}W_{12}$ and citric acid were dissolved in deionized water. All the metal nitrates were mixed continuously, stirred for 1 h, and then mixed with citric acid, followed by ethylene glycol. In this method, citric acid works as a fuel for combustion and also as a chelating agent, while ethylene glycol acts as a surfactant. After mixing of all the precursors, an ammonia solution is used to maintain a pH value at 7 and heated at 80 °C for the formation of gel. Further, the temperature of the magnetic stirrer was upsurged to 350 °C, leading to the self-ignition and combustion process. After combustion, the sample was ground to calcine at different temperatures in the range of 700 to 1000 °C to synthesize BYW:Eu^{3+} sample for further characterization.

2.2 Characterization instruments

Thermogravimetric analysis (TGA)–differential scanning calorimetry (DSC) of BYW powder sample were carried out by Setaram, (Model: LABSYS

evo) under inert atmosphere and heated from room temperature to 1000 °C. The phase as well as crystal structure of the synthesized samples have been identified at the scanning rate of 5°/minute with step size 0.04 using High-Resolution XRD (HR-XRD) (Rigaku-Smartlab) with Cu-Kα radiation (3 kilo Watt). The micrograph of the powder samples was perceived via FE-SEM (Zeiss, Model-Supra 55). For FE-SEM analysis, sample was coated with carbon material. Using a Xe lamp as the excitation source, luminescent characteristics of the samples were carried out using a spectrofluorophotometer (Shimadzu RF-5301PC) at room temperature. Edinburgh FLS900 was used for the decay measurements in which Xenon flash lamp is served as an excitation source.

3 Results and discussion

3.1 TGA–DSC investigation

The thermal behavior of $BiYWO_6$ with different synthesis routes was investigated by utilizing TGA and DSC profiles within the temperature range from 30 to 1000 °C, as shown in Fig. 1a–d. Figure 1a represents the thermal behavior of the precursor sample to be prepared via SSR method. The total weight loss from temperature 30 °C to 1000 °C is about 1.40%. First-stage weight loss 0.70% was observed from 30 to 500 °C due to the annealing process of the oxide compounds and adsorbed water molecules, which was also supported by the DSC curve. The observation of a plateau region reaching 700 °C signifies the end of volatile material evaporation. The peak observed at 725 °C in the SSR method in Fig. 1a is the indication

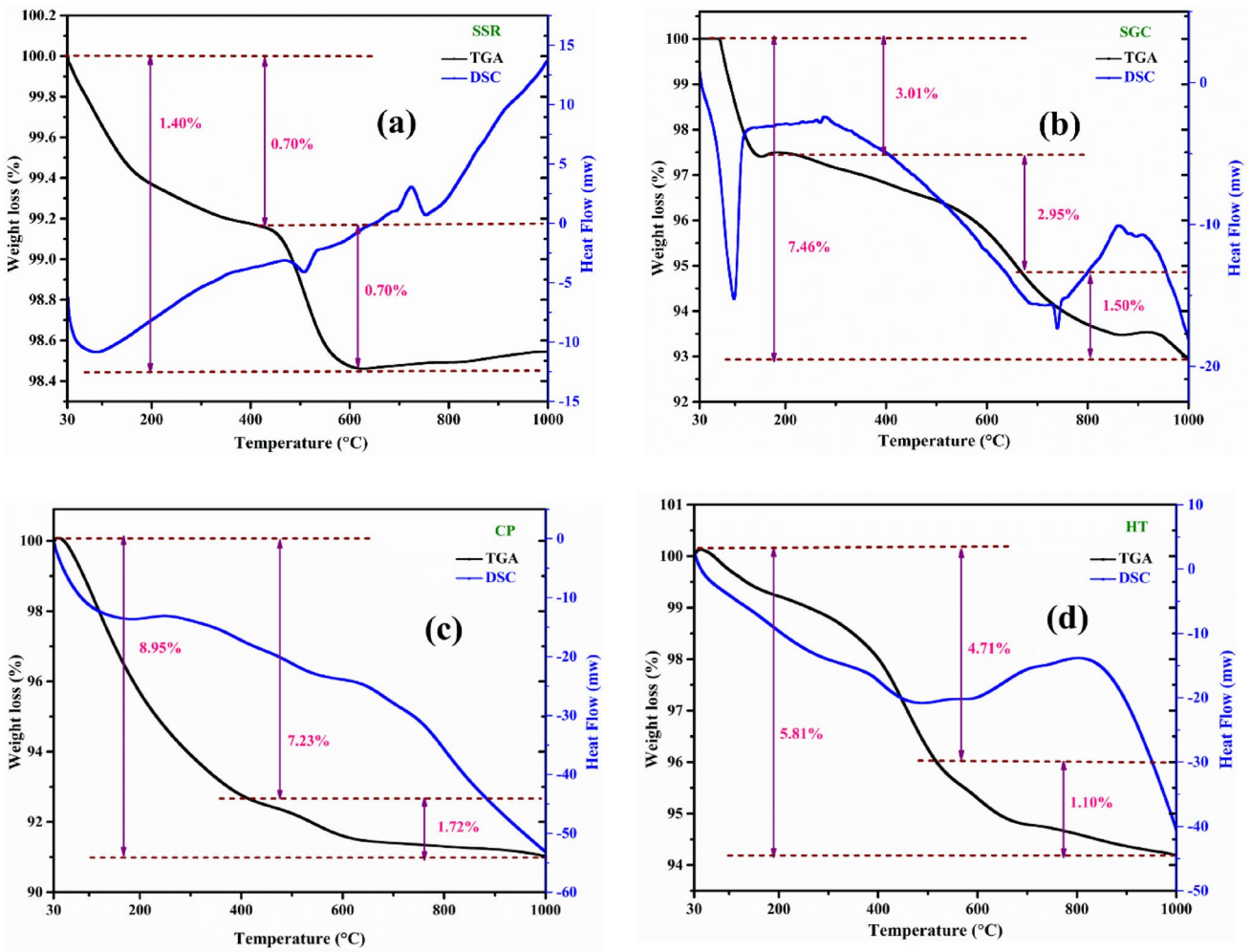


Fig. 1 TGA–DSC curves for $BiYWO_6$ powder materials synthesized via different routes **a** SSR **b** SGC **c** CP **d** HT

of the crystallization process for the $BiYWO_6$ sample. Figure 1b shows the TGA–DSC curve corresponding to the $BiYWO_6$ sample prepared via SGC method. The total weight loss was about 7.46% in the temperature range 30–1000 °C. In which, the first stage weight loss is about 3.01% around 30–150 °C, originated may be due to the evaporation of residual water and the corresponding DSC curve exhibits the endothermic peak. The second part of TGA curve exhibits weight loss of 2.95% in the temperature range 150–650 °C due to removal of some-NO_3 groups that could be related to one or more of the alkoxy nitrate species in the precursor present after the combustion [27].

Moreover, the mass loss was observed to be relatively low in the 650–1000 °C temperature range and is about 1.50% due to the pyrolysis of an organic compound and degradation of intermediate species formed during the combustion process [28]. This type of observation is reflected in the small endothermic peak at 740 °C of the DSC cure. After this temperature, the broad exothermic peak around 860 °C attributed to the crystallization process of the $BiYWO_6$ material. Figure 1c represents the thermal behavior of sample $BiYWO_6$ synthesized via CP method. The TGA plot exhibits a total weight loss of 8.95% in the temperature 30 to 1000 °C range. The total weight loss consists of two stages. The first stage represents 7.23% in the temperature range of 30–400 °C. The weight loss due to evaporation of absorbed moisture, ethanol, and ammonia content, which is correlated with the endothermic peak appeared within the 30 °C–200 °C range. The weight loss in the range of 200–600 °C, owing to the ejection of some nitrate group and ammonia decomposition, present in the precipitated material as small endothermic peaks reflected in the DSC curve. The second stage weight loss shows 1.72% in the temperature range of 400 °C–1000 °C, which is quite small in comparison to the first stage weight loss. Figure 1d presents the TGA–DSC curve of the $BiYWO_6$ sample prepared by employing HT method. The total weight loss consists of two stages in the temperature range 30–1000 °C that was found to be 5.81%. The first stage of weight loss was found to be 4.71% in the temperature range of 30–600 °C, due to the removal of absorbed moisture, ethanol, ammonia, and nitrate group present in the precursor, which is consistent with the small endothermic peak around 500 °C. The second stage weight loss was found to be about 1.10% in the temperature range of 400–1000 °C as reflected in the DSC curve related to a broad exothermic peak, which signifies the crystallization process of the hydrothermally prepared $BiYWO_6$ phosphor. As per the observed TGA/DSC curves, the $BiYWO_6$ sample was sintered at different temperatures (> 700 °C) to optimize the crystalline phase. The X-ray diffraction profiles recorded for all synthesized phosphors confirm the presence of pure phase in the material.

3.2 XRD structural analysis

Structural and phase purities of the prepared $BiYWO_6$ samples were analyzed with the help of diffraction patterns as shown in Fig. 2a–d. Figure 2a demonstrates the diffraction patterns of $BiYWO_6$ samples prepared via the SSR method calcinated at 1000 and 1100 °C. The diffraction pattern for $BiYWO_6$ prepared at 1100 °C was completely matched with the standard data without any impurity peak, while other diffraction patterns sintered at 1000 °C have little bit shifting of peaks as well as not analogous with the standard data.

Hence, the pure crystalline phase of the $BiYWO_6$ sample synthesized after calcination at 1100 °C via the SSR method has a monoclinic structure with the space group of P2/a (13) [29]. Figure 2b, c and d demonstrates XRD patterns for $BiYWO_6$ samples synthesized via SGC, CP, and HT methods, calcinated at numerous temperatures (i.e., 700, 800, 900, and 1000 °C), respectively. The samples initially exhibit broad crystalline peaks at 700 °C due to the formation of small crystallites and reduced fullwidth at half maxima (FWHM) value with increasing the calcination temperature beyond 700 °C due to enhancement in crystallite size as indicated in Table 1. Among all the patterns observed in Fig. 2b, c, and d, the diffraction peaks are completely analogous with the standard data (JCPDS-33-0224) when the samples were calcined at 900 °C.

Figure 2e represents the diffraction patterns of BYW: $x\mathrm{Eu}^{3+}$ (x = 1.0, 5.0, 20.0 mol%) samples synthesized via CP method at 900 °C calcination temperature. The ionic radii (Bi^{3+} = 1.03 Å, Y^{3+} = 0.90 Å, and Eu^{3+} = 0.947 Å) for coordination number (N = 6), which shows that the ionic radius of Eu^{3+} ion is smaller than the ionic radius of Bi^{3+} ion and having a similar ionic state. Hence, Eu^{3+} ions can effectively replace the Bi^{3+} ions usually. Further, it has been observed that Bi^{3+}/Y^{3+} have the same crystallite site in the crystal structure of $BiYWO_6$ derived from Bi_2WO_6 [15]. Thus, Eu^{3+} ions are expected to occupy preferably Bi^{3+} site, and there is a possibility of occupying Y^{3+} site in the case

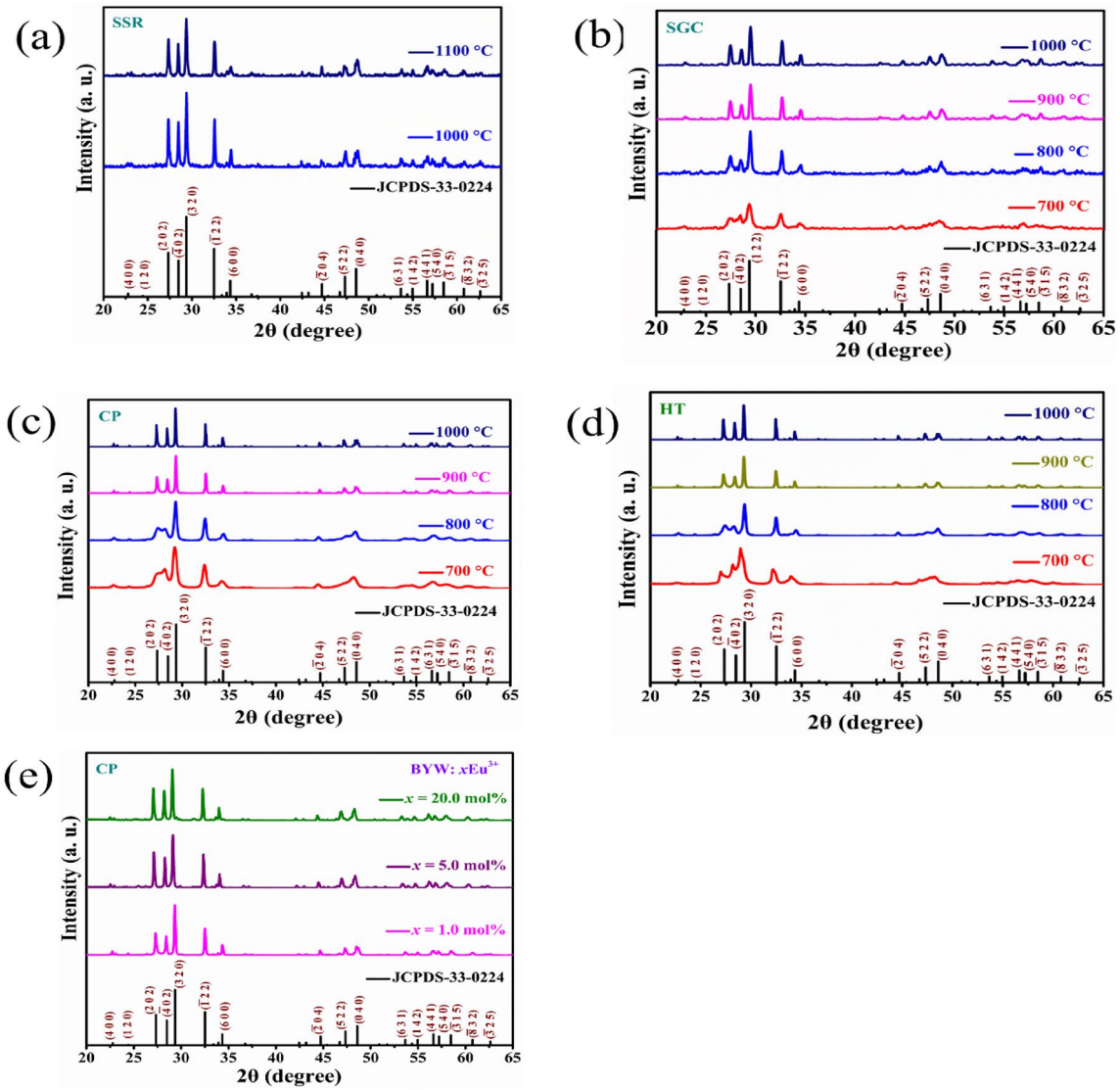


Fig. 2 X-ray diffraction patterns of $BiYWO_6$ via **a** SSR calcined at 1000 °C and 1100 °C **b** SGC route **c** CP route and **d** HT route sintered at various temperatures **e** Eu^{3+} (1.0, 5.0 and 20.0 mol%)-doped $BiYWO_6$ phosphor synthesized via CP route at 900 °C

Table 1 Crystallite size of BYW powder materials synthesized via different routes (SSR, CP, SGC, and HT methods)

Synthesis rotes	Crystallite size (nm)				
	700 °C	800 °C	900 °C	1000 °C	1100 °C
SSR	–	–	–	–	69.31
CP	12.44	14.21	46.15	60.10	-
HT	16.90	19.14	41.20	56.48	-
SGC	18.60	21.82	31.25	48.20	-

of higher doping concentrations of Eu^{3+} ions. All the patterns were compared with JCPDS data and are well matched with the observed peaks, which suggest the effective incorporation of Eu^{3+} ions in $BiYWO_6$ host sample. With increase in the concentration of the Eu^{3+} ion, the diffraction patterns slightly shifted towards lower angle due to increase in strain present in the crystal lattice [30, 31]. $BiYWO_6$ contains alternating Aurivillius-type $[BiYO_2]^{2+}$ layers inserted among $[WO_4]^{2-}$ perovskite layers. There are WO_6 octahedron connected and mixed edge chains. According to the described crystal structure, the Y^{3+}/Bi^{3+} ions are dispersed at random within the $BiYWO_6$ unit cell [32].

The average crystallite size (D) of the $BiYWO_6$ powder samples synthesized via various routes were calculated using Debye–Scherer's equation as follows [33]:

$$D = K\lambda/\beta Cos\theta, \quad (1)$$

where X-ray wavelength (λ = 1.54 Å Cu-Kα radiation), Bragg's diffraction angle (θ), shape factor (K = 0.94), and FWHM (β) were substituted in the above equation. The average crystallite size increased with the calcination temperature in various synthesis methods falls in the range from 12 to 70 nm.

Rietveld refinement has been performed via FullProf suite software for the prepared undoped and doped (1.0 mol% Eu^{3+}) $BiYWO_6$ samples. The resultant refinement patterns are shown in Fig. 3a and b, which depict the experimental and evaluated patterns along with difference between them. The assessed result discloses that the undoped and doped (1.0 mol% Eu^{3+}) $BiYWO_6$ samples having single monoclinic phase with the lattice parameters are a = 8.1244 Å, b = 3.7534 Å, c = 16.0345 Å and a = 8.1253 Å, b = 3.7539 Å, c = 16.0334 Å, respectively. The goodness of fit was found to be close to the unity denotes good refinement results. On the basis of refinement data, $BiYWO_6$ structure was prepared using VESTA software as demonstrated in Fig. 3c.

3.3 Morphological studies

The surface morphology of $BiYWO_6$ synthesized via different routes and calcinated at different temperatures with micro- to nanoresolution is shown in Fig. 4. The sample prepared via SSR gives the agglomerated, irregular shape and the average particle size (in micro range) of BYW at calcinated temperature 1100 °C as shown in Fig. 4a. The BYW sample synthesized via chemical routes clearly demonstrates the average particle size distribution in the nanorange (200–300 nm) as indicated in Fig. 4b, c, and d. The sample prepared via the SGC method displays inhomogeneous morphology as well as huge agglomeration when compared to the sample prepared using SSR, CP, and HT methods, which may be due to a calcination temperature of 900 °C [16]. In general, the SGC method is the function of the concentration of water and other reactants, including chelating agent and surfactant in the starting precursors, showing an improvement in grain growth of the morphology [34]. In comparison to the SSR, the reduced particle size and accumulation of particles have been observed while preparing the samples from SGC, HT, and CP methods. The micrographs also illustrate the non-uniform distribution of closely packed particles and the size of BYW particles enhanced with increasing calcination temperature as depicted in

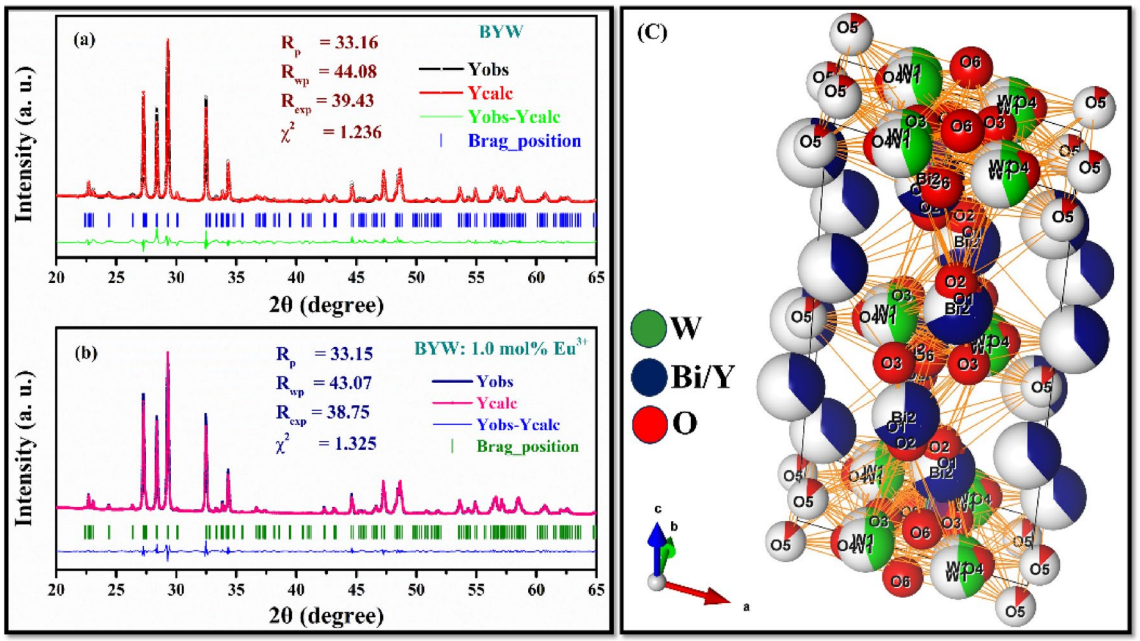


Fig. 3 The Rietveld refinement profiles of **a** $BiYWO_6$, **b** 1.0 mol% Eu^{3+}-doped $BiYWO_6$ phosphor synthesized via SSR route and **c** crystal structure of $BiYWO_6$

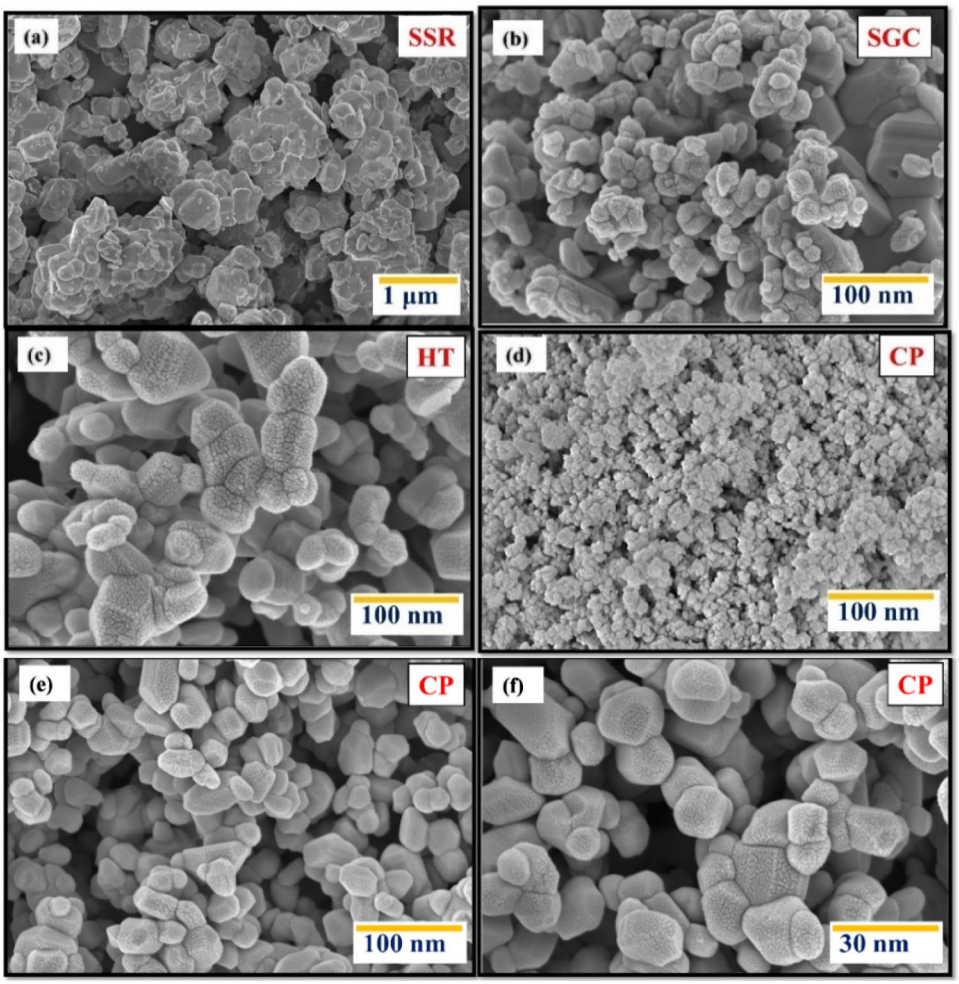


Fig. 4 FE-SEM images of 1.0 mol% Eu^{3+}-doped $BiYWO_6$ synthesized via **a** SSR at 1100 °C **b** SGC at 900 °C **c** HT at 900 °C and **d** CP at 700 °C and **e**, **f** CP at 900 °C with different magnifications

Fig. 4d, e and f, which was already reflected in the earlier section (XRD). The micro-to-nano meter crystalline particles may be appropriate to gain efficient luminescence and very much appropriate for solid-state lighting applications.

The EDS results fully disclose the uniform existence of numerous elements in the synthesized BYW phosphor via SSR and CP routes, calcined at 1100 °C and 900 °C, respectively, as indicated in Fig. 5a, b. The EDX profile confirms the existence of the compositional elements such as bismuth (Bi), yttrium (Y), tungsten (W), oxygen (O), and europium (Eu) in the prepared samples along with carbon (C) element.

The C element was observed because the sample was coated with carbon material for FE-SEM characterization. Additionally, the inset table of Fig. 5 represents the weight% and atomic% of each compositional element for the Eu^{3+}-doped $BiYWO_6$ sample, which confirms the presence all elements along with Eu-doping ions.

3.4 Photoluminescence characteristics

Figure 6a, b shows the excitation and emission spectrum of host material $BiYWO_6$ synthesized via solid-state reaction. Under 397 nm emission, the BYW host material shows a broad excitation band centered at 245 nm in the wavelength range from 225 to 260 nm, related to the $^1A_{1g} \rightarrow {}^1T_{1u}$ transition of WO_6^{-6}. The excitation peak at UV (245 nm) originated due to the host charge transfer band (CTB) of O^{2-}–W^{6+} in the W–O group and was used to measure emission profile of the BYW host sample [35, 36]. Under UV excitation, the BYW host material has been performed at ambient temperature in the wavelength range from 350 to 450 nm and a broad emission range has been observed at 362 to 440 nm, which is centered at 397 nm corresponding to the $^3T_{1u} \rightarrow {}^1A_{1g}$ transition of WO_6^{-6} group of the BYW host material [35, 36]. The results imply that the BYW host material exhibits intrinsic luminescence without any dopant.

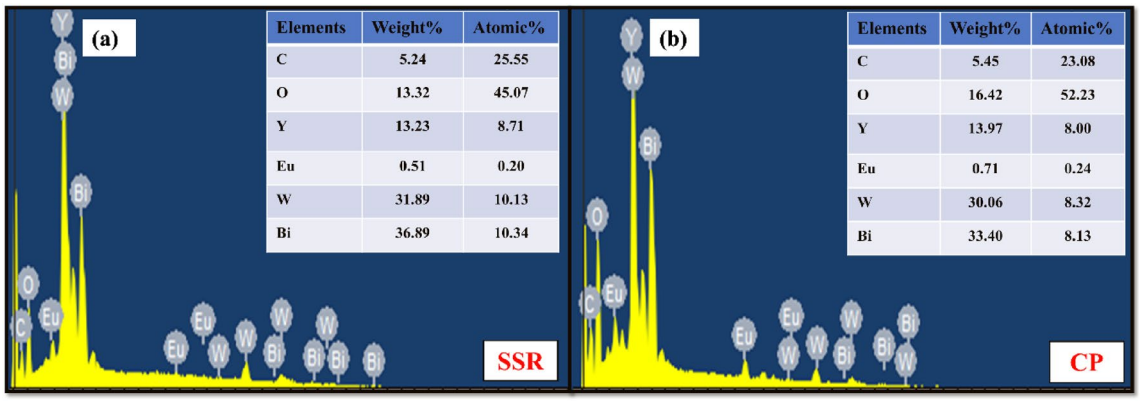


Fig. 5 EDX image of 1.0 mol % Eu^{3+}-doped $BiYWO_6$ (Inset shows the table of elemental composition) synthesized via **a** SSR at 1100 °C **b** CP at 900 °C

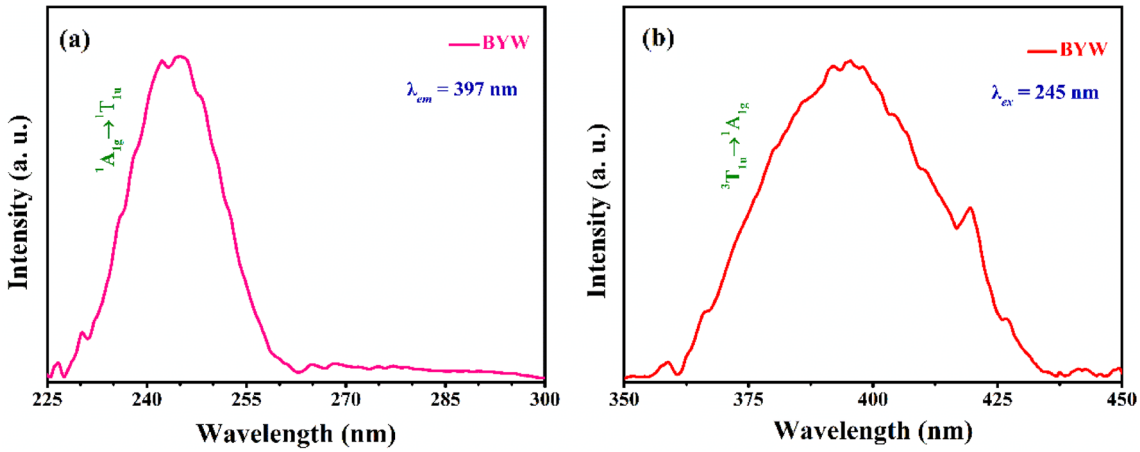


Fig. 6 PL **a** excitation **b** emission spectrum of host $BiYWO_6$

The PL characteristics of all Eu^{3+}-doped $BiYWO_6$ phosphors prepared via different synthesis routes were analyzed at room temperature. Figure 7a illustrates the PL excitation spectral profile of the BYW:xEu^{3+} (x = 1.0 mol%) phosphor synthesized via the SSR method at 1100 °C by monitoring the λ_{em} = 613 nm. The excitation spectrum exhibits various peaks in the range of 350–500 nm, with the most intense peak centered at 465 nm corresponding to the transition $^7F_0 \rightarrow {}^5D_2$ of Eu^{3+} ions. The inset plot clearly shows the other excitation peaks at 363, 382, 393, and 417, owing to the transitions from the $^7F_0$ level to numerous levels, as mentioned in Fig. 7a. An intense excitation wavelength at 465 nm is selected to record the emission spectra for all BYW:xEu^{3+} phosphors, which was well matched with the blue-emitting GaN-based LED chips. PL spectrum of BYW:xEu^{3+} (x = 1.0 mol%) phosphor consists of several peaks at λ_{ex} = 465 nm excitation, as shown in Fig. 7b. PL spectrum has five peaks situated at 579, 593, 613, 653, and 704 nm caused by the transition of ($^5D_0 \rightarrow {}^7F_0$), ($^5D_0 \rightarrow {}^7F_1$), ($^5D_0 \rightarrow {}^7F_2$), ($^5D_0 \rightarrow {}^7F_3$), and ($^5D_0 \rightarrow {}^7F_4$), respectively. The most intense emission peak situated at 613 nm is a forced electric dipole in nature corresponding to the transition $^5D_0 \rightarrow {}^7F_2$, which is hypersensitive and strongly influenced by the ligand ions in the crystal. On the other hand, the peak at 595 nm corresponds to the transition of $^5D_0 \rightarrow {}^7F_1$, showing the magnetic dipole in nature and regardless of the crystal environment of the host material [8, 37].

When Eu^{3+} occupies a site without inversion symmetry, the electric dipole transition ($^5D_0 \rightarrow {}^7F_2$) will become stronger than the magnetic dipole transition, whereas, if Eu^{3+} ions are placed at the

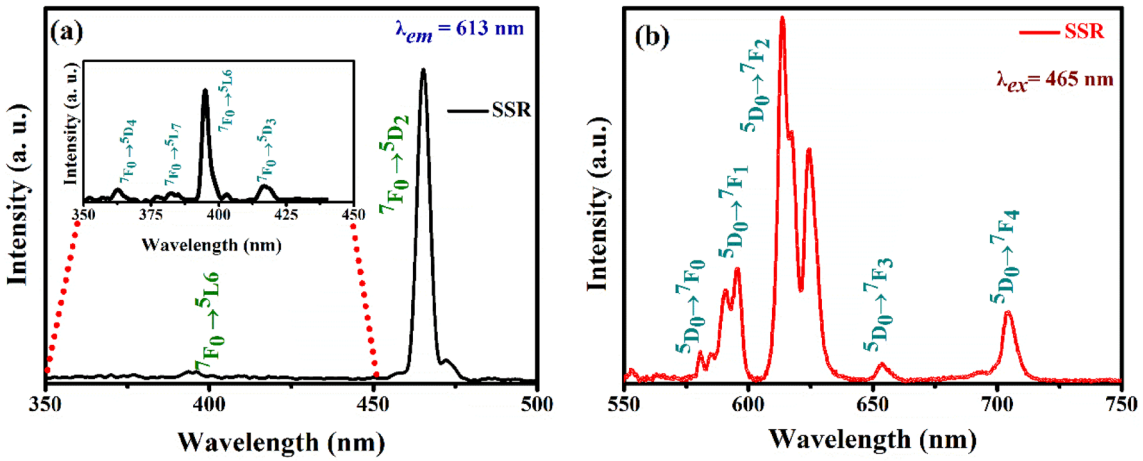


Fig. 7 **a** PLE spectrum of 1.0 mol% Eu^{3+}-doped $BiYWO_6$ phosphor. (The inset plot shows the magnifying of the PL excitation in the 350–450 nm wavelength range). **b** Emission spectra of 1.0 mol% Eu^{3+}-doped $BiYWO_6$ under an excitation wavelength of 465 nm

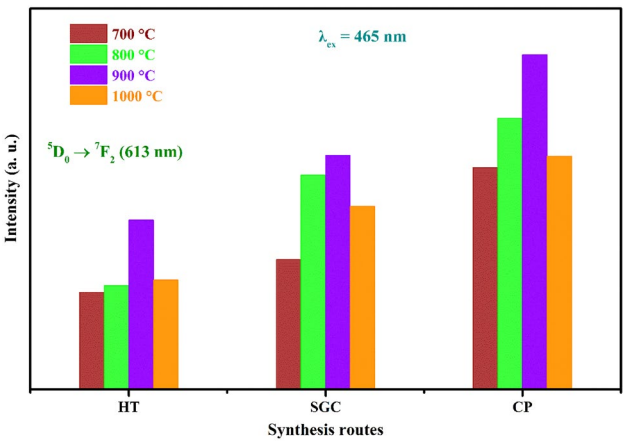


Fig. 8 Comparison of emission intensity of $^5D_0 \rightarrow {}^7F_2$ for 1.0 mol % Eu^{3+}-doped $BiYWO_6$, synthesized via CP, SGC, and HT at various sintering temperatures

inversion symmetry sites, the magnetic dipole transition ($^5D_0 \rightarrow {}^7F_1$) will predominate in the PL profile [38, 39]. In the BYW:xEu^{3+} phosphor, the electric dipole transition was stronger than the magnetic dipole, which discloses that the Eu^{3+} ions occupy without inversion symmetry site in the $BiYWO_6$ phosphor host.

For the comparison of the PL characteristics and to identify the best synthesis route along with the sintering temperature, BYW:xEu^{3+} phosphor (x = 1.0 mol%) was synthesized via HT, SGC, and CP method and sintered at different temperatures as depicted in Fig. 8. PL spectra reveal that the phosphor synthesized via CP method is efficient over the samples synthesized via SGC and HT method, and it exhibits the highest emission intensity at all sintering temperatures (i.e., 700, 800, 900, and 1000 °C).

Further, compared the PL characteristics of BYW:xEu^{3+} phosphor (x = 1.0 mol%) synthesized via CP method at various temperatures (700, 800, 900, and 1000 °C), as shown in Fig. 9a. The recorded PL spectra exhibit an increase in emission intensity with the rise in sintering temperature from 700 to 900 °C, and the intensity decreases for the phosphors sintered beyond 900 °C. Also, the inset plot of Fig. 9a discloses that the BYW:xEu^{3+} (x = 1.0 mol%) phosphor sintered at 900 °C has the highest emission intensity as compared with the phosphors sintered at 700, 800, and 1000 °C. On the basis of PL characteristics, it was concluded that the best way to synthesize BYW:xEu^{3+} phosphor is by employing the CP method at 900 °C sintering temperature.

Moreover, the PL spectra of the BYW:xEu^{3+} phosphor synthesized via CP method at 900 °C were compared with the same phosphor synthesized via SSR method at 1100 °C in Fig. 9b. The compared PL result indicates that the phosphor synthesized by CP method exhibits strongest emission and two times higher than the phosphor prepared by SSR method as shown in the inset of Fig. 9b. This is due to the CP chemical route producing the more homogeneous phosphor crystalline particle with uniform morphology, as seen in the FE-SEM micrograph. Accordingly, the CP approach has been chosen for further exploration, and it has been confirmed that this method is more efficient for producing BYW:xEu^{3+} phosphor with optimized emission intensity.

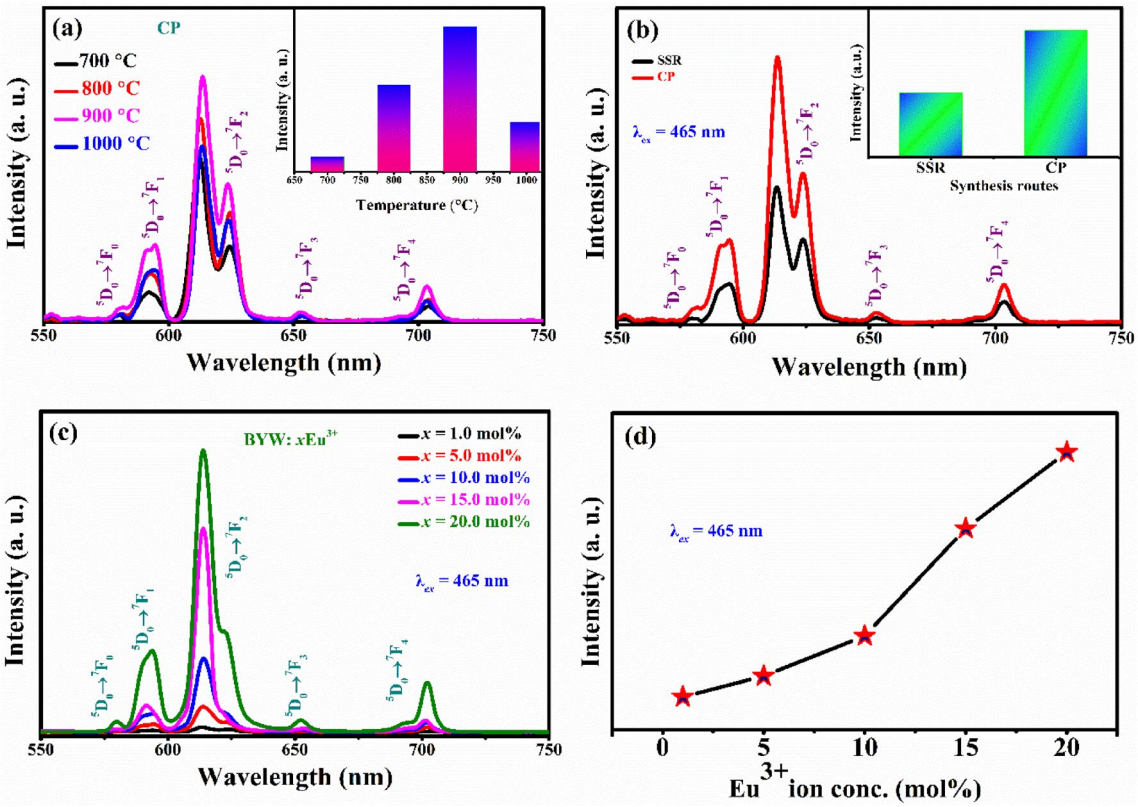


Fig. 9 **a** Comparison of emission spectra for 1.0 mol% Eu^{3+}-doped $BiYWO_6$ phosphor synthesized via co-precipitation route and sintered in the temperature range 700–1000 °C. **b** Comparison of emission spectra of BYW:Eu^{3+} phosphor synthesized via CP route at 900 °C and SSR route at 1100 °C. **c** Emission spectra of BYW:xEu^{3+} (x = 1.0, 5.0, 10.0, 15.0, and 20.0 mol %) synthesized via CP route and sintered at 900 °C. **d** The variation of emission intensity with varying the concentration of Eu^{3+} ions under 465 nm

To explore the effect of Eu^{3+} ion doping concentration on PL spectra of BYW:xEu^{3+} phosphor, PL spectra measured by varying dopant (Eu^{3+}) concentration from x = 1.0 to 20.0 mol% under an excitation wavelength of 465 nm as depicted in Fig. 9c. The emission intensity of the characteristic transitions of Eu^{3+} enhanced with an increase in activator ion concentration of Eu^{3+} from 1.0 to 20.0 mol% without any concentration quenching, as shown in Fig. 9d, which is another key feature of this host lattice having high rare-earth solubility.

To illustrate the luminescence phenomena involved in Eu^{3+}-doped $BiYWO_6$ phosphor, the simplified energy level diagram is displayed in Fig. 10. The population of Eu^{3+} ions finally relaxes to the $^5D_0$ level via non-radiatively after being excited by blue light at 465 nm (to the $^5D_2$ level) and subsequently exhibits radiative emission in the visible region starting from the $^5D_0$ level.

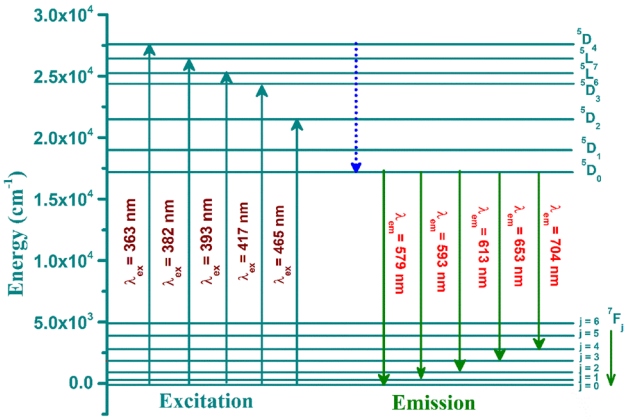


Fig. 10 Energy level diagram of Eu^{3+} in $BiYWO_6$ phosphor

In order, to further analyze the luminescence performance of BYW: xEu^{3+}, x = 20.0 mol% phosphor, its luminescence quantum yield has been assessed using

a $BaSO_4$-integrated sphere [40, 41]. The PL quantum yield for the BYW: xEu^{3+}, x = 20.0 mol% phosphor is 51.39%, which is higher than the previously reported phosphors [42, 43].

3.5 CIE coordinates

CIE color coordinates have been estimated using the recorded PL spectra of BYW: xEu^{3+} phosphor (x = 20.0 mol%). Figure 11 represents the CIE chromaticity diagram of the synthesized phosphor via CP method at 900 °C. The estimated coordinates (0.654, 0.345) were situated in the red region of the CIE diagram, which is very close to the standard data (National Television System Committee) for red light (0.67, 0.33) and comparable with the commercial red-emitting phosphor Y_2O_3:Eu^{3+} (0.645, 0.347) and Y_2O_2S:Eu^{3+} (0.647, 0.343) [44].

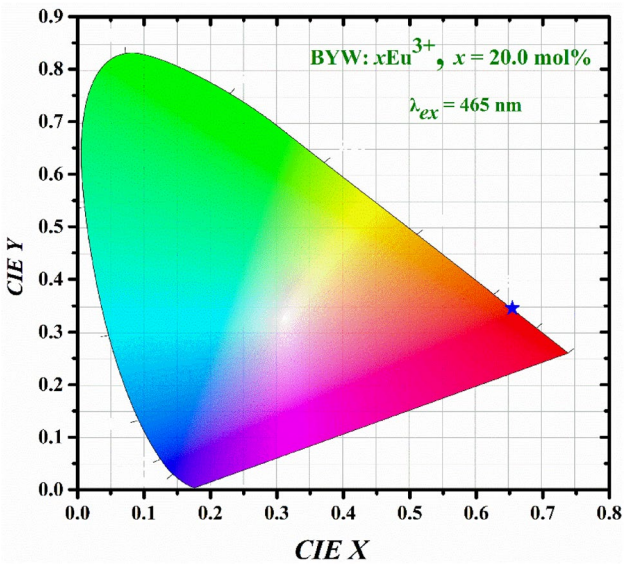


Fig. 11 CIE color coordinates of 20.0 mol% Eu^{3+}-doped $BiYWO_6$ phosphor synthesized via CP route

Further, the color purity was estimated on the basis of chromaticity coordinates for the same phosphor via the following equation:

$$\text{ColorPurity} = \sqrt{\frac{(x - x_{ee})^2 + (y - y_{ee})^2}{(x_d - x_{ee})^2 + (y_d - y_{ee})^2}}, \quad (2)$$

where, (x, y), (x_d, y_d), and (x_{ee}, y_{ee}) specify the CIE coordinates of the phosphor, dominant wavelength point, and standard white point, respectively. The color purity for BYW:xEu^{3+} phosphor (x = 20.0 mol%) was found to be greater than 99.0%, which is comparatively higher than the reported values in Table 2.

The aforesaid results suggested that BYW: xEu^{3+} phosphors could be used as an effective red-emitting constituent for the development of w-LEDs and other photonic device applications.

3.6 Lifetime analysis

By altering the Eu^{3+} ion concentration in $BiYWO_6$ phosphor at room temperature, the decay curves were measured in order to calculate the average lifetime for the excited level $^5D_0$, under 465 nm excitation as shown in Fig. 12. The decay curves were fitted using different exponential equations, and it was observed that the best fit was achieved with a bi-exponential function as depicted in Fig. 12. The bi-exponential equation is given as follows:

$$I(t) = I_0 + A_1\exp(-t/\tau_1) + A_2\exp(-t/\tau_2). \quad (3)$$

In this case, I(t) and I_0 stand for the luminescence intensities at time t and 0, respectively. τ_1 and τ_2 stand for exponential mechanism's quick and slow decay lifetimes, respectively. A_1 and A_2 are fitting constants. The average lifetime $\left(\tau_{avg}\right)$ values have been computed with the help of the following expression [50]:

Table 2 CIE coordinates and color purity of some reported Eu^{3+}-doped red phosphors

Sample	Color coordinates (x, y)	Color purity	Reference
Ca_2GdTaO_6:0.4Eu^{3+}	(0.6676, 0.332)	96.0%	[45]
$Na_5W_3O_9F_5$:0.25Eu^3	(0.6467, 0.3532)	97.44%	[46]
$Y_2Si_8WO_8$:0.12Eu^{3+}	(0.665, 0.335)	96.6%	[47]
$BaY_4Si_3O_{13}$:Eu^{3+}	(0.636, 0.362)	91.56%	[48]
Y_2O_3:Eu^{3+}	(0.61, 0.39)	83.0%	[18]
Y_2O_2S:Eu^{3+}	(0.636, 0.321)	71.0%	[49]
$BiYWO_6$:0.2Eu^{3+}	(0.654, 0.345)	99.0%	This work

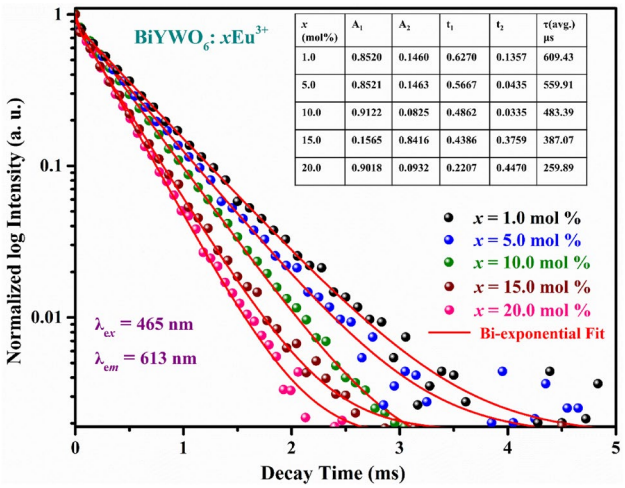


Fig. 12 Decay curves for Eu^{3+}-doped $BiYWO_6$ phosphors

$$\tau_{avg} = \frac{A_1\tau_1^2 + A_2\tau_2^2}{A_1\tau_1 + A_2\tau_2}. \quad (4)$$

The observed values of the slow lifetime component (τ_1), fast lifetime component (τ_2), and average lifetime (τ_{avg}) for the synthesized BYW:xEu^{3+} phosphor are tabulated in inset of Fig. 12. The τ_{avg} values of the synthesized BYW:xEu^{3+} phosphor (x = 1.0 to 20.0 mol%) were observed in the range of 609.43 to 259. 89 μs. The average lifetime values of the synthesized phosphor decline with an increase in the concentration of the dopant ions (Eu^{3+}). This decline in the average lifetime values is generally owing to a decrement in the distance between dopant ions (Eu^{3+}), which results in energy transfer from donor ions to acceptor ions.

4 Conclusions

In summary, various methods have been used to synthesize europium-activated $BiYWO_6$ phosphors, including SSR, SGC, CP, and HT. The investigation is focused on examining the thermal, structural, morphological, and luminescent characteristics of the as-synthesized phosphors. The crystalline monoclinic phase purity of Eu^{3+}-doped $BiYWO_6$ has been confirmed using X-ray diffraction. Rietveld refinement analysis also authorizes the formation of a single monoclinic phase of the BYW host and BYW:xEu^{3+}, x = 1.0 mol% phosphor. Under blue excitation (465 nm), all the synthesized Eu^{3+}-doped BYW phosphor samples exhibit intense emission peak in the red region. The BYW:Eu^{3+} phosphor sample attributed to the strong emission intensity when synthesized by CP technique at 900 °C and considered as an optimized synthesized procedure for this phosphor. The CIE chromaticity color coordinates were estimated for the phosphor synthesized via the optimized CP method, and it was concluded that the coordinates of all the samples were positioned in the red region along with high values of color purity. Furthermore, the PL quantum yield for the BYW:xEu^{3+}, x = 20.0 mol% phosphor has been observed to be 51.39%, which is higher than the reported one. The photoluminescence decay curve exhibits bi-exponential in nature for the optimized BYW phosphor with an average decay time in the microsecond range. The obtained results strongly recommend that Eu^{3+}-activated BYW phosphor can act as a blue excited potential red-illuminating participant for the applications in pc-converted w-LEDs, and other photonic devices.

Acknowledgements

The Authors (Mr. Indrajeet Maurya) are grateful to Delhi Technological University, Delhi, India for providing junior research fellowship.

Author contributions

Mr. Indrajeet Maurya: conceptualization, formal analysis, investigation, methodology, software, writing—original draft. Dr. M. Jayasimhadri: supervision, formal analysis, methodology, visualization, validation, writing—review and editing.

Data availability

Data will be made available on reasonable request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. C.M. Mehare, G. Mishra, N.S. Dhoble, S.J. Dhoble, Improvement of self-activated luminescence properties of $Ca_2KZn_2(VO_4)_3$ down-conversion materials by SSR method based on co-doped Eu^{3+}, Dy^{3+} rare earth ions concentrations. J. Mol. Struct. **1264**, 133250 (2022). https://doi.org/10.1016/j.molstruc.2022.133250
2. P. Dixit, V. Chauhan, P. Kumar, P.C. Pandey, Improvement in luminescence of thermally stable $CaMoO_4$:Tb^{3+} green phosphor by Bi^{3+} ions. Mater. Chem. Phys. **305**, 127913 (2023). https://doi.org/10.1016/j.matchemphys.2023.127913
3. V. Chauhan, P.K. Pandey, P. Dixit, P. Deshmukh, S. Satapathy, P.C. Pandey, Effect of Zn^{2+} co-doping on the luminescence of Sm^{3+} doped $SrMoO_4$ phosphor. J. Lumin. **248**, 118994 (2022). https://doi.org/10.1016/j.jlumin.2022.118994
4. A. Raja, G. Annadurai, V. Sivasubramani, K. Ramachandran, R. Kamesh, V. Govindan, P. Vijayakumar, V. Siva, P. Ramasamy, Synthesis, crystal structure, and luminescence properties of $RbCaF_3$: Eu^{3+} orange-red emitting phosphors for white light emitting diodes. Ceram. Int. **49**, 11393–11401 (2023). https://doi.org/10.1016/j.ceramint.2022.11.338
5. G.L. Bhagyalekshmi, D.N. Rajendran, Luminescence dynamics of Eu^{3+} activated and co-activated defect spinel zinc titanate nanophosphor for applications in WLEDs. J. Alloys Compd. **850**, 156660 (2021). https://doi.org/10.1016/j.jallcom.2020.156660
6. A.R. Kadam, S.J. Dhoble, Synthesis and luminescence study of Eu^{3+}-doped $SrYAl_3O_7$ phosphor. Luminescence **34**, 846–853 (2019). https://doi.org/10.1002/bio.3681
7. D. Haranath, P. Sharma, H. Chander, Optimization of boric acid content in developing efficient blue emitting, long persistent phosphor. J. Phys. D Appl. Phys. **38**, 371 (2005). https://doi.org/10.1088/0022-3727/38/3/003
8. A.K. Vishwakarma, M. Jayasimhadri, Significant enhancement in photoluminescent properties via flux assisted Eu^{3+} doped $BaNb_2O_6$ phosphor for white LEDs. J. Alloys Compd. **683**, 379–386 (2016). https://doi.org/10.1016/j.jallcom.2016.05.052
9. Z. Chen, Y. Pan, L. Xi, R. Pang, S. Huang, G. Liu, Tunable yellow-red photoluminescence and persistent afterglow in phosphors $Ca_4LaO(BO_3)_3$: Eu^{3+} and $Ca_4EuO(BO_3)_3$. Inorg. Chem. **55**, 11249–11257 (2016). https://doi.org/10.1021/acs.inorgchem.6b01786
10. K.N. Kumar, L. Vijayalakshmi, P. Hwang, A.D. Wadhwani, J. Choi, Bright red-luminescence of Eu^{3+}ion-activated $La_{10}W_{22}O_{81}$ microphosphors for noncytotoxic latent fingerprint imaging. J. Alloys Compd. **840**, 155589 (2020). https://doi.org/10.1016/j.jallcom.2020.155589
11. R. Nagaraj, A. Raja, S. Ranjith, Synthesis and luminescence properties of novel red-emitting Eu^{3+} ions doped silicate phosphors for photonic applications. J. Alloys Compd. **827**, 154289 (2020). https://doi.org/10.1016/j.jallcom.2020.154289
12. Z. Xia, W. Zhou, H. Du, J. Sun, Synthesis and spectral analysis of Yb^{3+}/Tm^{3+}/Ho^{3+}-doped $Na_{0.5}Gd_{0.5}WO_4$ phosphor to achieve white upconversion luminescence. Mater. Res. Bull. **45**, 1199–1202 (2010). https://doi.org/10.1016/j.materresbull.2010.05.018
13. G.N. Rocha, L.F.L. Melo, P.B.A. Fechine, P.V.S. Silva, M.A.S. Da Silva, A.S.B. Sombra (2011) Dielectric resonator antennas based in $BiYWO_6$ and operating at 33 GHz: Electrical properties study. In 2011 SBMO/IEEE MTT-S international microwave and optoelectronics conference. https://doi.org/10.1109/IMOC.2011.6169300
14. H. Ait Ahsaine, M. Ezahri, A. Benlhachemi, B. Bakiz, S. Villain, J.C. Valmalette, F. Guinneton, M. Arab, J.R. Gavarri, Structural, vibrational study and UV photoluminescence properties of the system $Bi_{(2-x)}Lu_{(x)}WO_6$ ($0.1 \leq x \leq 1$). RSC Adv. **5**, 96242–96252 (2015). https://doi.org/10.1039/c5ra19424e
15. S. Pasternak, D. Levy, Y. Paz, B. Pokroy, Structural analysis and optical properties of the $Bi_{2-x}Y_xWO_6$ system. CrystEngComm **18**, 6464–6470 (2016). https://doi.org/10.1039/c6ce01094f
16. B.V. Ratnam, M.K. Sahu, A.K. Vishwakarma, K. Jha, H.J. Woo, K. Jang, M. Jayasimhadri, Optimization of synthesis technique and luminescent properties in Eu^{3+}-activated $NaCaPO_4$ phosphor for solid state lighting applications. J. Lumin. **185**, 99–105 (2017). https://doi.org/10.1016/j.jlumin.2017.01.006
17. H. Kaur, M. Jayasimhadri, Optimization of structural and luminescent properties with intense red emitting thermally stable Sm^{3+} doped $CaBiVO_5$ phosphors for w-LED applications. Opt. Mater. **107**, 110119 (2020). https://doi.org/10.1016/j.optmat.2020.110119
18. S. Som, S. Das, S. Dutta, H.G. Visser, M.K. Pandey, P. Kumar, R.K. Dubey, S.K. Sharma, Synthesis of strong red emitting Y_2O_3: Eu^{3+} phosphor by potential chemical routes: comparative investigations on the structural evolutions, photometric properties and Judd-Ofelt analysis. RSC Adv. **5**, 70887–70898 (2015). https://doi.org/10.1039/c5ra13247a
19. 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). https://doi.org/10.1111/jace.16479

20. N. Xue, X. Fan, Z. Wang, M. Wang, Synthesis process and luminescence properties of Ln^{3+} doped $NaY(WO_4)_2$ nanoparticles. Mater. Lett. **61**, 1576–1579 (2007). https://doi.org/10.1016/j.matlet.2006.07.082
21. H. Gu, L. Yu, J. Wang, M. Ni, T. Liu, F. Chen, Tunable luminescence and enhanced photocatalytic activity for Eu(III) doped Bi_2WO_6 nanoparticles. Spectrochim. Acta Part A Mol. Biomol. Spectrosc. **177**, 58–62 (2017). https://doi.org/10.1016/j.saa.2017.01.034
22. R. Hao, Q. Meng, W. Liu, H. Liu, Synthesis and luminescent properties of Eu^{3+} doped Y_2WO_6 nanophosphors. J. Rare Earths **31**, 864–870 (2013). https://doi.org/10.1016/S1002-0721(12)60371-8
23. B. Yan, F. Lei, Molten salt synthesis, characterization and luminescence of ZnWO4: Eu^{3+} nanophosphors. J. Alloys Compd. **507**, 460–464 (2010). https://doi.org/10.1016/j.jallcom.2010.07.203
24. H. Zheng, B. Chen, H. Yu, J. Zhang, J. Sun, X. Li, M. Sun, B. Tian, S. Fu, H. Zhong, B. Dong, R. Hua, H. Xia, Microwave-assisted hydrothermal synthesis and temperature sensing application of Er^{3+}/Yb^{3+} doped $NaY(WO_4)_2$ microstructures. J. Colloid Interface Sci. **420**, 27–34 (2014). https://doi.org/10.1016/j.jcis.2013.12.059
25. R. Kamal, M. Saif, Barium tungstate doped with terbium ion green nanophosphor: low-temperature preparation, characterization, and potential applications. Spectrochim. Acta Part A Mol. Biomol. Spectrosc. **229**, 117928 (2020). https://doi.org/10.1016/j.saa.2019.117928
26. J. Huang, W. You, G. Gong, G. Liu, P. Liu, B. Wang, Electronic structure and photoluminescence of Dy^{3+} single-doped and Dy^{3+}/Tm^{3+} co-doped $NaBi(WO_4)_2$ phosphors. Opt. Mater. **88**, 534–539 (2019). https://doi.org/10.1016/j.optmat.2018.12.029
27. A. Boukerika, L. Guerbous, N. Brihi, Ce-doped YAG phosphors prepared via sol-gel method : effect of some modular parameters. J. Alloys Compd. **614**, 383–388 (2014). https://doi.org/10.1016/j.jallcom.2014.06.133
28. X. He, L. Xiaofang, R. Li, B. Yang, K. Yu, M. Zeng, R. Yu, Effects of local structure of Ce^{3+} ions on luminescent properties of $Y_3Al_5O_{12}$: Ce^{3+} nanoparticles. Sci. Rep. **6**, 1–11 (2016). https://doi.org/10.1038/srep22238
29. H. Liu, J. Yuan, W. Shangguan, Y. Teraoka, Visible-light-responding $BiYWO_6$ solid solution for stoichiometric photocatalytic water splitting. J. Phys. Chem. C **112**, 8521–8523 (2008). https://doi.org/10.1021/jp802537u
30. A. Oudhia, N. Shukla, P. Bose, R. Lalwani, A. Choudhary, Effect of various synthesis protocols on the doping profile of ZnO: Eu nanowires. Nano Struct. Nano Obj. **7**, 69–74 (2016). https://doi.org/10.1016/j.nanoso.2016.05.002
31. A. Dwivedi, M. Srivastava, A. Srivastava, C. Upadhyay, S.K. Srivastava, Tunable photoluminescence and energy transfer of Eu^{3+}, Ho^{3+}-doped $Ca_{0.05}Y_{1.93-x}O_2$ nanophosphors for warm white LEDs applications. Sci. Rep. **12**, 1–15 (2022). https://doi.org/10.1038/s41598-022-09630-x
32. S. Mamidi, R. Gundeboina, S. Kurra, R. Velchuri, V. Muga, Aurivillius family of layered perovskites, $BiREWO_6$ (RE = La, Pr, Gd, and Dy): synthesis, characterization, and photocatalytic studies. Comptes Rendus Chim. **21**, 547–552 (2018). https://doi.org/10.1016/j.crci.2018.01.011
33. B.P. Singh, A.K. Parchur, R.K. Singh, A.A. Ansari, P. Singh, S.B. Rai, Structural and up-conversion properties of Er^{3+} and Yb^{3+} co-doped $Y_2Ti_2O_7$ phosphors. Phys. Chem. Chem. Phys. **15**, 3480–3489 (2013). https://doi.org/10.1039/c2cp44195k
34. D.V. Deyneko, I.V. Nikiforov, D.A. Spassky, P.S. Berdonosov, P.B. Dzhevakov, B.I. Lazoryak, $Sr_8MSm_{1-x}Eu_x(PO_4)_7$ phosphors derived by different synthesis routes: Solid state, sol-gel and hydrothermal, the comparison of properties. J. Alloys Compd. **887**, 161340 (2021). https://doi.org/10.1016/j.jallcom.2021.161340
35. L. Li, J. Chen, X. Peng, S. Xu, R. Wei, H. Guo, Energy transfer and luminescent properties of Lu_2WO_6: Bi^{3+}, Eu^{3+} red emission phosphor. Ceram. Int. **49**, 25806–25814 (2023). https://doi.org/10.1016/j.ceramint.2023.05.127
36. M.N. Huang, Y.Y. Ma, F. Xiao, Q.Y. Zhang, Bi^{3+} sensitized Y_2WO_6: Ln^{3+} (Ln = Dy, Eu, and Sm) phosphors for solar spectral conversion. Spectrochim. Acta Part A Mol. Biomol. Spectrosc. **120**, 55–59 (2014). https://doi.org/10.1016/j.saa.2013.09.147
37. A.K. Vishwakarma, K. Jha, M. Jayasimhadri, A.S. Rao, K. Jang, B. Sivaiah, D. Haranath, Red light emitting $BaNb_2O_6$: Eu^{3+} phosphor for solid-state lighting applications. J. Alloys Compd. **622**, 97–101 (2015). https://doi.org/10.1016/j.jallcom.2014.10.016
38. Q. Ma, A. Zhang, M. Lü, Y. Zhou, Z. Qiu, G. Zhou, Novel class of aeschynite structure $LaNbTiO_6$-based orange-red phosphors via a modified combustion approach. J. Phys. Chem. B **111**, 12693–12699 (2007). https://doi.org/10.1021/jp0739162
39. Y.C. Chang, C.H. Liang, S.A. Yan, Y.S. Chang, Synthesis and photoluminescence characteristics of high color purity and brightness $Li_3Ba_2Gd_3(MoO_4)_8$: Eu^{3+} red phosphors. J. Phys. Chem. C **114**, 3645–3652 (2010). https://doi.org/10.1021/jp9084124
40. Y. Hua, W. Ran, J.S. Yu, Advantageous occupation of Europium(III) in the B site of double-perovskite $Ca_2BB'O_6$(B = Y, Gd, La; B′ = Sb, Nb) frameworks for white-light-emitting diodes. ACS Sustain. Chem. Eng. **9**,

7960–7972 (2021). https://doi.org/10.1021/acssuschemeng.1c02216

41. Y.D. Ma, X.S. Peng, M.Z. Fei, W.N. Zhang, L.M. Teng, F.F. Hu, R.F. Wei, H. Guo, Adjustable white luminescence and high thermal stability in Eu^{2+}/Eu^{3+}/Tb^{3+}/Al co-doped aluminosilicate oxyfluoride glass. J. Alloys Compd. **846**, 156435 (2020). https://doi.org/10.1016/j.jallcom.2020.156435
42. Y. Zhang, J. Xu, Q. Cui, B. Yang, Eu^{3+} -doped $Bi_4Si_3O_{12}$ red phosphor for solid state lighting: microwave synthesis, characterization, photoluminescence properties and thermal quenching mechanisms. Sci. Rep. **7**, 1–12 (2017). https://doi.org/10.1038/srep42464
43. J. Xiang, J. Yang, N. Luo, J. Zhu, S. Huang, Y. Mao, Optimized photoluminescence and electronic properties of europium doped phosphate red phosphor. Results Phys. **13**, 102258 (2019). https://doi.org/10.1016/j.rinp.2019.102258
44. S. Qi, D. Wei, Y. Huang, S. Il Kim, Y.M. Yu, H.J. Seo, Microstructure of Eu^{3+}-doped perovskites-type niobate ceramic $La_3Mg_2NbO_9$. J. Am. Ceram. Soc. **97**, 501–506 (2014). https://doi.org/10.1111/jace.12656
45. S. Wang, Q. Sun, B. Devakumar, J. Liang, L. Sun, X. Huang, Novel highly efficient and thermally stable Ca_2GdTaO_6: Eu^{3+} red-emitting phosphors with high color purity for UV/blue-excited WLEDs. J. Alloys Compd. **804**, 93–99 (2019). https://doi.org/10.1016/J.JALLCOM.2019.06.388
46. G. Wu, J. Xue, X. Li, Q. Bi, M. Sheng, Z. Leng, A novel red-emitting $Na_5W_3O_9F_5$: Eu^{3+} phosphor with high color purity for blue-based WLEDs. Ceram. Int. **49**, 10615–10624 (2023). https://doi.org/10.1016/j.ceramint.2022.11.249
47. S. Wang, C. Xu, X. Qiao, High thermal stability and color purity of red-emitting phosphor Y_2SiWO_8: Eu^{3+} for w-LEDs: synthesis and photoluminescence properties. Ceram. Int. **47**, 1063–1075 (2021). https://doi.org/10.1016/j.ceramint.2020.08.222
48. Q. Yu, J. Liu, P. Zhang, S. Hao, C. Zhu, C. Yang, Preparation and luminescence of a novel red phosphor $BaY_4Si_3O_{13}$: Eu^{3+} and effect of Li_2CO_3 doping. Opt. Mater. **149**, 115068 (2024). https://doi.org/10.1016/J.OPTMAT.2024.115068
49. Z. Wang, Y. Wang, J. Zhang, Y. Lu, The photoluminescence properties of Eu^{3+}, Bi^{3+} co-doped yttrium oxysulfide phosphor under vacuum ultraviolet excitation. Mater. Res. Bull. **44**, 1183–1187 (2009). https://doi.org/10.1016/j.materresbull.2008.11.003
50. V. Sangwan, M. Jayasimhadri, D. Haranath, Colour-tunable features in thermally stable Tb^{3+}/Eu^{3+} co-doped telluro tungstate glasses for photonic applications. J. Phys. D Appl. Phys. **57**, 195301 (2024). https://doi.org/10.1088/1361-6463/ad2475

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

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

Journal of Integrated
SCIENCE & TECHNOLOGY

Computation studies of phytocompounds of *Papaver somniferum* and *Boswellia serrata* for diabetes management

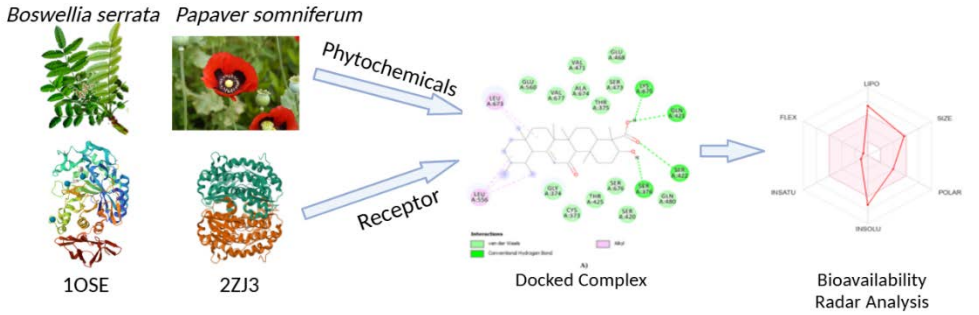
Madhura Roy[1+], Garvit Satija[2+], Ritika Luthra[3], Shreeja Datta[4], Arpita Roy[5*], Mohamed A.M. Iesa[6], Sarvesh Rustagi[7], Devvret Verma[8], Vaseem Raja[9], Cheng Wan Hee[10], Humaira Jan[11], Swetha Raj[12]

[1]Centre for Translational and Clinical Research, School of Chemical and Life Sciences, Jamia Hamdard University, India. [2]School of Pharmaceutical Education and Research, Jamia Hamdard University, Delhi, India. [3]Department of Chemical and Biomolecular Engineering, New York University Tandon School of Engineering, New York, USA. [4]Department of Biotechnology, Delhi Technological University, Delhi, India. [5]Department of Biotechnology, School of Engineering & Technology, Sharda University, Greater Noida, India. [6]Umm Al-Qura University, Faculty of Medicine at Al-Qunfudah, Department of Physiology. [7]School of Applied and Life Sciences, Uttaranchal University Dehradun, Dehradun, Uttarakhand, India.[8]Department of Biotechnology, Graphic Era Deemed to be Unversity, Dehradun, India 248002.[9]Department of Biotechnology, University Centre for Research and Development, Chandigarh University Gharuan, Mohali, Punjab- India,140413. [10]Faculty Health and Life Sciences, INTI International University, Persiaran Perdana BBN, Putra Nilai, Nilai 71800, Negeri Sembilan, Malaysia [11]School of Biological Sciences, University of Kashmir, Hazratbal, Srinagar, 190006, Kashmir, India. [12]Center for Global Health Research, Saveetha Medical College and Hospitals, Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai, India

Received on: 28-Aug-2024, Accepted and Published on: 15-Nov-2024

Article

ABSTRACT



Natural bioactive compounds derived from medicinal plants have been extensively studied for their anti-diabetic properties. *Papaver somniferum* and *Boswellia serrata* are one of important medicinal plants that have bioactive compounds which could be utilized against diabetes. Therefore, an in-silico investigation of their function as anti-diabetic agents was conducted in this study, against two receptors i.e. alpha-amylase and glutamine-fructose-6-phosphate transaminase 1 (GFAT-1). The phytocompound-receptor docked complex was evaluated with known inhibitors of both important proteins responsible for diabetes. The molecular binding energy and inhibition constant were calculated, as well as the pharmacodynamics and pharmacokinetics characteristics were evaluated. Selected forty compounds were studied via Lipinski's rule. Out of 40 compounds one compound i.e. 3-O-Acetyl-11-Keto-Boswellic Acid, was eliminated since it violated two of the parameters of Lipinski's rule. Molecular docking (MD) analysis were performed on 39 identified components where (+-)-Teframedine and Morphine were found to be the potential inhibitors of alpha-amylase and GFAT-1, respectively. Morphine and (+-)-teframedine also demonstrated a good potential in bioavailability analysis and they can be taken orally. However, the solubility and lipophilicity parameters were not followed by 11-keto-boswellic acid, ursane, 3-acetyl-boswellic acid, and -amyrin, suggesting that these 4 substances are less bioavailable when taken orally. Further, these compounds should undergo in-vitro studies to support these results.

Keywords: Diabetes mellitus, alpha-amylase, GFAT-1, Molecular Docking, health care, global health

INTRODUCTION

One of the most common metabolic diseases, DM is characterized by abnormally elevated blood glucose levels, accounting for about 80% of the total fatalities occurring every year, as per the World Health Organisation (WHO) data (who. int). According to the statistical data of 2020, 463 million people were diagnosed with diabetes worldwide, with around 77 million from India and the International Diabetes Federation (IDF) has anticipated that these figures would climb up to 642 million over the next two decades. Currently, the most common medications used to treat diabetes include insulin and oral hypoglycemic agents like biguanides, α-amylase inhibitors meglitinides, thiazolidinediones, sulfonylureas, and gliptins.[1] However, owing to the numerous undesirable side effects like hepatocellular damage, hypoglycemia, dizziness, neurological disorders, etc caused by

*Corresponding Author: Dr. Arpita Roy, Email: arbt2014@gmail.com
+Authors contributed equally

Cite as: *J. Integr. Sci. Technol., 2025, 13(2), 1040.*
DOI: 10.62110/sciencein.jist.2025.v13.1040

©Authors CC4-NC-ND, ScienceIN
http://pubs.thesciencein.org/jist

these drugs, novel entities with fewer or no adverse effects must be evaluated.[2]

Inhibiting carbohydrate-hydrolyzing enzymes such as amylases, lipases, and maltases in the gut reduces glucose and fat absorption, which is one of the most therapeutically important strategies for managing diabetes.[3] A vital enzyme for digestion in the human body is pancreatic alpha-amylase (EC 3.2.1.1). The hydrolysis reaction of alpha-1,4 glycosidic bonds found in glycogen, maltodextrins, amylose, starch, and amylopectin is catalyzed by EC 3.2.1.1 and results in the formation of oligosaccharides, which are then split into monosaccharides by -glucosidase.[4] Hence, inhibiting alpha-amylase may help individuals to manage their diabetes, by retarding the elevation of blood glucose levels after consuming a carbohydrate-rich diet.[5] In addition to the glycolysis pathway, the hexosamine biosynthetic pathway—which utilizes glutamine as an amino source is also very important. It converts fructose-6-phosphate into glucosamine-6-phosphate. Glutamine-fructose-6-phosphate amidotransferase (GFAT-1) catalyzes this reaction by regulating the glucose entering the hexosamine pathway and acts as the rate-limiting enzyme.[6] Exceptionally high levels of sugar in the blood lead to a higher concentration of fructose-6-phosphate flowing into the hexosamine pathway and may result in the development of diabetic complications.[7,8] In addition, elevated human GFAT1 activity has been associated with insulin resistance in both animal and cellular models, which is a feature of Type 2 DM.[9,10] Thus, alpha-amylase and GFAT1 can strongly be considered significant targets for the treatment of Type 2 DM.[11]

Natural compounds have the potential to act as drug candidate and they can be used for generation of novel therapeutic agents.[12] Utilizing herbal remedies is a cost-effective therapeutic option, with fewer side effects, for several ailments.[13,14] Plant-based drugs possess many advantages concerning selectivity, efficacy, and reduced off-target toxicities.[15,16] In the last few years, the utilization of phytocompounds has drastically enhanced due to the various adverse effects exhibited by chemically synthesized drugs. Moreover, the drugs that are currently available in the market exhibit high levels of toxicity and thus, there is a need to replace chemically synthesized drugs with herbal compounds that have lower toxicity.[17] Previously, some studies have been published, that assess these two proteins as an important target for the management of this disorder. A study reported potential inhibitors of α-amylase extracted from the *Leucas ciliata Benth* (Lamiaceae) and *Streptomyces longisporoflavus* and *L. ciliata*,[18] followed by MD studies. The results identified flavonoid and alkaloid compounds as the potential inhibitors of α-amylase which could help in development of potential α-amylase inhibitors against diabetes. Okechukwu et al., 2020, showed MD on palmatine with α-amylase, alpha-glucosidase, and gliptins.[19] Furthermore, the binding energies were evaluated with those of widely used medications such as sitagliptin and acarbose, and it was established that palmatine has antidiabetic properties.

Ukwenya et al., 2021, carried out an in silico analysis where compounds derived from *Anacardium occidentale* were used against GFAT1.[20] The findings revealed that 8 compounds that met the RO5 and were within the acceptable range for Absorption, Distribution, Metabolism, and Excretion (ADME) criteria, demonstrating their suitability for application in the production of anti-diabetic medications. Davella et al., 2019, conducted a study for the evaluation of compounds extracted from *Rumex vesicarius* for anti-Type 2 Diabetes Mellitus activity.[21] The results showed that Physcion out of all the compounds had a better docking score (-7.66 kcal/mol). These investigations suggest that there is a scope for improvement in finding better inhibitors for activity against diabetes.[22]

Apart from this, it has been shown that phytochemicals extracted from *Boswellia serrata* gum resin possess anti-inflammatory properties and these have been used to treat several chronic inflammatory conditions.[23] It has previously been demonstrated that administering *Boswellia serrata* gum resin to LADA (Latent Autoimmune Diabetes in Adults) patients decreased blood levels of IA2 antibodies, one of the markers associated with LADA autoimmune diabetes. According to experimental research, it has also been demonstrated that *Papaver somniferum* comprises critical characteristics of therapeutic significance. It has been used as a sedative, analgesic, narcotic stimulant, nutrition, and more. It is even useful for headaches, cardiac asthma, cough, biliary colic, and insomnia.[24]

Thus, in this study, the aim was to find potential natural inhibitors against the two major protein targets, alpha-amylase, and GFAT-1, using computational techniques for the potentail treatment of diabetes using the two selected medicinal plant compounds.

Results and Discussion

ADME Analysis

The 40 shortlisted phytocompounds were initially subjected to preliminary assessment using molecular characteristics of the ligand (Table 1). This was done to exclude some of the compounds that did meet the criteria for selection. When the drug likeliness was calculated using Lipinski's Rule of 5, one compound, 3-O-Acetyl-11-Keto-Boswellic Acid, was eliminated since it violated two of the characteristics taken into account. The remaining 39 compounds were considered as potential candidates and were further investigated using molecular docking to anticipate the interactions of the protein-ligand complex.

Molecular Docking

The goal of MD studies is to identify conformations of the docked complex that have the lowest binding affinities and to anticipate the optimal configuration of the identified ligand with the target protein. All 39 of the ligands that complied with Lipinski's five conditions were subjected to MD to the active sites of the enzyme alpha-amylase and the protein GFAT-1.

Alpha-amylase protein

In case of alpha-amylase, the binding energy of standard drug i.e. acarbose showed a minimum binding energy of -4.05kcal/mol. GLN63, TRP59, GLU240, GLY306 and ASP300 formed conventional H-bonds. ALA198 and LEU162 demonstrated alkyl interactions. There were 14 van der waal attractive interactions and carbon-hydrogen bonds (Figure 1). The complexes with lowest binding energy of *Papaver somniferum* and *Boswellia serrata* were -8.31kcal/mol and -10.15 kcal/mol, respectively (Table 2).

Table 1: ADME Analysis of selected ligands

Plant Name	Phytocompounds	PubChem ID	Mol wt	H-bond acceptors	H-bond donors	LogP (XLOGP3) <5	Molar Refractivity	Lipinski Violation
Papaver somniferum	Sanguinarine	5154	332.33 g/mol	4	0	4.45	94.68	0
	(+-)-Carnegine	442186	221.30 g/mol	3	0	2.14	68.48	0
	(+-)-Teframidine	436140	323.34 g/mol	5	0	2.94	90.15	0
	Berberine	2353	336.36 g/mol	4	0	3.62	94.87	0
	Kaempferol	5280863	286.24 g/mol	6	4	1.9	76.01	0
	Linolenic acid	5280934	278.43 g/mol	2	1	6.46	88.99	1
	Morphine	5288826	285.34 g/mol	4	2	0.76	82.27	0
	Papaverine	4680	339.39 g/mol	5	0	2.95	97.16	0
	Reticuline	439653	329.39 g/mol	5	2	3.01	97.01	0
	1-Benzylisoquinoline	23345	219.28 g/mol	1	0	4.06	71.2	0
	Xanthaline	96932	353.37 g/mol	6	0	3.66	97.59	0
	Oripavine	5462306	297.35 g/mol	4	1	1.87	86.53	0
	1,2-Dehydroreticuline	440930	328.38 g/mol	4	2	2.38	98.27	0
	14-Hydroxycodeinone	9926820	313.35 g/mol	5	1	1.12	86.98	0
	2-(2-Furanyl)-3-methyl-2-butenal	555644	150.17 g/mol	2	0	1.91	43.42	0
	2-(4-Hydroxyphenyl) acetaldehyde	440113	136.15 g/mol	2	1	0.53	38.44	0
	3,4-Dimethoxyphthalic acid	68209	226.18 g/mol	6	2	0.92	53.34	0
	4-hydroxybenzaldehyde	126	122.12 g/mol	2	1	1.35	33.85	0
	4-Hydroxybenzoic acid	135	138.12 g/mol	3	2	1.58	35.42	0
	4,5-Dimethoxyphthalic acid	290988	226.18 g/mol	6	2	0.92	53.34	0
Boswellia Serrata	Alpha boswellic acid	637234	456.70 g/mol	3	2	8.41	136.65	1
	α-Campholenic-Acid	117235	168.23 g/mol	2	1	1.92	49.11	0
	Euphane	12312921	414.75 g/mol	0	0	12.12	136.83	1
	(−)-Camphene	440966	136.23 g/mol	0	0	4.22	45.22	1
	(+)-α-Phellandrene	443160	136.23 g/mol	0	0	3.21	47.12	0
	(1S,2R,4S)-(−)-Bornyl acetate	442460	196.29 g/mol	2	0	4.3	56.33	0
	11-Keto-β-boswellic acid	9847548	470.68 g/mol	4	2	7.2	137.11	1
	3-Acetyl-β-boswellic acid	11386458	498.74 g/mol	4	1	8.29	146.65	1
	3-O-Acetyl-11-Keto-β-Boswellic-Acid	11168203	512.72 g/mol	5	1		146.85	2
	α-Terpinene	7462	136.23 g/mol	0	0	4.25	47.12	0
	β-Pinene	14896	136.23 g/mol	0	0	4.16	45.22	1
	l-Idose	11030410	180.16 g/mol	6	5	-3.24	35.74	0
	Myrcene	31253	136.23 g/mol	0	0	4.17	48.76	0
	P-Cymene	7463	134.22 g/mol	0	0	4.1	45.99	1
	Serratol	101618281	290.48 g/mol	1	1	5.85	95.92	1
	Ursane	9548870	412.73 g/mol	0	0	11.47	134.45	1
	α-Amyrin	73170	426.72 g/mol	1	1	9.01	135.14	1
	2,3-Dihydroxyurs-12-en-28-oic acid	155934	472.70 g/mol	4	3	6.37	138.08	1
	Beta boswellic acid	168928	456.70 g/mol	3	2	8.26	136.91	1
	11-Keto-β-boswellic acid methyl ester	101695788	468.71 g/mol	3	0	8.32	140.27	1

Table 2: Molecular docking studies of selected ligands against alpha-amylase

Plant Name	Phytocompound	PubChem ID	Binding Energy (ΔG) (Kcal/mol)	Ligand Efficiency	Inhibition constant (μM)	Intermolecular energy	Vdw H bond desolvation energy
Papaver somniferum	(+-)-Teframidine	436140	-8.31	-0.35	0.8168	-8.31	-6.57
	Sanguinarine	5154	-7.65	-0.31	2.47	-7.65	-7.83
	Oripavine	5462306	-6.93	-0.32	8.29	-7.53	-6.96
	Berberine	2353	-6.91	-0.28	8.67	-7.5	-7.41
	Kaempferol	5280863	-6.81	-0.32	10.28	-8.3	-7.75
	(+-)-Carnegine	442186	-6.8	-0.43	10.35	-7.4	-5.48
	Morphine	5288826	-6.72	-0.32	11.96	-7.31	-6.73
	14-Hydroxycodeinone	9926820	-6.17	-0.27	30.13	-6.76	-6.33
	Reticuline	439653	-5.81	-0.24	55.02	-7.6	-6.9
	Xanthaline	96932	-5.75	-0.22	60.63	-7.54	-7.38
	1,2-Dehydroreticuline	440930	-5.57	-0.23	83.17	-7.36	-7.25
	Linolenic acid	5280934	-5.31	-0.27	128.49	-9.48	-8.76
	1-Benzylisoquinoline	23345	-5.07	-0.3	192.38	-5.67	-5.62
	2-(2-Furanyl)-3-methyl-2-butenal	555644	-4.94	-0.45	241.11	-5.53	-5.51
	Papaverine	4680	-4.92	-0.2	247.06	-6.71	-6.68
	4-Hydroxybenzoic acid	135	-4.5	-0.45	503.8	-5.39	-4.64
	4-hydroxybenzaldehyde	126	-4.39	-0.49	609.36	-4.98	-4.78
	2-(4-Hydroxyphenyl) acetaldehyde	440113	-3.99	-0.4	1200	-4.88	-4.81
	3,4-Dimethoxyphthalic acid	68209	-3.7	-0.23	1930	-5.49	-5.21
	4,5-Dimethoxyphthalic acid	290988	-3.69	-0.23	1960	-5.48	-5.18
Boswellia Serrata	Ursane	9548870	-10.15	-0.34	0.03607	-10.15	-10.15
	α-Amyrin	73170	-9.89	-0.32	0.05675	-10.18	-10.16
	Euphane	12312921	-9.68	-0.32	0.08031	-11.17	-11.17
	11-Keto-β-boswellic acid methyl ester	101695788	-9.53	-0.28	0.104	-10.12	-10.06
	11-Keto-β-boswellic acid	9847548	-9.09	-0.27	0.21846	-9.98	-10.24
	Beta boswellic acid	168928	-8.69	-0.26	0.42423	-9.59	-9.91
	3-Acetyl-β-boswellic acid	11386458	-8.55	-0.24	0.54242	-9.74	-10.11
	Alpha boswellic acid	637234	-8.43	-0.26	0.65758	-9.33	-9.68
	2,3-Dihydroxyurs-12-en-28-oic acid	155934	-7.99	-0.24	1.39	-9.18	-9.32
	Serratol	101618281	-7.05	-0.34	6.83	-7.64	-7.47
	(1S,2R,4S)-(−)-Bornyl acetate	442460	-5.64	-0.4	73.97	-6.23	-6.12
	l-Idose	11030410	-5.17	-0.43	163.43	-6.96	-5.87
	(+)-α-Phellandrene	443160	-4.86	-0.49	274.07	-5.16	-5.15
	β-Pinene	14896	-4.86	-0.49	274.82	-4.86	-4.86

	α-Campholenic-Acid	117235	-4.85	-0.4	278.54	-5.74	-4.94
	α-Terpinene	7462	-4.72	-0.47	347.76	-5.02	-5
	(−)-Camphene	440966	-4.7	-0.47	360.85	-4.7	-4.7
	P-Cymene	7463	-4.4	-0.44	594.08	-4.7	-4.68
	Myrcene	31253	-3.9	-0.39	1380	-5.1	-5.08
Standard inhibitor	Acarbose	444254	-4.05	-0.09	1070	-10.61	-8.65

(+-)-Teframedine-alpha amylase complex demonstrated a minimum inhibition binding energy of -8.31 kcal/mol amongst all phytocompounds selected from *Papaver somniferum*. LEU673 and VAL677 exhibited alkyl interactions. There were 13 van der Waal attractive interactions present in the complex. ALA674 and SER422 formed conventional H-bonds (Figure 1(A)).

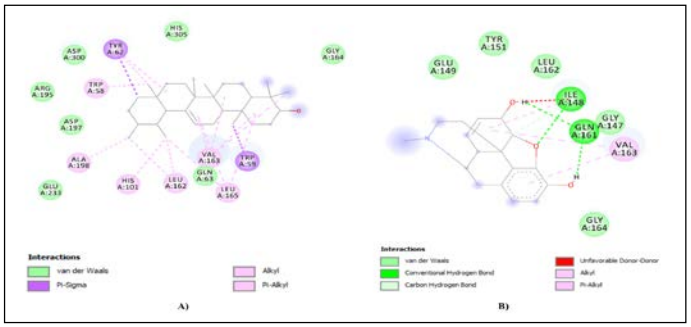


Figure 1. a) Interaction between (+-)-Teframedine & Alpha-amylase b) Interaction between Ursane & Alpha-amylase

In the case of *Boswellia serrata*, the Ursane-alpha amylase complex exhibited a minimum binding energy of -10.15 kcal/mol. HIS305, ASP300, ARG195, GLU233, ASP197, and GLN63 were shown to have van der Waal attractive interactions. There were 9 alkyl and pi-alkyl interactions present in the complex. TRP59 demonstrated a pi-sigma interaction (Figure 1(B)).

The second top-docked phytocompound of this plant was the α-Amyrin-alpha amylase complex, which demonstrated a minimum binding energy of -9.89 kcal/mol. As it can be seen in Figure 2(A), TYR62 and TRP59 exhibited pi-sigma bonds. TRP58, ALA198, HIS101, LEU162, VAL163, and LEU165 were shown to have alkyl and pi-alkyl interactions. There was a presence of 7 van der Waal attractive interactions in the complex.

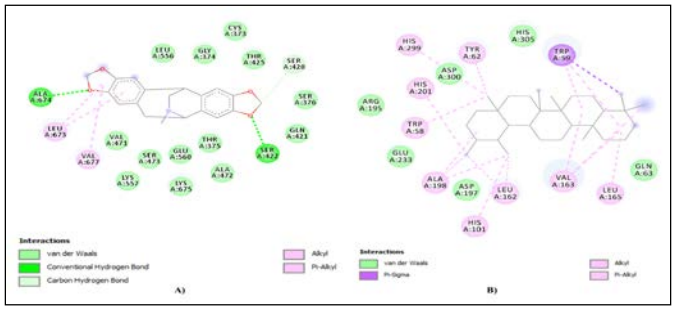


Figure 2. a) Interaction between α-Amyrin and Alpha-amylase b) Interaction between Morphine and GFAT1

GFAT1 protein

Metformin, a commonly used drug for GFAT-1, was taken as a standard inhibitor in this study. Metformin-GFAT1 complex was shown to have a most negative binding energy of -4.5 kcal/mol. VAL471 and ALA674 formed conventional H-bonds.LEU673, ALA472, SER473, LYS675, GLY374, and VAL677 were shown to have van der Waal attractive interactions. A salt bridge was observed between GLU560 and hydrogen.

The docking analysis exhibited the minimum binding energies of -8.01 kcal/mol and -9.97 kcal/mol among all the docked complexes of the two herbs, *Papaver somniferum* and *Boswellia serrata*, respectively (Table 3).

Table 3: Molecular docking studies of selected ligands against GFAT-1

Plant Name	Phytocompound	PubChem ID	Binding Energy (ΔG) (Kcal/mol)	Ligand Efficiency	Inhibition constant (μM)	Intermolecular energy	Vdw H bond desolvation energy
Papaver somniferum	Morphine	5288826	-8.01	-0.38	1.34	-8.61	-8.22
	Oripavine	5462306	-7.8	-0.35	1.92	-8.4	-7.74
	14-Hydroxycodeinone	9926820	-7.62	-0.33	2.58	-8.22	-7.95
	Xanthaline	96932	-7.43	-0.29	3.55	-9.22	-8.89
	(+-)-Teframidine	436140	-7.11	-0.3	6.17	-7.11	-6.66
	Sanguinarine	5154	-6.94	-0.28	8.22	-6.94	-6.75
	1,2-Dehydroreticuline	440930	-6.85	-0.29	9.55	-8.64	-8.48
	Reticuline	439653	-6.83	-0.28	9.85	-8.62	-7.7
	Kaempferol	5280863	-6.55	-0.31	15.78	-8.04	-7.76

	Berberine	2353	-6.45	-0.26	18.55	-7.05	-6.88
	Papaverine	4680	-6.21	-0.25	27.95	-8	-7.81
	1-Benzylisoquinoline	23345	-5.85	-0.34	51.76	-6.44	-6.38
	(+-)-Carnegine	442186	-5.47	-0.34	97.13	-6.07	-5.39
	4,5-Dimethoxyphthalic acid	290988	-5.18	-0.32	159.03	-6.97	-5.71
	3,4-Dimethoxyphthalic acid	68209	-5.1	0.32	182.42	-6.89	-5.74
	2-(2-Furanyl)-3-methyl-2-butenal	555644	-5.05	-0.46	199.4	-5.64	-5.5
	2-(4-Hydroxyphenyl) acetaldehyde	440113	-4.88	-0.49	266.24	-5.77	-5.66
	Linolenic acid	5280934	-4.87	-0.24	269.56	-9.05	-8.19
	4-Hydroxybenzoic acid	135	-4.81	-0.48	299.99	-5.7	-5.15
	4-hydroxybenzaldehyde	126	-4.73	-0.53	340.18	-5.33	-5.15
Boswellia Serrata	11-Keto-β-boswellic acid	9847548	-9.97	-0.29	0.04921	-10.86	-10.4
	3-Acetyl-β-boswellic acid	11386458	-9.79	-0.27	0.06629	-10.99	-10.56
	Beta boswellic acid	168928	-9.74	-0.3	0.07291	-10.63	-10.18
	Alpha boswellic acid	637234	-9.66	-0.29	0.08325	-10.55	-10.12
	Ursane	9548870	-9.11	-0.3	0.20913	-9.11	-9.11
	Euphane	12312921	-9.07	-0.3	0.22417	-10.56	-10.56
	11-Keto-β-boswellic acid methyl ester	101695788	-9.03	-0.27	0.24161	-9.62	-9.69
	α-Amyrin	73170	-8.9	-0.29	0.30115	-9.19	-9.05
	Serratol	101618281	-7.93	-0.38	1.55	-8.52	-8.52
	2,3-Dihydroxyurs-12-en-28-oic acid	155934	-7.65	-0.23	2.48	-8.84	-8.76
	(1S,2R,4S)-(−)-Bornyl acetate	442460	-6.53	-0.47	16.36	-7.13	-7.05
	α-Campholenic-Acid	117235	-5.4	-0.45	109.71	-6.3	-6
	l-Idose	11030410	-4.99	-0.42	220.59	-6.78	-6.54
	(+)-α-Phellandrene	443160	-4.9	-0.49	257.09	-5.2	-5.21
	β-Pinene	14896	-4.86	-0.49	276.08	-4.86	-4.85
	α-Terpinene	7462	-4.77	-0.48	320.18	-5.07	-5.07
	(−)-Camphene	440966	-4.73	-0.47	343.78	-4.73	-4.72
	P-Cymene	7463	-4.38	-0.44	615.47	-4.68	-4.68
	Myrcene	31253	-4.2	-0.42	828.66	-5.4	-5.39
Standard inhibitor	Metformin	4091	-4.5	-0.5	501.31	-4.5	-3.39

Amongst all the ten energy conformations obtained for the Morphine-GFAT1 complex from *Papaver somniferum*, -8.01 kcal/mol was the binding energy confirmation. GLU149, TYR151, LEU162, GLY147, and GLY164 demonstrated van der Waal interactions. ILE148 and GLN161 formed conventional H-bonds. Pi-alkyl and alkyl interactions were present (VAL163). Unfavorable donor-donor interactions were also observed (Figure 2(B)).

Figure 3(A) shows the best-docked complex from *Boswellia serrata*, i.e 11-Keto-β-boswellic acid, having a minimum binding energy of -9.97 kcal/mol. Three types of interaction can be seen i.e. conventional H-bond, van der Waal, and alkyl bonds. LYS675, GLN421, SER376, and SER422 were forming conventional H bonds. Alkyl interactions were observed with LEU673 and LEU556. Thirteen other van der Waal attractive interactions were found in the complex in addition to these interactions.

The second top-docked compound, 3-Acetyl-β-boswellic acid, which also belongs to *Boswellia serrata*, demonstrated the least binding energy of -9.79kcal/mol. LEU673 AND LEU556 were shown to have alkyl interactions. LYS675, GLN421, SER422, and GLN480 formed conventional H-bonds. Moreover, there were 15 van der Waal attractive interactions that were present in the complex (Figure 3(B)).

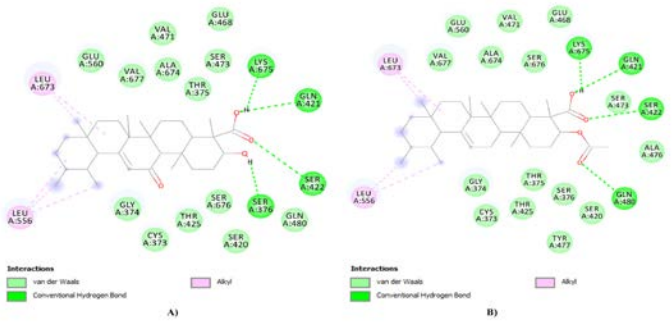


Figure 3. A) Interaction between 11-Keto-β-boswellic acid and GFAT1 B) Interaction between 3-Acetyl-β-boswellic acid and GFAT1

Bioavailability Radar

From *Boswellia serrata*, 11-keto-β-boswellic acid and 3-acetyl-β-boswellic acid were amongst the top two docked ligands against GFAT-1 protein that were studied. However, Morphine-GFAT-1 complex showed the minimum binding energy from a pool of phytocompounds obtained from *Papaver somniferum*. In case of alpha-amylase, ursane and α-amyrin of *Boswellia serrata* were amongst the top two obtained docked confirmations. (+-)-teframedine, On the other hand, exhibited the best results compared to all the phytocompounds chosen from *Papaver somniferum*. In order to conduct a thorough examination, bioavailability radar is a descriptive approach that takes 6 physical and chemical aspects into account when evaluating the drug-likeness of particular substances. It was noted that the ligand spectrum perfectly fit in the pink-colored areas in the cases of morphine and (+-)-teframedine, demonstrating that it is orally bioavailable. However, the solubility and lipophilicity parameters were not followed by 11-keto-boswellic acid, ursane, 3-acetyl-boswellic acid, and -amyrin, suggesting that these 4 substances are less bioavailable when taken orally. (Figure 4).

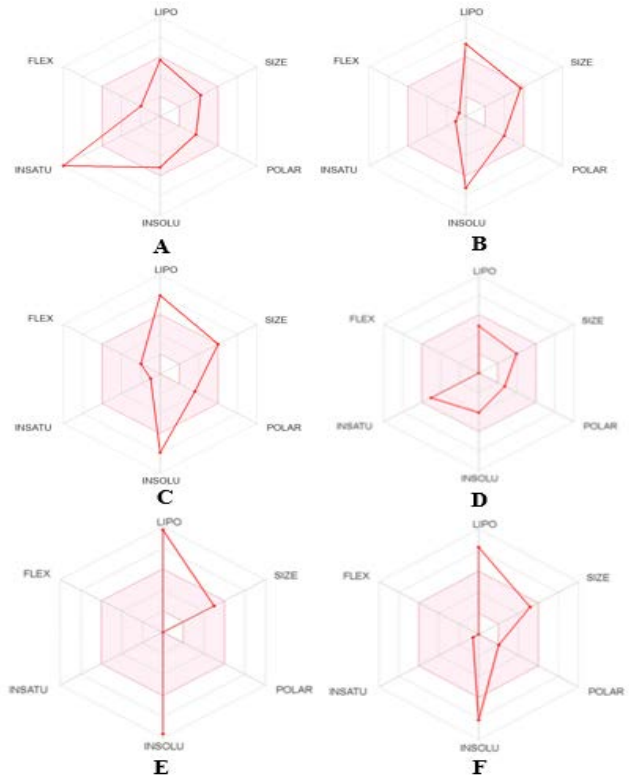


Figure 4. Bioavailability radars of (A) Morphine (B) 11-Keto-β-boswellic acid (C) 3-Acetyl-β-boswellic acid (D) (+-)-Teframedine (E) Ursane (F) α-Amyrin.

MATERIALS AND METHODS

Requirements for in silico investigations

IMPPAT database, PubChem, Protein Data Bank, Discovery studio visualizer Biovia, Alpha fold protein structure database, PyMOL, PyRx, SwissADME, and AutoDock v4.2.6 are the programs and datasets used in this investigation.[25]

Macromolecule

The 3D structures of alpha-amylase (PDB ID: 1OSE)[26] and GFAT-1 (PDB ID: 2ZJ3)[27], both being two major receptors involved in the progression of diabetes, were used for the investigation. Their structures were extracted from 'Protein Data Bank' in .pdb format.

Ligands

Two medicinal plants, *Papaver somniferum* (commonly known as opium poppy) and *Boswellia serrata*, owing to their therapeutic properties, were chosen for this study.[28] Fourty different phytocompounds present in the plants were selected using the IMPPAT (https://cb.imsc.res.in/imppat) database (Refer Table 1). To continue with docking studies, their 3-dimensional structures were downloaded from PubChem in the ".sdf" format and transformed into ".pdb" format using Biovia Discovery Studio's visualizer.[29]

ADME Analysis

The primary screening of the selected phytocompounds was done using the SwissADME (http://www.swissadme.ch/index.php) and their pharmacokinetic characteristics like ADME and toxicity were estimated using Lipinski's Rule of 5.[30] According to it, the molecular weight of the ligand should be <500 Dalton, H-bond donors should be <5, H-bond acceptors should be <10 in number, ligand should be highly lipophilic (LogP<5), and molar refractivity should be between 40-130.[31] Any ligand that exhibited more than 1 violation of Lipinski's rule was directly eliminated from further studies. Out of the forty phytocompounds, only 3-O-Acetyl-11-Keto-β-Boswellic-Acid was eliminated and remaining thirty-nine compounds were subject to MD.[32]

Molecular Docking

The MD of each selected ligand was carried out against alpha-amylase and glutamine-fructose-6-phosphate transaminase 1

(GFAT-1) proteins using Autodock v4.2.6. Water molecules were eliminated, and polar hydrogens and Kollmann charges were added.[33] Further, gasteiger charges were computed, followed by the removal of hetatoms from the macromolecules to complete the preparation of both proteins. For alpha-amylase, the active site residues were arranged in a grid box with dimensions 48x60x44 and with a spacing of 0.375, and for GFAT-1, 60x60x64 and 0.375. The docking studies resulted in Lamarckian GA output. For each ligand, the process was performed three times, and among the 10 possible conformations, the ideal conformation with the lowest binding energy was selected. Using Biovia Discovery Studio Visualizer v19.1.0.18287, it was then converted into a 2-D graphic showing the interaction between the ligand and the active site residue.[34]

Bioavailability Radar

The filtration of a potent ligand molecule was continued using a more reasonable evaluation of physicochemical characteristics. The SwissADME (http://www.swissadme.ch/index.php) web-based tool was used to create bioavailability radars of ligands that performed more effectively than the standard molecule. The compounds were evaluated using six parameters: size, lipophilicity, saturation, polarity, and flexibility. Non-oral bioavailability was suspected in ligands that deviated from the established values, thus they were excluded from further testing.[35,36]

CONCLUSION

In the present study, forty phytocompounds from two different medicinal plants, *Papaver somniferum* and *Boswellia serrata*, were chosen against two receptor, i.e. alpha-amylase and glutamine-fructose-6-phosphate transaminase 1 which are involved pathogenesis of diabetes. These were screened for their drug likeliness, out of which thirty-nine were further evaluated using molecular docking studies. The docked complexes underwent comparison with the standard inhibitors of both proteins. In the case of alpha-amylase, thirty-five phytocompounds possessed higher binding energies than the standard inhibitor acarbose. Similarly, thirty-seven phytocompounds exhibited higher binding energies against GFAT-1 compared to its standard drug metformin. For bioavailability radar testing, the top compounds from each of the two plants docked against each protein were selected. Morphine and (+-)-Teframedine of *Papaver somniferum* was found to be orally bioavailable. In conclusion this study suggests that *Papaver somniferum* and *Boswellia serrata*, are probably better anti diabetic agents than metformin and phythochemicals analysed are probably better drug candidiates than metformin. Further in-vitro research should be done to find an effective and reliable treatment for diabetes. These phytochemicals can be employed as possible plant-based inhibitors against the targeted proteins.

ACKNOWLEDGMENTS

The authors would like to thank their universities for neassary support.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

REFERENCES

1. K.N. Sangeetha, S. Sujatha, V.S. Muthusamy, S. Anand, K. Shilpa. Current trends in small molecule discovery targeting key cellular signaling events towards the combined management of diabetes and obesity. Bioinformation. **2017**,13(12), 394.
2. S. P. Gurjar, A. Roy, A. Gupta. Competitive inhibition of catechol from *Andrographis paniculata* in the complex of ERK2 in lung metastasis. *J. Integr. Sci. Technol.*. **2023**, 12(1), 719.
3. K. Hamden, K. Mnafgui, Z. Amri, A. Aloulou, A. Elfeki. Inhibition of key digestive enzymes related to diabetes and hyperlipidemia and protection of liver-kidney functions by trigonelline in diabetic rats [published correction appears in *Sci Pharm*. **2014**, 82(2), 449-50.
4. S. Lordan, T.J. Smyth, A. Soler-Vila, C. Stanton, R.P. Ross. The α-amylase and α-glucosidase inhibitory effects of Irish seaweed extracts. *Food Chem*. **2013**,141, 2170–6.
5. S. Poovitha, M. Parani. In vitro and in vivo α-amylase and α-glucosidase inhibiting activities of the protein extracts from two varieties of bitter gourd (*Momordica charantia L.*). BMC Complement Altern Med. **2016**, 16, 185.
6. K.C. Chou. Molecular therapeutic target for type-2 diabetes. *J Proteome Res*. **2004**, 3(6), 1284–1288.
7. H. Zhang, Y. Jia, J.J. Cooper, T. Hale, Z. Zhang, S.C. Elbein. Common variants in glutamine: fructose-6-phosphate amidotransferase 2 (GFPT2) gene are associated with type 2 diabetes, diabetic nephropathy, and increased GFPT2 mRNA levels. *J. Clin. Endocrinol. Metab.* **2004**, 89, 748– 755.
8. D. Beriault, Y. Shi, G. Werstuck. Investigating the role of the hexosamine biosynthesis pathway in diabetic atherosclerosis. *Can J Cardiol.* **2013**, 29, S254.
9. K.A. Robinson, D.A. Sens, M.G. Buse. Pre-exposure to glucosamine induces insulin resistance of glucose transport and glycogen synthesis in isolated rat skeletal muscles. *Diabetes*. **1993**, 42, 1333– 1346.
10. L. Rossetti, M. Hawkins, W. Chen, J. Gindi, N. Barzilai. In vivo glucosamine infusion induces insulin resistance in normoglycemic but not in hyperglycemic conscious rats. *J. Clin. Invest.* **1995**, 96, 132– 140.
11. L. Herbert, D.A. McClain. Overexpression of glutamine: fructose-6-phosphate amidotransferase in transgenic mice leads to insulin resistance. *J. Clin. Invest.* **1996**, 97, 1235– 1241.
12. R. Luthra, A. Roy, S. Pandit, R. Prasad. Biotechnological methods for the production of ginsenosides. *S. Afr. J. Bot.* **2021**, 141, 25-36.
13. R. Luthra, G. Satija, A. Roy. Centellosides: pharmaceutical applications and production enhancement strategies. *Plant Cell, Tissue Organ Cult.* **2021**, 1-15.
14. A. Roy, S. Datta. Medicinal Plants against Ischemic Stroke. *Curr. Pharm. Biotechnol.* **2021**, 22(10), 1302-1304.
15. R. Zainul, D. S. Novel, H. Satriawan, Goh, K. W., Jakhmola, V., Rebezov, M., Faridah. In Silico Gene Transcription of 4-Hydroxycinnamic Acid from Broccoli Fruit (Brassica oleracea var. italica) with Estrogen Receptor Beta Protein. *Pharmacognosy Journal*, **2024**, 16(4).131
16. H. Assaggaf, N. El Hachlafi, M. El fadili et.al. GC/MS profiling, in vitro antidiabetic efficacy of Origanum compactum Benth. essential oil and in silico molecular docking of its major bioactive compounds. *Catalysts*, **2023**, 13(11), 1429.
17. Z. Rabiei, M.R. Bigdeli, Z. Lorigooini. A review of medicinal herbs with antioxidant properties in the treatment of cerebral ischemia and reperfusion. *J. Babol Univ. Medical Sci.* **2015**, 17(12), 47-56.
18. J. V. Akshatha, H. S. Santosh Kumar, H. S. Prakash, M.S. Nalini. In silico docking studies of α-amylase inhibitors from the anti-diabetic plant *Leucas ciliata Benth.* and an endophyte, *Streptomyces longisporoflavus*. *3 Biotech.* **2021**, 11(2), 51.
19. P. Okechukwu, M. Sharma, W.H. Tan, H.K. Chan, K. Chirara, A. Gaurav, M. Al-Nema, In-vitro antidiabetic activity and in-silico studies of the binding energies of palmatine and the standard compounds with the three receptors of alpha amylase, alpha glucosidase, and DPP-IV enzyme. *Pharmacia.* **2020**, 67(4), 363-371.
20. V.O. Ukwenya,; S.A. Adelakun,; O.O. Elekofehinti. Exploring the antidiabetic potential of compounds isolated from *Anacardium occidentale*

using computational aproach: ligand-based virtual screening. *In silico pharmacology*. **2021**, 9(1), 25.
21. R. Davella, E. Mamidala. In silico Molecular Docking Studies of compounds from *Rumex vesicarius* against GFAT1. *Biolife*. **2019**, 6(3), 6302.
22. S. Datta, R. Luthra, N. Bharadvaja. Medicinal Plants for Glioblastoma Treatment. *Anticancer Agents Med Chem.* **2022**, 22. 4739
23. Z. Franić, Z. Franić, N. Vrkić, N.N. Gabaj, I. Petek. Effect of extract from *Boswellia serrata* gum resin on decrease of GAD65 autoantibodies in a patient with Latent Autoimmune Diabetes in Adults. *Altern Ther Health Med.* **2020**, 26(5), 38-40.
24. M. Masihuddin, M. Jafri, A. Siddiqui, S. Chaudhary. Traditional uses, phytochemistry and pharmacological activities of *Papaver somniferum* with special reference of unani medicine an updated review. JDDT. **2018**, 8(5), 2069
25. S.P. Gurjar, A. Roy, A. Gupta. Understanding Pharmacokinetics, Bioavailability Radar, and Molecular Docking Studies for Selected Medicinal Plants Against Lung Cancer Receptors. In Harnessing Medicinal Plants in Cancer Prevention and Treatment. *IGI Global.* **2024**, 343-388.
26. C. Gilles, J. Astier, G. Marchis-Mouren, C. Cambillau, F. Payan. Crystal Structure of Pig Pancreatic α-amylase Isoenzyme II, in Complex with the Carbohydrate Inhibitor Acarbose. European Journal of Biochemistry, **1996**, 238(2), 561–569.
27. Y. Nakaishi, M. Bando, H. Shimizu, K. Watanabe, F. Goto, H. Tsuge, K. Kondo, M. Komatsu. Structural analysis of human glutamine:fructose-6-phosphate amidotransferase, a key regulator in type 2 diabetes. FEBS Letters. **2008**, 583(1), 163–167.
28. B. Haque, A. Gupta, A. Roy, A. Malik, & A. A. Khan. Green fabrication of Ag–Ni–Mn-Zn nanoparticles from watermelon peels and its antioxidant, dye degradation and molecular docking studies. Clean Technologies and Environmental Policy, **2024**, 1-28.
29. A. Roy, N. Sharma, R. Luthra. In-silico analysis of antiviral fungal inhibitors against Mpro receptor protein. *Vegetos*. **2024**. 1-14
30. H. Kashyap, G. Sekhon, A. Varshney, M. Khatri. In silico screening of potential antidiabetic phytochemicals from Tephrosia purpurea: A polypharmacology approach. *Biomed. Ther. Lett.* **2024**, 11 (2), 911.
31. A. Daina, O. Michielin, V. Zoete. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep.* **2017**, 7(1), 42717.
32. A. Gupta, A. Roy, S. Pandit, N. Pandey, Rustagi, S. Evaluation of bioactive compounds obtained from *Ginkgo biloba* against crystal structure of myelin oligodendrocyte glycoprotein (MOG). *OBM Neurobiology*, **2023**, 7(4), 1-34.
33. A. Roy, A. Gupta, D. Verma. Evaluation of different phytochemicals from *Cannabis sativa* against Myelin oligodendrocyte glycoprotein receptor. *Vegetos.* **2023**. 1-11.
34. A. Gupta, A. Roy, V. Raja, S. Rustagi, S. Malik, & D. Verma. In-silico studies of phytoconstituents of *Bacopa monnieri* and *Centella asiatica* with Crystal structure of Myelin Oligodendrocyte Glycoprotein against primary demyelination in Multiple Sclerosis. *J. Integr. Sci. Technol.*, **2023**, 12(3), 764.
35. S.P. Gurjar, A. Gupta, A. Roy. Molecular docking studies of phytocompounds from *Artemisia monosperma* against ERK2 kinase in lung cancer. *J. Mol. Chem.*, **2023**, 3(2), 591-591.
36. U. Kaushik, S.C. Joshi. Formulation and evaluation of Momordica charantia fruit extract based transdermal drug delivery against diabetes. *Biomed. Ther. Lett.* **2024**, 11 (1), 905.

Physica Scripta

INCLUSIVE PUBLISHING TRUSTED SCIENCE

PURPOSE-LED PUBLISHING™

ACCEPTED MANUSCRIPT

Design of a Low-Power Dynamic Latched Comparator for Biomedical Applications

To cite this article before publication: Mohit Tyagi *et al* 2024 *Phys. Scr.* in press https://doi.org/10.1088/1402-4896/ad9b55

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© 2024 IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.

As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript will be available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 14.139.251.98 on 09/12/2024 at 06:48

Design of a Low-Power Dynamic Latched Comparator for Biomedical Applications

Mohit Tyagi[a,b], Poornima Mittal[a], and Parvin Kumar[b]

[a]Department of Electronics and Communication Engineering, Delhi Technological University, Delhi-110042 (India)
[b]Department of Electronics and Communication Engineering, KIET Group of Institutions, Delhi-NCR, Ghaziabad, 201206 (India)
Email id: mohit.tyagi@kiet.edu, poornima2822@gmail.com, parvin.kaushik@kiet.edu
*Corresponding author: poornima2822@gmail.com

Abstract
The latched comparator is a fundamental component of all ADC topologies. Thermal noise, kickback noise, and offset voltage impact the energy efficiency of the comparator significantly. In cardiac IMD ADCs, latched comparator kickback noise may impact resolution, accuracy, and settling time. This paper describes the design and development of an a low-power dynamic latched comparator optimised for biomedical applications and running at 1V. The suggested comparator is achieved on a 45 nm CMOS technology node, having the main goal of reducing kickback noise. The main concept of the recommended comparator has been utilizing the novel resetting approach and kickback noise reduction approach to preserve the charge and kickback noise, respectively. This approach aids in minimizing delay, power consumption, and kickback noise using shared charging logic. The charging-sharing technique involves using a single-pass transistor positioned across two output nodes. The pass transistor distributes the charge evenly across two output nodes throughout the reset phase. Due to the shared charging of the load capacitances, the voltage at the output nodes would not drop under the threshold value. As a result, the input signal could be evaluated more quickly during the regenerative stage. The comparator's delay is calculated using a rigorous statistical analysis considering the circuit's random factors. The offset of the proposed comparator is verified using thorough Monte Carlo simulations. The comparator reduces power while preserving noise. Also, the measured simulation results of recommended comparator outperform state-of-the-art comparators regarding kickback noise, delay, power consumption, and power delay product (PDP). The input voltage inversely affects the comparator's delay. Further, the simulation shows that the comparator consumes 9.36μW at 1 V and 1GHz sampling frequency. The simulation results confirm that the suggested comparator effectively reduces power consumption by atleast 66.98%. The proposed comparator has less PDP value by 44.44%. Also, suggested comparator effectively reduces offset voltage by 7.6% but with a corresponding increase in the area by 5%. Additionally, the comparator successfully reduces kickback noise by 44.6%. The obtained results are not silicon proven and bound to be in the pre-silicon stage. The low-power behaviour of the suggested technique is validated using analytical derivations, PVT corner analysis, and post-layout simulation.

Keywords*: Latched Comparator, ADC topologies, cardiac IMDs,45-nm CMOS technology, power delay product (PDP), biomedical applications*

1. INTRODUCTION

The semiconductor sector has experienced substantial expansion and is currently extensively utilised in several domains owing to its unique circuitry, which offers solutions for a diverse array of biomedical applications. In the present electronic era, a multitude of applications incorporates an analog to digital converter for transforming analog signal information into its equivalent digital code, thereby facilitating the transmission and storage of digital data. The growing demand for electronic devices that consume less power has led to the development and advancement of CMOS-based innovations and architecture [1, 2]. The need for designing an optimized analog to digital converter has become important due to its major application in wearables and implant devices in the biomedical field [3].

The simplified representation of a bio-signal acquisition device [1] is depicted in Figure 1. The sensor receives an analog signal from the input and amplifies and processes it to produce an output. In addition, the input signal is fed to an ADC, which then multiplies it by a clock signal generated by an oscillator, leading to the development of a digitised output. As a consequence, a digital code was created and then communicated via an antenna [4,5]. Different architectures of ADCs like flash, delta-sigma are commonly utilised in the mixed signal sensor applications, but biomedical systems require medium resolution and medium speed analog to digital converters with low power consumption along with noise reduction is the critical stringent requirements in biomedical scenarios [6]. So, compared to existing ADC architectures, SAR-ADC is widely acknowledged as a highly desirable ADC due to its capability to efficiently function at low power consumption with medium resolution and medium speed applications [7,8] which are highly desirable for biomedical applications in wearable or implantable devices.

Accepted Manuscript

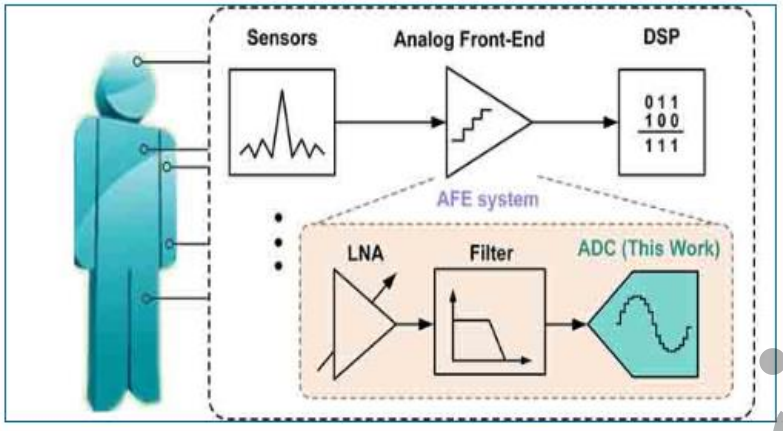


Figure 1: Chain of analog front end for bio-medical applications

The SAR ADC consist of four essential sub-blocks: comparator, the sample and hold (S/H) circuit, digital-to-analog converter (DAC), and control logic. Design of optimized comparator with reduced kickback noise, low power consumption along with reduced input referred offset voltage is one of the major designing assignments required for low power comparator module in SAR ADC which is considered in the present work. Also, it is worth mention that the frequency range of biomedical signals comes under the range of 0.5 Hz to 10 KHz which suggests that the comparator designed for biomedical applications is sufficient to work at medium speed and medium resolution but low noise along with low power consumption is very important [9,10]. The low-power double-tail latched comparator has been proposed by the authors in reference [9]. The design successfully mitigates kick-back noise by the implementation of a self-neutralization mechanism but with a substantially elevated input parasitic capacitance.

In [11] authors have introduced a dynamic comparator that incorporates an offset cancelling circuit with the capacity to detect minute fluctuations in voltage, but this comparator is linked to a comparatively elevated level of power consumption. Numerous power-saving methodologies have been proposed in scholarly works, including the imposition of voltage limitations at the preamplifier's output [12] and the constraint of the preamplifier's output swing [13] has also been considered in the literature. The phenomenon of kick-back noise, which refers to the detectable nonlinearity effect caused by distortion, is a commonly recognised occurrence [11-14]. Further, comparator common mode variation effect analysis is well presented in [15]. Authors have also analysed the effect of differential voltage gain and gm/Id of input transistors to handle Vcm variations in a dynamic comparator. A thorough analysis was conducted to examine the different configurations of dynamic latch comparators in terms of power consumption, kick-back noise, input-referred offset voltage, speed [16]. When power consumption and delay are used as major criteria, it is clear that the double-tail dynamic latched comparator has less energy usage and delay than other designs [16]. In [17,18] charge sharing method to improve delay time by preserving the change at nodes is well presented. Common mode component of kickback noise in transconductance enhanced latch and strong-ARM latch is analysed in [19]. Comparator with the methodology of dynamic floating inverter amplifier is well presented in [20,28]. Various dynamic comparison methodologies are discussed in [34-37] for ultra-low power comparator design for multiple applications.

In this paper, novel dynamic latched comparator is designed using 45 nm CMOS technology node, having the main goal of reducing kickback noise, low power consumption, reduced delay and reduction of input referred offset voltage. Also, proposed comparator mitigates the effect of common mode voltage (Vcm) variations by optimizing differential voltage gain (G) for the proposed comparator in the pre-amplification phase. Further, it is worth mention that the recommended comparator utilizes the novel resetting approach and kickback noise approach to preserve the charge and kickback noise, respectively. Also, statistical analysis of input referred offset voltage (σ_{os}) and analysis of root mean square value of comparator noise (σ_n) is carried out. The proposed design aids in minimizing delay, power consumption, input referred offset voltage, power delay product (PDP) and kickback noise. The introduction of shared charging during the reset phase makes the proposed design to operate effectively without the need for extensive transistor stacking or voltage boosting, hence enabling its functionality at low supply voltages.

Further, the paper is structured into five sections. Section 2 provides an overview of the existing conventional single and double-tail comparators. Section 3 offers a comprehensive explanation of the approach as well as the implementation of the proposed comparator. Simulation results are presented in Section 4. Finally, Section 5 concludes the paper with highlighting of the optimized parameters.

Accepted Manuscript

2. EXISTING CONVENTIONAL COMPARATORS

Contemporary bio-medical systems incorporate the low voltage and low power devices in the circuit design. The optimization of power consumption, area and delay in a comparator can be achieved by eliminating pre-amplification performed by the pre-amplification stage of the comparator. Also, the system speed is improved significantly with the elimination of pre-amplification stage. Further, this section presents different existing architectures of dynamic latched comparators with varying types.

2.1 Single Tail Dynamic Latched Comparator: STDLC

Figure 2 illustrates the architecture of a single-tail dynamic latched comparator [28]. The circuit operates in two separate stages, which are dependent on the clock signal (clk). When the value of the clk signal is '0', the comparator operates in the reset phase. In the given condition, transistor N1 will remain in a non-conduction state, leading to the elimination of a discharge path from the outputs to the ground. Moreover, in this operating condition, transistors P1 and P4 are in the conductive state, and the voltage level of the comparator outputs will be equal to V_{DD}. When the clk signal is set to a logically high value of 1, the comparator gets into the comparison phase. In the given conditions, the transistor N1 will operate in an active state, establishing a discharge pathway from the outputs of the comparator to the ground. The rate at which outputs are discharged depends on the gate voltages of the input transistors of the comparator, N2 and N3. These transistors have their gate terminals connected to the inn and inp ports through which input voltages V_{inn} and V_{inp} to be compared are applied, respectively. According to reference [13], if V_{inn} is greater than V_{inp}, it can be deduced that the discharge of output V_{outn} is going to take place at a faster rate compared to that of output V_{outp}, and vice versa.

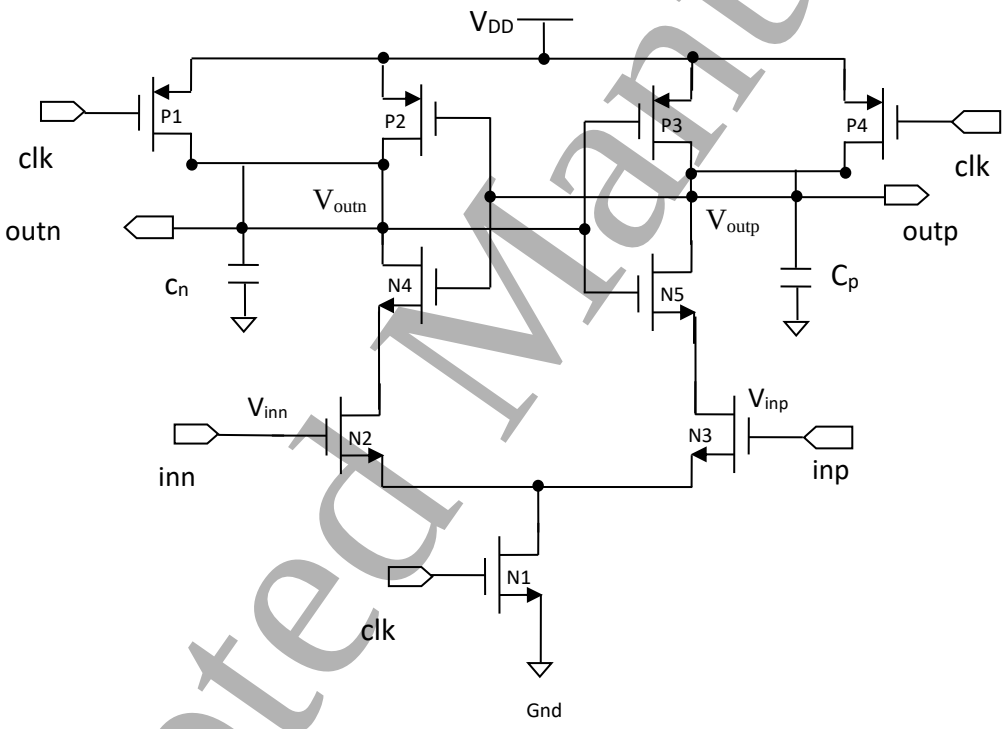


Figure 2: Schematic Diagram of STDLC

2.2 Double-Tail Dynamic Latched Comparator: DTDLC

Figure 3 depicts the circuit of double tail dynamic latched comparator (DTDLC) [32]. The discussed topology has a lower stacking level than the typical dynamic comparator, also known as the single-tail dynamic comparator. The use of a double-tail architecture allows the achievement of two key aims. For starters, it enables the generation of a significant electric current throughout the decision stage, allowing for quick latching regardless of the input common-mode voltage. In addition, it enables a low starting current by using a lower P1, which helps decrease offset [14,21]. During the reset & decision-making phases, DTDLC performs similarly to a single-tail dynamic comparator. The leakage current of the double-tail comparator design is lower than that of the traditional single-tail dynamic comparator. Furthermore, as previously shown [21, 23], a double tail in the assessment step allows for a substantial flow of current, minimising offset voltage. The comparator delay is the amount of time it takes for the resulting differential signal to reach $V_{DD}/2$. The total delay caused by the comparator may be determined by taking two factors into account: the delay of the latch, that has been affected by the cross-coupled inverters (N4/P4, N5/P5), along with

Accepted Manuscript

the delay associated with charging the load capacitance's output terminal till the initially activated MOS transistors (N1/P1) gets activated [24,25].

The typical double-tail comparator's total delay may be expressed analytically by eq. (1-5) as follows [22]:

$$t_{delay} = t_0 + t_{latch} \tag{1}$$

$$t_0 = \frac{V_{THN} C_{out}}{I_p} = \frac{2V_{THN} C_{out} u_p C_{ox} \frac{W_p}{L_p}}{g_{mP}^2} \tag{2}$$

where u_p is the mobility of holes, V_{THN} represents threshold voltages of the transistors N_4 and N_5, I_p is the current through P_4, $C_{out\text{-}load}$ denotes the capacitance of the comparator output nodes, g_{mp} is the P_4 transistor's transconductance, Lp, W_p are lengths and widths of transistor P4 respectively.

The time delay associated with the latch can be determined using eq. (3), which is provided below:

$$t_{latch} = \frac{C_L}{g_{m,eff}} . \ln\left(\frac{\Delta V_{out}}{\Delta V_0}\right) \tag{3}$$

$$= \frac{C_L}{g_{m,eff}} . \ln\left(\frac{V_{DD}/2}{\Delta V_0}\right) \tag{4}$$

$$t_{delay} = \frac{2V_{THN} C_{out} u_p C_{ox} \frac{W_p}{L_p}}{g_{mP}^2} \ldots + \frac{C_L}{g_{m,eff}} . \ln\left(\frac{V_{DD}/2}{\Delta V_0}\right) \tag{5}$$

ΔV_0 constitutes a variable that represents the variation in the latch's initial voltage at the output at the commencement of the regeneration procedure. Throughout the evaluation, both cross-coupled inverters remain within the ON state.

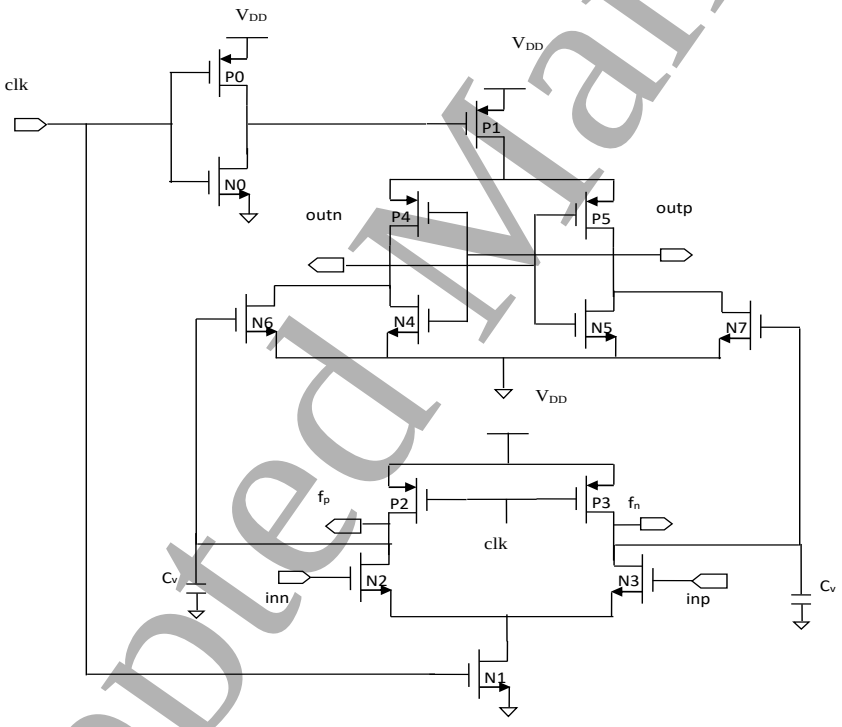


Figure 3: Schematic Diagram of DTDLC

The comparator's operational delay is significantly affected by the latch output nodes' capacitive load, the voltage variation at the input, and the transconductance (g_m) for the input and intermediate transistors, as demonstrated in eq. (5). An increase in ΔV_0 causes a reduction in latency.

3. PROPOSED COMPARATOR

3.1 Circuit Description

Dynamic latched comparators have shown superior performance compared to other comparator designs in terms of optimizing speed, power consumption, delay and input referred offset voltage [27,43]. Moreover, the circuit's

Accepted Manuscript

architectural design offers enhanced isolation between the input and output, along with reduced undesirable noise, making it highly suitable for low-voltage applications within the realm of biomedical implants. The comparator being examined has been constructed using a double-tail comparator, which exhibits superior performance compared to a single-tail dynamic latched comparator. Figure 4 displays the suggested comparator's architectural design, which leverages shared charged circuitry and incorporates the kickback noise reduction mechanism with the inclusion of two clocked tranistors M8 and M9 to ground the input pair source transitors. The suggested comparator successfully resolves the issue of kickback noise by reducing the effect of switching of fn and fp node voltages on the input voltage. The comparator that has been given has two distinct modes of operation: reset mode and comparison/evaluation mode. During the reset step, dynamic latched comparators use two unique reset techniques to guarantee that the right logical level is maintained throughout the evaluation stage. The reset technique involves either charging or draining the output nodes to VDD or ground. The output terminals are electrically charged to the voltage level represented as V_{DD} in the context of the STDLC arrangement. In contrast, the output nodes in the DTDLC architecture are discharged to ground. Following the comparator's evaluation step, one of the outputs is at a reference potential, also known as ground potential, while the other output node is at the supply voltage level, designated as V_{DD}.

During the reset step, the comparator's output nodes are either entirely discharged to ground or possibly charged to V_{DD} in all the mentioned topologies. Following that, the next round of assessment commences, which might begin at either zero or source voltage levels. This results in a longer time period necessary to get the final assessment result, as well as greater energy usage. Once again, the process of emptying energy from one node or refilling energy in another terminal causes an increase in power consumption during the following reset phase. The proposed comparator's core premise is the use of a unique resetting mechanism and kickback noise approach to preserve the charge and kickback noise, respectively. This approach makes it easier to reduce latency, power consumption, and kickback noise. The distinct technique mentioned in this context is known as shared charging logic. The charging-sharing approach employs a single-pass transistor strategically positioned between two output nodes. During the reset phase, the pass transistor guarantees that charge is distributed uniformly across two output nodes. The voltage at the output nodes will stay over the chosen threshold value due to the collective charging of the load capacitances. As a result, the input signal may be evaluated more efficiently during the regeneration phase which speeds up the process. The implementation results show a significant improvement in power and latency as a consequence of using this approach. The use of sample switches is a very effective approach for reducing kickback noise and improving comparator operating performance.

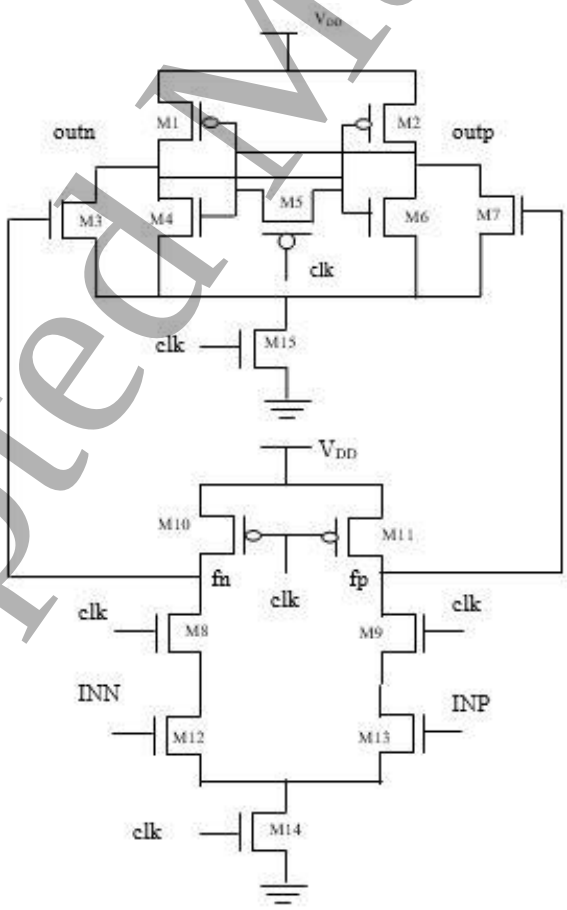


Figure 4: Architecture of Proposed Low Power Kickback Noise Reduced Comparator

Accepted Manuscript

The clk signal is linked to ground during the reset or recharge phase due to which tail transistors M14 and M15 are disabled, reducing static power consumption. During this phase, transistors M10 and M11 boost the voltage levels of nodes fn and fp to the V_{DD} potential. As a result, it starts the activation of transistors M3 and M7. During the reset phase, the shared transistor M5 is used to link the outputs outp and outn, allowing it to serve as a shared charging logic transistor. Following the previous assessment step, charge is distributed between two nodes when one node is at the voltage V_{DD} and the second node is at the ground potential. Deactivating the kickback noise mitigation transistors M8 and M9 causes the comparator's input and output terminals to become disconnected. The latch outputs had no effect on the transistors M12 and M13, which were coupled to the input signals. Once the clk signal attains the V_{DD} supply voltage threshold, the circuit undergoes a transition into the evaluation phase. This transition triggers the activation of transistors M8 and M9, which are integral components of the Kickback noise mitigation system. The activation of transistors M8-M9 contributes to the mitigation of kickback noise. The differential voltage generated at the inputs is subsequently routed to cross-coupled latches based on inverters, which facilitate the required decision-making process. During this phase, transistors M14 and M15 are in an active state, whereas reset transistors M10 and M11 are in a deactivated state. Assume that $V_{inn}>V_{inp}$, then the potential of fp lowers at a higher rate than fn. The results of the implementation demonstrate a notable enhancement in power and latency as a result of employing this technique. The utilization of sample switches is a highly efficacious strategy for mitigating kickback noise and augmenting the operational efficacy of the comparator. During the reset or recharge phase, the clk signal is connected to the ground potential, and the tail transistors M14 and M15 are deactivated, therefore mitigating static power consumption. In this phase, the voltage levels of nodes fn and fp are raised to the V_{DD} potential via transistors M10 and M11. Consequently, it initiates the activation of transistors M3-M7. During the reset phase, the shared transistor known as CSL forms a connection between the outputs outp and outn, allowing it to function as a shared charging logic transistor.

The distribution of charge occurs between two nodes when one node is at a voltage of V_{DD} and the other node is at the ground potential, subsequent to the preceding evaluation stage. The nature of this occurrence would be altered if we were to take into account the inverse state of the input voltage. The sole distinction in functioning within this design lies in the fact that, during the assessment phase, the decision does not stem from two voltage levels situated at opposite extremes. In contrast, it can be observed that the voltage level at both output terminals is approximately half of the level of the source voltage. When the output terminal is initially set to a voltage equal to half of the supply voltage, the evaluation time is reduced, and power consumption is decreased. The delay analysis provides empirical evidence that aligns with this theoretical concept. The main benefit of the proposed comparator design is in its only reliance on a single regulating clock signal. The efficiency of the circuit is boosted by removing the need for synchronization among many control signals. The circuit under consideration additionally provides reduced latency and power optimization. The tail transistors possess dimensions that are sufficiently small to provide the diminished tail current required for the differential input transistor stage. The circuit complexity has been reduced due to the absence of any requirement for additional transistors to configure or reset the device. In the process of regeneration, it is imperative to ensure the isolation of the drain terminals of the transistors comprising the input differential pair from the regeneration nodes. This approach is applied in order to mitigate the incidence of kickback noise that arises from transients at the regeneration nodes. In order to execute this procedure, the regeneration node is linked to the drains of the input differential pair through the utilization of switching transistors M8 and M9. Switches that are intended to operate in a condition of strong inversion have been developed with enlarged dimensions in order to achieve a higher transconductance, thereby mitigating the impact of thermal noise. Therefore, in the process of regeneration, the isolation of the regeneration node and the drains of the input differential pair is achieved by employing M8-M9. In order to mitigate the effects of mismatch and manufacturing uncertainty, the design process incorporates customized transistor sizing and matching methodologies aimed at reducing offset voltage.

3.2 DESIGN CONSIDERATIONS IN PROPOSED DESIGN

The following design considerations are carried out while implementing the proposed Comparator:

3.2.1 Region of Operation

In order to achieve a reduction in both supply voltage and power consumption, the present design bias the employed MOSFETs in the subthreshold operation region. The development of comparator within ultra-deep submicrometric CMOS technologies is hindered by low supply voltages, mainly due to the lack of proportional scaling between the threshold voltages of the devices and the supply voltages of contemporary CMOS processes [26]. Therefore, designing comparators for low power analog to digital converters becomes increasingly difficult as the supply voltage VDD decreases. Further, the transistors are usually operated in weak/moderate inversion regions to manage low-power circuits. The relationship between gate-to-source voltage and drain current is defined by given eq.(6) as shown:

Accepted Manuscript

$$I_D = \frac{W}{L} I_{DO} e^{\frac{qV_{GS}}{nKT}} \quad (6)$$

where W/L is the ratio of width to length of the transistor, I_D is drain current, I_{DO} is the reverse saturation current, n denotes slope factor, V_{GS} represents gate to source voltage, K is the Boltzmann constant and T is room temperature. The topic under examination involves a unified all-regional framework, commonly called the "EKV Model," which explores the feasibility of transistor operation within an inversion region. The model above is employed to optimize comparator noise along with reduction in power dissipation [29-32]. The following formulas for the MOS device compute the EKV configuration.

$$Ic = \frac{I_D}{I_S} \quad (7)$$

$$I_s = 2\,\mu C_{ox} \frac{W}{L} \emptyset_t^2 \quad (8)$$

$$g_m = \frac{I_D}{\eta \Phi_t} \frac{1 - \exp\left(-\sqrt{Ic}\right)}{\sqrt{Ic}} \quad (9)$$

where, I_S is the normalization current, I_D denotes drain current, ŋ is the slope factor, which is taken as 1,C_{ox} is oxide capacitance, t_{ox} is oxide layer's thickness, ϕ_t denotes thermal voltage, Ic is for inversion coefficient which is chosen as 0.1 in this work. The transistors utilised in this study function within the weak inversion region. The dimensions of the transistor are determined based on the model provided. The design procedure of the proposed comparator requires careful consideration of several design variables. The determination of transistor sizes in the comparator depends on the EKV model, as illustrated in Table 1.

Table 1. Transistor sizes for the Proposed Comparator

Name of the Transistor	W/L ratio
M3-M4-M5-M6-M7-M12-M13	2/1
M1- M2-M10- M11	4/1
M14- M15	4/1
M8-M9	4/1

3.2.2 Kickback noise

The input voltage in CMOS latched comparators can be adversely affected by the output voltage variation connected to the input transistors. The reduction of kickback noise is necessary in order to achieve a higher level of precision. The input of the comparator is represented by a thevenin equivalent circuit, whereby the resistor value is 8 kΩ. The comparator under consideration exhibits significant kickback noise due to its resemblance to the single-tail comparator's construction. Furthermore, as stated in the study conducted by reference [21], the increased speed of output settling in dynamic comparators results in a more significant disruption at the input nodes. As a result, kickback noise in high-speed comparators is unavoidable. Specific noise reduction approaches, such as neutralization, as proposed in reference [21], have the potential to mitigate the impact of kickback noise significantly. Typically, the reference nodes exhibit low impedance characteristics, resulting in fewer voltage disturbances on the reference voltages than the inputs. Furthermore, the primary disruption occurs during the initiation of the amplification stage, allowing the reference voltages sufficient time to stabilize at their standard level before the commencement of the comparison stage [36-37]. The proposed comparator demonstrated a reduced level of kickback noise mainly the due to the utilization of auxiliary clocked transistors M8 and M9. Utilizing these transistors served the purpose of isolating the output nodes, thus reducing the mismatch effects which also influence alongside the transistor's parasitic capacitance in the circuit design stage. Further, offset reduction in the feedback loop is achieved through appropriate transistor sizing that operates in the weak inversion region and works for the optimization of mismatch.

Further, common mode component of kickback noise occurs when switching of clock signal (clk) affects the input signal due to coupling with parasitic capacitances as stated in figure: From [16] It is confirmed that to reduce the common mode component of kickback noise, slew rate of source and drain terminal voltages of input transistor pair is required to be reduced. For this inclusion, we may reduce the slew rate of clk signal which in turn reduce the slew rate of source and drain terminal of input transistor pair. This can be done by incorporating a R-C circuit just after the generation of clock signal to increase its rise and fall time which in turn reduces the slew rate of node voltages fn and fp which decreases the kickback noise current to reduce the kickback noise current. Resistance (R) can be created bypass transistor. Also, it is worth mentioned that this technique will result in increase in offset voltage of the proposed design.

Accepted Manuscript

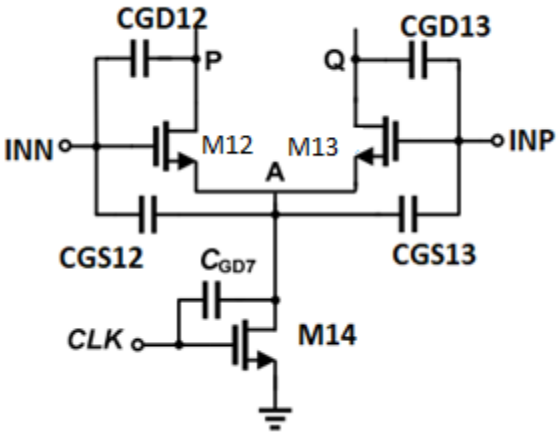


Figure 5: Kickback Noise Paths

The initial phenomenon often occurs when fn (labelled as P) and fp (labelled as Q) in fig. 5 reduced towards the ground potential at different rates due to different INN and INP. This switching of voltages at P and Q node affects the input voltages INN and INP due to the coupling of these nodes with the inputs through parasitic capacitances CGD12 and CGD13. The significance of this effect increases when M12 and M13 enter the triode region, leading to an increase in their gate-drain capacitances and causes differential kickback noise but due to inclusion of M8 and M9, direct connection between the transiting nodes fn(P) and fp(Q) with the input is avoided which results in reduced kickback noise current. Further, the common-mode kickback noise currents exhibit much greater intensity and occur during the transition of clk signal from a low to high state, coinciding with the activation of M14 in figure 5. To clarify, at the start of the "Evaluation" stage. Currently, M14 is extracting drain current from CGS12 and CGS13. Furthermore, when the clk signal transitions to a low state, the change in clk level is sent by CGD7 to both CGS12 and CGS13 and results in change in input voltage and this is avoided by increasing rise and fall time of clock signal which in turn reduces the slew rate of node fn(P) and fp(Q). Pass-transistors are utilized to increase the rise and fall time of clk signal. However, this will increase mismatch in the proposed design and will lead to increase in the input referred offset voltage.

3.3 Derivation of Time Constant for the Cross-Coupled Inverter Latch

The time constant (time delay of latch regeneration) of cross-coupled inverter latches is computed in the subsequent part. The latch has been generated by a back-to-back connection between two inverters. Figure 6 depicts a circuitry comprised of two cross-coupled inverters.

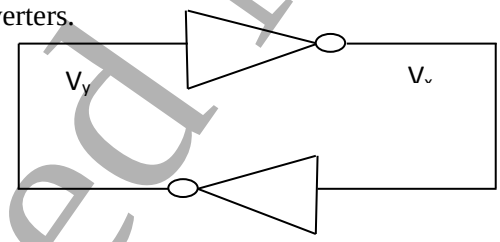


Figure 6: Cross-Coupled Inverter Latch

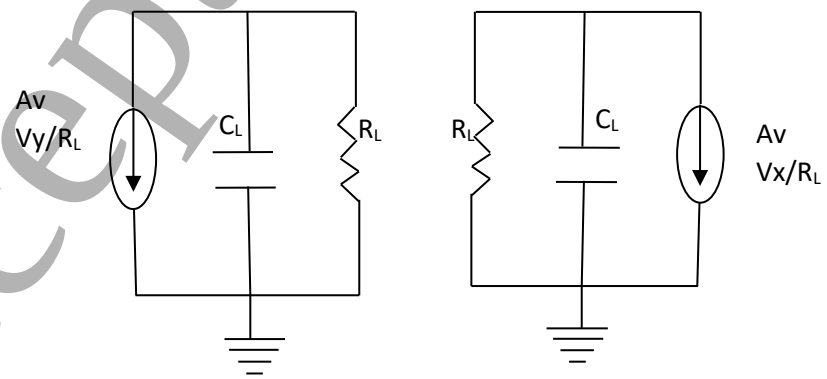


Figure 7: Simplified Equivalent Circuit of a Latch.

The small signal analysis of the cross-coupled latch is illustrated in Figure 7; the outputs of the inverters are very close to each other at the start of the comparison phase and offer a voltage controlled current source with a RC load. A_V is the low-frequency gain of the inverter. The inverter transconductance is given approximately as Gm = A_V/R_L.
For this circuit model,

Accepted Manuscript

$$\frac{A_V}{R_L}.V_y = -C_L.\left[\frac{dV_x}{dt}\right] - \left[\frac{V_x}{R_L}\right] \quad (10)$$

$$\frac{A_V}{R_L}.V_x = -C_L.\left[\frac{dV_y}{dt}\right] - \left[\frac{V_y}{R_L}\right] \quad (11)$$

When eq. (10) and (11) are multiplied by R_L and afterwards rearranged, the result is

$$\tau.\left[\frac{dV_x}{dt}\right] + V_x = -A_V.V_y \quad (12)$$

$$\tau.\left[\frac{dV_x}{dt}\right] + V_x = -A_V.V_y \quad (13)$$

Where $\tau = R_L C_L$ is the time constant at each inverter output node. Subtracting the Equation (11) from (10) and the differential voltage will be given by eq. (14) as:

$$\left[\frac{\tau}{A_V - 1}\right].\left[\frac{d\Delta V}{dt}\right] = \Delta V \quad (14)$$

Here $\Delta V = V_x - V_y$ is the difference in output voltages of the cross-coupled inverters. Eq. (14) represents a first-order mathematical differential equation with no force function with its solution in given in eq. (15) as:

$$\Delta V = \Delta V_o.e^{(A_V - 1).t/\tau} \quad (15)$$

Where, ΔV_o is the initial differential voltage generated during the start of the evaluation phase.

Thus, for a particular time constant, this voltage differential rises exponentially over time, as demonstrated by

$$\tau_{latch} = \frac{\tau}{A_V - 1} \quad (16)$$

$$\tau_{latch} \cong \frac{R_L.C_L}{A_V} \quad (17)$$

$$\tau_{latch} = \frac{C_L}{G_m} \quad (18)$$

Where A_V is gain and G_m is the transconductance of the inverter. Typically, the load present at the output node of MOS devices is directly proportional to the capacitance of the gate-source in a transistor as given by eq. (19) as:

$$C_L = K_1\, W\, L\, C_{ox} \quad (19)$$

K_1 is proportionality constant, between 1to 2. In addition, the Equation demonstrates the transconductance of inverter is directly proportional to the transistor's transconductance:

$$G_m = K_2 g_m = \mu_n.C_{ox}.\frac{W}{L}.V_{eff} \quad (20)$$

Here, *the* K_2-proportionality constant is between 0.5 to 1. Substituting eq. (12) and eq. (13) in eq. (11) will result in:

$$\tau_{latch} = \frac{K_1}{K_2}.\frac{L^2}{\mu_n.V_{eff}} = K_3\frac{L^2}{\mu_n.V_{eff}} \quad (21)$$

Where K_3 proportionality is constant, between 2 to 4, eq. (21) indicates that τ_{latch} mostly depends on technology rather than design. For the CMOS technology stated, eq. (21) is highly beneficial for estimating the maximal clock frequency of a latch contained in a dynamic comparator.
Using eq. (14) & eq. (17), the appropriate time for a voltage differential ΔV_{logic} may be calculated as

$$T_{latch} = \frac{C_L}{G_m}.\ln\left[\frac{\Delta V_{logic}}{\Delta V_o}\right] \quad (22)$$

$$= \frac{L^2}{\mu_n.V_{eff}}.\ln\left[\frac{\Delta V_{logic}}{\Delta V_o}\right] \quad (23)$$

$$= \frac{L^2}{\mu_n.V_{eff}}.\ln\left[\frac{\Delta V_{logic}}{\Delta V_o}\right] \quad (24)$$

Accepted Manuscript

The preceding equations reveal that the outcome of a latch varies exponentially over time. The outcome of latched inverter rises exponentially using a time constant determined by G_m/C_L and a unity gain frequency. When the applied input voltage is extremely low, the issue of metastability arises. The simplified formula for the delay of an inverter latch consisting of an output swing proportional to V_{DD}, is as follows:

$$T_{latch} = \frac{C_L}{G_m}.\ln\left[\frac{V_{DD}/_2}{\Delta V_o}\right] \quad (25)$$

It shows that the latch delay is also a function of input supply voltage.

3.4 Analysis of Delay in Proposed Comparator

A mathematical equation for the delay has been calculated to illustrate the impact of the improved architecture, which incorporates a double-tail dynamic latched comparator as well as shared charged logic. The analytical method closely resembles that of the typical double tail dynamic latched comparator. A theoretical analysis was conducted to demonstrate the impact of various design factors on the functioning of the dynamic latched comparator [32, 33]. In the suggested comparator design, two parameters related to delay are being enhanced: ΔV_0 at the commencement of the regeneration stage and operational transconductance, in order to enhance delay. The overall delay is further subdivided into two components: t_0 and t_{latch}. The value of t_0 can be determined using eq. (26). The delay t_0 corresponds to the time it takes for the load capacitance to reduce until the initial nMOS (N1or N2) is activated. Therefore, t_0 has been determined by eq. (26) as:

$$t_0 = \frac{C_L\,(V_{thn})}{I_{IB1}} = \frac{C_L\,(V_{DD}-2V_{thp})}{I_{\text{current in the tail2}}} \quad (26)$$

Where in, I_{B1} represents the current in the drain terminal of either N1 or N2, which is equivalent to $I_{tail2}/2$. The t_{latch} delay has been generated by the latch delay of two cross-coupled inverters. Consider the output voltage swing as V_{DD} derived from the differential input voltage. Latch delay is given as:

$$t_{latch} = \frac{C_L}{g_{m,eff}}.\ln\left(\frac{\Delta V_{out}}{\Delta V_0}\right) \quad (27)$$

$$= \frac{C_L}{g_{m,eff}}.\ln\left(\frac{V_{DD}}{\Delta V_0}\right) \quad (28)$$

Where $t_{latch}\frac{C_L}{g_{m,eff}}$ is the transconductance of the latch. In dynamic Comparator, ΔV_0 expression is given as follows:

$$\Delta V_0 = 2.\left(V_{DD} - V_{thp}\right)\frac{\Delta V_{fn(fp)}\, g_{mMR1,2}}{I_{\text{current in the tail2}}} \quad (29)$$

Where in $g_{mMR1,2}$ is transconductance of the transferring phase is determined by the transistors MR2, MR1. $\Delta V_{fn(fp)}$ represents the differential voltage which occurs during the first stage at time t_0 between the regenerative nodes fn and fp. From eq. (25) and eq. (28), $\Delta V_{fn/fp}$ can be determined as:

$$\Delta V_{fn/fp} = \frac{C_L\,(V_{DD}-2V_{thp})}{I_{\text{current in the tail2}}} \cdot \frac{g_{mn1,2}\ \Delta Vin}{C_L\ fn(fp)} \quad (30)$$

Where $g_{mn1,2}$ is the transconductance of input transistor pair. By substituting eq. (29) into eq. (30), the initial value of voltage difference is determined as:

$$\Delta V_0 = \frac{2.\ (V_{DD}-2V_{thp})\ (V_{DD}-V_{thp})}{I_{\text{current in the tail2}}^{\ 2}} \quad \frac{g_{mMR1,2}\ g_{mn1,2}\ C_L\ \ \Delta Vin}{C_L\ ,fn(fp)} \quad (31)$$

Combining the eq. (25), (29) and (30), the total delay can be determined as

$$t_{totaldelay} = \frac{C_L\,(V_{DD}-2V_{thp})}{I_{\text{current in the tail2}}} + \frac{C_L}{g_{m,eff}}.\ln\left(\frac{V_{dd}\ \ C_L\ ,fn(fp)\ I_{\text{current in the tail2}}^{\ 2}}{4.\ (V_{DD}-2V_{thp})\ (V_{DD}-V_{thp})\ g_{mMR1,2}\ g_{mn1,2}\ C_L\ \ \Delta Vin}\right) \quad (32)$$

The influence of different design factors on total delay time may be observed in eq. (32). Based on this study, the recommended delay time of the comparator is normally directly proportional to the load capacitance (C_L) of the comparator and inversely associated to the differential voltage (ΔV_{in}) between the two inputs.

Accepted Manuscript

3.5 Statistical Analysis of noise and input referred offset voltage due to common mode voltage variation (Vcm)

keeping consideration, the effect of variation of input common mode voltage on noise, proposed circuit is less sensitive to V_{cm} variations as gm/I_D of input transistors M12 and M13 is increased with the biasing of transistors in weak inversion region. An input referred comparator noise generated due to threshold voltage ($VTn_{12,13}$) mismatch of transistors M12 and M13 is inversely proportional to the gm/I_D of these transistors as given by the eq. (33) [20]:

$$\sigma_{n,int}^2 \approx \frac{I_D}{g_m} \cdot \frac{4kT_\gamma}{V_{THN} C_X} \tag{33}$$

From equation (1) it is verified that the noise is inversely proportional to the gm/I_D of input transistors and with proper biasing of input transistors in weak inversion region, this ratio is increased to limit the effect of noise. Also, it is worth mentioned that in conventional strong-ARM latch-based comparators, the gm/I_D ratio of around 20/V remains constant but in the proposed comparator due to tail capacitor C_{tail} potential of source voltage Vs is increased which in turn reduces the overdrive voltage and increased value of gm/I_D is obtained.

Further, pre-amplification gain G affects the input referred offset voltage of the comparator. Gain (G) defines the differential voltage gain of comparator when the pre-amplification phase ends. Input referred voltage in preamplification phase can be modelled as $\sigma_{os,preamp}$ which is mainly due to threshold mismatch $V_{Tn12,13}$ of input transistors M12 and M13 and $\sigma_{os,latch}$ is the input referred offset due to latch regeneration phase caused by the $V_{Tn1,2}$ mismatch of M1 and M2. Now, noise due to total input referred offset [15] is given by σ_{OS} given by eq. (34)

$$\sigma_{os} = \sqrt{\sigma_{os,preamp}^2 + \frac{\sigma_{os,latch}^2}{G^2}} \tag{34}$$

Equation (34) shows that increased value of pre-amplification gain G attenuates the offset voltage due to latch regeneration phase to a greater extent which in turn reduces the dependence of proposed comparator on common mode voltage variation.
Also, from [15] root mean square value of comparator noise (σ_n) is given by eq. (35) as:

$$\sigma_n = \sqrt{\sigma_{n,preamp}^2 + \frac{\sigma_{n,latch}^2}{G^2}} \tag{35}$$

Keeping in view these considerations, pre-amplification gain G which is the differential voltage gain at end of pre-amplification phase which affects the comparator offset and noise is optimized for the improved performance of proposed comparator in biomedical scenarios. Gain (G) is given by eq. (36) as:

$$G \approx \frac{\Delta V_{out}}{\Delta V_{in}} \tag{36}$$

At the end of pre-amplification phase maximum charge variation occurs at outp and outn node is shown by eq. (37) as given by:

$$G \approx \frac{\frac{\Delta Q}{C_O}}{\Delta V_{in}} \tag{37}$$

$$\approx \frac{g_m}{I_D}\left\{V_{Tp1,2} + \frac{C_X}{C_O}\left(V_{Tp1,2} + V_{Tn4,6}\right)\right\} \tag{38}$$

Equation (38) depicts that Gain G also depends on the gm/I_D which is a function of common mode voltage variation. Also, preamplification Gain G affects the offset voltage and noise in the comparator. Capacitor Ctail has given the increased value of differential gain by increasing the source potential of transistors M12 and M13 which in turn also improve the gm/I_D of input transistors M12 and M13 in the proposed circuit.

4. SIMULATION RESULTS

Figures 8 and 9 illustrate the transient simulation response of STDLC and DTDLC respectively. In STDLC, an input voltage of 600mV is applied at the V_{inn} node, while the input voltage at the V_{inp} node is 500mV. Clock signal with a period of 1ns is applied. When clk signal is in low state, comparator is in reset mode and it can be observed that the output nodes, namely outp and outn gets charged to the voltage level of V_{DD}. If the clk input is in a high state, comparator moves to comparison phase and output nodes outn and outp starts discharge at varying rate depending on the input voltages provided. As $V_{inn} > V_{inp}$, the voltage outn goes to zero while outp tends to V_{DD}. The delay is measured in between inputs to the outputs of the comparator using of threshold voltage as 0.6V at either of the rising or falling edges.

Accepted Manuscript

Figure 8: Transient analysis simulation of STDLC

In DTDLC also, V_{inn} of 600 mV is applied, and 500 mV is applied as $V_{inp.}$, clk signal of period 1 ns is given, when the clock input of the latched comparator is equal to zero (clk = 0), the output nodes outp and outn are connected to the 0V. During the comparator's comparison phase, when the clock input equals V_{DD}, the outputs that are pre-charged to V_{DD} , discharge at various speeds depending on the input voltages provided. As $V_{inn} > V_{inp}$, the voltage outp goes to zero while outn tends to V_{DD}.

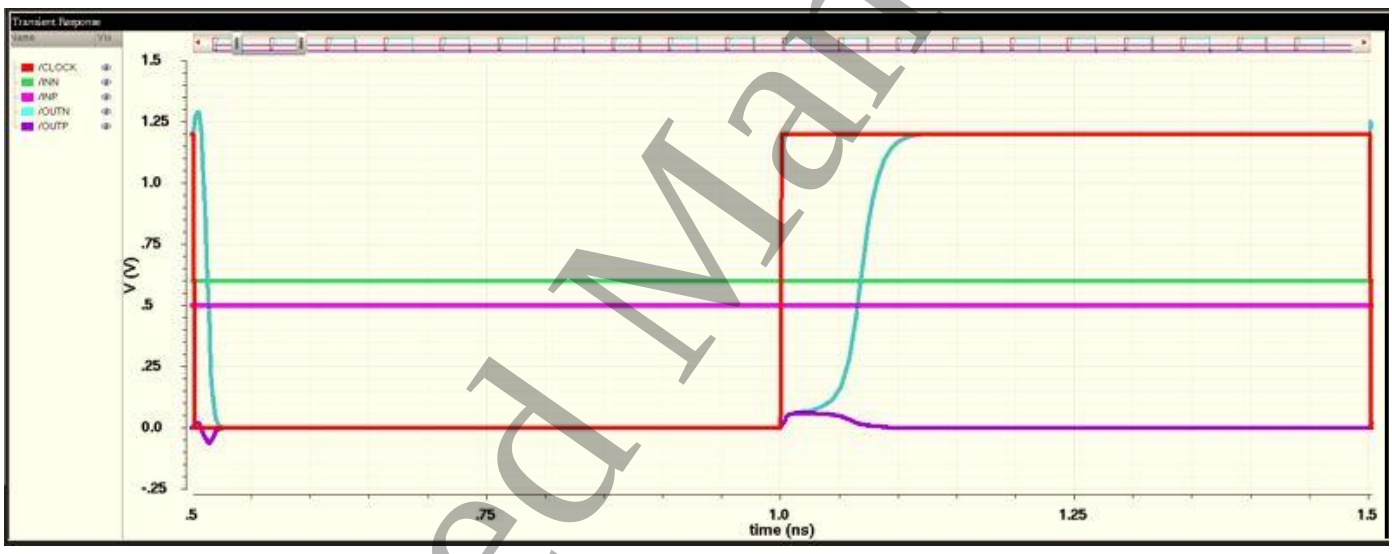


Figure 9: Transient analysis simulation of DTDLC

A comparative analysis of the proposed comparator and the existing comparators has been carried out. All circuits are simulated using 45 nm CMOS technology with V_{DD} of 1 V, a differential input voltage $V_{diff\ equals}$ to 30 mV, and clock signal (clk) of 1 GHz. The corresponding transient response of proposed comparator is shown in Figure 10. The comparator output terminals get connected to V_{DD} whenever clk is set to logic '0.' Further, When the clk is set to logic '1,' the comparator's output node outp switches to zero, and outn pulls up to V_{DD} for the input voltages condition of Vinp>Vinn, and vice versa on Vinp<Vinn. The comparator circuit optimization has been done in terms of the physical dimensions also for the employed transistors with proper adjustment to achieve a consistent offset standard deviation of 0.5 V at the input common-mode voltage.

Accepted Manuscript

Figure 10: Transient analysis Simulation of Proposed Comparator

Proposed comparator handles common mode voltage (Vcm) variations as gm/I_D of input transistors M12 and M13 is increased with the biasing of transistors in weak inversion region with the help of optimized voltage differential gain (G).

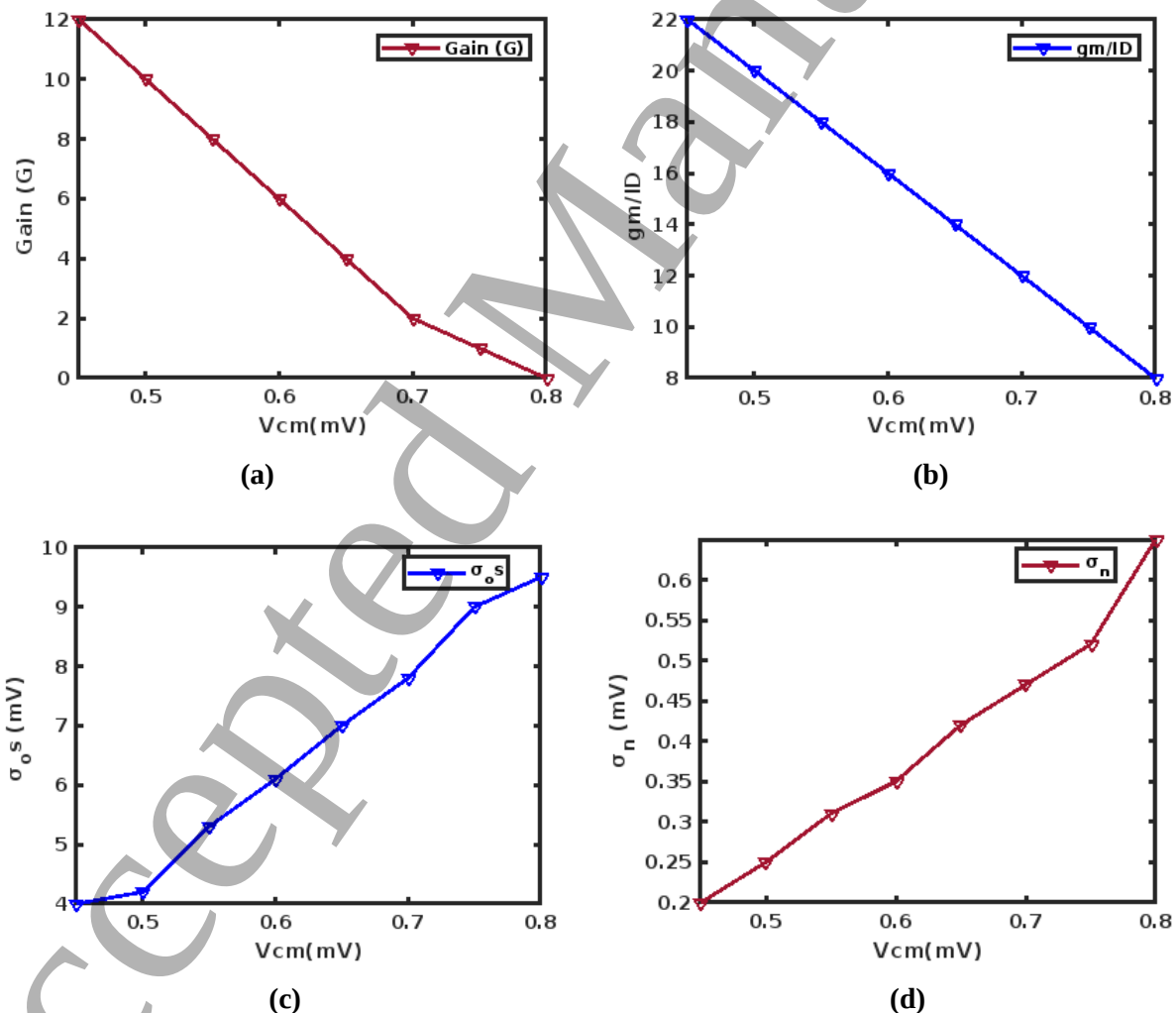


Figure 11: (a) Gain (G) vs V_{cm} (b) gm/I_D vs V_{cm} (c) σ_{os} vs V_{cm} (d) σ_n vs V_{cm} of proposed comparator

With increased value of differential gain and gm/I_D , the input referred offset voltage gets reduced along with increase in the noise rejection capability of the proposed design as also verified by the equation. Figure 11 shows the Gain(G), gm/I_D, rms value of input referred offset voltage (σ_{os}) and rms value noise (σ_n) variations with respect to the common mode voltage.

Accepted Manuscript

4.1 Pre-layout Simulation of Proposed Comparator

Table 2 comprises of summary of the results obtained for various parameters related to the proposed comparator with the given specifications as Vcm=0.7V, Vdiff=20mV, clk= 1GHz, V_{DD}=1V.

Table 2: Performance of the Proposed Comparator

Parameter	Value
Supply voltage (V)	1V
Technology(nm)	45
Operating frequency **(**GHz)	1
Power dissipation (μW)	9.36
Offset voltage (mV)	7.8
Delay(ps)	48.2
Power delay product (PDP) (fJ)	0.451
Kickback Noise(mV)	0.47

Table 3: Comparison of Proposed Comparator with the existing comparators using pre-layout simulation

Comparator Properties	[38]	[39]	[40]	[41]	[42]	Proposed comparator
Technology (nm)	45	45	45	45	45	45
Power dissipation(μW)	8.41	26.62	72.92	18.32	15.63	9.36
Delay (ps)	77.7	66.4	71.5	57.6	73.4	48.2
Kickback noise (mV)	1.1	0.72	1.19	0.93	0.68	0.47
Offset voltage(mV)	8.4	9.4	11.8	10.4	8.9	7.8
Power Delay Product **(**PDP) (fJ)	0.653	1.766	5,213	1.055	1.147	0.451

In contrast, Table 3 assess performance metrics for the frameworks mentioned in the proposed comparator, including delay, power consumption, and PDP. The performance enhancement of the suggested comparator design is shown in Table 3, where a comparison is made with the currently available comparators in terms of power consumption, latency, kickback noise, offset voltage, and power delay product. Based on the data shown in Table 3, it can be noticed that the suggested comparator has a lower latency in comparison to the current systems. In the context of double tail topologies, it is observed that the impact of changes in the input common mode voltage on delay during the comparison phase is quite minimal compared to other topologies.

Therefore, a wider input common mode voltage range is chosen. Simulations of the proposed comparator are carried out as a function of different input common mode and differential voltages with V_{DD} of 1V and clock frequency of 1 GHz. The power consumption of the proposed Comparator is slightly higher when compared to the comparator in [38]. Apart from this suggested comparator effectively reduces power consumption by atleast 66.98% when compared with [42]. But the proposed comparator has less PDP value by 44.44%. Also suggested comparator effectively reduces offset voltage by 7.6% when compared with [31] but with a corresponding increase in the area by 5%. Additionally, the comparator successfully reduces kickback noise by 44.6% when compared to [32] The low-power behavior of the suggested technique is validated using analytical derivations, PVT corner analysis, and post-layout simulation.

The pre-layout simulation results for the delay and power of the proposed comparator in addition to the referenced existing comparators measured about the input differential voltage V_{diff} are presented in Figure 12 and 13, respectively. As observed by the results shown in Figure 12, the delay exhibited by the proposed comparator is comparatively lower than that of other referenced comparators across the entire range of input differential voltage V_{diff}. The findings presented in Figure 13 show that the power consumption of proposed comparator remain consistently minimal across all ranges. In the context of double-tail current topologies, it is generally observed that the comparator delay is less susceptible to variations in input common-mode voltage when compared to the conventional dynamic topology. As a result, the former exhibits a more comprehensive applicable common-mode voltage range. Also, it is noticed from

Accepted Manuscript

Figure 14, the power delay product of the proposed comparator is low when compared to the other architectures across all the differential input voltages. Further, simulation and parametric analysis of proposed comparator is conducted to analyse the parameters of interest as a function of input differential voltage V_{diff} with varying common-mode voltage levels. Also, supply voltage V_{DD} is kept at 1 V and clock frequency of 1 GHz is maintained.

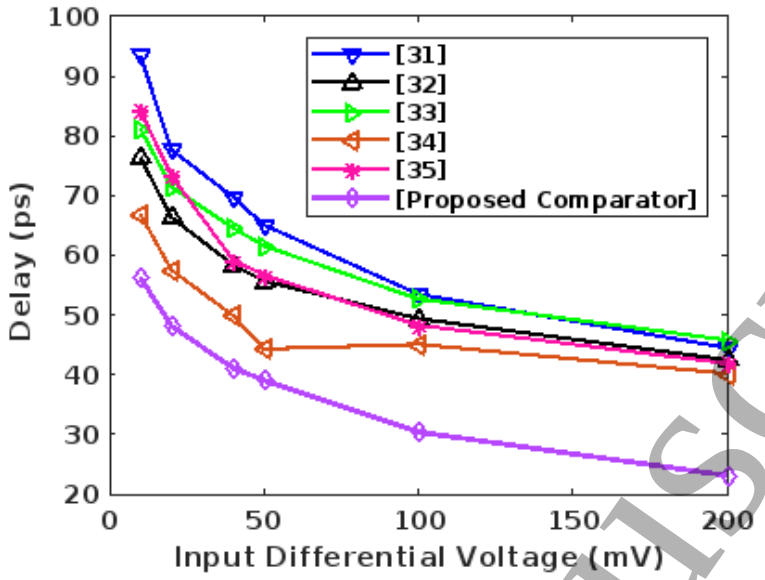


Figure 12: Delay versus input differential voltage (V_{diff}) of the proposed comparator

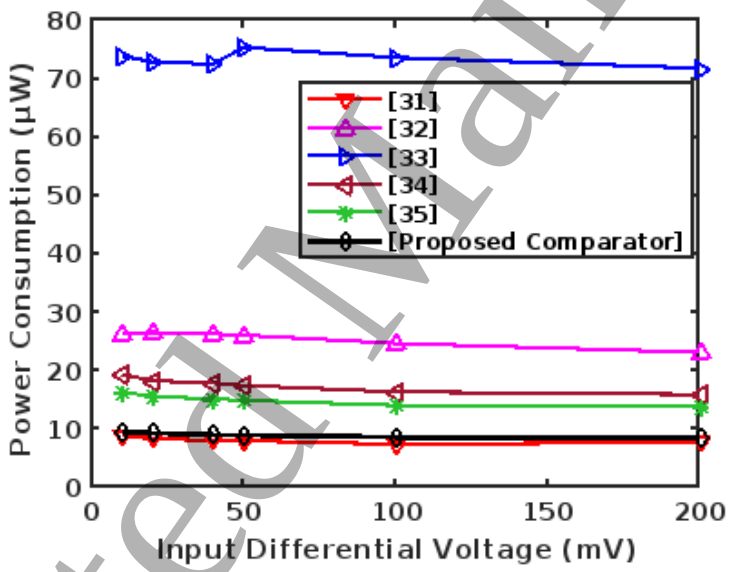


Figure 13: Power consumption versus input differential voltage (V_{diff}) of the proposed comparator

Accepted Manuscript

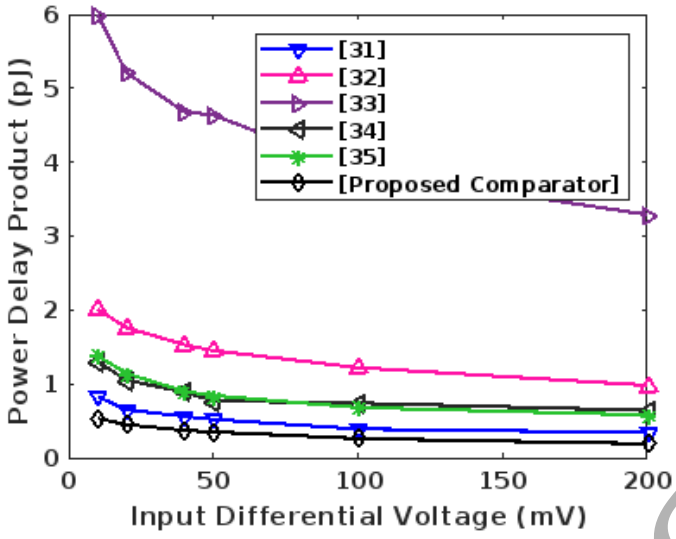


Figure 14: Power Delay Product (PDP) versus input differential input Voltage (V_{diff}) of the proposed comparator

The graphical representation in Figure 15 illustrates the correlation between the delay of the comparator under consideration and the input differential voltage levels with varying common mode voltages. At a specified V_{DD} of 1 V, the delay is found to be 48.2 ps when the differential input voltage level (V_{diff}) is maintained at 20 mV for a common mode voltage of 0.7V. The observed delay exhibits a reduction from 56.3 ps to 23.17 ps as the input differential voltage (V_{diff}) ranges from 10 mV to 200 mV as depicted in Figure 15 for varying common-mode voltage levels (Vcm). The delay is found to follow a decreasing behaviour with larger differential voltage difference. The impact of the common-mode voltage (Vcm) on power dissipation at different values of V_{diff} has been simulated, and the result outcomes have been depicted in Figure 16. The observation can be made that power dissipation shows a decreasing behaviour as the value of Vcm rises from 0.5 to 1 V for a specific value of input differential voltage V_{diff}.

Nevertheless, when the Vcm swing ranges from 0.7 to 1 V, there is an increase in power dissipation although still lower than that observed for the preceding amplitude. When the differential input voltage variation (V_{diff}) is 20 mV, the power dissipation decreases from 9.36 µW at a common-mode voltage (V_{cm}) of 0.7 V to 10.06 µW at a V_{cm} of 0.8 V. As illustrated in Figure 16, the proposed comparator under consideration exhibits reduced dependence on V_{cm} owing to the lack of static power dissipation.

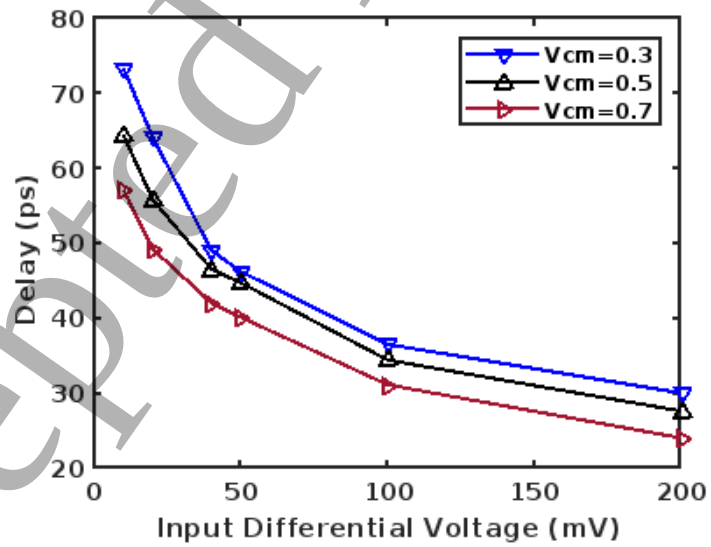


Figure 15: Delay versus input differential voltage (V_{diff}) of the proposed comparator at different V_{cm}values

Accepted Manuscript

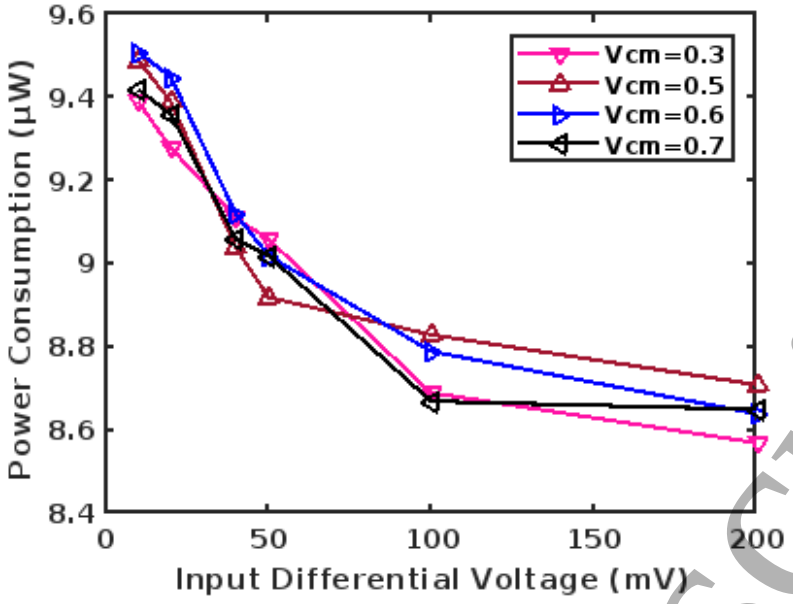
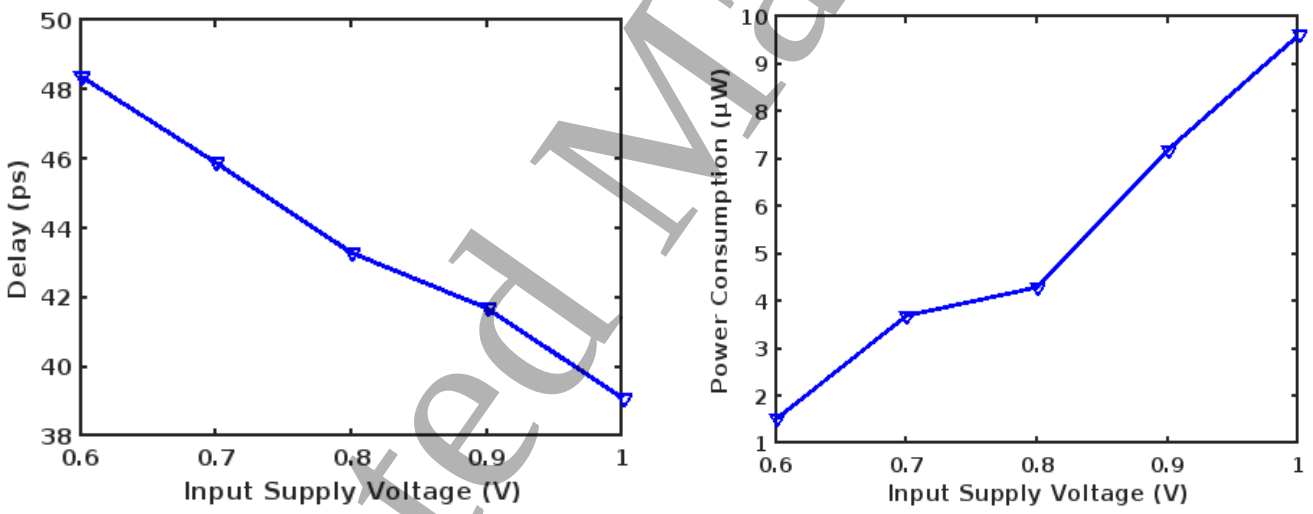


Figure 16: Power consumption versus input differential voltage (V_{diff}) of the proposed comparator at different V_{cm} values

The simulation examines the impact of varying supply voltages V_{DD} on power dissipation and delay. The result outcomes are depicted in Figure 17 that demonstrates a significant relationship between the supply voltage and the delay exhibited by the circuit. It is observed that a rise in the V_{DD} leads to the decrease in delay exhibited by the circuit. In this work, the supply voltage V_{DD} is varied from 0.6V to 1V.



(a) **(b)**

Figure 17: (a) Delay versus input supply voltage (b) Power consumption versus input supply voltage at Vcm = 0.5 V, clk = 1GHz

Variation of kickback noise w.r.t Input Differential Voltage:

The relationship between increasing input differential voltage V_{diff} and reduced impact of kickback noise in our system is evident from the figure 18. During the pre-charged period, the input differential voltage can be applied and amplified without any occurrence of disruption as the process of regeneration has not yet commenced. During the first stage of the comparison stage, the activation of these auxiliary transistors facilitates the commencement of the decision-making and subsequent regeneration process. Additionally, the transistors operating in the triode region are connected in parallel with the input differential pair transistors. This arrangement aids in maintaining a reduced voltage variation at the drain of the input transistors, hence resulting in decreased kickback noise.

Accepted Manuscript

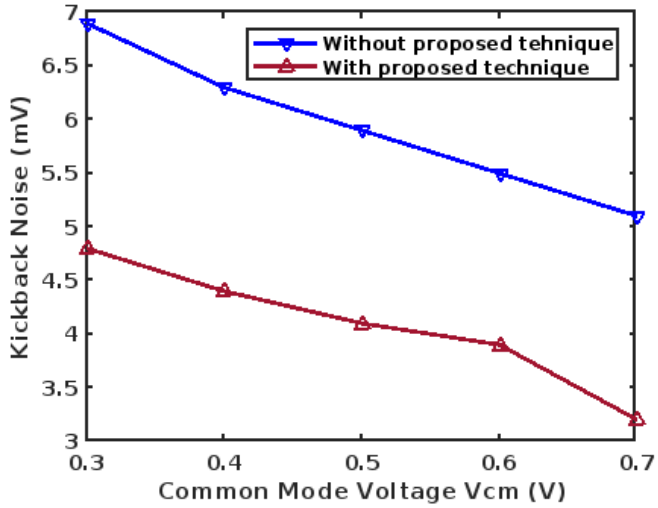
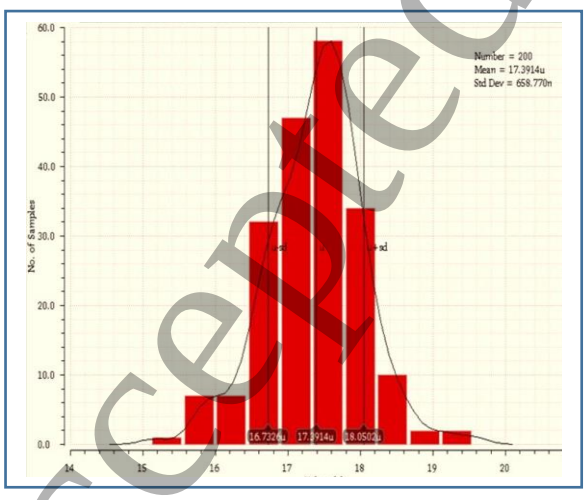


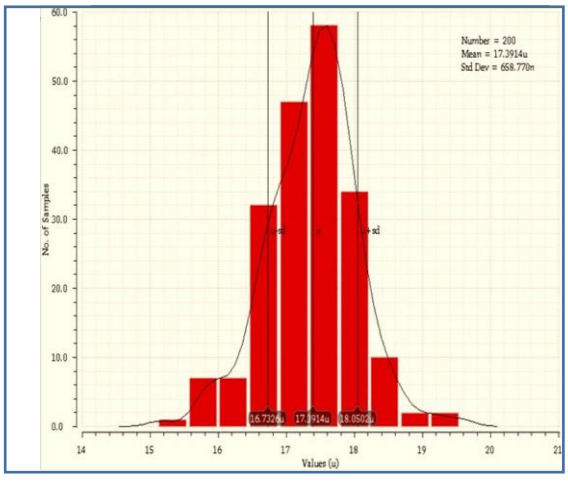
Figure 18: Variation of Kickback noise w.r.t differential Voltage

Comparators operating at fast speeds and exhibiting high power efficiency tend to emit a more significant amount of kickback noise [34]. Figure 18 illustrates the relationship between the kickback noise and the input common mode voltage difference (V_{cm}) for the suggested comparator. An increase in V_{cm} leads to a reduction in the noise impact in our design. As the input Common mode voltage increases, there is a reduction in the kickback noise due to the impact of less voltage variations on the input transistors of the comparator. The input differential voltage can be applied and amplified during the pre-charged period without disturbance, as the regeneration process has not yet commenced. During the first stage of the comparison phase, the activation of auxiliary transistors facilitates the initiation of the decision-making and subsequent regeneration process [1,33]. Additionally, the transistors operating in the triode region are connected in parallel with the input differential pair transistors. This configuration aids in maintaining a lower voltage variation at the drain of the input transistors, hence reducing the occurrence of kickback noise. The proposed comparator demonstrated reduced kickback noise by employing auxiliary transistors M8 and M9 to separate the output nodes. Additionally, this approach effectively mitigated the mismatch effect and minimized the impact of parasitic capacitance associated with the transistor design.

Further, monte carlo simulation of proposed design is performed all corners like tt, ff, ss, sf, fs. Also, the monte carlo simulation plots for power consumption of the proposed design at multiple corners are given in below fig. 19. In ff process corner the mean power consumption is 17.39 µW with standard deviation of 658.77n as shown in fig. 19 (a). In fs process corner the mean power consumption is 9.929 µW with standard deviation of 658.77n as shown in fig.19 (b).

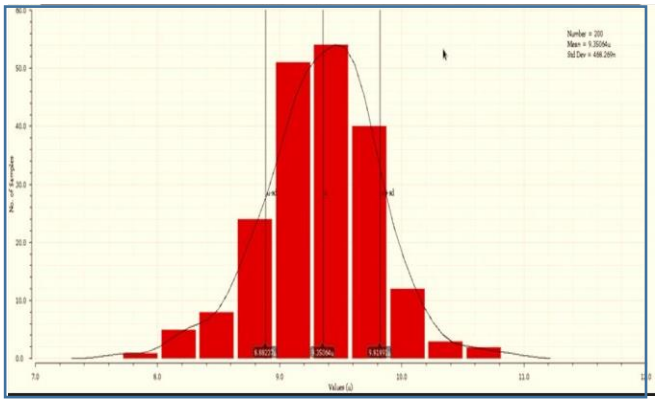


(a) ff corner

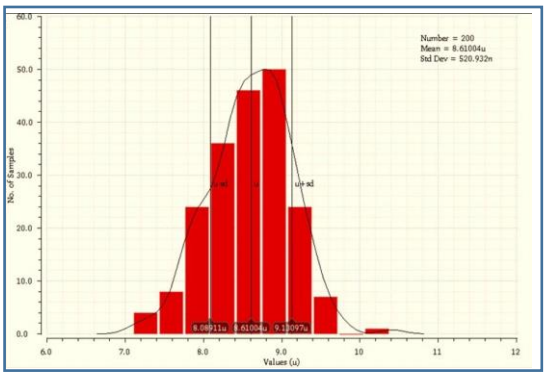


(b) fs corner

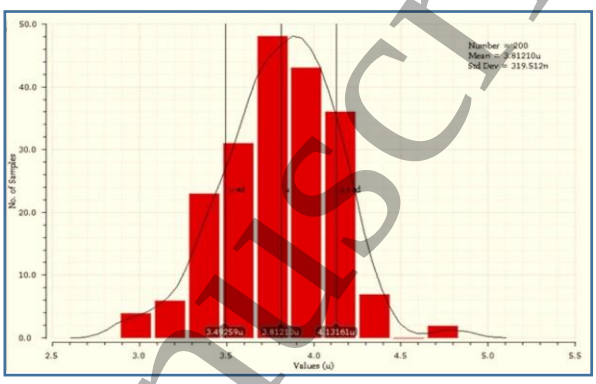
Accepted Manuscript



(c) tt corner



(d) sf corner



(e) ss corner

Figure 19: Monte carlo plot for power at multiple process corners

Table: 4 Summmary of monte carlo power analysis at multiple corners

	tt	ss	sf	fs	ff
Power Consumption (μW)	9.5604	3.8121	8.6104	9.929	17.3914
Standard deviation	468.26n	319.51n	520.92n	658.77n	658.77n

Similarly, the monte carlo power analysis results for other corners like tt, sf and ss are summarised in table, so for biomedical applications, under tt corner analysis the mean power of propsoed design is 9.56 μW.

Monte Carlo simulations for Delay

Similarly, the monte carlo analysis is carried out for delay. Summary of monte carlo analysis is with mean delay in ps and standard deviation is presented in table 5.

Table: 5 Summmary of monte carlo delay analysis at multiple corners

	Ss	**sf**	**tt**	**fs**	**ff**
Delay (ps)	107.67	101.54	48.2	44.27	40.229

Further, simulation is carried out to demonstrate the offset voltage exhibited by the designed comparator circuit. Figure 20 illustrates that standard deviation associated with the offset of the comparator proposed in this work, measuring at 7.8 mV. The data was acquired via Monte Carlo simulations consisting of 300 samples.

Accepted Manuscript

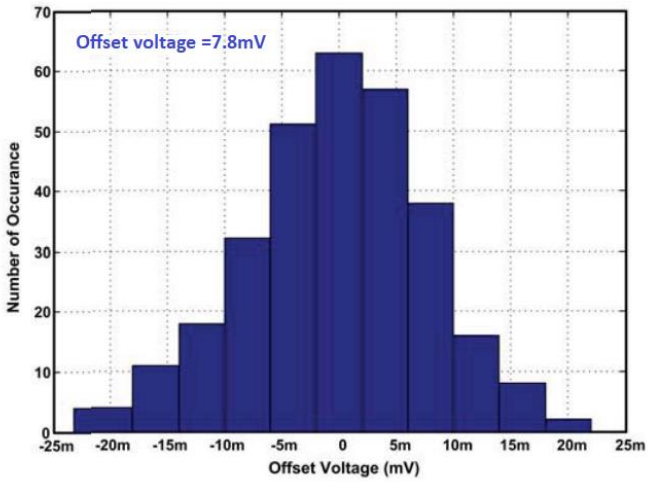


Figure 20: Monte Carlo Analysis of Offset Voltage for 300 Samples

In order to assess the efficacy of the proposed comparator, pre-layout and post-layout computations were conducted for both conventional and recommended designs. The placement of components was carefully considered to minimise any potential effect on the comparator's latency and power.

4.2 Post –Layout Simulation of Proposed Comparator

In order to provide additional validation of the advantages of the suggested approach, the standard comparator that was proposed was implemented. The schematic layout of the proposed comparator is illustrated in Figure 21. The layout is designed with a high degree of symmetry in order to reduce the adverse impact of parasitic elements on the performance of the comparator. Table 6 displays a comparison between the simulation results obtained before and after the layout design. The findings indicate that the estimated area demand of the suggested comparator is either similar or minimal overall when Vcm = 0.7V, V_{DD}=1V, and V_{diff}=20mV.

Table 6: Summary of the Comparator's Performance with respect to Pre-layout and post- layout simulations

Comparator Properties		**STDLC**	**DTDLC**	**Proposed comparator**
Power dissipation (μW)	Pre-Layout	8.41	26.62	9.36
	Post –Layout	10.53	28.77	11.93
Delay (ps)	Pre-Layout	77.7	66.4	48.2
	Post –Layout	79.37	67.39	49.3
Area		7.37 μm X 7.75 μm	7.47μm X 7.5μm	7.18 μm X 8.22 μm

Accepted Manuscript

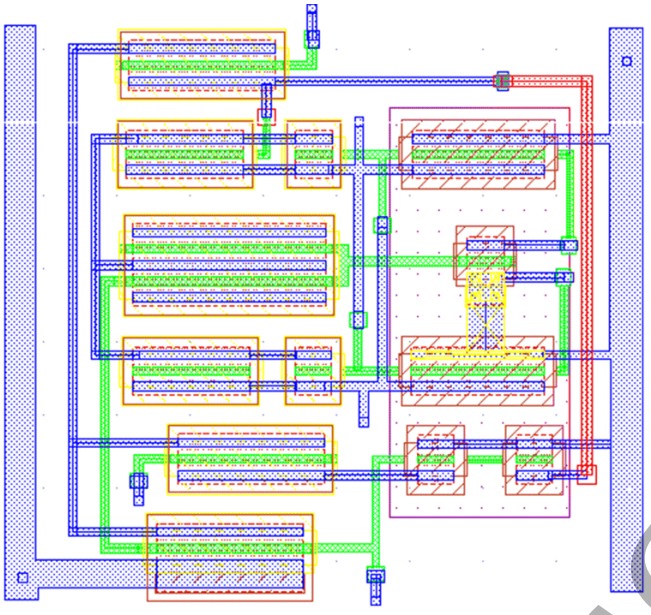
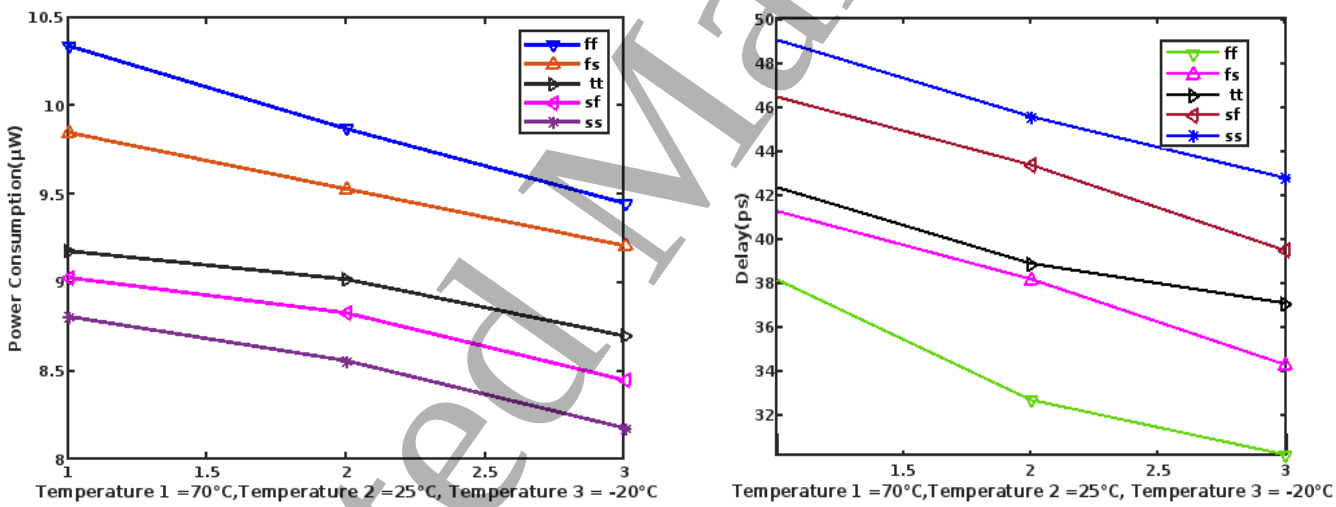


Figure 21: Layout of proposed comparator

Corner Analysis:

This section presents corner simulations to validate the efficacy of the proposed strategy across several corners, as depicted in Figure 22. Simulations were conducted using extreme process, volume, and temperature (PVT) conditions. For this test set up has been used, the entire simulation time has been set to 1000 ns and for each 50 ns a test sequence has been given as input.



(a) (b)

Figure 22: Process Corner Variations on (a) Power Consumption and (b) Delay for the Proposed Comparator

The inputs inn and inp are given such that the differential input voltage lies in boundary conditions of either very small or very high i.e 300 mV and 300 μV. In corner simulation PVT values are set to different process (ff, fs, ss, sf, tt), and temperature ranging from -20°C to 80 °C (-20 °C, 30 °C and 80 °C). First, the circuit was simulated for slow-slow (ss) process at T (-20 °C, 30 °C and 80 °C, then followed by process variations of slow-fast (SF), typical -typical (TT), slow-fast (SF), fast-fast (FF) at (-20 °C, 30 °C and 80 °C). Also, Figure 22 demonstrates that at ff, there is a significant increase in power consumption and a decrease in delay, but at ss, there is a decrease in power consumption and an increase in delay. Figure 23 demonstrates an inverse correlation between the delay of the suggested comparator in post layout state and input differential voltage for multiple Vcm.

Accepted Manuscript

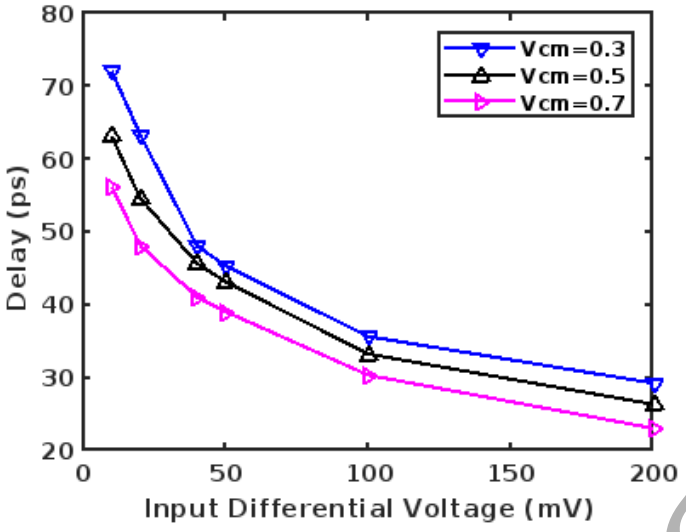


Figure 23: Variation of Delay versus input differential voltage of proposed comparator post layout for various Vcm

5. Conclusion

A novel comparator with reduced power consumption and minimized kickback noise has been suggested for bio-medical applications. Current research introduces a technique for reducing the adverse effects of kickback noise in dynamic comparators. The suggested technique is characterized by its ease of implementation and lack of necessity for an additional design process. The initial step involves the design of a latched comparator. Subsequently, a proposed technique for minimizing kickback noise is implemented by incorporating sampling switches between the latch and the input transistors, employing a straightforward sizing approach. The main concept of the recommended comparator has been utilizing the novel resetting approach and kickback noise approach to preserve the charge and kickback noise, respectively. This approach aids in minimizing delay, power consumption, and kickback noise. The name of this novel method is shared charging logic. The charging-sharing technique involves using a single-pass transistor positioned across two output nodes. The pass transistor distributes the charge evenly across two output nodes throughout the reset phase. Due to the shared charging of the load capacitances, the voltage at the output nodes would not drop under the threshold value. As a result, the input signal could be evaluated more quickly during the regenerative stage. The comparator's delay is calculated using a rigorous statistical analysis considering the circuit's random factors. The offset of the proposed comparator is verified using thorough Monte Carlo simulations. The comparator reduces power while preserving noise. Also, the recommended comparator outperforms state-of-the-art comparators regarding kickback noise, delay, power consumption, and power delay product (PDP). The input voltage inversely affects the comparator's delay. The simulation results confirm that the suggested comparator effectively reduces power consumption and offset voltage by 3%, but with a corresponding increase in the area by 5%. Additionally, the comparator successfully reduces kickback noise by 70% and exhibits a reduction in speed of approximately 15%. The low-power behavior of the suggested technique is validated using analytical derivations, PVT corner analysis, and post-layout simulation. The recommended comparator is yet to be fabricated on silicon and simulation results of pre-silicon stage are proposed to be tested and compared with the post-silicon fabricated comparator in future.

Also, the low-power behavior of the suggested technique is validated using analytical derivations, PVT corner analysis, and post-layout simulation. Further work is suggested to be done on bi-directional dynamic comparators to make amplification process independent of PVT and limiting of common mode voltage variations in the design of ultra-low power comparator.

References

[1] T.-Y. Wang, M.-R. Lai, C. M. Twigg, S.-Y. Peng, “A fully reconfigurable low-noise biopotential sensing amplifier with 1.96 noise efficiency factor,” IEEE Trans. Biomed. Circuits Syst., vol. 8, no. 3, pp. 411–422, 2014.

[2] S.-Y. Peng, L.-H. Liu, P.-K. Chang, T.-Y. Wang, H.-Y. Li, “A power-efficient reconfigurable output-capacitor-less low-drop-out regulator for low-power analog sensing front-end,” IEEE Trans. Circuits Syst. I, Fundamental Theory Applications, vol. 64, no. 6, pp. 1318–1327, Jun. 2017.

[3] S.-Y. Peng, Y.H. Lee, T.Y. Wang, H.C. Huang “A power-efficient reconfigurable OTA-C filter for low-frequency biomedical applications,” IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 65, no. 2, pp. 543–555, Feb. 2018.

Accepted Manuscript

[4] M. Yip, A. P. Chandrakasan, “A resolution-reconfigurable 5-to-10-bit 0.4-to-1 V power scalable SAR ADC for sensor applications,” IEEE J. Solid-State Circuits, vol. 48, no. 6, pp. 1453–1464, 2013.
[5] F. M. Yaul A. P. Chandrakasan, "A 10-bit SAR ADC with data-dependent energy reduction using LSB-first successive approximation," IEEE J. Solid-State Circuits, vol. 49, no. 12, pp. 2825–2834, 2014.
[6] H. Tang, Z. C. Sun, K. W. R. Chew, L. Siek, “A 1.33 μW 8.02- ENOB 100 kS/s successive approximation ADC with supply reduction technique for implantable retinal prosthesis,” IEEE Trans. Biomed. Circuits Syst., vol. 8, no. 6, pp. 844–856, Dec. 2014.
[7] D. Gangopadhyay, E. G. Allston, A. M. R. Dixon, K. Natarajan, S. Gupta, D. J. Allston, "Compressed sensing analog front-end for bio-sensor applications," IEEE J. Solid-State Circuits, vol. 49, no. 2, pp. 426–438, Feb. 2014.
[8] Z. Zhu , Y. Liang, "A 0.6-V 38-NW 9.4-ENOB 20-kS/s SAR ADC in 0.18-μm CMOS for medical implant devices," IEEE Trans. Circuits Syst. I, Fundamental Theory Applications, vol. 62, no. 9, pp. 2167–2176, Sep. 2015.
[9] P. Harpe, E. Cantatore, A. van Roermund, “A 10b/12b 40 kS/s SAR ADC with data-driven noise reduction achieving up to 10.1b ENOB at 2.2 fJ/conversion-step,” IEEE J. Solid–State Circuits, vol. 48, no. 12, pp. 3011–3018, 2013.
[10] M. Ahmadi, W. Namgoong, “Comparator power minimization analysis for SAR ADC using multiple comparators,” IEEE Trans. Circuits Syst. I, Fundamental Theory Applications, vol. 62, no. 10, pp. 2369–2379, Oct. 2015.
[11] M. Elzakker, E. vanTuij, P. Geraedts, D. Schinkel, E. A. M. Klumperink, B. Nauta, “A 10-bit charge-redistribution ADC consuming 1.9uWat 1MS/s,” IEEE JSSC, vol. 45, no. 5, pp. 1007–1015, May 2010.
[12] M. Liu, K. Pelzers, R. van Dommel, A. van Roermund, and P. Harpe, "A 106nW 10b 80kS/s SAR ADC with duty-cycled reference generation 65 nm CMOS," IEEE JSSC, vol. 51, no. 10, pp. 2435–2445, Oct. 2016.
[13] P. Harpe, E. Cantatore, A. V. Roermund, "A 2.2/2.7fJ/conversion step 10/12b 40kS/s SAR ADC with data-driven noise reduction," IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers, San Francisco, CA, USA, pp. 270–271, 2013.
[14] Z. Zhu, Z. Qiu, M. Liu, R. Ding, “A 6-to-10-Bit 0.5 V-to-0.9 V reconfigurable 2 MS/s power scalable SAR ADC in 0.18 μm CMOS,” IEEE Trans. Circuits Syst. I, Fundam. Theory Appl., vol. 62, no. 3, pp. 689–696, 2015.
[15] L. Chen, A. Sanyal, J. Ma, X. Tang N. Sun, "Comparator common-mode variation effects analysis and its application in SAR ADCs," IEEE International Symposium on Circuits and Systems (ISCAS), Montreal, QC, Canada, pp. 2014-2017, 2016.
[16] M. Yousefirad, M. Yavari, “Kick back noise reduction and offset cancellation technique for dynamic latched comparator”, 29th Iranian conference on electrical engineering, pp. 149-153, 2021.
[17] N. Rai, A. Yadav, S. Wairya, "Design of a High Speed and Low Power Charge Shared Based Dynamic Comparator for ADC Application," First International Conference on Advances in Computing and Future Communication Technologies (ICACFCT), Meerut, India, pp. 125-130,2021.
[18] S. A. Halim, N. A. N. B. Z. Abidin, A. A. A. Rahim, "Low power CMOS charge sharing dynamic latch comparator using 0.18μm technology, IEEE Regional Symposium on Micro and Nano Electronics, Kota Kinabalu, Malaysia, pp. 156-160, 2011.
[19] Diaz-Madrid J-A, Domenech-Asensi G, Hauer J, Mateu L, “ A low kickback fully differential dynamic comparator for pipeline analog-to-digital converters,” Engineering Reports, pp. 12055, 2019.
[20] X. Tang, L. Shen, B. Kasap, X. yang, W.Shi, A. Mukherjee, D.Z. Pan, N. Sun., "An Energy-Efficient Comparator with Dynamic Floating Inverter Amplifier," in IEEE Journal of Solid-State Circuits, vol. 55, no. 4, pp. 1011-1022, 2020.
[21] P.-C. Lee, J.-Y. Lin, C.-C. Hsieh, “A 0.4 V 1.94 fJ/conversion-step 10 bit 750 kS/s SAR ADC with input-range-adaptive switching,” IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 63, no. 12, pp. 2149–2157, 2016.
[22] M.tyagi, P.Mittal, P.Kumar, “Design of 8-bit low power SAR ADC in 45nm for biomedical implants,” Physica Scripta, IOP science, vol. 98, 116101, 2023.
[23] Y. Song, Z. Xue, Y. Xie, L. Geng, S. Fan, "A 0.6-V 10-bit 200-kS/s fully differential SAR ADC with the incremental converting algorithm for energy-efficient applications," IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 63, no. 4, pp. 449–458, 2016.
[24] T. Kobayashi, K. Nogami, T. Shirotori, Y. Fujimoto, "A current-mode latch sense amplifier and a static power-saving input buffer for low-power architecture", VLSI Symposium, pp. 28–29, 1992.
[25] B. Razavi, “ The Strong-ARM Latch [A Circuit for All Seasons]”, in IEEE SSC Magazine, vol. 7, no. 2, pp. 12-17, 2015.
[26] D. Schinkel, E. Menin, E. Klumperink, E. van Tuijl, B. Nauta, “A double-tail latch-type voltage sense amplifier with 18ps setup+holdtime,” in IEEE Int. Solid-State Circuits Conf.(ISSCC) Dig. Tech. Papers, San Francisco, CA, USA, pp. 314–605. 2007.
[27] H. S. Bindra, C. E. Lokin, A.-J. Annema, B. Nauta, “A 30fJ/comparison dynamic bias comparator,” in Proc. 43rd IEEE Eur.Solid State Circuits Conf. (ESSCIRC), pp. 71–74, 2017.

Accepted Manuscript

[28] X. Tang, B. Kasap, L. Shen, X. Yang, W. Shi, N. Sun, "An energy-efficient comparator with dynamic floating inverter pre-amplifier," in Proc. IEEE Symp. VLSI Circuits (VLSI), pp. C140–C141, 2019.
[29] B. Razavi, "The design of a comparator [the analog mind]", IEEE Solid-State Circuits Mag., vol. 12, no. 4, pp. 8-14, 2020.
[30] N. Mishra, P. Mittal, B. Kumar, "Analytical Modeling for Static and Dynamic Response of Organic Pseudo All-p Inverter Circuits", Journal of Computational Electronics, vol. 18, no. 4, pp. no. 1490–1500, 2019.
[31] H. S. Bindra, J. Ponte, and B. Nauta, "A 174μVRMS input noise 1 G8/s comparator in 22nm FDSOI with a dynamic-bias preamplifier using tail charge pump and capacitive neutralization across the latch", IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers, pp. 62-63, Feb. 2022.
[32] Babayan-Mashhadi, and Lofti., " Analysis and Design of a Low-Voltage Low-Power Double-Tail Comparator", IEEE Transactions on Very Large-Scale Integration (VLSI) Systems, vol. 22, no. 2, 2014.
[33] Z. Zhu ,Y. Liang, " A 0.6-V 38-NW 9.4-ENOB 20-kS/s SAR ADC inv 0.18 um- CMOS for Medical Implant Devices", IEEE Journal of Solid-State Circuits, vol. 62, no. 9, 2015.
[34] X. Xin, J. Cai, R. Xie, " Ultra-low power comparator with dynamic offset cancellation for SAR ADC", Electron. Lett.,53, (24), pp. 1572–1574, 2017.
[35] Y. H. Hwang, D.K. Jeong.," Ultra-low-voltage low-power dynamic comparator with forward body bias scheme for SAR ADC", Electronics Letters vol. 54 No. 24 pp. 1370–1372, 2018.
[36] X. Tong, W. Jin, C. Zhang, S. Dong, Q. Li, "A 0.6-V 12-bit 13.2-fJ/conversion-step SAR ADC with time-domain VCDL-based comparator and metastability immunity technique", Microelectronics Journal, Volume - 122, 105406, 2022.
[37] S. Nejadhasan, , M. Moghadam, Z. Abiri, "Design of a Dynamic ADC Comparator with Low Power and Low Delay Time for IoT Application", Wireless Personal Communication 123, 1573–1591, 2022.
[38] S. B. Mashhadi, R. Lotfi, "Analysis and design of a low-voltage low-power double-tail comparator" IEEE Transactions on Very Large-Scale Integration(VLSI)Systems,22, 343–352, 2014.
[39] V. Savani, N.M. Devashree, " Design and analysis of low-power high-speed shared charge reset technique based dynamic latch comparator" Microelectronics Journal, vol. 74, pp. 116–126, 2018.
[40] K. Dubey, R. K. Nagaria, "Low-power high-speed CMOS double tail dynamic comparator using self-biased amplification stage and novel latch stage," Analog Integrated Circuits and Signal Processing, vol. 101, (2), pp. 307-317, 2019.
[41] H. Zhuang, H. Tang, X. Liu. "Voltage Comparator With 60% Faster Speed by Using Charge Pump." IEEE Transactions on circuits and systems—II: Express Briefs, vol. 67, no. 12, 2020.
[42] M. Cao, W. Tang, "The High-Speed Low-Power Dynamic Comparator", Journal of Physics: Conference Series, 2113, 012064, 2021.
[43] X. Tang, J. Liu, Y. Shen, S. Li, L. Shen, A. Sanyal, K. Ragab, N. Sun, "Low-Power SAR ADC Design: Overview and Survey of State-of-the-Art Techniques," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 69, no. 6, pp. 2249-2262, 2022.

Accepted Manuscript

Digital Transformation: Using Technology to Reshape Industries

Dr. Arjun Reddy
Department of Computer Science and Engineering
Delhi Technological University (DTU), India

Abstract

Digital transformation has become a powerful force reshaping industries across the globe, fundamentally altering how organizations operate, compete, and deliver value. This in-depth analysis explores the wide-ranging impact of digital transformation, focusing on key technological drivers such as real-time analytics, edge computing, automation, and cloud infrastructure. It examines the transformative effects of these innovations across various sectors, with particular attention to energy, manufacturing, and healthcare, where traditional processes and operational models have been revolutionized. The paper also addresses the implications for the workforce, highlighting changes in job markets, new opportunities, and the urgent need for reskilling in the digital era. Looking ahead, it explores the future of business competitiveness, the integration of sustainability, and offers strategic recommendations for organizations navigating the digital landscape. The findings underscore that digital transformation is not just about adopting new technologies but also about reimagining business models, enhancing customer engagement, and creating new value propositions.

Keywords: Digital Transformation, Industry Innovation, Workforce Development, Cloud Computing, Sustainability Integration

I. Introduction

A significant change in how businesses use technology to drastically increase performance and reach is known as digital transformation. The IEEE Digital Reality Initiative claims that this change goes much beyond simple digitalization and includes a thorough rethinking of value generation models, consumer experiences, and corporate processes [1].

The combination of cutting-edge technology and changing business requirements is causing this revolution to reshape the fundamental basis of contemporary industry. Three crucial aspects are where the fundamental effects of digital transformation are evident. First, in company operations, where data-driven decision-making and intelligent automation are transforming conventional procedures. The observable advantages of this technical advancement are demonstrated by the operational efficiency gains of up to 40% recorded by organizations executing digital transformation efforts [2].

Second, dynamic, real-time insights have replaced static analysis in data interaction. 5G networks and smart city projects have combined to provide previously unheard-of data processing capabilities that let businesses analyze enormous amounts of data quickly and accurately. The way that companies gather, examine, and use data to inform strategic choices has been drastically altered by this technical breakthrough.

Third, the rise of new digital business models and platforms for consumer involvement has changed the way value is delivered. The IEEE Digital Reality Initiative emphasizes how businesses are embracing platform-based strategies more and more, establishing ecosystems that enable smooth communication between suppliers and customers. With businesses reporting improvements in customer satisfaction measures of up to 30% thanks to digital

engagement platforms, this change has resulted in the creation of new income streams and improved customer experiences.

II. Key Technological Drivers Real-Time Analytics and Edge Computing

Real-time analytics skills are at the heart of today's digital transformation, transforming how businesses handle and respond to data. Recent IEEE study on AI edge analytics indicates that operational efficiency has increased by up to 60% for firms using real-time analytical systems [3]. This development is especially noticeable in data-driven decision making, where edge computing allows massive data streams to be processed instantly at the source.

AI-enabled edge analytics has revolutionized operational monitoring by enabling companies to identify and address irregularities instantly. In manufacturing settings, where predictive maintenance solutions have decreased equipment downtime by up to 45% [3], this feature has proven very useful. Similar benefits have been seen in market responsiveness, where businesses use real-time data to dynamically modify price and inventory levels in response to current market circumstances.

One of the most important uses of real-time analytics is environmental impact optimization. Sensor networks and analytics platforms are being used by organizations more and more to track and lower their carbon footprint. Some have even been able to cut their energy use by up to 25% by using real-time optimization techniques.

Innovation in Processes and Automation Technologies Industry-wide changes have been significantly accelerated by the use of automation technology. Operational efficiency has significantly increased as a result of process simplification through intelligent automation; companies have reported processing times for common processes being reduced by up to 35%. With some firms reporting defect rate reductions of up to 90%, automated quality control and verification systems have greatly improved output quality by reducing errors.

Scalability and Cloud Infrastructure

Cloud computing, which offers previously unheard-of scale and flexibility, has become the foundation of digital transformation projects. Strategic cloud adoption can result in cost reductions of 25–45% for enterprises, according to IEEE's complete framework for cloud computing infrastructure [4]. Because cloud solutions provide dynamic resource allocation based on real consumption patterns, this cost-effectiveness is especially noticeable in data processing and storage capacities. Cloud services have made technology more accessible and leveled the playing field for companies of all sizes. Without requiring a large initial investment, small and medium-sized businesses may now access enterprise-grade computing capabilities, fostering innovation and giving them a competitive edge. Cloud infrastructure has made it possible for businesses to expand internationally while upholding constant security and performance requirements.

Key benefits of cloud-based infrastructure include:

- Reduced capital expenditure through pay-as-you-go models
- Enhanced data security through distributed architecture
- Improved collaboration capabilities across geographic boundaries

Technology Area	Key Performance Indicator	Improvement Rate
Real-Time Analytics	Operational Efficiency	Up to 60%
Edge Computing	Equipment Downtime Reduction	Up to 45%
Environmental Analytics	Energy Consumption Reduction	Up to 25%
Process Automation	Processing Time Reduction	Up to 35%

Quality Control Automation	Defect Rate Reduction	Up to 90%
Cloud Computing	Cost Reduction Through Adoption	25-45%

Table 1: Digital Transformation Impact Metrics Across Technologies [3, 4]

III. Cross-Industry Applications Energy Sector Evolution through Digital Innovation

Digital technologies have significantly changed the energy business, especially in the areas of production optimization and monitoring. Recent IEEE study on energy production monitoring claims that real-time analytics and sophisticated sensor networks have made it possible to see energy creation processes with never-before-seen clarity [5]. Businesses using this technology have claimed a 25% decrease in unscheduled downtime and a 30% increase in production efficiency.

By combining artificial intelligence and smart grid technology, energy consumption optimization has advanced to a more complex level. These systems automatically modify power generation in response to current demand, improve distribution networks, and continually study consumption trends. Major utilities have seen average energy savings of 15-20% as a consequence of the installation of such systems.

Digital monitoring and control technologies have greatly improved environmental impact reduction efforts. Energy businesses have been able to maintain production levels while reducing their carbon footprint by up to 40% because to real-time emissions tracking and predictive analytics. This accomplishment highlights how important digital transformation is to the generation of sustainable energy.

Manufacturing Sector Advancement

Automation of assembly lines and sophisticated robots have brought about a paradigm change in the industrial industry. Production processes have been transformed by the integration of IoT sensors and real-time quality control systems, which has increased product quality and decreased waste. Businesses that have used this technology have claimed up to 35% gains in productivity.

Innovation in Healthcare and Digital Integration

With automated data input technologies radically altering the way patient information is maintained, the healthcare industry has emerged as one of the main beneficiaries of digital transformation. Contextual data input methods have improved accuracy by 60% and decreased documentation time by up to 45%, according to IEEE study on healthcare data systems [6]. Healthcare professionals may now devote more time to patient care rather than administrative duties thanks to this efficiency boost.

AI-powered diagnostic advancements have increased the precision and speed of medical diagnosis. Healthcare professionals can gain complete insights from these systems' simultaneous analysis of test findings, patient histories, and medical imaging. Putting such procedures in place has resulted in:

- 40% reduction in diagnostic errors
- 30% decrease in time-to-diagnosis
- 25% improvement in patient outcomes

Administrative optimization through digital workflows has streamlined healthcare operations, reducing wait times and improving patient satisfaction. Integration of electronic health records with automated scheduling and billing systems has resulted in: ● 50% reduction in administrative processing time

- 35% decrease in billing errors
- 40% improvement in appointment scheduling efficiency

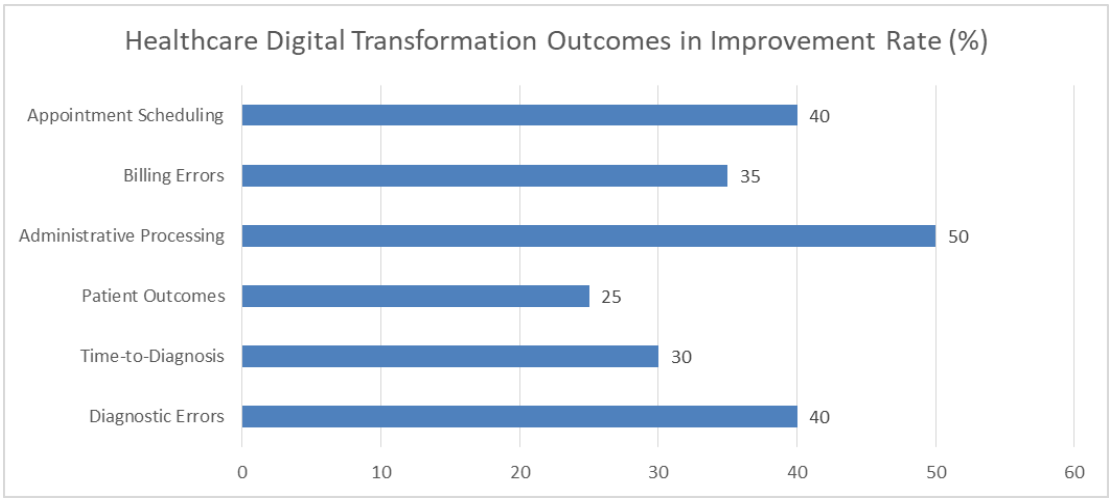


Fig 1: Ideal for a radar chart showing the comprehensive impact of digital transformation in healthcare in Improvement Rate (%) [5, 6]

IV. Workforce Impact Emerging Job Opportunities in the Digital Era

The work market has seen a dramatic change as a result of the digital transition, with previously unheard-of opportunities in cutting-edge technology disciplines. Job postings for specialist digital professions have increased by 300% between 2019 and 2021, per current IEEE study on the demands of the ICT labor market [7]. The increasing need for knowledge in crucial areas of digital transformation is reflected in this spike. Organizations are actively looking for experts that can convert complicated data into useful insights, making data science one of the most sought-after areas. Since 2012, the need for data scientists has increased by 650%, and throughout the last five years, median pay have increased by 40%. Specialists in cloud computing are also in great demand; job listings for cloud architects and engineers have increased by 200% in the past year.

With businesses making significant investments in machine learning and artificial intelligence capabilities, AI specialization has become more and more important. AI-related job advertisements have increased by 71% in the market, especially in industries like manufacturing, healthcare, and finance. With a predicted 350% rise in employment prospects by 2025, cybersecurity specialists are in more demand than ever before due to the development of digital infrastructure and the complexity of cyberthreats.

Labor Market Transformation and Future of Work

Recent events throughout the world have expedited the labor market's transition, radically altering the way work is handled and carried out. 85% of companies have permanently changed their hiring procedures to allow for remote employment, according to an analysis of the U.S. IT labor market during the epidemic [8]. The necessary abilities have evolved significantly as a result of this change, with a focus on remote project coordination, virtual team management, and digital collaboration.

Enabling remote work has become commonplace; since 2019, companies have reported a 400% rise in remote job advertisements. This change has

- Expanded access to global talent pools
- Reduced geographical barriers to employment
- Enhanced work-life balance opportunities
- Decreased operational costs for organizations

Global talent access has been revolutionized through digital platforms and remote work technologies. Organizations can now tap into expertise regardless of location, leading to:

- 60% increase in international hiring
- 45% reduction in time-to-hire
- 30% improvement in candidate quality
- 25% decrease in recruitment costs

Reskilling requirements have become increasingly important as technologies evolve. Organizations are investing heavily in digital training programs, with: ● 70% increase in online learning platform adoption

- 85% of employees requiring significant reskilling by 2025
- 40% of workforce participating in upskilling programs annually
- 55% of organizations implementing continuous learning initiatives

Category	Metric	Percentage (%)	Impact Area
Remote Work Adoption	Organizations Modified Hiring	85	Work Structure
Global Hiring	International Recruitment Increase	60	Talent Access
Recruitment Efficiency	Time-to-Hire Reduction	45	Process Improvement
Talent Quality	Candidate Quality Improvement	30	Hiring Outcomes
Recruitment Costs	Cost Reduction	25	Financial Impact
Learning Platforms	Online Platform Adoption	70	Skills Development
Workforce Reskilling	Employees Needing Reskilling	85	Future Readiness
Current Upskilling	Workforce Participation	40	Skills Enhancement
Organizational Learning	Continuous Learning Programs	55	Development

Table 2: Digital Transformation Impact on Workforce Development [7, 8]

V. Future Implications Evolving Business Competitiveness Landscape

Initiatives for digital transformation are drastically altering how competitive businesses will be in the future. Organizations that have completely embraced digital transformation are seeing a 45% greater rise in their market share than their rivals, citing current IEEE study on corporate information systems [9]. Improved client experiences, increased operational efficiency, and data-driven decision-making skills are the main sources of this competitive advantage. Modern companies are using AI and sophisticated analytics more and more to:

- Predict market trends with 85% accuracy

- Reduce operational costs by 30-40%
- Improve customer retention rates by 25%
- Accelerate product development cycles by 50%

Sustainability Integration and Environmental Impact

The integration of sustainability initiatives with digital transformation has become a critical factor in longterm business success. Recent studies on environmental sustainability in digital operations show that smart integration of renewable technologies and digital systems can reduce operational carbon footprints by up to 60% [10]. Organizations are increasingly adopting: ● Smart grid technologies for optimal energy management

- AI-driven resource optimization systems
- Digital twin solutions for environmental impact modeling
- Blockchain-based sustainability tracking

Workforce Development and Future Skills

The workforce of the future will require a unique blend of technical expertise and soft skills. Organizations are preparing for this shift by:

- Investing 15-20% of annual budgets in employee upskilling
- Creating hybrid learning environments
- Developing AI-assisted training programs
- Implementing virtual reality-based skill development

Industry Innovation and Digital Ecosystems

Innovation in various industries is accelerating through digital transformation, creating new business models and opportunities. Key trends include:

- Platform-based business models showing 300% higher growth rates
- Cross-industry collaboration through digital platforms
- Emergence of industry-specific digital marketplaces
- Integration of IoT and edge computing solutions

Strategic Recommendations for Organizations

To maintain competitiveness in this evolving landscape, organizations should:

1. Develop comprehensive digital transformation roadmaps
2. Invest in emerging technologies while maintaining cybersecurity
3. Foster a culture of continuous innovation and learning
4. Build sustainable and resilient digital infrastructure
5. Prioritize environmental sustainability in digital initiatives

Challenges and Opportunities

While the future presents significant opportunities, organizations must navigate:

- Rapidly evolving technology landscapes
- Increasing cybersecurity threats
- Regulatory compliance requirements
- Talent acquisition and retention challenges

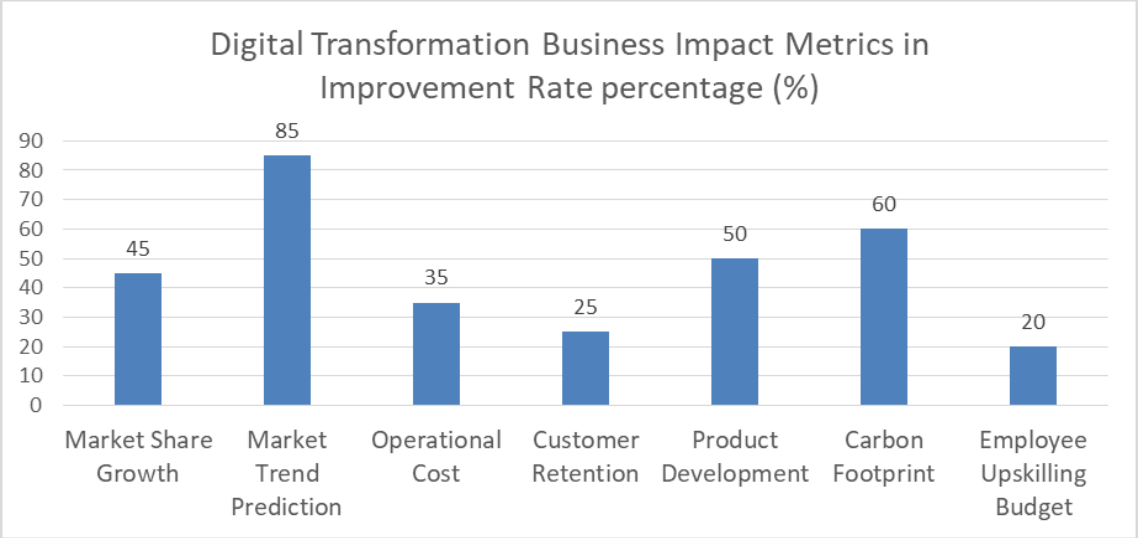


Fig 2: Combination bar and line chart is to show improvement rates and their corresponding business benefits in Improvement Rate percentage (%) [9, 10]

Conclusion

A thorough examination of the effects of digital transformation on several industries shows how fundamentally it has changed the face of modern business. Advanced technology integration has been shown to have revolutionary implications on value generation models, customer engagement, and operational efficiency. In addition to improving corporate skills, the development of real-time analytics, automation, and cloud infrastructure has made cutting-edge technical solutions more accessible to businesses of all sizes. The applications across industries demonstrate how digital transformation requires a fundamental rethinking of business procedures and service delivery methods rather than just a technical advancement. The consequences for the workforce highlight how crucial it is to keep learning and adapting in the digital era, and the sustainability integration shows how important digital technologies are to reaching environmental objectives. The capacity to strike a balance between technology innovation, human capital development, sustainable practices, and adaptable business models is becoming more and more important as businesses continue to navigate this digital transition. In order to build resilient and future-ready enterprises, digital transformation will need careful integration of technology with workforce development, environmental awareness, and business strategy.

References

1. L. Alboul et al., "Digital Transformation - IEEE Digital Reality Initiative White Paper," IEEE Digital
2. Reality Initiative, 2021.
3. M. Tealab, A. Hassebo, A. Dabour, M. AbdelAziz, "Smart Cities Digital Transformation and 5G –
4. ICT Architecture," IEEE Conference Publication, 2020.
5. Vassilis Tsakanikas, Tasos Dagiuklas, "Enabling Real-Time AI Edge Video Analytics," 2021 IEEE
6. International Conference on Communications (ICC), 2021.

7. Jonas Repschlaeger, Stefan Wind; Ruediger Zarnekow, Klaus Turowski, "A Reference Guide to Cloud
8. Computing Dimensions: Infrastructure as a Service Classification Framework," 2012 45th Hawaii
9. International Conference on System Sciences (HICSS), 2012.
10. Aidong Yan, Paul Ohodnicki, Michael Buric, Shiwoo Lee, Ming-Jun Li, Kevin Chen, "In-vivo monitoring of energy chemistry and energy production with high spatial resolution," IEEE Xplore, 2023.
11. Maxime Baas, Stéphanie Bernonville, Christophe Kolski, "Contextual data entry system design in the healthcare domain," IEEE Xplore, 2022.
12. Ioannis Apatsidis, Konstantinos Georgiou, Nikolaos Mittas, Lefteris Angelis, "A Study of Remote and
13. On-site ICT Labor Market Demand using Job Offers from Stack Overflow," 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), 2021.
14. Xuan Wang; Yaojie Li, Thomas F. Stafford, Daqi Xin, "The IT Labor Market Amid the Pandemic: The Case of the United States," IEEE Engineering Management Review, Vol. 49, No. 4, 2021.
15. Serafeim A. Triantafyllou, "Use of Business Information Systems to Achieve Competitiveness," 2022 13th National Conference with International Participation (ELECTRONICA), 2022.
16. Lorenzo Bartolucci, Stefano Cordiner, Vincenzo Mulone, Marina Santarelli,Fernando Ortenzi, Manio
17. Pasquali, "Optimal integration of Renewables and Second-Life batteries to improve the environmental sustainability of Electric Vehicle Fleets," 2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2021.

Knowledge and Information Systems
https://doi.org/10.1007/s10115-024-02273-6

REGULAR PAPER

Digital twin: securing IoT networks using integrated ECC with blockchain for healthcare ecosystem

Vikas Sharma[1] · Akshi Kumar[2] · Kapil Sharma[1]

Received: 5 May 2024 / Revised: 22 July 2024 / Accepted: 16 October 2024
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

Abstract

Digital twin (DT) innovation is becoming increasingly prevalent in numerous areas. This is often particularly genuine in healthcare, which depends increasingly on Internet of Things (IoT) systems. But this combination brings up enormous issues with safety, security, and being able to develop. Since information is private, it is exceptionally critical to keep it secure from individuals who shouldn't have gotten to it and from being stolen. This consideration proposes a modern framework that combines elliptic curve cryptography (ECC) with blockchain innovation to bargain with these issues. The system is implied to ensure IoT systems by making demonstrative apparatuses in healthcare more secure. A genetic algorithm-based random forest (GAO-RF) demonstration is additionally utilized to form include choice work way better, which makes it beyond any doubt that information taking care of and analysis go easily. The GAO-RF show includes the body of unused thoughts by progressing the method of choosing highlights, which is exceptionally critical for proficiently overseeing colossal sums of private information. The show that was put into activity works exceptionally well, as appeared by its execution measurements as F1-score of 97.3%, an exactness of 98.4%, an exactness of 97.3%, a review of 97.4%, an MCC of 97.69%, and a kappa measurement (KS) of 97.31%. These results show an exceptionally strong framework that can handle and protect private health data. The safety and security of understanding information in IoT systems have enormously progressed by including ECC and blockchain in the DT system. A genetic algorithm has been shown to work well in the random forest model for feature selection, which has led to better security methods. This study has big effects on the healthcare field because it gives us a strong way to keep patient information safe. This method creates trust in the healthcare system by making sure that private data are handled in an honest and safe way. It could change how patient data are protected in this digital age.

✉ Kapil Sharma
kapil@ieee.org

Vikas Sharma
vikassharma12387@gmail.com

Akshi Kumar
akshi.kumar@gold.ac.uk

[1] Department of Information Technology, Delhi Technological University, New Delhi, India

[2] Department of Computing, Goldsmiths, University of London, London, UK

Published online: 18 November 2024

Keywords Cybersecurity · Digital twin (DT) · Elliptical curve cryptography (ECC) · Internet of Things (IoT)

1 Introduction

As a key part of business 4.0, digital twin (DT) technology is quickly improving and getting a lot of attention from both university experts and people who work in the business. A digital twin is a computer model of a real product that carefully copies every step of its existence, from planning and making the product to fixing it and giving it new life [1]. By combining DT with smart algorithms, it is easier to keep an eye on, control, improve, and guess how long a product will last, how much it will cost, and how efficient it will be [2]. DT technology is used in healthcare to make virtual copies of people, medical equipment, and operating processes. This makes personalized healthcare services a lot better. But these apps deal with private and sensitive data, and there is a big chance that private patient data could be leaked [3–5]. To protect against misuse and illegal access, it is important to make sure that strong methods are in place for handling, storing, and sending these data securely within DT structures. Additionally, adding DT to IoT settings, which naturally joins things like cloud services, industries, and smart devices, makes them more vulnerable to cyber dangers and data breaches. Some of these things can make virtual copies less secure and private [6–8]. The Internet of Things (IoT) changes everything by putting sensors in everyday things that let them connect to the Internet and talk to each other using standard protocols [4]. Healthcare, supply chain, and RFID systems areas can improve using this network, benefiting individuals and society [9, 10]. However, many IoT apps use centralized data systems, raising serious security and privacy issues [11–13]. Moving from controlled storage to decentralized models can help, and blockchain is key in this shift by providing a secure and unchangeable record of data started by peer nodes. Blockchain also enhances privacy, changing how businesses manage data. Distributed ledgers serve as decentralized storage [14], where all transactions need a network majority. Bitcoin is a famous example of blockchain use. When combined with IoT, blockchain enables real-time data sharing across devices without relying on central servers, reducing third-party risks and single points of failure [15, 16]. However, this IoTblockchain integration has also led to increased cyberattacks like Distributed Denial of Service (DDoS), which overwhelms the network. This has made cybersecurity a priority, and Intrusion Detection Systems (IDS) are crucial here. IDS monitors network traffic and detects threats [17–19], using either signature-based detection (finding known patterns) or anomaly-based detection (spotting unusual behavior) to identify both known and unknown threats.

Digital twin (DT) utilization in healthcare represents a dynamically advancing domain, holding significant potential to transform patient care, medical research, and healthcare resource management through secure information transmission within the IT network [20]. Digital twins (DT), virtual replicas of physical systems, are increasingly finding applications in security and privacy domains. DT of industrial equipment can monitor performance, predict failures, and identify security vulnerabilities before they become critical. This helps prevent downtime, data breaches, and physical damage [21]. DT can create anonymized representations of real-world data, protecting sensitive information while enabling valuable insights, usually significant for applications dealing with individual or private information [22]. DT of cities can coordinate information from different sources (cameras, sensors,

etc.) to analyze activity designs, identify suspicious movement, and optimize crisis reaction. This could improve open security regarding personal protection [23]. DT of people can be utilized to oversee individual security and security settings over diverse gadgets and administrations. This enables clients to control their information and tailor security based on their needs [24]. The potential applications of DT for security and security are continually advancing. This consideration [25] proposes a system for interruption location in industrial control system (ICS) by leveraging computerized twins. It conducts recreations of different assault scenarios on a computerized twin of a mechanical filling plant, counting command infusion, DoS, and estimation adjustment. Actualizing a stacked gathering classifier for real-time interruption discovery accomplishes tall precision and F1-score. This ponder [26] centers on improving security analytics in IoT frameworks by utilizing advanced twins. The proposed system coordinates advanced twins with a security information and event management (SIEM) framework, making a cohesive approach. The system assembles information from IoT gadgets, utilizes the advanced twin for investigation, and hence produces security alarms. A down-to-earth exhibit of the framework's adequacy is showcased through a case considered on a shrewd framework, outlining its potential for supporting security in IoT situations. This ponder [27] analyzes computerized twins and their part in cybersecurity. It delves into the flexible applications of computerized twins in cybersecurity assignments, including interruption location, peculiarity location, and occurrence reaction. The discourse moreover investigates the cooperative energy between computerized twins and artificial intelligence (AI), emphasizing how AI upgrades the capabilities of computerized twins for cybersecurity purposes. Moreover, the audit sheds light on the challenges confronted in this space and diagrams potential future headings for investigation, contributing important bits of knowledge to the advancing cybersecurity scene. This examination [28] dives into the application of advanced twins for associated therapeutic gadgets with a particular center on security and security contemplations. The dialog addresses potential dangers to security and security, such as unauthorized get to and information breaches, related to utilizing computerized twins. To relieve these dangers, the ponder proposes a system for creating secure and privacy-preserving computerized twins for associated therapeutic gadgets. Tending to these security and protection concerns is underscored, emphasizing the need to guarantee the fruitful and dependable selection of computerized twins within the healthcare segment.

IDS techniques have progressed from methods like k-NN [29], SVM [30], and K-means [31] to deep learning models such as CNN [32], LSTM [33], and RNN [34], which require large datasets. To improve accuracy, feature selection has been enhanced by bio-inspired optimizations like GA [35], PSO [36], Aquila Optimizer [37], GWO [38], and CSA [39].

Detailed research is essential in this vast and linked world of IoT, blockchain, security, and digital twin technology. To handle these complex issues, authors secure IoT devices, use blockchain technology for security, optimize intrusion detection systems, and leverage digital twin technology for enhanced monitoring and simulation.

1.1 Motivations

Improving IoT security through new research methods is driven by three main goals: scalability, openness, and accuracy. All three of these goals are very important in the fields of IoT and network security.

- Scale Smart: This reason makes it even more important to test suggested security methods in big IoT networks. The goal is to make sure that these methods work in a wide range of real-world settings, not just in theory. It is important to test on large networks to find

and fix any possible growth problems that may arise as networks get bigger and more complicated. This method helps to avoid the usual problems that happen when systems are put into use and work well in controlled or small-scale situations but not when they are scaled up.
- Transparency Triumph: The goal of transparency triumph is to improve the reliability and usefulness of research by being open about how machine learning models are built and how the hyperparameters are tuned. By sharing the specifics of how the model is set up, experts make it possible for other people to repeat their work and help the community make it even better. This transparency not only builds faith in the results but also motivates people to work together to make things better and come up with new ideas, which moves the field forward.
- Fuse for Accuracy: This reason encourages looking into mixed models that combine several methods to make IoT breach detection better. These models try to use the best parts of different methods to make up for the bad parts of others. This fills in gaps in the current research and makes IoT networks safer overall. Researchers can make stronger systems that can handle the complex and changing threats that come up in IoT settings by mixing different monitoring methods.

These reasons work together to make a complete plan for improving IoT security research that focuses on real-world applications, openness in methods, and new ways to improve accuracy. The goal of this strategy framework is not only to make security solutions work better but also to make study more open and collaborative.

1.2 Objectives

- A Digital Twin (DT) framework and a machine learning-based intrusion detection model GAO-RF are developed.
- A blockchain-based elliptic curve cryptography (ECC) is created to improve IDS security and privacy.
- The genetic algorithm (GAO) improved the detection system performance in this experiment.

The GAO-RF model has several key benefits. It works well with authentication and communication systems, is faster, has fewer errors, is more accurate, can classify multiple types of attacks, and is more reliable.

This paper makes healthcare IoT networks safer by introducing a few important new ideas:

- Development of a Digital Twin (DT) Framework: Sets up a DT framework that is designed for managing and tracking IoT networks in healthcare in real time.
- The GAO-RF interruption discovery demonstration presents a hereditary algorithm-optimized irregular timberland (GAO-RF)-based machine learning-based interruption location demonstration that makes it simpler to discover dangers and react to them.
- Integration of blockchain and elliptic curve cryptography (ECC):

A blockchain structure and ECC are employed to secure information exchanges and keep basic understanding information private and secure.

- Superior security and security for IDS:

Decentralized blockchain innovation is used to form interruption discovery frameworks more securely and more privately, bringing down their dependence on centralized frameworks

- Genetic algorithm performance improvement: This method uses genetic algorithms to improve the random forest model's feature selection, which makes recognition much more accurate and the system more reliable.

By combining advanced encryption and machine learning technologies, these advances make hospital IoT systems much safer, more private, and more efficient.

1.3 Paper organization

The rest of this article is structured as follows: Sect. 2 discusses the related work based on intrusion detection systems. Section 3 represents the digital twin framework. Section 4 describes the materials and methods of the proposed model. Section 7 highlights the experimental result. Section 8 presents the overall discussion of the results. Lastly, Sect. 9 concludes the proposed work and future work.

2 Related work

The authors noticed various studies focused on identifying cyberattacks on the Internet of Things and Industrial Internet of Things (IoT) networks as authors investigated recent research. When developing intrusion detection systems for these diverse networks, each research takes a different approach despite having the same objective. The next part provides a synopsis of these significant studies, considering the difficulties brought about by the growing number of assaults and threats. In the face of constantly increasing cybersecurity threats, performing intrusion detection measurements in communication networks in a manner that is both accurate and timely continues to be a complex problem.

A few imaginative techniques that utilize digital twin (DT) innovation have been displayed in later inquiries about how to move forward with the security and execution of different frameworks. A security system based on the advanced twin was displayed by Seba et al. [40] in 2022. Khraisat et al. [41] created an include determination approach based on data picked up to recognize IoT highlights that result within the most prominent sum of differing qualities among the organized activity. It too addresses the zero-day assaults with tall discovery precision. This system secured real-time mechanical control frameworks and utilized administered machine learning strategies. The centralization of control is foremost to radically progress the security of industrial control system (ICS). This crossover demonstration for intrusion detection system (IDS) in computerized twins was displayed by Swati et al. [42]. Ashraf et al. [43] displayed an interruption discovery framework (IDS) for IoT healthcare. It utilized the BOTIoT dataset and utilized fake neural systems and unified learning. Their IDS's essential center was information security and receiving a unified learning design. The BoT-IoT dataset was utilized for the introduction. A solidified procedure is being taken to secure advanced twin frameworks from expanding cyber dangers. Thakkar et al. [44] suggested a DNN-based intrusion detection system (IDS) with statistical feature selection and used the CICIDS, NSL-KDD, and UNSWNB- 15 datasets in their research. Their strategy is geared toward enhancing the performance of DNN-based intrusion detection systems with centralized design. Huan et al. [45] recommended an advanced twin approach for recognizing the behavior of Internet of Things assaults. This strategy utilizes datasets such as UNSW-NB15 and CICIDS2017. Convolutional neural networks (CNN), bidirectional long short-term memory (BiLSTM), and deep neural networks (DNN) are the three sorts of neural systems that are utilized in their centralized strategy to identify cyberattack

propensities precisely. A digital twin method and an opensource UAV ambush dataset were utilized by Benjamin et al. [46] to explore the security of unmanned aerial vehicles (UAVs) within the year 2021. Their centralized demonstration, which makes utilization of machine learning (ML) and deep learning (DL), explores the modern cyber dangers that unmanned aerial vehicles (UAVs) are up against. Kumar et al. [9] presented a blockchain-based IoT architecture with fog computing for detecting distributed denial of service attacks using the BOT-IoT dataset. Blockchain and random forest are their approaches, which emphasize the decentralization of intrusion detection systems via federated learning. In 2021, Qinghua and colleagues [47] concentrated on computerized twin-based inconsistency discovery for cyber-physical frameworks. It made utilize of datasets such as SWaT, Aqueduct, and BATADAL. Their centralized strategy, which employs generative adversarial network (GAN), is planned to distinguish variations from the norm interior cyber-physical frameworks. The aggregate discoveries of these ponder illustrate the flexibility of computerized twin innovation over different ranges, with specific accentuation on utilizing centralized control to make strides in security and combat unused dangers. NSLKDD dataset, [48] displayed a network-based IDS based on distinctive machine learning procedures: choice tree, irregular timberland, and Xgboost on the NSL-KDD dataset for an SDN. Their proposed demonstration centered on identifying peculiarities and defending the SDN stage from real-time assaults. Abdullah Ayub Khan and his colleagues have recently published papers that look into new ways to keep the IoT and advanced network designs safe. Khan et al. [49] describe a new video streaming compression model for IoT settings that uses GANs and fuzzy logic to improve the efficiency of sending multimedia [49]. Their study includes using blockchain to improve security for serverless computing in fog and edge settings, which gives us strong options for protecting infrastructure [50]. The group also works on managing data from drones using metaheuristic algorithms and blockchain for safe fog settings [51] and looks at the latest developments in IoT security made stronger by blockchain technology [52]. In addition, they suggest a design that uses machine learning to make nextgeneration radio access networks work better in factories [53] and a way to keep remote sensing data safe in smart towns [54]. These additions make it much easier to use safe and effective technologies in IoT and network systems. Utilizing the CIC-IDS dataset and LSTM calculations, [55] displayed a blockchain-based, questioned, and unified learning approach. It gives differential security to guarantee information security, permitting collaborative preparation. Furthermore, to identify the assaults within the IoT environment, their accentuation on securing the protection and security of their information is reliable with a unified learning procedure. Through the use of the KDDCUP- 99 dataset and support vector machines (SVM), [56] presented a non-symmetric deep autoencoder for a network intrusion detection system (NIDS). The authors validate the effectiveness and robustness of the proposed model using different metrics. Their primary goal is to enhance the efficacy and robustness of the NIDS via a centralized design. They [50] developed a DL-based novel model, Bloc-Net, to handle the dataset's imbalance-ness for underrepresented instances. It combined different sampling techniques to marinate the dataset's integrity. This IDS included deep learning methods and took a centralized approach to solving the issue of dataset imbalance. These studies highlight the evolution of IDS methods, addressing specific cybersecurity needs and advancing federated learning for better privacy and security.

Many researchers have found ways to find attacks and strange things happening in networks. One exciting method combines centrality measures with deep learning algorithms, feature extraction, classification strategies and hierarchical grouping. Every study adds to what we know about intrusion detection systems. The study aims to add to what has already been done by finding the most critical problems from previous research. Table 1 shows the issues and restrictions when finding network threats and strange behavior in IoT and IIoT

Table 1 Literature review of the most recent studies on intrusion detection systems (IDS)

S. No	References	Summary	Dataset	Techniques	Focus	Security and privacy	Scalability
1	Khraisat et al. [41]	IDS for IoT attacks	BoT-IoT	SVM, SIDS, and AIDS	Detecting attacks in IoT environment,	No	No
2	Alzahrani et al. [48]	IDS for software-defined networks	NSL-KDD	Decision tree, Random Forest, and Xgboost	Anomaly detection, SDN security	No	No
3	Benjamin et al. [57]	Security of unmanned aerial systems using digital twin	open-source UAV attack dataset	ML and DL	UAV modern cyber threats are explored	Yes	No
4	Qinghua et al. [47]	Digital twin-based anomaly detection	SWaT, WADI, BATADAL	GAN	Anomaly detection for cyber-physical systems	Yes	No
5	He et al. [55]	IDS for UAV networks	CIC-IDS	LSTM	Data security, privacy	Yes	Yes
6	Ashraf et al. [43]	IDS for IoT healthcare	BoT-IoT	ANN, Federated Learning	Data privacy	Yes	No
7	Kumar et al. [9]	Blockchain-based IoT framework with fog computing	BoT-IoT	Blockchain, Xgboost, RF	IDS to detect DDoS attack	Yes	No
8	Imran et al. [56]	Non-symmetric deep autoencoder for NIDS	KDD-CUP-99	Deep Autoencoder, SVM	NIDS effectiveness and robustness	No	No
9	Seba et al. [40]	Digital twin-based security framework for industrial control systems	Real-Time	Supervised Machine Learning Algorithms	Enhancing ICS security	Yes	No
10	Bowen et al. [58]	IDS for imbalanced datasets	CIC-IDS, NSL-KDD, IoT-23	Deep Learning Techniques	Handling imbalanced datasets	No	Yes

Table 1 (continued)

S. No	References	Summary	Dataset	Techniques	Focus	Security and privacy	Scalability
11	Thakkar et al. [44]	DNN-based IDS with statistical feature selection	CIC-IDS, NSL-KDD, UNSW-NB-15	Deep Learning, Statistical Feature Selection	Performance improvement of DNN-based IDS	No	No
12	Swati et al. [42]	Hybrid model for IDS in digital twins	BoT-IoT	DL-RF	Securing digital twin systems against cyber threats	Yes	No
13	Huan et al. [45]	Digital twin method for IoT attack behavior identification	UNSW-NB15, CICIDS2017	CNN, BiLSTM, DNN	Cyberattack behavior identification	Yes	No

settings. Some of these problems are not enough scalability analysis, not enough work on integrating deep learning models, not having a decentralized storage module, not being able to compare well enough with the latest methods, and not doing enough experiments. The study aims to solve these problems by checking for scalability, comparing with the most up-to-date techniques, using thorough evaluation metrics, and carefully analyzing the results.

3 Digital twin system framework

Blockchain and artificial intelligence (AI) in intrusion detection systems have received attention in healthcare digital twin technologies. There is still a knowledge vacuum on effectively combining blockchain technology with artificial intelligence (AI) to identify network breaches in the healthcare industry, despite individual studies into AI-driven intrusion detection systems and security protocols. This synergistic integration has the potential to significantly elevate the precision, resilience, and interpretability of intrusion detection systems within healthcare ecosystems. A thorough investigation of their integration is required to design a state-of-the-art framework that can detect and mitigate attacks while protecting data privacy, guaranteeing scalability, and maintaining real-time performance. The literature currently in publication only offers a partial understanding of frameworks that combine blockchain with AI to create effective and dependable intrusion detection systems.

Combining blockchain and optimization procedures creates a secure system for detecting attacks on IoT devices. This system monitors devices, collects data, and detects threats through five main steps: data collection, analysis, prediction, key creation, and blockchain tracking. Using data from over ten IoT devices and categorizing it into five threat types, the model cleans and processes data, then splits it for training and testing. An analysis module detects threats with genetic algorithms, and visual results are shown. Encryption protects data using ECC, and the blockchain stores data securely on a shared ledger, though it has some limitations.

Blockchain approaches and optimization techniques are used in constructing Internet of Things (IoT)-based intrusion detection systems in the healthcare industry. Identifying network threats inside healthcare infrastructures is made possible by continuously monitoring IoT devices. The digital twin framework for securing IoT transmission using ECC and blockchain is shown in Fig. 1 and is used in healthcare systems to ensure patient information and streamline workflows. With IoT devices, data analytics, blockchain, and ECC, the proposed DT framework may improve healthcare centers and facilitate efficient collaboration between healthcare authorities. As seen in Fig. 1, IoT wearable sensors collect and transmit real-time data from patients or healthcare centers. The transmitted data will be cleaned, preprocessed, and converted to utilize for data analytics and prediction. This framework allows for the constant monitoring of data transmission status and the early detection of anomalies. Beginning with a reasonable dataset obtained from IoT devices/sensors separated into bunches, the preprocessing arrangement included normalization, name encoding, and information cleaning. The dataset is at that point isolated into prepare and test sets. The examination and expectation module employed labeled information for patterns indicating interruptions or noxious action. This module employed a wellness function-based developmental calculation to progress highlights, coming about information visualization and execution assessment measurements expectation. To ensure IoT gadgets and their associations, the information reaction and basic era stage employed circular bend cryptography to scramble information and create and trade mystery keys. At long last, despite issues with adaptability, keenness

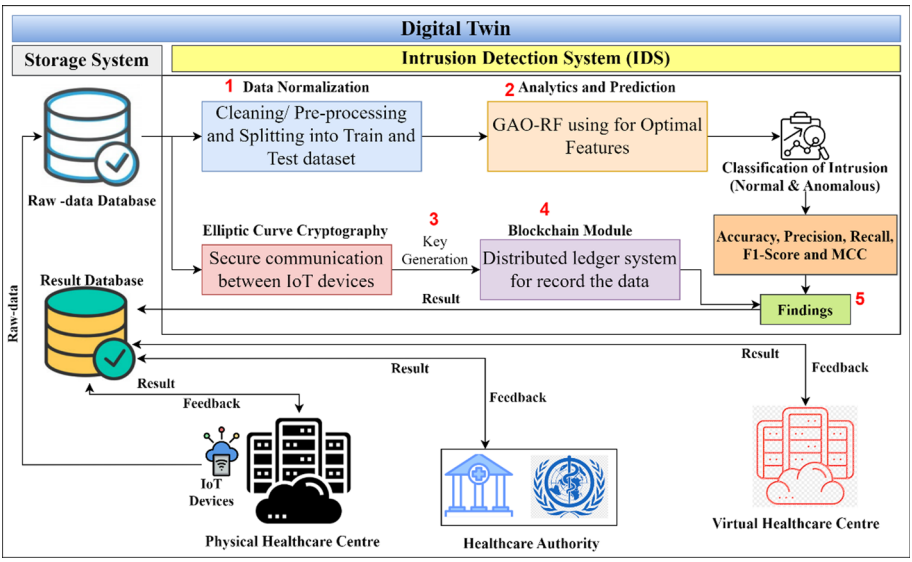


Fig. 1 System architecture

and assault rates, the blockchain module stores therapeutic information on a cloud-distributed record to protect it from any assaults.

4 Materials and methods

The proposed e-healthcare framework depends on a combination of elliptic curve cryptography (ECC) and blockchain innovation to keep private well-being information secure. Each innovation has its possess qualities that make it valuable for information assurance. Elliptic curve cryptography (ECC) is utilized to keep discussions between IoT gadgets in a clinic arranged and secure. Since it can give taller levels of security with smaller key sizes than standard cryptographic strategies, ECC works particularly well in places with constrained assets, like IoT gadgets. ECC makes beyond any doubt that private understanding information remains private, whereas it's being sent between gadgets and the advanced twin framework by scrambling the information.

The independent and unchangeable record in blockchain innovation makes information indeed more secure. A blockchain keeps track of all exchanges that happen with information, such as changes to information, getting to logs and sharing of understanding data. This makes beyond any doubt that all information trades are clear and can be followed. Once information is recorded on the blockchain, it can't be changed without arranged endorsement. This makes it much less likely that information will be altered. You'll be able to entirely control who can get to and alter blockchain information with savvy contracts. These contracts set strict rules around who can get to and alter information and under what circumstances.

Even with these strong security steps, there are a few things that could go wrong and make data less private:

- Key Management Vulnerabilities: It is very important to handle keys well. A breach can happen when bad habits are used, like sending or storing secret keys without proper security.

This risk can be reduced by using safe key management tools and conducting regular checks.

- Smart Contract Vulnerabilities: Smart contracts are used to enforce security rules on the blockchain. It can be broken if it is poorly designed or implemented. It is very important that these contracts are tested thoroughly and have security checks done.
- Endpoint Security: There must be strong physical or virtual security for IoT devices because they can be attacked directly. Endpoint security needs to be improved by using strong login, regular updates, and physical safeguards.
- Scalability and efficiency Problems: As the size of a network grows, it can be hard to find a good balance between security and system efficiency. These problems can be fixed by making the blockchain design better and using strong encryption methods like ECC.
- Quantum Computing Threats: As quantum computing gets better, it's possible that ECC and other encryption systems will be broken. It is very important to keep up with changes in quantum-resistant security and get ready for future changes.

To secure IoT and IIoT systems, a new method uses blockchain, genetic algorithms (GA), and the random forest model. Figure 2 illustrates the GAO-RF proposed model. Data is encrypted with elliptic curve cryptography (ECC) and stored in blockchain blocks. GA monitors network activity, while real-time datasets help keep the system updated. Data is encrypted with ECC and added to the blockchain for protection against threats. This method strengthens security and helps detect intrusions by combining encryption and machine learning.

GAO-RF Proposed Model:

Step 1: Initialization.

Initialize a population of genetic algorithms (GA) with random feature subsets:

$$P(t) = \{\mathrm{F1},\ \mathrm{F2}, ...,\ \mathrm{Fn}\}$$

- where P(t) represents the population at generation t, and Fi are the individual feature subsets.

Step 2: Fitness Evaluation.

Evaluate each subset using the random forest (RF) classifier to determine its fitness based on prediction accuracy:

$$\mathrm{Fitness(Fi)} = \frac{1}{(1 + \ \mathrm{Error(RF(Fi))})} \tag{1}$$

- where Error(RF(Fi)) is the classification error of the RF model trained with features Fi.

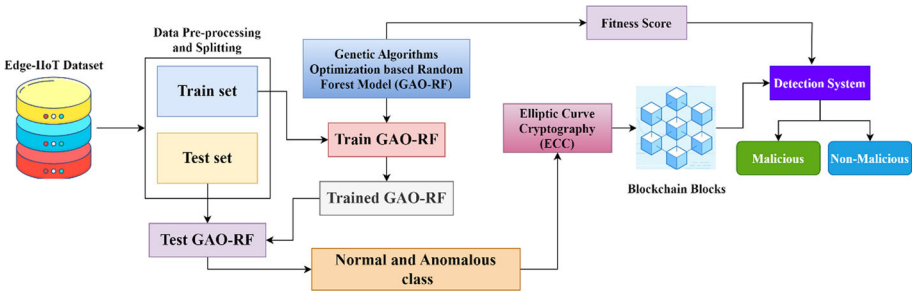


Fig. 2 GAO-RF proposed model

Step 3: Selection.
Select feature sets with higher fitness scores for reproduction using tournament selection:

$$\text{F_selected} = \text{Tournament}(P(t)) \tag{2}$$

- where the tournament function selects the best feature set from a randomly sampled subset of the population.

Step 4: Crossover and Mutation.
Selected feature sets undergo crossover and mutation to generate new feature sets for the next generation:

$$F_{\text{new}} = \text{Crossover}(F_{\text{selected}}) \tag{3}$$

$$F_{\text{mutated}} = \text{Mutation}(F_{\text{new}}, \text{mu}) \tag{4}$$

- where mu is the mutation rate, affecting the probability of altering each feature in the feature set.

Step 5: Model Update and Termination.
Update the population with new feature sets, and iterate the process until a termination criterion is met (often a fixed number of generations or a convergence threshold):

$$P(t+1) = \{F_{\text{mutated}}\} \cup \{\text{Best of} P(t)\} \tag{5}$$

$$\text{Terminate if} \max(\text{Fitness}(P(t+1))) \geq \text{Threshold} \tag{6}$$

This formulation allows the GAO-RF model to systematically optimize the feature selection process, enhancing the effectiveness and reliability of the intrusion detection system.

4.1 Dataset (Edge-IIoT)

Internet of Things (IoT) and Industrial Internet of Things (IIoT) applications are the focus of the Edge-IIoT [59]. The dataset was developed to train machine learning-based intrusion detection systems in cybersecurity. This dataset contains information from more than ten sensors, including but not limited to temperature and humidity sensors, pH meters, ultrasonic sensors, heart rate sensors, water level detectors, soil moisture sensors, and flame sensors. Edge-IIoT-2022 authors simulated 14 attacks, classifying them into five categories: (i) DoS/DDoS, (ii) information gathering, (iii) man in the middle (MITM), (iv) injection, and (v) malware. We create multi-class intrusion detection systems (IDSs) using the Edge-IIoT-2022 dataset. In this case, there are fifteen classes: 14 classes that represent each attack and one normal class, as shown in Table 2. The Edge-IIoT-2022 dataset specifies the data points as vectors of 61 features, of which 43 are numeric, and the other features are nominal and string. *Attack_label* and *Attack_type* are two extra-label features. In the multi-class setup, *Attack_type* is used as the class label. There are a total of 1,176 variables or characteristics that are included in the dataset, and out of them, 61 features and characteristics indicate significant association. The dataset has a total of 1,909,671 records, consisting of 1,363,998 normal instances and 545,673 instances of attacks. It has been separated into an 80% training set (1,527,736 samples) and a 20% testing set (381,935 samples) for consistent evaluation across 15 distinct classes.

Table 2 Types of cyberattacks in Edge-IIoT

S. No	Attacks	Data Record	S. No	Attacks	Data Record
1	Normal	1,091,198	9	Password	39,946
2	DDoS_ICMP	54,351	10	Backdoor	19,221
3	DDoS_UDP	97,253	11	Ransomware	7751
4	DDoS_TCP	40,050	12	Fingerprinting	682
5	DDoS_HTTP	38,835	13	Vulnerable_Scanner	40,021
6	SQL_Injection	40,661	14	Port_Scanning	15,982
7	XSS	12,058	15	MITM	286
8	Uploading	29,446			

4.2 Elliptic curve cryptography (ECC)

Elliptic curve cryptography (ECC) is an exciting and powerful way to protect privacy in a world where digital security constantly changes. This incredible cryptography offers vigorous security, doing so rapidly and efficiently, making it pivotal for keeping advanced communications and exchanges secure. Investigate the world of ECC with us, where the magnificence of elliptic curves meets the necessity for security, marking a paradigm shift in protecting digital spaces. Before information is stored on the blockchain or any other secure storage system, it is encrypted using ECC, ensuring that the data remains secure and confidential, even in the event of unauthorized access. Figure 3 illustrates the implementation of ECC in the proposed model. Within blockchain blocks, encrypted data is securely stored. Each block includes a timestamp, cryptographic key, and the data's cryptographic hash. The data is tamper-proof, with access restricted by the cryptographic key, thanks to the blockchain's distributed ledger technology. Since ECC can provide robust security with very small key sizes, it is highly efficient and ideally suited for resource-constrained environments, such as those encountered in Internet of Things (IoT) devices. ECC strength is typically expressed

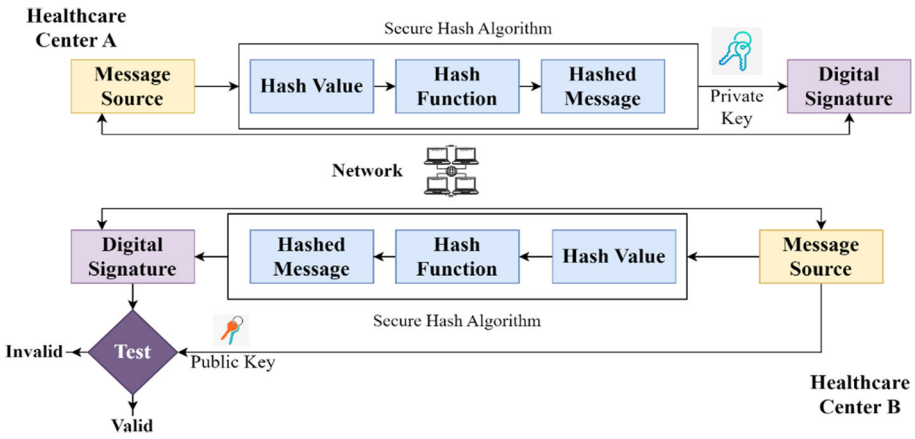


Fig. 3 Implementation of ECC in the proposed model

in bits. For example, a 3072-bit RSA key is roughly equivalent in strength to a 256-bit ECC key, illustrating how effective ECC is at providing strong security with shorter keys.

Elliptic curve cryptography (ECC) is a widely used public-key cryptography algorithm that provides strong security with relatively shorter key lengths compared to other public-key algorithms like RSA. ECC is particularly well-suited for resource-constrained environments, making it popular for securing communications in various applications, including security and privacy.

4.2.1 Initialization

- Choose an elliptic curve E defined over a finite field F_P.
- Select a base point G on the curve with prime order n.
- Choose cryptographic parameters for ECC operations.

$$\text{Choose } E(F_P),\ G,\ n$$

Choose cryptographic parameters

4.2.2 Key generation for IoT security

- Each IoT device generates a unique key pair.
- Generate a private key d_i as a cryptographically secure random number.
- Compute the corresponding public key Q_i using scalar multiplication: $Q_i = d_i \times G$.

$$d_i \xleftarrow{\$} [1, n-1]$$

$$Q_i = d_i \times G \tag{7}$$

4.2.3 Secure communication between IoT devices

- Devices exchange public keys ($Q_i and Q_j$) Openly.
- Calculate shared secrets (Sij) using private keys and public keys as follows:

$$\text{Sender } S_{ij} = d_i \times Q_j \tag{8}$$

$$\text{Receiver } S_{ij} = d_j \times Q_i \tag{9}$$

- Derive a symmetric encryption key from the shared secret for secure data exchange.

4.2.4 Data encryption and blockchain integration

- Encrypt the IoT data using the derived symmetric encryption key.
- Prepare data for blockchain storage:
- Create a block with data, a timestamp and a cryptographic hash of the data.
- Include the public key (K) and digital signature (Sig) for data integrity and authenticity.

- Encrypt the block using ECC, producing the encrypted block ($EncBlock$).

$$C = \text{AES}(S_{\text{ij}}, IoTData) \tag{10}$$

$$Block = \{C, Timestamp, Hash(Data), K, Sig\} \tag{11}$$

$$EncBlock = ECC_Encrypt(Block) \tag{12}$$

4.3 Intrusion detection system (IDS)

Calculation 1 gives more points of interest approximately the usage of RF within the GA. This handle separates datasets A and B into preparing (Atrain, Btrain) and testing (Atest, Btest) sets. A irregular woodland show is initialized and prepared utilizing Atrain and Btrain and learning designs and connections inside the information. The model's execution is at that point assessed on Atest, and in this way, expectations are produced for Btest. At last, a wellness score is computed based on the model's execution on the Btest, giving a quantitative degree of how well the irregular woodland show performs on the given assignment. This approach makes a difference survey the model's capacity to generalize to unused information and make precise expectations. The foremost ideal show is one that yields the most elevated precision score.

Algorithm 1: Fitness function computation

Input: A, B; the input data frame and output series.

Output: Fitness score obtained by the random forest model.

Step 1: Data splitting.

Split data into training and testing sets using a complex stratified split to ensure uniform distribution of classes:

$$A_{\text{train}}, A_{\text{test}}, B_{\text{train}}, B_{\text{test}} = \text{split}(A, B, \text{test}_{\text{size}} = 0.3, \text{stratify} = B)$$

- This function partitions the dataset into training and testing subsets, maintaining the proportion of classes across them.

Step 2: Model Initialization.

Initialize the random forest model with a high number of trees and depth to enhance the learning capability:

$$\text{RF} = \text{Random Forest Classifier}(n_\text{estimators} = 100, \text{max_depth} = \text{None}, \text{ random_state} = 42)$$

- Here, 'n_estimators' represents the number of trees in the forest, 'max_depth' allows the trees to grow until all leaves are pure, and 'random_state' ensures reproducibility.

Step 3: Model Fitting.

Fit the random forest model using the training data with bootstrapping and feature selection:

$$\text{RF.fit}(A_\text{train}, B_\text{train})$$

- This step involves building trees where each tree is trained on a bootstrapped sample of the data, and at each node, a subset of features is randomly selected to determine the split.

Step 4: Model Evaluation and Fitness Computation.

Evaluate the model on the testing set and compute the prediction accuracy with a detailed error matrix:

$$\text{predictions} = \text{RF.predict}(A_\text{test})$$

$$\text{Fitness Score} = \left(\frac{1}{\left(1 + \text{sum}\left(\frac{\text{abs}(\text{predictions} - B_{\text{test}})}{\text{len}(B_{\text{test}})}\right)\right)} \right) * 100 \tag{13}$$

- This complex equation for the Fitness Score is an adaptation of the mean absolute error, inversely transformed to reflect higher scores for better performance, normalized to a percentage scale.

The genetic algorithm 2 for feature selection on the Edge-IIoT dataset begins by initializing a binary-encoded population, each representing a feature subset. The algorithm iteratively evolves the population through crossover and mutation, aiming to improve the fitness of individuals based on a predefined fitness function. The best individual, denoted as Gbest, is updated throughout the process. The evolution loop continues for a specified number of iterations (Mbest), and the selected feature subset is recorded in Elist. The algorithm terminates when reaching a predefined convergence criterion or the maximum number of iterations. The ultimate objective is to identify a subset of features that optimally contribute to the model's performance on the Edge-IIoT dataset.

Algorithm 2: Genetic algorithm for feature selection on Edge-IIoT dataset

Prerequisites:

Dataset: D, Edge-IIoT Dataset.

Feature Names Array: F

Target Domain Value: T

Empty Feature Subset List: E_{list}.

Maximum Iterations (Best Features): M_{best}.

Step 1: Initialization.

Initialize a population of individuals, where each individual represents a potential solution (i.e., a subset of features):

$$P(0) = \{x1, x2, ..., xn\} \tag{14}$$

- where P(t) is the population at generation t, and xi are binary strings where each bit represents the presence (1) or absence (0) of a feature in the dataset.

Step 2: Fitness Evaluation.

Evaluate the fitness of each individual using an integral over the data distribution to assess predictive performance:

$$f(\text{xi}) = \int \text{Accuracy}(M(\text{xi}, D))\text{dD} \tag{15}$$

- where M(xi, D) represents the model trained with features xi on dataset D, and Accuracy is the performance metric.

Step 3: Selection.

Select individuals for reproduction using tournament selection:

$$x_\text{selected} = \text{Tournament}(P(t), f) \tag{16}$$

- where Tournament is a function that selects the individual with the highest fitness from random subsets of the population.

Step 4: Crossover.
Apply a crossover operation to generate new offspring from selected individuals:

$$x_\text{offspring} = \text{Crossover}(x_{\text{selected}}) \tag{17}$$

- This function swaps features between pairs of individuals at randomly chosen crossover points.

Step 5: Mutation.
Introduce mutations with a small probability to maintain genetic diversity:

$$x_\text{mutated} = \text{Mutation}\left(x_{\text{offspring}},\ \text{mu}\right) \tag{18}$$

- where mu represents the mutation rate, determining the probability of a feature being toggled.

Step 6: Update and Termination.
Update the population with the new generation and check for termination conditions:

$$P(t+1) = x_{\text{mutated}} \cup \text{Best of} P(t) \tag{19}$$

$$\text{if } \max(f(P(t+1))) > \text{threshold or} t = t_{\max} : \text{Terminate} \tag{20}$$

This model outlines a comprehensive genetic algorithm process tailored for optimizing feature selection in an Edge-IIoT dataset, considering the computational and data complexities typical in industrial IoT applications. The GA modified for case study's main stages is shown in Fig. 4. The simplified form of the pseudo-code implementation used in Algorithms 1 and 2 is shown in this flowchart.

Combining different methods to improve detection is what hybrid models for better IoT breach detection are all about. A mixed model could combine rule-based systems with machine learning techniques to make a stronger framework. Here is a step-by-step mathematical example of how this kind of mixed model could be set up:

Hybrid Model for IoT Intrusion Detection:
Step 1: Data Collection.
Collect data from various IoT devices and sensors:

$$D = \{\text{d1}, \text{d2}, \ldots, \text{dn}\} \tag{21}$$

- where D represents the dataset collected, and di represents data points from IoT devices.

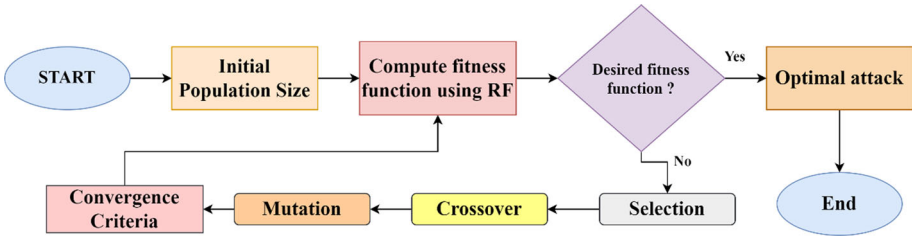


Fig. 4 Flowchart of GA-fitness function in RF

Step 2: Data Preprocessing.
Normalize and feature-engineer the data to prepare for analysis:

$$D_\text{processed} = f_{\text{preprocess(D)}} \tag{22}$$

- where f_preprocess is a function that includes normalization, handling missing values, and feature extraction.

Step 3: Rule-Based Filtering.
Apply rule-based filters to quickly eliminate known benign behaviors:

$$D_\text{filtered} = \{d \quad \text{in} \quad D_\text{processed}|\text{rules}(d) = \text{true}\} \tag{23}$$

- where rules(d) are predefined conditions that data points must satisfy to be considered for further analysis.

Step 4: Feature Selection Using Genetic Algorithm.
Optimize feature selection using a genetic algorithm to reduce dimensionality and enhance model performance:

$$S = \text{GA}(D_{\text{filtered}}, \text{fitness}) \tag{24}$$

- where S is the subset of selected features, GA represents the genetic algorithm, and fitness is a function evaluating the effectiveness of the feature subset.

Step 5: Machine Learning Model Training.
Train a machine learning model using the selected features:

$$M = \text{train}(S, D_{\text{train}}) \tag{25}$$

- where M is the trained model, and D_train is the training subset of D_filtered.

Step 6: Anomaly Detection.
Deploy the model to detect anomalies in new data:

$$y_{\text{pred}} = M(x_{\text{new}}, S) \tag{26}$$

- where y_pred is the predicted outcome (anomalous or not), x_new is new incoming data, and S is the set of optimized features used by model M.

This six-step approach integrates rule-based filtering and advanced machine learning techniques, optimized by genetic algorithms, to develop a robust system for detecting intrusions in IoT networks. Each step is designed to refine the data and model progressively, focusing on enhancing detection capabilities with accuracy and efficiency.

5 Experimental result

The consideration evaluated the proficiency of the proposed show for producing intrusion detection system (IDSs) in IoT gadgets, utilizing the Edge-IIoT dataset. A multi-class test was conducted to classify activity into different assault sorts. For the binary-class situation, the Attack_label highlight was utilized, whereas the Attack_type highlight was utilized for the multi-class situation. The tests were carried out on a portable workstation prepared with an

Intel 11th Gen Center i7-4510U CPU processor (2.0 GHz, 4 centers, 8 coherent processors), 16 GB of Smash, and a 64-bit Windows 11 working framework. Python, alongside libraries such as Sklearn, Tensorflow, Keras, Pandas, and Numpy, was utilized for the tests to help in information preprocessing. The proposed show was prepared utilizing 80% of the dataset, with a encourage division to refine the model's hyperparameters. This part comes about in a preparing set comprising 80% of the first information and approval set of 20%. The choice to partition the dataset this way was based on the direction from [60], pointing to adjust the preparing handle and avoid overfitting, as proposed in [61]. This information division procedure was significant for assessing the model's execution dependably and heartily. Six diverse models were created to realize the most noteworthy classification exactness. This segment presents the parameters, execution measurements, and assessment that comes about for each show. The utilize of these numerous models was basic to distinguish the best-performing one, guaranteeing a comprehensive evaluation of the proposed IDS's adequacy in identifying different assault sorts in IoT situations.

Confusion Matrix: The disarray lattice may be a visual tool with different assessment parameters. To disentangle, we'll center on the binary-class perplexity network, overlooking the multi-class adaptation, which is an expansion of the same concept. The four assessment measurements are True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), outlined in Fig. 5. The cleared out inclining speaks to accurately classified information focuses (genuine), whereas the correct inclining appears type-1 (FP) and type-2 (FN) mistakes. This clear portrayal helps in understanding the model's classification execution (Figs. 6 and 7).

Accuracy: The concept of accuracy in intrusion detection refers to the efficacy of a model in properly recognizing both positive (intrusion) and negative (non-malicious) occurrences.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100 \tag{27}$$

The proposed GAO-RF model achieved accuracy values of 98.4% for ten epochs, 98.1% for twenty epochs, 98.3% for thirty epochs, 98.2% for forty epochs, and 98.4% for fifty epochs, as shown in Table 3. These values were achieved during a short period. In contrast to other models or methods, the model that was presented constantly outperformed them, highlighting the usefulness of the model as well as the solid security that it provides (Figs. 8 and 9).

Precision: Precision in intrusion detection measures a model's consistency and repeatability in properly recognizing positive incursions, calculating the percentage of correctly labeled

Fig. 5 Binary-class confusion matrix

		Predicted Class	
		Normal	Anomalous
Actual Class	Normal	TN	FP
	Anomalous	FN	TP

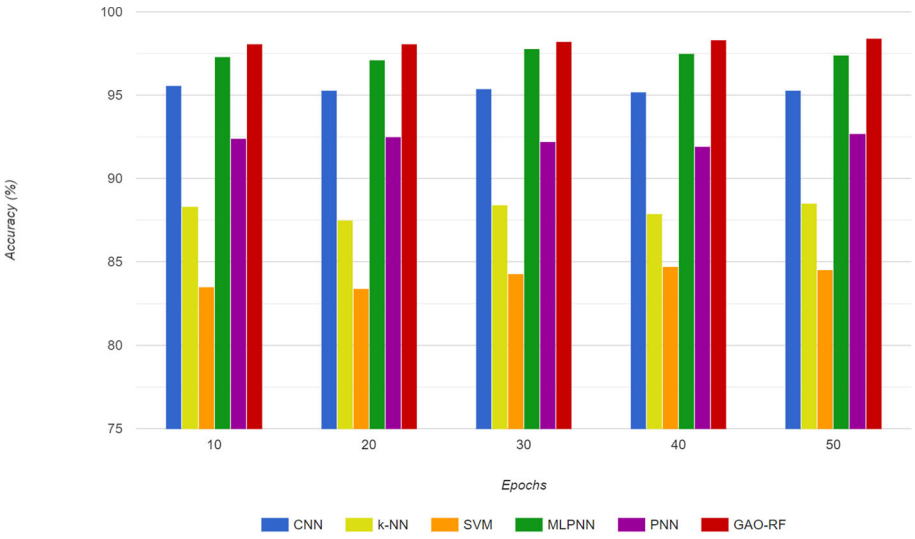


Fig. 6 Accuracy (%)

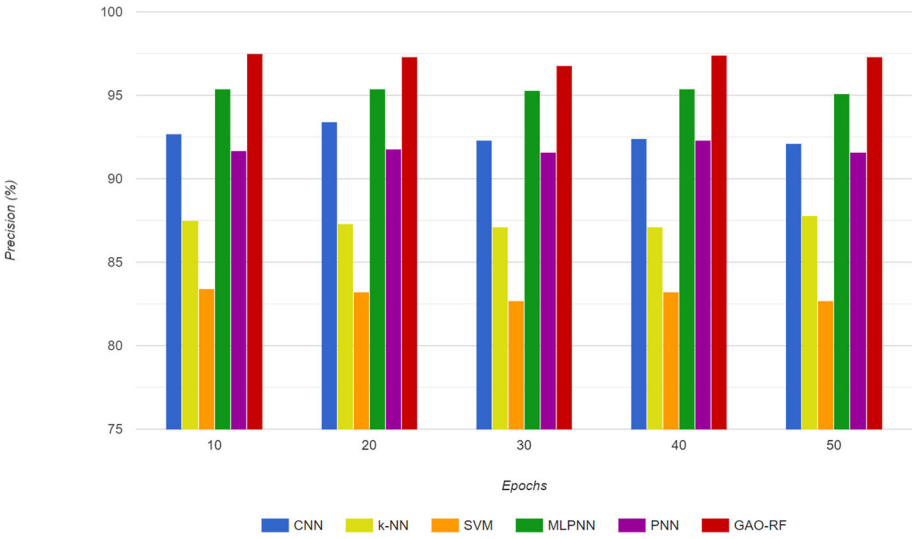


Fig. 7 Precision value (%)

positive events among all positive predictions produced by the model.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100 \tag{28}$$

The proposed GAO-RF model achieved precision values of 97.5% for ten epochs, 97.3% for twenty epochs, 96.8% for thirty epochs, 97.4% for forty epochs, and 97.3% for fifty epochs, as shown in Table 4. These values were achieved during a short period. In contrast

Table 3 Accuracy (%): proposed versus ML/DL model with different epochs

Epochs	CNN	k-NN	SVM	MLPNN	PNN	GAO-RF
10	95.6	88.3	83.5	97.3	92.4	98.1
20	95.3	87.5	83.4	97.1	92.5	98.1
30	95.4	88.4	84.3	97.8	92.2	98.2
40	95.2	87.9	84.7	97.5	91.9	98.3
50	95.3	88.5	84.5	97.4	92.7	**98.4**

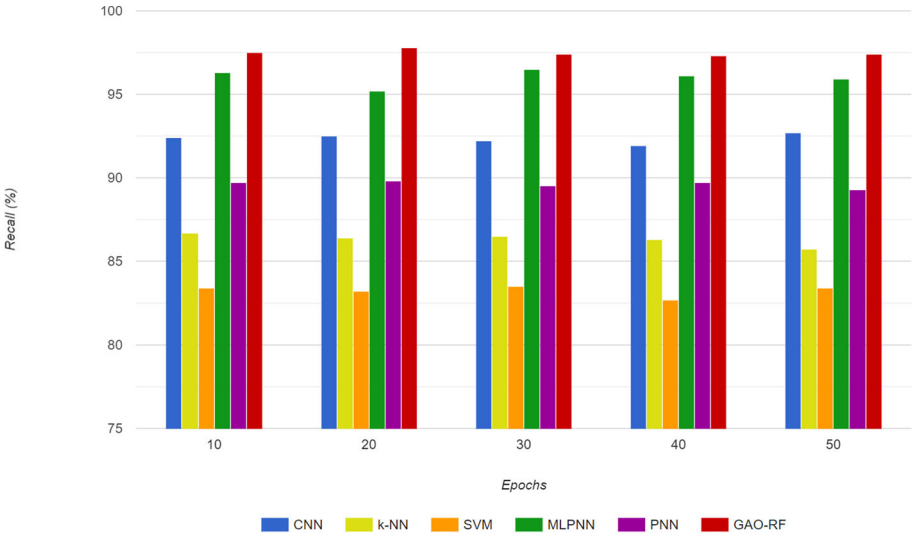


Fig. 8 Recall value (%)

to other models or methods, the model that was presented constantly outperformed them, highlighting the usefulness of the model as well as the solid security that it provides (Figs. 10 and 11).

Recall: The model's capacity to accurately identify and detect positive network intrusion is evaluated via a recall function. This function computes the proportion of true positives that were accurately expected.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \tag{29}$$

The proposed GAO-RF model achieved recall values of 97.5% for ten epochs, 97.8% for twenty epochs, 97.4% for thirty epochs, 97.3% for forty epochs, and 97.4% for fifty epochs, as shown in Table 5. These values were achieved during a short period. In contrast to other models or methods, the model that was presented constantly outperformed them, highlighting the usefulness of the model as well as the solid security that it provides.

F1-score: The definition of this term is the weight of the harmonic mean of the recall and precision test measures. The calculation is based on recall and precision of measurement to

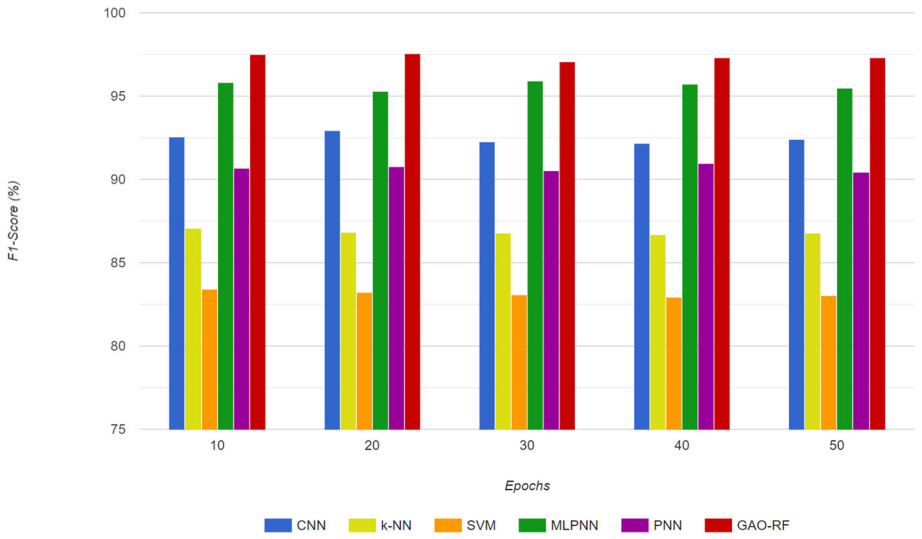


Fig. 9 F1-score (%)

Table 4 Precision (%): proposed versus ML/DL model with different epochs

Epochs	CNN	k-NN	SVM	MLPNN	PNN	GAO-RF
10	92.7	87.5	83.4	95.4	91.7	97.5
20	93.4	87.3	83.2	95.4	91.8	97.3
30	92.3	87.1	82.7	95.3	91.6	96.8
40	92.4	87.1	83.2	95.4	92.3	97.4
50	92.1	87.8	82.7	95.1	91.6	**97.3**

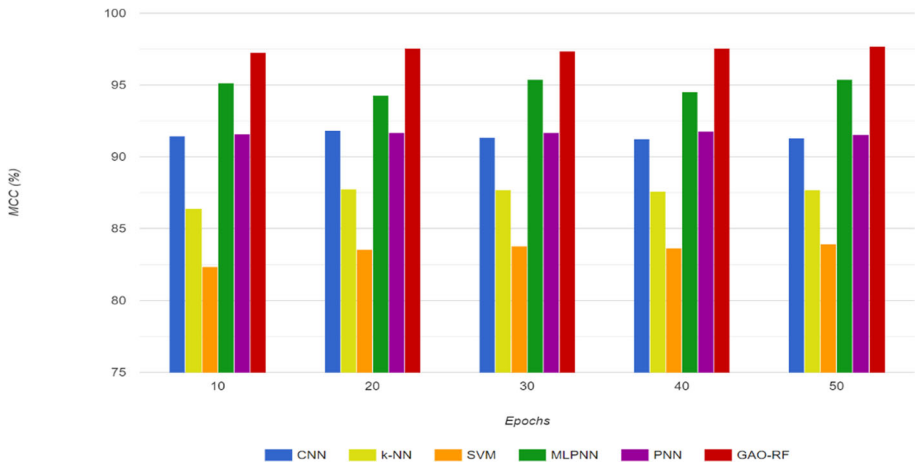


Fig. 10 MCC value (%)

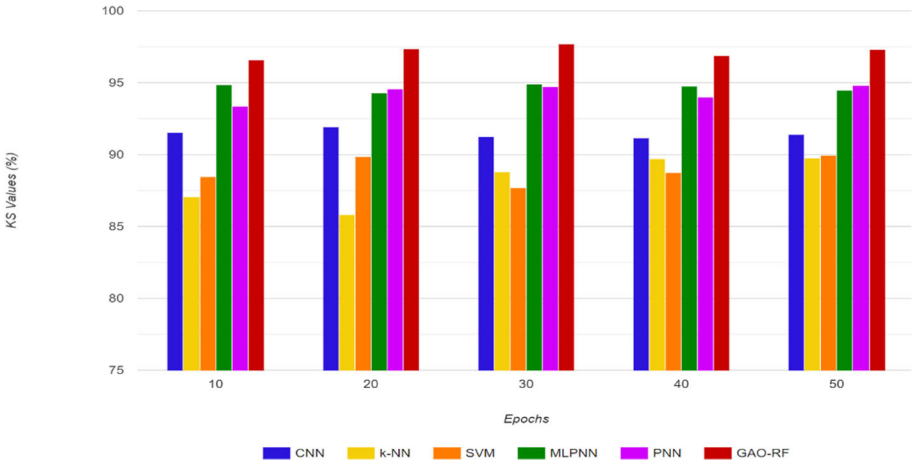


Fig. 11 KS value (%)

Table 5 Recall (%): proposed versus ML/DL model with different epochs

Epochs	CNN	k-NN	SVM	MLPNN	PNN	GAO-RF
10	92.4	86.7	83.4	96.3	89.7	97.5
20	92.5	86.4	83.2	95.2	89.8	97.8
30	92.2	86.5	83.5	96.5	89.5	97.4
40	91.9	86.3	82.7	96.1	89.7	97.3
50	92.7	85.7	83.4	95.9	89.3	**97.4**

determine how effective intrusion detection is.

$$\text{F1 score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{30}$$

The proposed GAO-RF model achieved F1-score values of 97.5% for ten epochs, 97.54% for twenty epochs, 97.09% for thirty epochs, 97.3% for forty epochs, and 97.3% for fifty epochs, as shown in Table 6. These values were achieved during a short period. In contrast

Table 6 F1-score (%): proposed versus ML/DL model with different epochs

Epochs	CNN	k-NN	SVM	MLPNN	PNN	GAO-RF
10	92.54	87.09	83.4	95.84	90.68	97.5
20	92.94	86.84	83.2	95.29	90.78	97.54
30	92.24	86.79	83.09	95.89	90.53	97.09
40	92.14	86.69	82.94	95.74	90.98	97.3
50	92.39	86.78	83.04	95.49	90.43	**97.3**

to other models or methods, the model that was presented constantly outperformed them, highlighting the usefulness of the model as well as the solid security that it provides.

Matthews correlation coefficient (MCC): Pearson's correlation coefficient has a discrete case known as the MCC. MCC is a valuable metric for evaluating the reliability of binary classification. The formula determines the worst possible prediction, MCC = -1, while MCC = + 1 indicates the best prediction.

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}} \tag{31}$$

The proposed GAO-RF model achieved MCC values of 97.27% for ten epochs, 97.54% for twenty epochs, 97.37% for thirty epochs, 97.57% for forty epochs, and 97.69% for fifty epochs, as shown in Table 7. These values were achieved during a short period. In contrast to other models or methods, the model that was presented constantly outperformed them, highlighting the usefulness of the model as well as the solid security that it provides.

Cohen's kappa statistics (KS): Cohen's kappa statistics (KS) is a performance analysis indicator used in classifier performance analysis. The Kappa statistic determines the agreement between a dataset's expected and actual values.

$$k_s = \frac{p_0 - p_c}{1 - p_c} \tag{32}$$

where p_o is the total agreement probability, and p_c is the hypothetical probability of chance agreement.

The proposed GAO-RF model achieved KS values of 96.57% for ten epochs, 97.35% for twenty epochs, 97.67% for thirty epochs, 96.87% for forty epochs, and 97.31% for fifty epochs, as shown in Table 8. These values were achieved during a short period. In contrast

Table 7 MCC (%): proposed versus ML/DL model with different epochs

Epochs	CNN	k-NN	SVM	MLPNN	PNN	GAO-RF
10	91.43	86.39	82.34	95.14	91.59	97.27
20	91.84	87.74	83.57	94.29	91.69	97.54
30	91.34	87.69	83.79	95.39	91.67	97.37
40	91.24	87.59	83.64	94.54	91.76	97.57
50	91.29	87.68	83.94	95.39	91.52	**97.69**

Table 8 KS value (%): proposed versus ML/DL model with different epochs

Epochs	CNN	k-NN	SVM	MLPNN	PNN	GAO-RF
10	91.54	87.09	88.45	94.84	93.38	96.57
20	91.94	85.84	89.84	94.29	94.58	97.35
30	91.24	88.79	87.69	94.89	94.73	97.67
40	91.14	89.69	88.74	94.74	93.98	96.87
50	91.39	89.78	89.94	94.49	94.83	**97.31**

to other models or methods, the model that was presented constantly outperformed them, highlighting the usefulness of the model as well as the solid security that it provides.

6 Discussions

In this part, we discuss the results of suggested GAO-RF model using a blockchain framework. The proposed model performed well in accuracy, precision, recall, and F1-score for detection. To check how effective and efficient model is, we compared it with popular machine learning (ML) and deep learning (DL) algorithms using the Edge-IIoT dataset. We simulated these algorithms in a Python environment. Table 9 and Fig. 12 show that model outperforms the other algorithms, showing better intrusion detection abilities and proving its effectiveness in spotting intrusions.

Table 9 Overall performance evaluation

Models	Accuracy	Precision	Recall	F1-score	MCC	KS
CNN	95.3	92.1	92.7	92.39	91.29	91.39
k-NN	88.5	87.8	85.7	86.78	87.68	89.78
SVM	84.5	82.7	83.4	83.04	83.94	89.94
MLPNN	97.4	95.1	95.9	95.49	95.39	94.49
PNN	92.7	91.6	89.3	90.43	91.52	94.83
GAO-RF	98.4	97.3	97.4	97.3	97.69	97.31

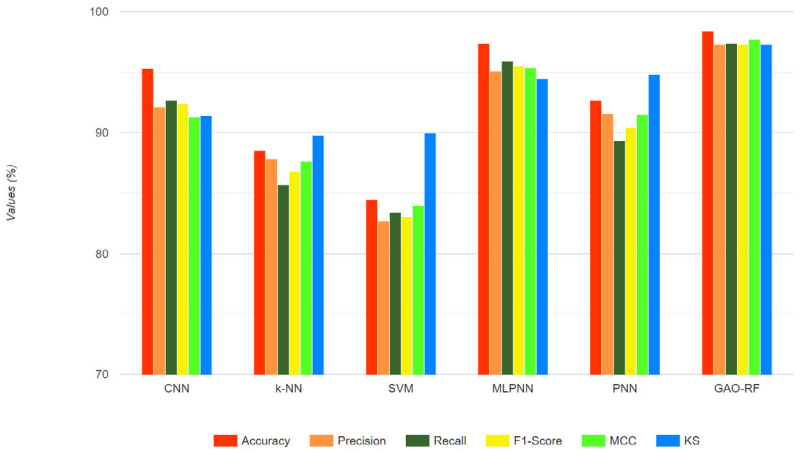


Fig. 12 Overall performance of the proposed model

6.1 Comparison of the proposed method

Compared to previous research, the suggested approach produced significant results, obtaining an accuracy level of 98.4% with the Edge-IIoT dataset. The current research used an innovative digital twin (DT) framework and GAO-RF model to improve accuracy, unlike previous studies that relied on traditional methodologies. This part compares accuracy with some of the most advanced IDSs for IIoT and IoT systems. We focus closely on the datasets used, accuracy metrics, number of classes in the classification task, and models employed, as shown in Table 10.

Scalability is very important for the suggested e-healthcare system, especially as more people use it and more IoT devices are included. Combining blockchain and elliptic curve cryptography (ECC), two technologies that are naturally scalable, this system uses a digital twin design. As the network grows, Blockchain's speed stays high because it is not controlled and can handle more transactions without a single point of failure. For example, ECC helps because it offers strong security with smaller key sizes than standard methods. This requires less computing power and works well for expansive networks. The system can also automatically adjust to changes in data trends as more devices connect because it uses a genetic algorithm-optimized random forest (GAO-RF) model for intruder detection. Depending on the current network load and user behavior, this model chooses the best features. It can handle a lot of login requests without slowing down or compromising security. It can support a growing network of devices and users easily because it uses advanced cryptography and machine learning methods.

7 Conclusion and future work

This study presents a novel GAO-RF model for intrusion detection in the context of digital twins in healthcare. Using the dataset, the authors tested the effectiveness of six distinct data analytics techniques for multi-class intrusion detection. The authors present a unique approach that integrates genetic algorithms, blockchain, ECC encryption, and machine learning to provide a dependable and efficient security solution for information transfer in the healthcare ecosystem. To illustrate where the examined algorithms might be placed on the enormous IoT/IIoT IDS map, the authors provided in-depth comparisons of cutting-edge solutions. Real-time data assessment shows remarkable performance with 98.4% of detection accuracy, 97.3% of precision, 97.4% of recall, 97.3% of F1-score, 97.69% of MCC, and 97.31% of KS. The thorough comparative analysis result demonstrates that, compared to other well-known models, the suggested model outperformed them on various performance parameters. It offers scalable and efficient defense against constantly changing threats and attacks, marking a significant advancement in the security of the Internet of Things environments.

In further research, authors want to look at how various hybrid optimization approaches can be integrated for improved performance; proxy servers will be implemented for re-encryption. A wide range of researchers, academics, data scientists and practitioners in the IoT/IIoT cybersecurity field will find value in the findings from study. A potential application of this study is that industries must protect IoT networks against cyberattacks. Continuous monitoring provides real-time surveillance to protect data and streamline processes. IoT security and privacy must be improved to prevent cyberattacks. Bio-inspired algorithms increase intrusion detection accuracy and adaptability, especially in changing industrial landscapes.

Table 10 Comparison of the proposed method with previously published work

S. No	References	Domain	Dataset	Learning approach	Features	Model	Accuracy (%)
1	Zolanvari et al. [62]	IoT/ IIoT	WUSTL-IIoT	Centralized	41	RF	99.99
						DT	99.98
						k-NN	99.98
						LR	99.90
						SVM	99.64
						ANN	99.64
						NB	97.48
2	Koroniotis et al. [63]	IoT	BoT-IoT	Centralized	46	SVM	99.98
						RNN	97.70
						LSTM	98.05
3	Vaccari et al. [64]	IoT	MQTTset	Centralized	33	RF	99.43
						NN	99.32
4	Al-Hawawreh et al. [65]	IoT/ IIoT	X-IIoTID	Centralized	59	DT	99.45
						NB	47.08
						k-NN	98.21
						SVM	98.14
						LR	96.61
						DNN	98.39
						GRU	99.46
5	Ferrag et al. [59]	IoT/ IIoT	Edge-IIoT	Centralized and federated learning	61	DT	67.11
						RF	80.83
						SVM	77.61
						k-NN	79.18
						DNN	94.67
						DT	97.80
						NB	98.29
6	Badawi et al. [66]	IoT/ IIoT	Edge-IIoT	Centralized and federated learning	61	J48	92.92
						PART	92.80
						BN	90.86
						AB	86.29
						LB	85.40
						ASC	90.43
7	Md Mamunur Rashid et al. [67]	IoT/ IIoT	Edge-IIoT	Centralized and federated learning	61	CNN	93
						RNN	94
8	Our Proposed Model	**IoT/ IIoT**	**Edge-IIoT**	**Centralized and federated learning**	**61**	**CNN**	**95.3**
						k-NN	**88.5**

Table 10 (continued)

S. No	References	Domain	Dataset	Learning approach	Features	Model	Accuracy (%)
						SVM	**84.5**
						MLPNN	**97.4**
						PNN	**92.7**
						GAO-RF	**98.4**

Genetic algorithms optimize detection systems to reduce false alarms in varied network environments.

Author contributions VS contributed to implementation, experimentation, manuscript writing, writing—original draft, conceptualization, methodology, and visualization. KS contributed to conceptualization, methodology, writing—review and editing, supervision, and investigation. AK writing—review and editing, supervision, and investigation.

Data availability No datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Tao F, Qi Q (2019) Make more digital twins. Nature 573(7775):490–491
2. Mao W, Tian S, Fan J, Liang X, Safian A (2020) Online detection of bearing incipient fault with semi-supervised architecture and deep feature representation. J Manuf Syst 55:179–198
3. Gopal L, Singh H, Mounica P, Mohankumar N, Challa NP, Jayaraman P (2023) Digital twin and IOT technology for secure manufacturing systems. Meas. Sensors 25:100661. https://doi.org/10.1016/j.measen.2022.100661
4. Pavlov V (2022) Security aspects of digital twins in IoT platform. MS thesis. University of Twente
5. Stergiou CL, Koidou MP, Psannis KE (2023) IoT-based big data secure transmission and management over cloud system: a healthcare digital twin scenario. Appl Sci. https://doi.org/10.3390/app13169165
6. Akash SS, Ferdous MS (2022) A blockchain based system for healthcare digital twin. IEEE Access 10:50523–50547
7. Zheng Y, Lu R, Guan Y, Zhang S, Shao J (2021) Towards private similarity query based healthcare monitoring over digital twin cloud platform. In: 2021 IEEE/ACM 29th International symposium on quality of service (IWQOS), IEEE, pp. 1–10
8. Okegbile SD, Cai J, Yi C, Niyato D (2022) Human digital twin for personalized healthcare: vision, architecture and future directions. IEEE Network 37(2):262–269
9. Kumar R, Kumar P, Tripathi R, Gupta GP, Garg S, Hassan MM (2022) A distributed intrusion detection system to detect DDoS attacks in blockchain-enabled IoT network. J Parallel Distrib Comput 164:55–68. https://doi.org/10.1016/j.jpdc.2022.01.030
10. Kumar P, Gupta GP, Tripathi R (2021) Design of anomaly-based intrusion detection system using fog computing for IoT network. Autom Control Comput Sci 55(2):137–147. https://doi.org/10.3103/S0146411621020085
11. Mothukuri V, Khare P, Parizi RM, Pouriyeh S, Dehghantanha A, Srivastava G (2022) Federated-learning-based anomaly detection for IoT security attacks. IEEE Internet Things J 9(4):2545–2554. https://doi.org/10.1109/JIOT.2021.3077803

12. Kumar P, Tripathi R, Gupta GP (2021) P2IDF: a privacy-preserving based intrusion detection framework for software defined Internet of Things-Fog (SDIoT-Fog). ACM International conference proceeding series, pp. 37–42
13. Belhadi A, Djenouri Y, Srivastava G, Djenouri D, Lin JCW, Fortino G (2021) Deep learning for pedestrian collective behavior analysis in smart cities: a model of group trajectory outlier detection. Inf Fusion 65:13–20. https://doi.org/10.1016/j.inffus.2020.08.003
14. Latif SA et al (2022) AI-empowered, blockchain and SDN integrated security architecture for IoT network of cyber physical systems. Comput Commun 181:274–283. https://doi.org/10.1016/j.comcom.2021.09.029
15. Kumar R, Kumar P, Tripathi R, Gupta GP, Kumar N, Hassan MM (2022) A privacy-preserving-based secure framework using blockchain-enabled deep-learning in cooperative intelligent transport system. IEEE Trans Intell Transp Syst 23(9):16492–16503. https://doi.org/10.1109/TITS.2021.3098636
16. Kumar P, Gupta GP, Tripathi R (2021) TP2SF: a trustworthy privacy-preserving secured framework for sustainable smart cities by leveraging blockchain and machine learning. J Syst Archit. https://doi.org/10.1016/j.sysarc.2020.101954
17. Kumar P, Kumar R, Gupta GP, Tripathi R (2021) A dis0074ributed framework for detecting DDoS attacks in smart contract-based blockchain-IoT systems by leveraging fog computing. Trans Emerg Telecommun Technol 32(6):e4112
18. Alazzam H, Sharieh A, Sabri KE (2020) A feature selection algorithm for intrusion detection system based on pigeon inspired optimizer. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2020.113249
19. Zhao L, Li H, Lin N, Lin M, Fan C, Shi J (2022) Intelligent content caching strategy in autonomous driving toward 6G. IEEE Trans Intell Transp Syst 23(7):9786–9796. https://doi.org/10.1109/TITS.2021.3114199
20. Sharma V, Akshi K, Kapil S (2024) Digital twin application in women's health: Cervical cancer diagnosis with CervixNet. Cogn Syst Res 87:101264
21. Damjanovic-Behrendt V (2018) A digital twin architecture for security, privacy and safety. ERCIM NEWS 115:25–26
22. Yi H (2023) Improving cloud storage and privacy security for digital twin based medical records. J Cloud Comput. https://doi.org/10.1186/s13677-023-00523-6
23. Grübel J, Thrash T, Aguilar L, Gath-morad M, Chatain J (2022) The Hitchhiker's guide to fused twins: a review of access to digital twins in situ in smart cities. Remote Sens 14:3095
24. Minerva R, Lee GM, Crespi N (2020) Digital twin in the IoT context: a survey on technical features, scenarios, and architectural models. Proc IEEE 108:1785–1824. https://doi.org/10.1109/JPROC.2020.2998530
25. Varghese SA, Ghadim AD, Balador A, Alimadadi Z, Papadimitratos P (2022) Digital twin-based intrusion detection for industrial control systems. In: 2022 IEEE international conference on pervasive computing and communications workshops and other affiliated events (PerCom Workshops), IEEE, pp. 611–617.
26. Empl P (2023) Digital-twin-based security analytics for the Internet of Things. Information 14(2):95
27. Sancho N (2023) A review of digital twins and their application in cybersecurity based on artificial intelligence. arXiv:2311.01154
28. Chen J, Yi C, Okegbile SD, Cai J, Shen XS (2023) Networking architecture and key supporting technologies for human digital twin in personalized healthcare: a comprehensive survey. IEEE Commun Surv Tutor 26(1):706–746
29. Liao Y, Vemuri VR (2002) Use of k-nearest neighbor classifier for intrusion detection. Comput Secur 21(5):439–448
30. Gu J, Wang L, Wang H, Wang S (2019) A novel approach to intrusion detection using SVM ensemble with feature augmentation. Comput Secur 86:53–62
31. Kumar GR, Mangathayaru N, Narasimha G (2015) An improved k-means clustering algorithm for intrusion detection using Gaussian function. In: Proceedings of the the international conference on engineering & MIS 2015, pp. 1–7
32. Riyaz B, Ganapathy S (2020) A deep learning approach for effective intrusion detection in wireless networks using CNN. Soft Comput 24(22):17265–17278. https://doi.org/10.1007/s00500-020-05017-0
33. Imrana Y, Xiang Y, Ali L, Abdul-Rauf Z (2021) A bidirectional LSTM deep learning approach for intrusion detection. Expert Syst Appl 185:115524. https://doi.org/10.1016/j.eswa.2021.115524
34. Kasongo SM (2023) A deep learning technique for intrusion detection system using a recurrent neural networks based framework. Comput Commun 199:113–125. https://doi.org/10.1016/j.comcom.2022.12.010
35. Kunhare N, Tiwari R, Dhar J (2022) Intrusion detection system using hybrid classifiers with meta-heuristic algorithms for the optimization and feature selection by genetic algorithm. Comput Electr Eng 103:108383

36. Ahmad J, Shah SA, Latif S, Ahmed F, Zou Z, Pitropakis N (2022) DRaNN_PSO: a deep random neural network with particle swarm optimization for intrusion detection in the industrial internet of things. J King Saud Univ Inf Sci 34(10):8112–8121
37. Perumalla S, Chatterjee S, Kumar APS (2023) Modelling of oppositional Aquila Optimizer with machine learning enabled secure access control in Internet of drones environment. Theor Comput Sci 941:39–54. https://doi.org/10.1016/j.tcs.2022.08.019
38. Otair M, Ibrahim OT, Abualigah L, Altalhi M, Sumari P (2022) An enhanced grey wolf optimizer based particle swarm optimizer for intrusion detection system in wireless sensor networks. Wirel Netw 28(2):721–744
39. Khanna A, Rani P, Garg P, Singh PK, Khamparia A (2022) An enhanced crow search inspired feature selection technique for intrusion detection based wireless network system. Wirel Pers Commun 127(3):2021–2038
40. Varghese SA, Dehlaghi Ghadim A, Balador A, Alimadadi Z, Papadimitratos P (2022) Digital twin-based intrusion detection for industrial control systems. In: 2022 IEEE international conference on pervasive computing and communications workshops and other affiliated events (PerCom Workshops) pp. 611–617.
41. Khraisat A, Gondal I, Vamplew P, Kamruzzaman J, Alazab A (2019) A novel ensemble of hybrid intrusion detection system for detecting Internet of Things attacks. Electron. https://doi.org/10.3390/electronics8111210
42. Lipsa S, Dash RK (2023) A novel intrusion detection system based on deep learning and random forest for digital twin on IOT platform. Int J Sch Res Eng Technol 2(1):51–64. https://doi.org/10.56781/ijsret.2023.2.1.0020
43. Ashraf E, Areed NF, Salem H, Abdelhay EH, Farouk A (2022) Fidchain: federated intrusion detection system for blockchain-enabled iot healthcare applications. Healthcare 10(6):1110
44. Thakkar A, Lohiya R (2023) Fusion of statistical importance for feature selection in deep neural network-based intrusion detection system. Inf Fusion 90:353–363. https://doi.org/10.1016/j.inffus.2022.09.026
45. Wang H, Di X, Wang Y, Ren B, Gao G, Deng J (2023) An intelligent digital twin method based on spatio-temporal feature fusion for IoT attack behavior identification. IEEE J Sel Areas Commun 41(11):3561–3572. https://doi.org/10.1109/JSAC.2023.3310091
46. Fraser B, Al-Rubaye S, Aslam S, Tsourdos A (2021) Enhancing the security of unmanned aerial systems using digital-twin technology and intrusion detection. In: 2021 IEEE/AIAA 40th Digital avionics systems conference (DASC) pp. 1–10
47. Xu Q, Ali S, Yue T (2021) Digital twin-based anomaly detection in cyber-physical systems. In: 2021 14th IEEE conference on software testing, verification and validation (ICST), pp. 205–216
48. Alzahrani AO, Alenazi MJF (2021) Designing a network intrusion detection system based on machine learning for software defined networks. Futur Internet. https://doi.org/10.3390/fi13050111
49. Khan AA, Laghari AA, Elmannai H, Shaikh AA, Bourouis S, Hadjouni M, Alroobaea R (2023) GAN-IoTVS: a novel internet of multimedia things-enabled video streaming compression model using GAN and fuzzy logic. IEEE Sens J 23:29434–29441
50. Khan AA, Laghari AA, Baqasah AM, Alroobaea R, Almadhor A, Sampedro GA, Kryvinska N (2024) Blockchain-enabled infrastructural security solution for serverless consortium fog and edge computing. Peer J Comput Sci. https://doi.org/10.7717/peerj-cs.1933
51. Khan AA, Laghari AA, Gadekallu TR, Shaikh ZA, Javed AR, Rashid M, Estrela VV, Mikhaylov A (2022) A drone-based data management and optimization using metaheuristic algorithms and blockchain smart contracts in a secure fog environment. Comput Electr Eng 102:108234
52. Khan AA, Laghari AA, Shaikh ZA, Dacko-Pikiewicz Z, Kot S (2022) Internet of Things (IoT) security with blockchain technology: a state-of-the-art review. IEEE Access 10:122679–122695
53. Khan AA, Laghari AA, Baqasah AM, Alroobaea R, Gadekallu TR, Sampedro GA, Zhu Y (2024) ORAN-B5G: a next generation open radio access network architecture with machine learning for beyond 5G in industrial 5.0. IEEE Trans Green Commun Netw
54. Khan AA, Laghari AA, Alroobaea R, Baqasah AM, Alsafyani M, Bacarra R, Alsayaydeh JAJ (2024) Secure remote sensing data with blockchain distributed ledger technology: a solution for smart cities. IEEE Access 12:69383–69396
55. He X, Chen Q, Tang L, Wang W, Liu T (2023) CGAN-based collaborative intrusion detection for UAV networks: a blockchain-empowered distributed federated learning approach. IEEE Internet Things J 10(1):120–132. https://doi.org/10.1109/JIOT.2022.3200121
56. Imran M, Haider N, Shoaib M, Razzak I (2022) An intelligent and efficient network intrusion detection system using deep learning. Comput Electr Eng 99:107764. https://doi.org/10.1016/j.compeleceng.2022.107764
57. Shabbir M et al (2021) Enhancing security of health information using modular encryption standard in mobile cloud computing. IEEE Access 9:8820–8834. https://doi.org/10.1109/ACCESS.2021.3049564

58. Bowen B, Chennamaneni A, Goulart A, Lin D (2023) BLoCNet: a hybrid, dataset-independent intrusion detection system using deep learning. Int J Inf Secur 22(4):893–917. https://doi.org/10.1007/s10207-023-00663-5
59. Ferrag MA, Friha O, Hamouda D, Maglaras L, Janicke H (2022) Edge-IIoTset: a new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning. IEEE Access 10:40281–40306
60. Géron A (2022) Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media Inc., Sebastopol
61. Guyon I (1997) A scaling law for the validation-set training-set size ratio. AT&T Bell Lab 1(11)
62. Zolanvari M, Teixeira MA, Gupta L, Khan KM, Jain R (2019) Machine learning-based network vulnerability analysis of industrial Internet of Things. IEEE Internet Things J 6(4):6822–6834
63. Koroniotis N, Moustafa N, Sitnikova E, Turnbull B (2019) Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-IoT dataset. Futur Gener Comput Syst 100:779–796
64. Vaccari I, Chiola G, Aiello M, Mongelli M, Cambiaso E (2020) MQTTset, a new dataset for machine learning techniques on MQTT. Sensors 20(22):6578
65. Al-Hawawreh M, Sitnikova E, Aboutorab N (2021) X-IIoTID: a connectivity-agnostic and device-agnostic intrusion data set for industrial Internet of Things. IEEE Internet Things J 9(5):3962–3977
66. Al Nuaimi T et al (2023) A comparative evaluation of intrusion detection systems on the edge-IIoT-2022 dataset. Intell Syst Appl 20:200298
67. Rashid MM, Khan SU, Eusufzai F, Redwan MA, Sabuj SR, Elsharief M (2023) A federated learning-based approach for improving intrusion detection in industrial Internet of Things networks. Network 3(1):158–179. https://doi.org/10.3390/network3010008

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

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

Vikas Sharma is a research scholar in the Department of Information Technology at Delhi Technological University, Delhi, India. He received his M. Tech & B. Tech degree in Information Technology from SRM Institute of Science and Technology, Chennai. His research interests are the Digital Twin, Cloud computing, Malware Analysis, Blockchain, and IoT security.

Akshi Kumar is currently working as a senior lecturer (associate professor) and director-postgraduate researcher (PGR) in the Department of Computing at Goldsmiths, University of London, UK. She has pivotal positions at Manchester Metropolitan University, UK, NSUT, and DTU in New Delhi, India. She has received international accolades, including being named in Stanford's "Top 2% Scientist" list in 2021, 2022, 2023, and 2024. His research interests include Affective Computing, Social-media & Network analytics, Natural Language Processing (NLP), and AI-driven healthcare.

Kapil Sharma is currently working as a Professor at Delhi Technological University, Delhi, India, in the Department of Computer Science and Engineering. He has published more than 150 papers in refereed conferences and journals. His research interests include Data Science, Big Data analytics, Web Mining, Social Networks Analysis, Machine Learning, Data Mining, system design, pattern recognition, computer vision, and soft computing.

Photonic Network Communications (2025) 49:2
https://doi.org/10.1007/s11107-024-01026-3

ORIGINAL PAPER

Dual hop Malaga underwater wireless optical communication system with pointing errors for IoUT

Dushyant Singh Chauhan[1] · Gurjit Kaur[1] · Dinesh Kumar[1]

Received: 18 October 2022 / Accepted: 15 November 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
The Internet of Underwater Things (IoUT) has attracted a lot of researchers due to its numerous applications for aquatic life monitoring, underwater surveillance like disaster prevention, and even military applications. But most underwater sensors/devices operate on batteries. So, we propose a dual-hop system that reduces the power requirements at a single node and also helps in increasing the coverage area in underwater wireless optical communication (UWOC). In this paper, the misalignment issue due to water waves is considered using pointing errors and utilizing both optical heterodyne detection (OHD) and intensity modulation/direct detection (IM/DD). Further, the effect of various turbulence conditions is discussed and analytically we present unified exact closed-form expressions for the channel capacity, average bit error rate, and outage probability in terms of the Meijer-G function.

Keywords Internet of Underwater Things (IoUT) · Malaga distribution · IM/DD · Bit error rate (BER) · Outage probability

1 Introduction

In recent years the Internet of things (IoT) has gained a huge attention. As this technology advances it gives rise to other similar terminologies like the IoUT and the network requirements for IoT and IoUT are a bit different as IoUT utilizes battery-operated sensors/devices [1]. Due to UWOC advantages, like high bandwidth, low power, low latency, and high security, it is attracting the researchers, indicating that it has potential for environmental monitoring, underwater surveillance like disaster prevention, and even military surveillance applications [2]. Optical waves, on the other hand, are significantly affected by a variety of inherent disturbances, such as absorption, scattering, and oceanic turbulence, which are caused by density, salinity, and temperature variations in the underwater medium [3].

The challenges for IoUT systems have been discovering ways to analyze the impacts of underwater turbulence on an optical beam and suggesting appropriate communication mechanisms. The impact of underwater turbulence on the system performance has recently been explored by looking at variations in the strength of optical waves. In an oceanic scenario, beam spreading due to turbulence on flat, spherical, asymmetrical Gaussian beam waves has been thoroughly investigated [4]. These studies have also made it possible to analyze and understand the nature of oceanic turbulence. The scintillation index of the optical waves has been investigated further under various turbulence conditions using the extended Huygens–Fresnel principle and Rytov's approximate theory [5]. The effect of underwater turbulence on the optical intensity variations, for example, the multi-Gaussian Schell-model vortex beam, has been investigated analytically [6]. In anisotropic oceanic turbulence, the polarization of an optical wave, structure–function, and the scintillation index (SI), has also been investigated [7]. With an asymmetrical Gaussian beam, the performance of a UWOC system in underwater anisotropic turbulence has been studied and analytically proved that higher bit error rate (BER) can be achieved by increasing the beam size [7]. The effect of underwater turbulence on the channel capacity under strong turbulence has also been investigated, with the faded channel modeled using a gamma-gamma (GG) distribution. The authors [8] developed the scintillation index (SI) of spherical and plane waves traveling through moderate-to-strong

✉ Gurjit Kaur
gurjeet_kaur@rediffmail.com

Dushyant Singh Chauhan
dsc.niet@gmail.com

Dinesh Kumar
dineshkumar@dtu.ac.in

[1] Department of ECE, Delhi Technological University, Delhi 102045, India

Published online: 20 December 2024

underwater turbulence, and a GG-distribution model has been used to examine the average BER (ABER) of the UWOC systems. Furthermore, [9] several modulation and coding schemes have been used, such as phase-shift keying-subcarrier intensity and pulse position modulation, to lower the BER and increase the channel capacity. For weak oceanic turbulence, a log-normal based model has been utilized to determine the channel capacity of UWOC systems with the effect of pointing error taken into account [10]. The authors [11] established the multi-hop relay strategy to increase the UWOC system's system performance.

Furthermore, the impacts of underwater turbulence on plane waves, Laguerre– Gaussian waves, and other types of waves have been extensively researched; however, the effects on spherical waves have received little attention. Many studies have investigated the channel capacity and BER of UWOC systems, but another crucial statistic, the outage probability, which signals when the communication system's signal-to-noise ratio (SNR) falls below a predetermined threshold SNR, has yet to be thoroughly investigated.

The Malaga-M statistical function is a flexible distribution that can represent a large number of optical variations and has mathematical tractability that is appealing for performance investigation. UWOC link outage and failure can be caused by pointing errors and fading channels [12]. Dual-hop systems have been examined in several recent types of research employing amplify-and-forward (AF) and decode-and-forward (DF) routing protocols in an attempt to overcome the listed limits in the FSO environment [13]. To the best of the author's knowledge, mostly all current research either characterizes a weak (lognormal) or a strong (GG) turbulence model [14]. Also due to misalignment of the transceiver, the maximum data may not be able to transfer this is called pointing errors. These errors arise in the underwater channel due to water waves inside oceans [15].

Our goal is to design an underwater channel model that best characterizes the ocean turbulence and compute performance metrics that handle the issues that arise in the high data transmission with low latency as a result of the turbulence-induced fading. Specifically, we want to increase the coverage while lowering the power requirements at a single node. To achieve this goal:

- We propose a model which works best under weak and strong turbulence-induced fading conditions with the help of Malaga distribution.
- We have applied dual-hop relaying to increase the coverage area while reducing the power requirements in the IoUT systems.
- Due to water waves present in oceans, we also take care of misalignment issues in the UWOC system.
- On–Off Keying (OOK) modulation has been employed in a lot of research, we investigated the UWOC system in both OHD and IM/DD employing frequency and phase shift keying (FSK and PSK) modulation.
- Using analytical methods, we derived the expressions for ABER, outage probability (OP), and ergodic capacity (EC), which best represent the UWOC system's performance.

We think that our research will support the performance assessment of IoUT systems. The rest of this paper is arranged in the following manner. In Section II, we describe the UWOC system and channel modeling utilizing Dual Hop relaying while considering the effect of pointing errors. Also, analytically we derived the performance indicators such as EC, ABER, and OP. The Meijer-G function is used to evaluate closed-form equations analytically. The numerical results in terms of EC, ABER, and OP are presented in Section III. Finally, Section IV brings the paper to a conclusion.

2 System and channel model

Figure 1 shows a schematic representation of an IoUT system in which two sensors are communicated using a relay node to improve the communication range and data rate. For a typical optical wireless system, the laser sources and photodiodes are located at the transmitter and receiver sides, respectively. The optical wave is first emitted and then transmitted over a transmission distance (*L)* through the underwater medium from sensor-1. Due to sun rays the temperature of water varies which results in a variation of refractive index. One other factor which indirectly varies the refractive index (RI) of water is the salinity fluctuations in an underwater environment. Both the factors will result in the optical intensity to fluctuate significantly. The photodiode on the relay node will then receive and collect the scattered signal before transmitting it to sensor-2.

As the author [8] presented the power spectrum (ϕ_n^{ocean}) of RI using the classical Kolmogorov model and various parameters used to calculate the power spectrum are shown in Table 1.

$$\phi_n^{\text{ocean}}(k) = k^{-11/3}\chi_T \frac{\varepsilon^{-1/3}}{2.577\times 10^8} \times \left(e^{-A_T\delta} - 2\tau^{-1}e^{-A_{TS}\delta} + \tau^{-2}e^{-A_S\delta}\right) \times \left[1 + 2.35(k\eta)^{2/3}\right] \tag{1}$$

where $\delta = 8.284(k\xi)^{4/3} + 12.978(k\xi)^2$, $A_T = 1.863\times 10^{-2}, k = 2\pi/\lambda$, $A_{TS} = 1.9\times 10^{-4}$, $A_S = 9.41\times 10^{-3}$.

The SI of the optical waves has been investigated further under various turbulence conditions using the Rytov variance model. So, variance based on this model is given by

$$\sigma_R^2(z,\lambda) = 78.87zK^2 \int_0^1 \int_0^\infty k\phi_n^{\text{ocean}}(k)\left[1 - \cos\left(\frac{zk^2}{K}\xi\right)\right] \mathrm{d}k\mathrm{d}\xi \tag{2}$$

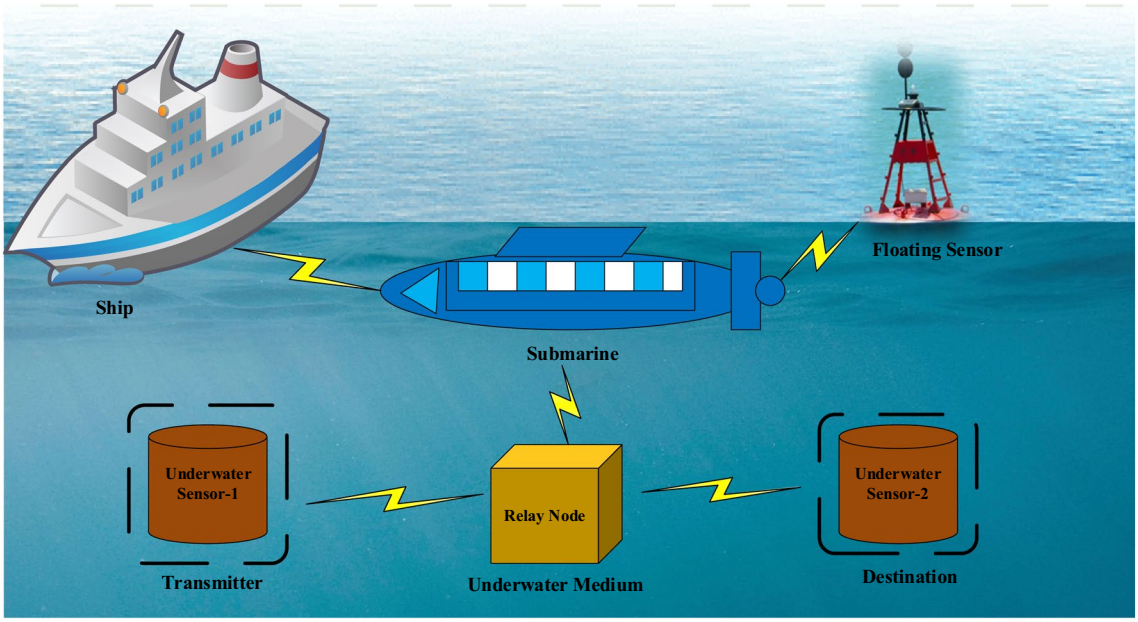


Fig. 1 Dual Hop scheme for UWOC system

Table 1 Parameters associated with the calculate the power spectrum of RI

Symbol	Parameter	Values
k	Spatial frequency	15 MHz at $\lambda = 0.417$ µm
η	Kolmogorov scale	$10^{-3}m$
χ_T	Dissipation rate (temperature)	10^{-10} to $10^{-2}\mathrm{K}^2/\mathrm{s}$
τ	Range of temperature and salinity-induced RI	[−5 and 0]
ϵ	Dissipation rate of the turbulent kinetic energy	$10^{-1}\mathrm{m}^2/\mathrm{s}^3$ to $10^{-10}\mathrm{m}^2/\mathrm{s}^3$

where σ_R^2 is known as Rytov's variance, also α and β are parameters associated with the effective underwater conditions in [16] as $\alpha = \left\{ \exp\left[\frac{0.49\sigma_R^2}{\left(1+1.11\sigma_R^{12/5}\right)^{7/6}} \right] - 1 \right\}^{-1}$ and $\beta = \left\{ \exp\left[\frac{0.51\sigma_R^2}{\left(1+0.69\sigma_R^{12/5}\right)^{5/6}} \right] - 1 \right\}^{-1}$.

Further, α, β has been calculate with the help of Rytov variance [i.e., $(\alpha = 2.296, \beta = 2), (\alpha = 4.2, \beta = 3)$, and $(\alpha = 8, \beta = 4)$] for strong, moderate, and weak turbulence conditions, respectively] [12].

Malaga distribution function has been used to quantify turbulent channel fading conditions under the scenario of weak-to-strong underwater turbulence. The fading model based on this distribution has been found to accurately describe the optical behavior when propagating in a homogenous medium with modest to heavy turbulence. The probability distribution function (PDF) of generalized Malaga distribution model's for underwater optical turbulence (UOT) is expressed in [16] as

$$f_a(I_a) = \sum_{i=1}^{2} A_i \sum_{k=1}^{\beta_i} a_{i,k} I_a K_{\alpha_i - k}\left(2\sqrt{\frac{\alpha_i \beta_i I_a}{g_i \beta_i + \Omega_i'}} \right), \tag{3}$$

and,

$$A_i = \frac{2\alpha_i^{\alpha_i/2}}{g_i^{1+\alpha_i/2}\Gamma(\alpha_i)} \left(\frac{g_i \beta_i}{g_i \beta_i + \Omega_i'} \right)^{\beta_i + \alpha_i/2}, \tag{4}$$

$$a_{i,k} = \binom{\beta_i - 1}{k-1} \frac{(g_i \beta_i + \Omega_i')^{1-k/2}}{(k-1)!} \left(\frac{\Omega_i'}{g_i} \right)^{k-1} \left(\frac{\alpha_i}{\beta_i} \right)^{k/2} \tag{5}$$

where K_x is the modified Bessel function of 2nd kind,x is the order, α_i, β_i, g_i and Ω_i are the fading parameters, $\Omega_i' = \Omega_i + 2b_o\rho_i + 2\sqrt{2b_o\rho_i\Omega_i}\cos(\phi_A + \phi_B)$, ϕ_A, ϕ_B are the deterministic phases of the LOS and the coupled-to-LOS scatter terms, respectively,$g_i = 2b_0(1-\rho_i)$ and A_i is a variable. Now the parameters $(\alpha = 2.296, \beta = 2), (\alpha = 4.2, \beta = 3)$, and $(\alpha = 8, \beta = 4)$] are set for strong, moderate, and weak turbulence conditions, respectively [12] along with

$\phi_A + \phi_B = 90^o$ in A_i, which is used in calculating performance matrices like ABER, EC and OP.

Now the intensity fluctuations due to pointing error impairments can be considered and the PDF is given by [17] as

$$f_p(I_p) = \frac{\xi^2}{A_o^{\xi^2}} I_p^{\xi^2-1}, 0 \le I_p \le A_o \tag{6}$$

where ξ is the ratio between the equivalent beam radius at the receiver and the pointing error displacement standard deviation at the receiver, A_o is a constant term that defines the pointing loss. Now the joint probability distribution of $I = I_a I_p$ can be calculated by $f_{I_a,I_p}(I_a, I_p) = f_a(I_a) f_I(I/I_a)$. Now using Eq. (7) in Eq. (3) along with random variable transformation in Eq. (6) the PDF of the intensity variation due to both turbulence and pointing error impairments is given in [17] as Eq. (8)

$$K_{\alpha_i-k}\left(2\sqrt{\frac{\alpha_i\beta_i I_a}{g_i\beta_i + \Omega_i\prime}}\right) = \frac{1}{2} G_{0,2}^{2,0}\left(\frac{\alpha_i\beta_i I_a}{g_i\beta_i + \Omega\prime}\middle| \begin{matrix} - \\ \frac{\alpha_i-k}{2}, \frac{\alpha_i-k}{2} \end{matrix}\right) \tag{7}$$

$$f_I(I) = \sum_{i=1}^{2} \frac{\xi_i^2 A_i}{2I} \sum_{k=1}^{\beta_i} b_{i,k} G_{1,3}^{3,0}\left[\frac{\alpha_i\beta_i}{g_i\beta_i + \Omega\prime}\frac{I}{A_o}\middle| \begin{matrix} \xi_i^2+1 \\ \xi_i^2, \alpha_i, k \end{matrix}\right] \tag{8}$$

where $b_{i,k} = a_{i,k}[\alpha_i\beta_i/(g_i\beta_i + \Omega\prime)]^{-(\alpha_i+k)/2}$,$G_{p,q}^{m,n}[.]$ is the Meijer-G function defined in Kilbas et al. [18].

In the UWOC system, we now use the Malaga-distribution function to model the turbulence-induced fading channel. It can be helpful in establishing different performance parameters analytically. With a variety of modulation formats, we presented two detection techniques known as OHD and IM/DD. The absorption/scattering, as well as their impact such as inter-symbol interference, are less as compared to UOT so more emphasis is given to the fading characteristics of the underwater medium. The UWOC suffers from both scattering and absorption as well as UOT. The phase and the intensity of the transmitted optical beam gets distorted, leading to significant signal distortions and impairments such as fading, scintillation, and beam wander, resulting in the degradation of system performance parameters like Bit error rate, Outage probability and Ergodic capacity [19]. Turbulence has a greater impact on underwater optical communications compared to scattering and absorption due to its influence on refractive index variations, affecting received intensity and power loss [20]. Also, the salinity and temperature variation results in the fading of optical signals. These fluctuations occur dynamically and can induce rapid fluctuations in the optical channel, leading to severe signal distortions.

In contrast, absorption and scattering, while prevalent, are often more static and predictable. Overall, the dynamic and unpredictable nature of optical turbulence, combined with its significant impact on signal fidelity, beamforming, and tracking, makes it a major challenge in underwater optical wireless communication, often outweighing the effects of absorption and scattering [21]. So, our proposed model focused more on the fading of the optical signals due to UOT.

For the UWOC system, we first derive the general form of PDF and CDF for Malaga- distribution for both detection techniques. As a result, these relations can be used to evaluate closed-form expressions for performance indicators. Taking into account both types of detection schemes IM/DD and OHD. In [22] demonstrated that IM/DD techniques is cost-effective and has low complexity. In comparison to the previous detection technique, OHD has a higher capacity. Furthermore, at low SNR the gap between capacities of both detection techniques is constant, but it increases at high SNR.

Instantaneous SNR can be given as $\gamma = \frac{(\eta I)^{r_i}}{N_0}$ where r is the parameter associated with the type of detection is shown in Table 2,ith UWOC link where $i = 1$ corresponds to UWOC link established between sensor-1 and relay node and $i = 2$ is the link between the relay node and sensor-2 as shown in Fig. 1 and N0 is the additive white Gaussian noise sample. A general expression for average electrical SNR is $\mu_{r_i} = \frac{(\eta E[I])^{r_i}}{N_0}$. So, the general expression relating to average electrical SNR under OHD ($r = 1$) scheme is given by the relation $I = A_o\xi_i^2(g_i + \Omega\prime)/[\mu_{1_i}(\xi_i^2 + 1)]$. Also, the general expression relating to average electrical SNR and average SNR under OHD ($r = 1$) scheme is given as $\mu_{1_i} = \overline{\gamma}$. The μ_{1_i} is the average electrical SNR for heterodyne detection technique for i$^{\text{th}}$ hop. On substituting μ_{1_i} in Eq. (8), the PDF of the proposed system under OHD scheme for both turbulence and pointing errors is given as

$$f_\gamma(\gamma) = \sum_{i=1}^{2} \frac{\xi_i^2 A_i}{2\gamma} \sum_{k=1}^{\beta_i} b_{i,k} G_{1,3}^{3,0}\left[B_i\left(\frac{\gamma}{\mu_{1_i}}\right)\middle| \begin{matrix} \xi_i^2+1 \\ \xi_i^2, \alpha_i, k \end{matrix}\right], \tag{9}$$

where $B_i = \xi_i^2\alpha_i\beta_i(g_i + \Omega\prime)/[(\xi_i^2 + 1)(g_i\beta_i + \Omega\prime)]$.

Similarly, under IM/DD ($r = 2$) scheme average electrical SNR is given by $\mu_{2_i} = \frac{(\eta E[I])^2}{N_0}$ and can be calculated as

Table 2 Parameter values for both detection techniques

Parameter (r)	Detection technique
1	Optical heterodyne detection (OHD)
2	Intensity-modulated direct detection (IM/DD)

$$\mu_{2_i} = \frac{\alpha_i \xi_i^2 (\xi_i^2+1)^{-2}(\xi_i^2+2)(g_i+\Omega_i\prime)}{(\alpha_i+1)[2g_i(g_i+2\Omega_i\prime)+\Omega_i\prime^2(1+1/\beta_i)]}\mu_{1_i}, \quad (10)$$

Now substituting Eq. (10) in Eq. (8), the PDF under IM/DD scheme is given as

$$f_\gamma(\gamma) = \sum_{i=1}^{2} \frac{\xi_i^2 A_i}{2^2\gamma} \sum_{k=1}^{\beta} b_{i,k} G_{1,3}^{3,0}\left[B_i \left(\frac{\gamma}{\mu_{2_i}}\right)^{1/2} \middle| \begin{matrix} \xi_i^2+1 \\ \xi_i^2, \alpha_i, k \end{matrix} \right], \quad (11)$$

Using (9) and (11), the unified PDF expression for the proposed system reduces to

$$f_\gamma(\gamma) = \sum_{i=1}^{2} \frac{\xi_i^2 A_i}{2^{r_i}\gamma} \sum_{k=1}^{\beta} b_{i,k} G_{1,3}^{3,0}\left[B_i \left(\frac{\gamma}{\mu_{r_i}}\right)^{1/r_i} \middle| \begin{matrix} \xi_i^2+1 \\ \xi_i^2, \alpha_i, k \end{matrix} \right], \quad (12)$$

Further this unified PDF expression helps in obtaining the analytical expressions for the performance evaluation under both types of detection schemes.

Any UWOC system's performance can be mainly evaluated using performance metrics such as EC, ABER, and OP. Average (i.e., Ergodic) capacity is a key parameter that describes the performance of any communication system. As a result, it is defined as the maximum rate at which information bits can be transferred over a UWOC channel with an arbitrarily low probability of error. The UWOC channel is considered to be memoryless and stationary, with the assumption that perfect channel state information (CSI) is available at both the receiving and transmitting ends. Now, the ergodic channel capacity with an average value is given in [19] as

$$\overline{C} \triangleq E[\log_2(1+\tau\gamma)] \quad (13)$$

where τ is constant. For OHD ($\tau=1$) and ($\tau=e/2\pi$) for IM/DD. On substituting Eq. (12) in Eq. (13), and utilizing $\ln(1+\gamma) = G_{2,2}^{1,2}\left[\gamma \middle| \begin{matrix} 1,1 \\ 1,0 \end{matrix}\right]$, along with Eq. (1.1.3) of Kilbas et al. [18], we obtained a generalized expression for ergodic channel capacity as Eq. (14). As it can be observed as average SNR increases, channel ergodic capacity increases as well

$$\overline{C} = \sum_{i=1}^{2} \frac{D_i}{\ln(2)} \sum_{k=1}^{\beta_i} c_{i,k} G_{r_i+2,3r_i+2}^{3r_i+2,1}\left[\left(\frac{E_i}{\mu_{r_i}}\right) \middle| \begin{matrix} 0,1,x_{i1} \\ x_{i2},0,0 \end{matrix} \right], \quad (14)$$

ABER is also an important performance parameter that reflects the effectiveness of a UWOC system. To learn more about the performance of dual-hop UWOC systems the ABER is expressed in [23] for various modulation techniques as

$$P_e = \frac{\delta}{2\Gamma(p)} \sum_{k=1}^{n} \int_0^\infty \Gamma(p, q_k\gamma) f_\gamma(\gamma) d\gamma, \quad (15)$$

where p, q_k are the parameters that specify the type of modulation and $\Gamma(p, q_k\gamma)$ is the incomplete gamma function. Table 3 presents the various parameter values associated with the modulation format [19].

On substituting Eq. (12) in Eq. (15), the ABER for CSI-assisted fixed gain relaying UWOC system employing Malaga distribution is given as

$$P_e = \sum_{i=1}^{2} \sum_{j=1}^{n} \frac{\delta D_i}{2\Gamma(p)} \sum_{k=1}^{\beta_i} c_{i,k} G_{r_i+2,3r_i+1}^{3r_i,2}\left[\left(\frac{E_i}{\mu_{r_i} q_k}\right) \middle| \begin{matrix} 1-p,1,x_{i1} \\ x_{i2},0 \end{matrix} \right], \quad (16)$$

where $G_{p,q}^{m,n}[.]$ is Meiger-G function,$D_i = \xi_i^2 A_i/[2^{r_i}(2\pi)^{r_i-1}]$, $c_{i,k} = b_k r_i^{\alpha_i+k-1}$, $E_i = B^{r_i}/r^{2r_i}$. Both x_{i1} and x_{i2} are the intermediate variables depending upon the detection schemes, pointing errors, fading conditions and $x_{i1} = \frac{\xi_i^2+1}{r_i}, ..., \frac{\xi_i^2+r_i}{r_i}$ consists of r_i terms and $x_{i2} = \frac{\xi_i^2}{r_i}, ..., \frac{\xi_i^2+r_i-1}{r_i}, \frac{\alpha_i}{r_i}, ..., \frac{\alpha_i+r_i-1}{r_i}, \frac{k}{r_i}, ..., \frac{k+r_i-1}{r_i}$ consists of $3r_i$ terms.

Under various turbulence conditions the OP is an important performance indicator for evaluating the performance of the UWOC system. And it is defined as the probability of the instantaneous SNR falling below a specified threshold value γ_{th}, mathematically OP is expressed in [19] as

$$P_{out} = P_r(\gamma \le \gamma_{th}) = F_\gamma(\gamma_{th}), \quad (17)$$

where $F_\gamma(\gamma)$ is the CDF under Malaga fading channel. On substituting Eq. (12) in $F_\gamma(\gamma) = \int_0^\gamma f_\gamma(\gamma)d\gamma$, the unified CDF can be evaluated as

Table 3 Parameters associated with various modulation formats

Modulation	δ	p	q_k	n	Detection type
On–off keying (OOK)	1	0.5	0.25	1	IM/DD
Binary phase shift keying (BPSK)	1	0.5	1	1	OHD
M-ary phase shift keying (M-PSK)	$\frac{2}{\max(\log_2 M,2)}$	0.5	$\sin^2\left(\frac{(2k-1)\pi}{M}\right)$	$\max\left(\frac{M}{4},1\right)$	OHD
M-ary quadrature amplitude modulation (M-QAM)	$\frac{4}{\log_2 M}\left(1-\frac{1}{\sqrt{M}}\right)$	0.5	$\frac{3(2k-1)^2}{2(M-1)}$	$\frac{\sqrt{M}}{2}$	OHD

$$F_\gamma(\gamma) = \sum_{i=1}^{2} D_i \sum_{k=1}^{\beta_i} c_{i,k}.G_{r_i+1,3r_i+1}^{3r_i+1,1}\left[E_i\left(\frac{\gamma}{\mu_{r_i}}\right)\middle|\begin{matrix}1, x_{i1}\\ x_{i2}, 0\end{matrix}\right], \quad (18)$$

Substituting Eq. (18) in Eq. (17), the closed-form expression for OP is given as

$$P_{out} = F_\gamma(\gamma_{th}) = \sum_{i=1}^{2} D_i \sum_{k=1}^{\beta_i} c_{i,k}G_{r_i+1,3r_i+1}^{3r_i+1,1}\left[E_i\left(\frac{\gamma_{th}}{\mu_{r_i}}\right)\middle|\begin{matrix}1, x_{i1}\\ x_{i2}, 0\end{matrix}\right], \quad (19)$$

In the next section, we will observe the performance metrics of our proposed system with varying pointing errors under various turbulent conditions. Also, we compare the capacity of our proposed system with the model presented in [12].

3 Numerical results

The impact of channel and system characteristics on the ABER, OP, and EC of the UWOC system is explored in this section. Every figure is labeled with the type of detection technique along with the values of turbulence and pointing errors. Note that these parameters provide a weak-to-strong turbulence scenario that the Málaga distribution fading model can accurately characterize.

Under fixed pointing error, the ABER performance of the proposed system under both types of detection techniques operated under various turbulence conditions is shown in Fig. 2. Another significant result is that the OHD outperforms the IM/DD scheme, even though it is a more complex method of detection.

Under fixed pointing effect the ABER of a CSI-assisted relaying UWOC Malaga-M fading system under weak ($\alpha = 8, \beta = 4$) and strong ($\alpha = 2.296, \beta = 2$) turbulence scenarios (Guanjun et al. 2020) operated under various modulation schemes are shown in Fig. 3. Various modulation techniques of OHD schemes are compared where BPSK outperforms higher modulation schemes of QAM because BPSK is more robust to signal distortions and fluctuations induced by UOT. In 16-PSK modulation, only the phase of the carrier signal is varied to represent different signals whereas in 16-QAM modulation both amplitude and phase are varied to represent different signals. So, under weak turbulence conditions, the effect on optical signal distortion is minimal and hence both 16-PSK and 16-QAM show similar performance, whereas under high underwater turbulence conditions, there is a significant variance in their performance because in 16-PSK, the phase of the transmitted signal undergoes rapid and unpredictable changes, which result in phase errors. On the other side, the 16-QAM depends both on phase and amplitude variations but mainly relies on amplitude variations to convey

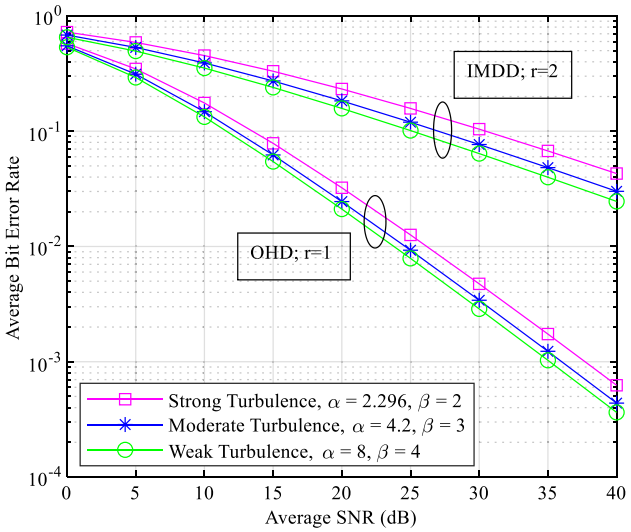


Fig. 2 ABER vs average SNR operated under both detection schemes for various turbulence conditions

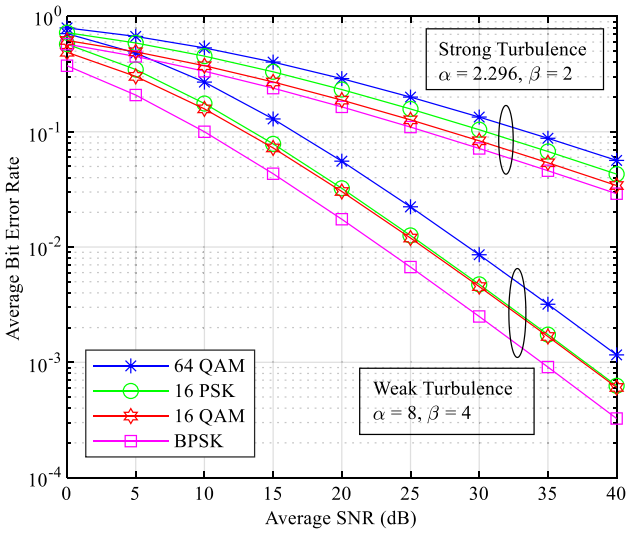


Fig. 3 ABER vs average SNR under strong and weak turbulence conditions for various modulation techniques operated under OHD scheme

symbol information, so it is more robust to phase distortions leading to a better performance under strong turbulence conditions.

Figure 4 shows the ABER performance for various turbulence conditions under the varying effect of pointing errors (i.e., ($\xi_1 = 1.1, \xi_2 = 0.8$) for strong and ($\xi_1 = 6, \xi_2 = 6.7$) for weak) using 16-PSK modulation. It can be observed that at low SNR up to 5 dB the effect of noise is high which overshadowed the effect of turbulence and hence shows similar performance, whereas at high SNR, and system operated under strong pointing errors are due to the misalignment of

the transmitted signal and receiver's aperture, causing signal loss which degrade the UWOC system performance.

Under strong pointing errors ($\xi_1 = 1.1, \xi_2 = 0.8$), the ABER performance of the proposed system operated under various turbulence conditions is shown in Fig. 5. Here mixed detection scheme is employed, and it shows that OHD/OHD scheme outperforms OHD/IMDD scheme. In OHD, the incoming signal is mixed with a local oscillator frequency, resulting in a beat frequency that is much easier to detect even in the presence of noise. This enhanced SNR helps maintain signal integrity over longer transmission distances and through multiple hops the performance of OHD/OHD is better than OHD/IMDD.

Figures 6 and 7 show the EC of the proposed system under the various level of turbulence. As the effect of turbulence and/or severe pointing condition, the EC deteriorates. For fixed pointing error ($\xi = 1$) under both detection schemes, the OHD technique outperforms IM/DD technique. It is shown in Fig. 6 that at higher SNR the transmitted signal becomes more robust against noise so the capacity improves drastically.

Figure 7a presents the EC under the strong-pointing error condition ($\xi_1 = 1.1, \xi_2 = 0.8$). As underwater turbulence (α, β) increases, the proposed system's EC decreases. Also, with a higher SNR, the signal power is much greater than the noise power in the channel. Consequently, the impact of noise on the received signal is reduced, leading to an increase in EC. We also compare the EC of the proposed system with [12]'s) system. We employed dual-hop relaying, whereas [12] used single relaying. In dual-hop data transmission, data transmission takes place with the help of a relay node. By splitting the total path into two hops, we mitigate scattering and signal degradation, thereby improving the signal quality and enhancing EC.

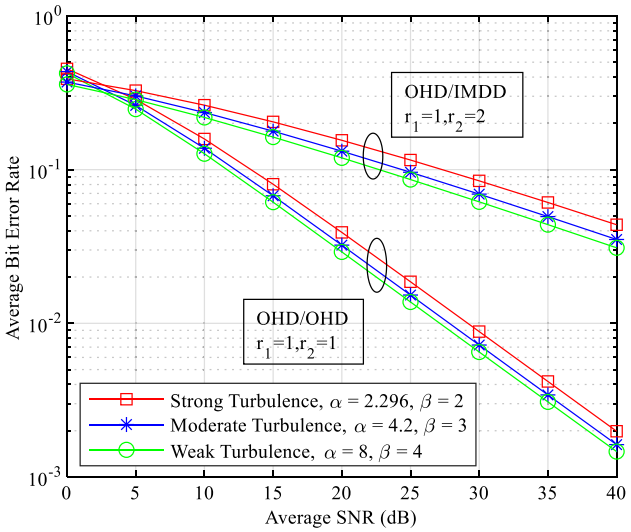


Fig. 5 ABER vs average SNR for various turbulence conditions operated under mixed detection schemes

Figure 7b shows the ergodic capacity under the weak pointing error conditions ($\xi_1 = 6, \xi_2 = 6.7$). The signal loss due to underwater turbulence is minimal, so there is an increase in EC. At low SNR, the noise is dominant in the signal, so the EC of dual-hop relaying and single relaying becomes comparable. Noise overshadows the advantages of dual-hop relaying at low SNR values. When the SNR is high, the difference in EC between dual-hop and single-relaying increases due to the variations in SNR distributions and diversity gains. Dual-hop

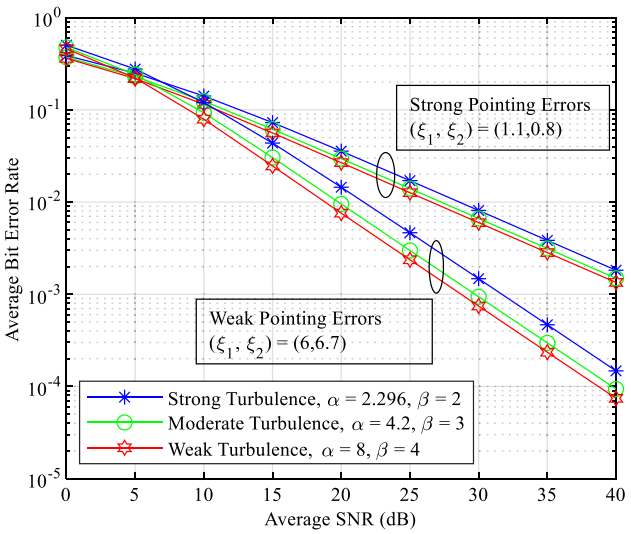


Fig. 4 ABER vs average SNR for various turbulence conditions under weak and strong pointing errors using 16-PSK modulation

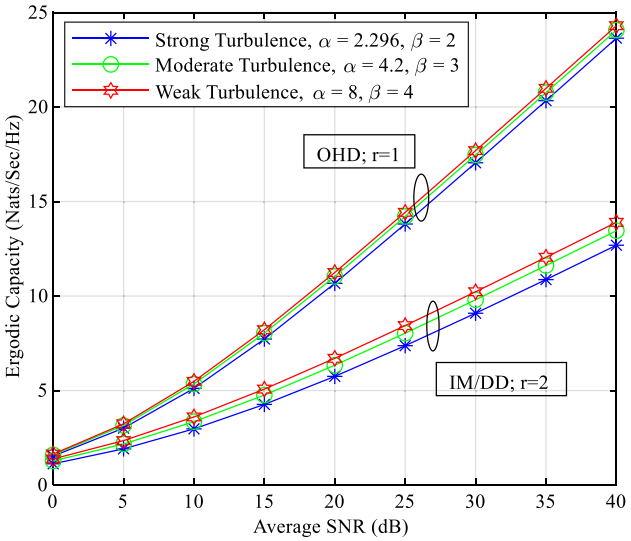
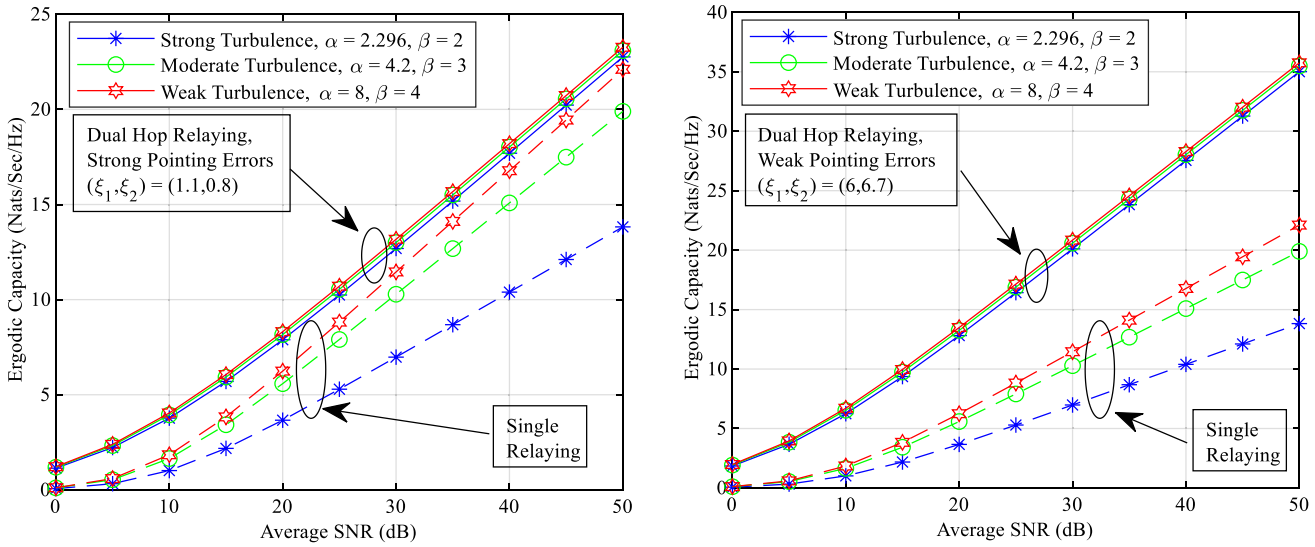


Fig. 6 Ergodic capacity vs average SNR for various turbulence conditions operated under both detection schemes



(a) EC under strong pointing conditions $(\xi_1 = 1.1, \xi_2 = 0.8)$ (b) EC under weak pointing conditions $(\xi_1 = 6, \xi_2 = 6.7)$

Fig. 7 Ergodic capacity vs average SNR for various turbulence conditions

relaying involves amplifying the signal and reducing noise at the relay node, resulting in better performance as compared to single-hop relaying.

In Fig. 8, the OP of the UWOC system is plotted as a function of average SNR for three different values of α and β[i.e., $(\alpha = 2.296, \beta = 2)$, $(\alpha = 4.2, \beta = 3)$, and $(\alpha = 8, \beta = 4)$] are set for strong, moderate, and weak turbulence conditions, respectively [12]. It is self-evident that the OP is substantially influenced by underwater turbulence conditions. As a result, the OP in weak turbulence is substantially lower than in moderate and heavy turbulence. It's worth noting that similar conclusions have been reached for optical waves in air turbulence. The OP degrades as expected when the pointing error displacement standard deviation is reduced, i.e., for smaller ξ, or the turbulence fading parameter is reduced, i.e., for smaller α and β. Results show that the performance degradation caused by strong turbulence is more severe under weak pointing error situations than under strong pointing error levels.

Using fixed-gain relaying, the OP for various UWOC detection algorithms (i.e., IM/DD and OHD) and pointing error severity are shown in Fig. 9. Using IM/DD detection in a single UWOC link, regardless of pointing defects, results in a significant performance reduction when compared to all OHD-based UWOC transmissions since OHD uses complex coherent receivers.

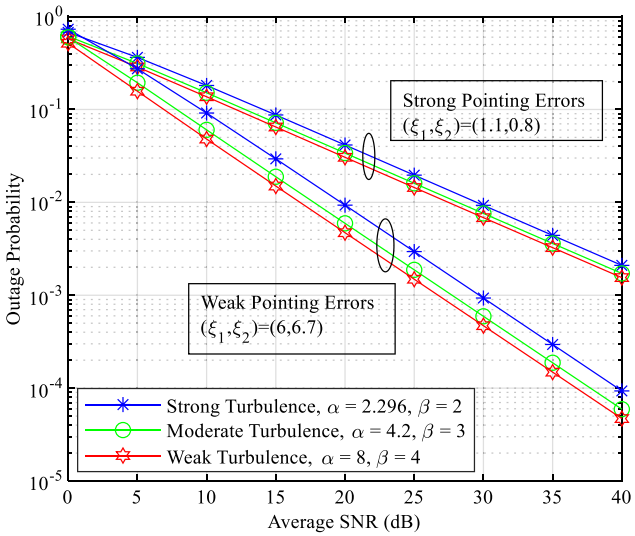


Fig. 8 Outage probability vs average SNR for various turbulence conditions operated under variable pointing errors

4 Conclusion

In this paper, we present a model in which underwater turbulence-induced fading is characterized by Malaga-M distribution. We applied dual-hop communication under fixed gain relaying to increase the coverage while reducing the power requirements at a single node in the IoUT systems. We analyzed the system under different turbulence conditions using both types of pointing errors. In addition to this, we investigated the UWOC system in both IM/DD and OHD techniques employing various modulation schemes. Using the Meijer-G function analytically closed-form expressions have been derived for ABER, OP, and EC, which best represent the UWOC system's performance.

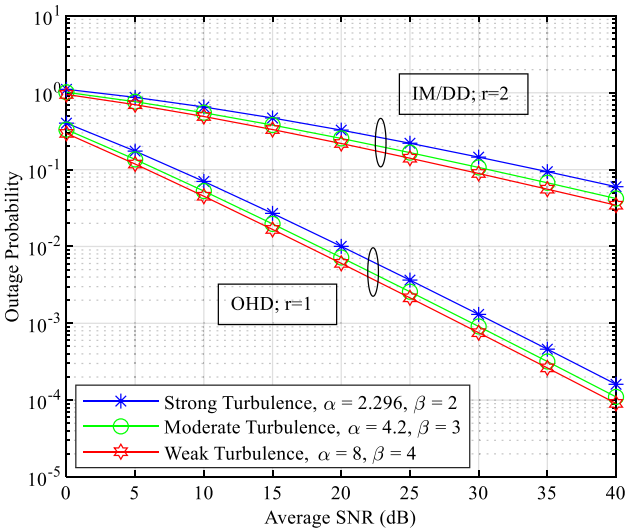


Fig. 9 Outage probability vs average SNR for various turbulence conditions operated under both detection schemes

Author contributions All authors contributed to the study conception and design. Material preparation, modeling and analysis were performed by Dushyant Singh Chauhan, Gurjit Kaur and Dinesh Kumar. The first draft of the manuscript was written by Dushyant Singh Chauhan, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding No funding is involved in the presented research work.

Data availability Data sharing is not applicable to this article as no data were created or analyzed in this study.

Declarations

Conflict of interest We declare that the authors have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Ethical approval Not applicable.

References

1. Kaushal, H., Kaddoum, G.: Underwater optical wireless communication. IEEE Access **4**, 1518–1547 (2016)
2. Gabriel, C., Khalighi, M. A., Bourennane, S., Léon, P., Rigaud, V.: Investigation of suitable modulation techniques for underwater wireless optical communication. In: 2012 International Workshop on Optical Wireless Communications (IWOW). Oct 22;1–3: IEEE (2012).
3. Zuo, Z., Wang, Y., Mao, Y., Ruan, X., Guo, Y.: Security of quantum communications in oceanic turbulence. Phys. Rev. A **104**(5), 052613 (2021)
4. Yu, L., Zhang, Y.: Beam spreading and wander of partially coherent Lommel-Gaussian beam in turbulent biological tissue. J. Quant. Spectrosc. Radiat. Transfer **217**, 315–320 (2018)
5. Charnotskii, M.: Extended Huygens-Fresnel principle and optical waves propagation in turbulence: discussion. JOSA A **32**(7), 1357–1365 (2015)
6. Zhang, Y., Liu, L., Zhao, C., Cai, Y.: Multi-Gaussian Schell-model vortex beam. Phys. Lett. A **378**(9), 750–754 (2014)
7. Baykal, Y.: Effect of anisotropy on intensity fluctuations in oceanic turbulence. J. Mod. Opt. **65**(7), 825–829 (2018)
8. Ata, Y., Baykal, Y.: Scintillations of optical plane and spherical waves in underwater turbulence. JOSA A **31**(7), 1552–1556 (2014)
9. Faridzadeh, M., Gholami, A., Ghassemlooy, Z., Rajbhandari, S.: Hybrid pulse position modulation and binary phase shift keying subcarrier intensity modulation for free space optics in a weak and saturated turbulence channel. JOSA A **29**(8), 1680–1685 (2012)
10. Dong Y, Liu J.: On BER performance of underwater wireless optical MISO links under weak turbulence. In Oceans 2016-Shanghai,1–4, IEEE (2016).
11. Gupta, A., Sharma, N., Jain, M., Garg, P.: Multihop underwater optical wireless communication system for internet of underwater things applications. Int. J. Commun. Syst. **35**, e5041 (2022)
12. Xu, G., Song, Z., Zhang, Q.: Outage probability and channel capacity of an optical spherical wave propagating through anisotropic weak-to-strong oceanic turbulence with Málaga distribution. JOSA A **37**(10), 1622–1629 (2020)
13. Jin, M., Liu, W., Hao, Y., Wu, R., Wei, Z., Deng, D., Liu, H.: Hybrid dual-hop RF/FSO terrestrial-deep space communication system under solar scintillation during superior solar conjunction. Appl. Sci. **12**(2), 619 (2022)
14. Gerçekcioğlu, H.: Bit error rate of focused Gaussian beams in weak oceanic turbulence. JOSA A **31**(9), 1963–1968 (2014)
15. Lu, H., Chen, W., Jiang, M.: Deep learning aided misalignment-robust blind receiver for underwater optical communication. IEEE Wirel. Commun. Lett. **10**(9), 1984–1988 (2021)
16. Boluda-Ruiz, R., Salcedo-Serrano, P., Castillo-Vázquez, B., García-Zambrana, A., Garrido-Balsells, J.M.: Capacity of underwater optical wireless communication systems over salinity-induced oceanic turbulence channels with ISI. Opt. Express **29**(15), 23142–23158 (2021)
17. Ansari, I.S., Yilmaz, F., Alouini, M.S.: Performance analysis of free-space optical links over málaga turbulence channels with pointing errors. IEEE Trans. Wirel. Commun. **15**(1), 91–102 (2015)
18. Kilbas, A., Saigo, M.: H-Transforms: Theory and Applications (Analytical Method and Special Function), 1st edn. CRC Press, Boca Raton (2004)
19. Zedini, E., Oubei, H.M., Kammoun, A., Hamdi, M., Ooi, B.S., Alouini, M.S.: Unified statistical channel model for turbulence-induced fading in underwater wireless optical communication systems. IEEE Trans. Commun. **67**(4), 2893–2907 (2019)
20. Vali, Z., Gholami, A., Ghassemlooy, Z., Michelson, D.G., Omoomi, M., Noori, H.: Modeling turbulence in underwater wireless optical communications based on Monte Carlo simulation. J. Opt. Soc. Am. A Opt. Image Sci. Vis. **34**(7), 1187–1193 (2017)
21. Ata, Y., Kiasaleh, K.: Analysis of optical wireless communication links in turbulent underwater channels with wide range of water parameters. IEEE Trans. Veh. Technol. **72**(5), 6363–6374 (2023)
22. Yu, Y., Che, Y., Bo, T., Kim, D., Kim, H.: Low-complexity, loop-unrolled decision-feedback equalizer for IM/DD system using PAM formats. In Optical Fiber Communication Conference, Th1A-44. Optical Society of America (2021).
23. Chauhan, D.S., Kaur, G., Kumar, D.: Design of novel MIMO UOWC link using gamma–gamma fading channel for IoUTs. Opt. Quant. Electron. **54**(8), 1–5 (2022)

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

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

Mr. Dushyant Singh Chauhan is currently pursuing his Ph.D. degree in Department of Electronics & Communication Engineering, Delhi Technological University, Delhi, India. His research interest focuses on the modelling of underwater channel. He has received his Master's degree (M.E.)in Electronics from Punjab Engineering College, Chandigarh, India. He has received her Bachelor's degree (B.Tech.) in Electronics & Communication Engineering from Dr. APJ Abdul Kalam Technical University, Lucknow, India. He has qualified GATE & UGC-NET in 2018. He has received Commendable Research Award for excellence in 2022 from Research and Council Board of DTU, Delhi. He has published several SCI/SCIE/Scopus Indexed Journals, book chapters, patents and presented his research work in several international conferences.

Dr. Gurjit Kaur worked as a Professor in the Department of Electronics & Communication Engineering at the Delhi Technological University (DTU), Delhi, India. She has been a topper throughout her academic education. As a testimonial to the same, she has been awarded by Chief Minister S. Prakash Singh Badal for being the topper in Punjab state. After that, she was awarded a 'Gold Medal' by the former President of India Dr. A P J Abdul Kalam for being the overall topper of the Punjab Technical University, Jalandhar in the B.Tech. program. She also received an Honour from the Guru Harkrishan Education Society for being a topper among all the colleges and all the disciplines of PTU, Jalandhar. She also worked as a convener for two international conferences i.e., International ICIAICT 2012 which was organized by CSI, Noida Chapter, and international conference EPPICTM 2012 which was held in collaboration with M TMI, USA, University of Maryland Eastern Shore, USA, and Frostburg State University, USA at School of Information and Communication Technology, Gautam Buddha University, Greater Noida, India.

Prof. Dinesh Kumar is working as Faculty in the Department of Electronics and Communication Engineering of Delhi Technological University (DTU) (Erstwhile Delhi College of Engineering). Prior to his joining DTU in 2018, he served Guru Jambheshwar University of Science and Technology (GJUST) (A Haryana State Government University), Hisar, Haryana in various capacities as Reader (1997–2005), Associate Professor (2006–2009), and Professor (2009–2018). He was with Deenbandhu Chhotu Ram University of Science and Technology (DCRUST) (A Haryana State Government University) (Erstwhile C. R. State College of Engineering), Murthal, Sonipat, Haryana and worked there as a Lecturer (1992-1997). Before his joining DCRUST, he also served, though for a very short span of time, the National Institute of Technical Teachers Training and Research (NITTTR) (Erstwhile Technical Teachers' Training Institute), Chandigarh as a Lecturer (1992). He did his B.Tech. ECE (Hons.) from the National Institute of Technology (NIT) (Formerly Regional Engineering College), Kurukshetra in 1990 and M. Tech. (ECE) (Gold Medal) from the same institute in 1992. He did his Ph.D. at Guru Gobind Singh Indraprastha University (GGSIPU), Delhi in 2009. His research work includes Image Processing, Biometrics, Computer Vision, Medical Imaging, Natural Language Processing, Clustering, and Soft Computing.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3523182

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Effect of Number of Probes on Spherical Near-field Measurement System

DONGMING WEI[1], JUNWEI DONG[2], RAVI KUMAR ARYA[2], YIDA FAN[3], SHIYUAN KONG[2], AND LIJUAN LI[2]

[1]School of Electronic Information Engineering, Changchun University of Science and Technology, Changchun, Jilin, China (e-mail: 17261854769@163.com)
[2]Xiangshan Laboratory, Zhongshan Institute of Changchun University of Science and Technology, Zhongshan, Guangdong, China (e-mail: dongjunwei@163.com, raviarya@cust.edu.cn)
[3]School of Optoelectronic Engineering, Changchun University of Science and Technology, Changchun, Jilin, China (e-mail: FanYD_cust@outlook.com, 184217177@qq.com, custjuan@126.com)

Corresponding author: Junwei Dong (e-mail: dongjunwei@163.com)

This work is funded by the Zhongshan Social Public Welfare Technology Research Project (2022B2004).

ABSTRACT Multi-probe spherical near-field measurement systems offer significant advantages in antenna characterization but are constrained by a fixed number of probes. Determining the optimal probe count for such systems is challenging, as it directly impacts measurement accuracy, system complexity, and cost-effectiveness. This work explores using a robotic arm to replicate a multi-probe spherical near-field measurement system and demonstrates a method for determining the ideal probe count during system design.

INDEX TERMS multi-probe spherical near-field measurement system, near field, far field, electromagnetics, antennas

I. INTRODUCTION

Near-field and far-field testing are fundamental methods used to characterize electromagnetic fields, each applicable in different spatial regions relative to the source. The near field [1], encompassing the region close to an antenna or electromagnetic source, typically extends within a distance of about 3∼8 wavelengths from the source. Conversely, the far field extends far away from the source, generally the distance satisfies $R > 2D^2/\lambda$, where D is the maximum size of the antenna aperture and λ is the working wavelength of the antenna. Near-field testing involves directly probing the electromagnetic field close to the source. In contrast, far-field testing entails capturing the radiated field using antennas and receivers placed at a distance from the source.

While mathematically intensive, near-field testing can yield superior accuracy to far-field measurements due to its reduced susceptibility to environmental factors such as reflections, scattering, and atmospheric conditions. The proximity of near-field testing to the source renders them less vulnerable to such environmental effects, ensuring more precise and reliable measurements, especially in complex scenarios.

Within near-field testing, three primary geometries — planar, cylindrical, and spherical are commonly utilized [2]–[5]. Among these, Spherical Near Field Testing (SNFT) is the most complex due to its mathematical and implementation complexity. SNFT belongs to a way of antenna near-field measurement technology where the probe moves on the entire spherical scanning surface during measurement and can collect data in all directions, therefore, spherical near-field testing is suitable for medium gain antennas, including fan-beam and wide-beam antennas. In SNFT, an antenna probe scans fields over a spherical surface but being a single probe and a large area to scan, it is time-consuming for each single measurement.

To solve the problem of low efficiency of single-probe spherical near-field measurement, the French company, Satimo (now called MVG, the Microwave Vision Group) [6] proposed a multi-probe spherical near-field measurement system in the 1990s. The system's working principle is still based on the basic spherical near-field measurement theory. Still, the implementation method is innovative, using probe scattering modulation technology and array switch control [7] and replacing turntable rotation sampling with multi-probe electronic sampling in one dimension of the scanning surface, which realizes a significant improvement in measurement efficiency.

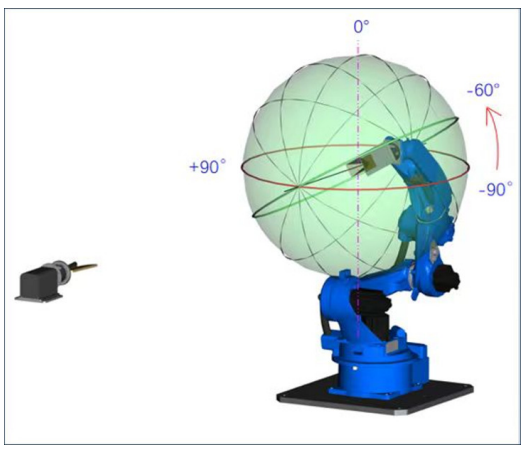
While a multi-probe spherical near-field measurement system offers significant advantages, it also suffers from the disadvantages of a fixed number of probes, limited bandwidth, and number of spherical modes. This limitation limits the

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

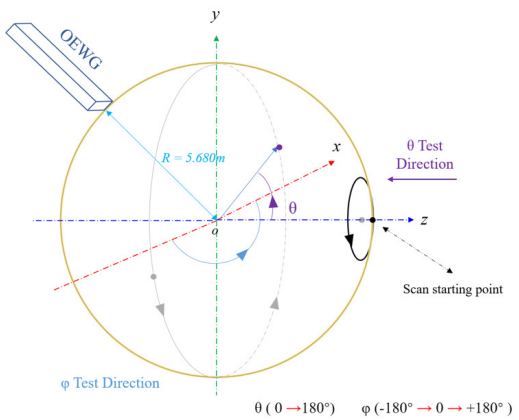
This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3523182

system's ability to adjust the number of probes to specific test requirements, creating challenges for both the user and the manufacturer. Determining the optimal number of probes for such a system can be a daunting task, as it involves considerations in terms of measurement accuracy, system complexity, and cost. In the current work, this issue is explored. The impact of moving the probe to different angular positions to replicate similar behavior as a multi-probe system is studied. Hence, manufacturers of multi-probe systems in this way can know how many probes they may need without sacrificing accuracy, while also keeping the system simple and cost-effective.

In this work, a novel approach to address this challenge is proposed by utilizing a robotic arm to replicate a multi-probe spherical near-field measurement system (see Fig. 1). By leveraging the flexibility and versatility of robotic automation, it is demonstrated how the number of probes can be systematically determined during the design phase of such a system. The proposed methodology aims to optimize the probe configuration to meet specific testing objectives while maximizing measurement accuracy and efficiency.



(a) Spherical near-field measurement system



(b) Spherical field scanning method

FIGURE 1: Spherical near-field measurement system

To evaluate the feasibility and performance of the proposed system, comprehensive comparisons between the replicated system's antenna measurements and simulations, as well as far-field measurements are conducted. These comparisons serve to validate the accuracy and reliability of the replicated system, providing valuable insights into its potential as a viable alternative for multi-probe spherical near-field testing.

II. TRADITIONAL SPHERICAL NEAR-FIELD ANTENNA MEASUREMENT SYSTEM

According to the number of probes used, the spherical near-field testing system can be divided into two types: single-probe near-field test system and multi-probe near-field test system. The traditional single-probe spherical near-field test system is composed as in Fig. 2a where the transmitter is mounted by a fixed bracket with a dual-polarization standard probe antenna, the antenna under test is erected on the azimuth turntable, and the motion controller controls the azimuth turntable and the pitch pivot respectively to complete sampling [8]. For the test, the antenna under test must be rotated in two dimensions, first, from $0°$ to $360°$ in azimuth and then from $0°$ to $180°$ in elevation. To determine the field on a spherical surface surrounding the antenna, several factors play a role and can affect the measurement accuracy. The large number of sampling points with a single probe, the long measurement time, the drift of the signal, the stability of the instrument, and the temperature change are a few of such factors.

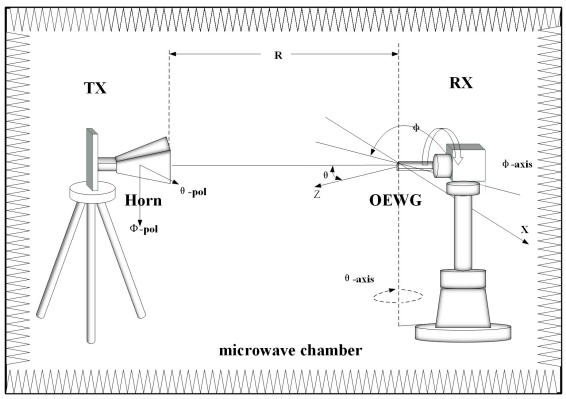
On the other hand, the multi-probe spherical near-field test system (see Fig. 2b) arranges multiple broadband and low-reflection probes in the pitch direction for electronic scanning so that the measured antenna only needs to rotate $180°$ in the azimuth direction, which increases the antenna measurement speed by multiple times due to the instantaneous completion of electronic scanning in one go. The principle of a multi-probe spherical near-field antenna testing system is to evenly arrange several orthogonal dual-polarized probes at specific angle intervals on a circular scanning bracket in a microwave anechoic chamber according to the sampling theorem. The probe antenna is connected to the rear electronic switching matrix switch through an RF cable, and the tested antenna is mounted on a turntable or pole, ensuring that its phase center is at the center of the scanning bracket. During testing, the antenna to be tested serves as the transmitting antenna, and the electronic switch sequentially switches the receiving probe to the vector network analyzer to complete data acquisition for each receiving channel. The motion controller of the system drives the rotary shaft through a stepper motor, which drives the antenna turntable to rotate the tested antenna and complete the collection of three-dimensional spherical near-field data [9]–[11].

III. FAR-FIELD ANTENNA MEASUREMENT SYSTEM

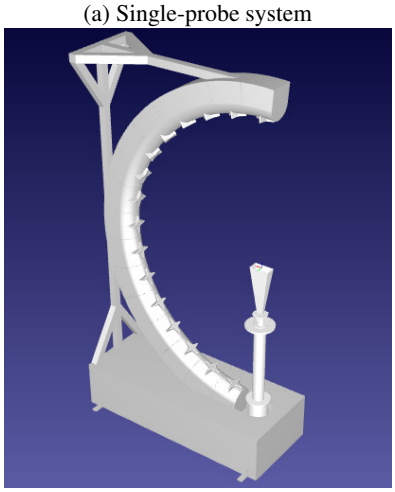
The modern far-field antenna measurement setup comprises several key components: a microwave anechoic chamber, a vector network analyzer, and a versatile robotic arm equipped

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3523182



(a) Single-probe system



(b) Multi-probe system

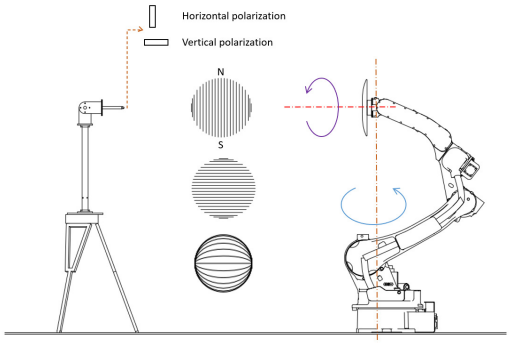
FIGURE 2: Spherical near-field test systems

with six degrees of freedom on each side, as illustrated in Fig. 3.

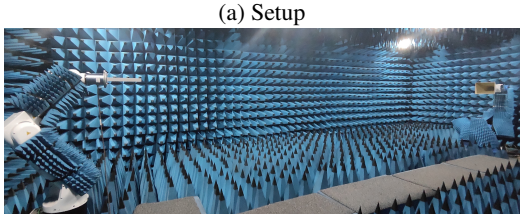
In such a system, the transmitter utilizes either an open waveguide probe or a standard horn antenna for signal transmission. During measurements, the robotic arm dynamically adjusts the polarization direction, aligning it horizontally or vertically as per the testing requirements. At the receiving end, the antenna under examination is securely mounted. Controlled by the robotic arm, the movement trajectory of the test antenna is precisely managed, ensuring optimal positioning throughout the measurement process. This synchronized motion allows for effective interaction between the transmitting antenna and the test antenna. Data collection occurs through various scanning methods, with parameters tailored to the specific testing scenario [12].

IV. MATHEMATICAL ANALYSIS

In this section, the mathematical framework underpinning spherical near-field measurements is explained. The positioning and number of probes are pivotal in capturing accurate near-field data, and this analysis highlights the key principles governing these factors.



(a) Setup



(b) Anechoic chamber used for this work

FIGURE 3: Typical modern far-field measurement systems

The theory behind spherical near-field measurements relies on the accurate representation of the electromagnetic fields over a spherical surface surrounding the antenna under test (AUT). The electric field components E_θ and E_ϕ on this spherical surface can be expanded using spherical harmonics [13]. The spherical harmonics provide a complete orthogonal basis set for representing functions defined on the surface of a sphere. These harmonics are mathematically expressed as [14]:

$$E_\theta(\theta,\phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} a_{nm} Y_{nm}(\theta,\phi) \tag{1}$$

$$E_\phi(\theta,\phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} b_{nm} Y_{nm}(\theta,\phi) \tag{2}$$

where a_{nm} and b_{nm} are the expansion coefficients, and $Y_{nm}(\theta,\phi)$ are the spherical harmonics, defined as:

$$Y_{nm}(\theta,\phi) = \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}} P_{nm}(\cos\theta) e^{im\phi} \tag{3}$$

where P_{nm} represents the associated Legendre functions, and θ and ϕ are the spherical coordinates.

The sampling theorem for spherical harmonics dictates that the number of sampling points required on a spherical surface to accurately reconstruct the field depends on the maximum spherical harmonic degree N_m [15]. Specifically, the relationship between spherical wave coefficient and maximum spherical harmonic is given by the following equation:

$$N_m = 2(N_n + 1) \tag{4}$$

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3523182

The number of samples N_θ (number of probes) required is defined as:

$$N_\theta = N_n + 3 \tag{5}$$

For a given frequency and electrical size of the AUT, N_m can be estimated as:

$$N_m \approx kR \tag{6}$$

where k is the wavenumber ($k = \frac{2\pi}{\lambda}$), and R is the radius of the spherical measurement surface.

In multi-probe systems, the angular spacing between probes $\Delta\theta$ and azimuth rotation spacing $\Delta\phi$ are crucial in determining the overall number of probes and their arrangement. For efficient and accurate sampling, the spacing $\Delta\theta$ and $\Delta\phi$ should satisfy [16]–[18] :

$$\Delta\theta \leq \frac{\pi}{N_m} \tag{7}$$

$$\Delta\phi \leq \frac{2\pi}{N_m} \tag{8}$$

Ensuring that the probe spacing adheres to this criterion helps minimize interpolation errors and enhances measurement precision.

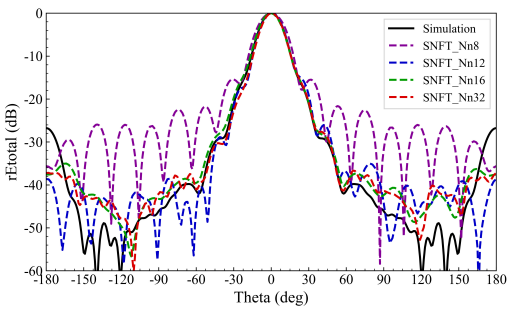
The impact of probe placement on the accuracy of spherical near-field measurements is significant. By systematically varying the angular positions of the probes and analyzing the resulting data, it is possible to determine the optimal probe configuration. The subsequent sections discuss the implementation of a robotic arm-based system for dynamically adjusting probe positions and present experimental results validating the proposed approach.

V. PROPOSED MEASUREMENT SYSTEM

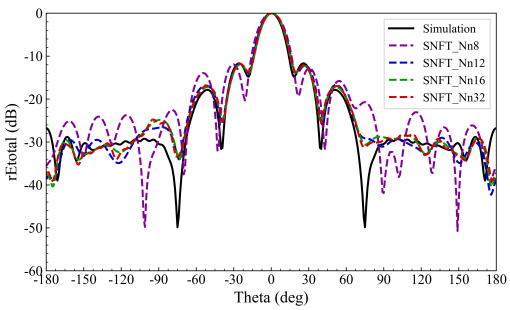
To check the accuracy of the proposed system, two sets of measurements were carried out, namely, the far-field antenna measurement and the spherical near-field antenna measurement. A standard open-ended waveguide probe (for frequency 3.94-5.99 GHz) was used as the transmitting antenna at the transmitter end, and the 3.94-5.99 GHz standard gain horn antenna (with a gain of 20dBi) was used as the antenna to be tested at the receiving end, as shown in Fig. 1, wherein, the aperture dimension of the horn (denoted by D) is 345mm, and the far-field testing distance requirement is $R > 2D^2/\lambda$ = 4.76m. The proposed testing method uses a far-field testing distance of 5.68m which is the distance from the open waveguide antenna at the transmitting end to the phase center of the horn antenna at the receiving end. In single-probe spherical near-field measurement technology, if the object being measured is relatively small, its measurement distance (R) may satisfy the conditions for far-field measurement. The transmission power for both far-field antenna measurement and spherical near-field antenna measurement is set to 16dBm.

Based on the relationship between spherical wave coefficients N_n and the maximum spherical harmonic number N_m detailed in Section IV, one can independently determine the number of sampling points and the scanning range for both azimuth and elevation angles. To find the optimum value of N_n, the N_n value is increased from 8 to 32 and radiation patterns for each value of N_n are tracked. N_n values of 8, 12, 16, and 32 are used for this study. For accuracy check, the far-field and spherical near-field antenna measurements are conducted at 4 frequency points of 4, 4.5, 5, and 5.5 GHz. The background noise of the vector network analyzer is about -110dBm, the S_{21} value after the alignment of the transmitting and receiving antenna is found to be -50dBm and the dynamic range of the system meets the measurement requirements.

After the far-field antenna measurements, the 2D radiation pattern (normally as sections of $\phi = 0°$ and $\phi = 90°$) or 3D radiation pattern can be directly obtained. The measurement results of different spherical wave coefficients (N_n = 8, 12, 16, or 32) at 4.5 GHz and 5.5 GHz are shown in Figs. 4 and 5. It is evident from the figures that the larger the value of N_n, the more points are collected which leads to the corresponding higher accuracy. However, it also leads to a more complex and expensive system. From the figure, it is found that for N_n=8 and N_n=12, the pattern has more sidelobes, however, the main lobe is consistent. It is also evident that with the increase of N_n, the measurement curve gradually approaches the curve corresponding to the simulation values.



(a) $\phi = 0°$



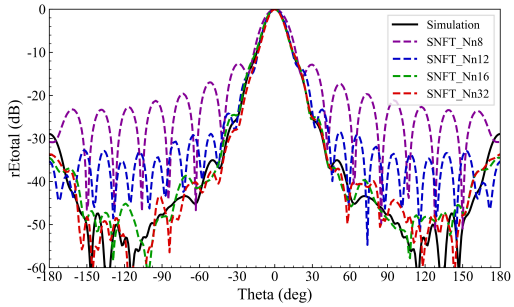
(b) $\phi = 90°$

FIGURE 4: Comparison of measurement results for different N_n orders at 4.5 GHz

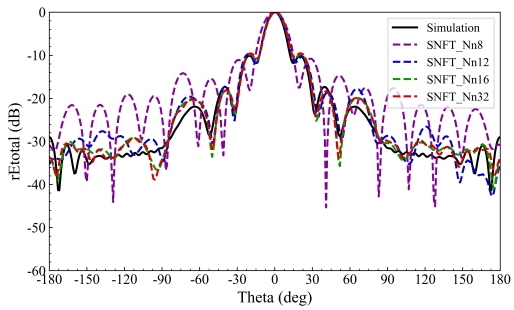
The standard gain horn of the 48SGAH20N model is

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3523182



(a) $\phi = 0°$



(b) $\phi = 90°$

FIGURE 5: Comparison of measurement results for different N_n orders at 5.5 GHz

modeled through FEKO simulation, as shown in Fig. 6. Its operating frequency range is from 3.94 GHz to 5.99 GHz, with the horn aperture facing the Z-axis. Taking the operating frequency of 4.5 GHz as an example, the near-field and far-field radiation patterns of the simulated antenna are shown in Fig. 7.

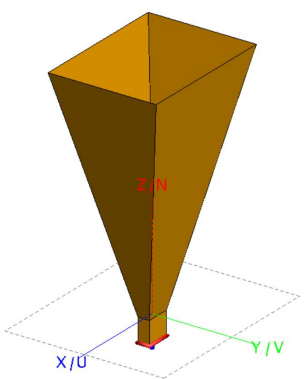
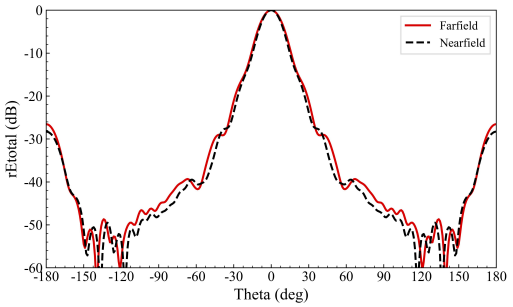
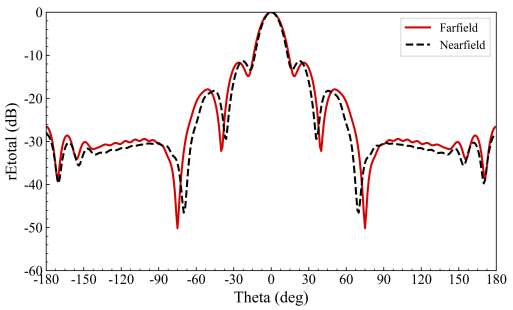


FIGURE 6: A FEKO simulation model of standard gain horn antenna

Taking $\Delta\theta = 12°$ and $\Delta\phi = 13°$, the far-field data obtained from FEKO software simulation is used as a reference. Figs. 8 and 9 respectively show the comparison of spherical near-field and far-field measurement and simulation data at the same frequency of 4.5GHz and different cutoff frequencies for spherical wave coefficients N_n = 16 and N_n = 32.

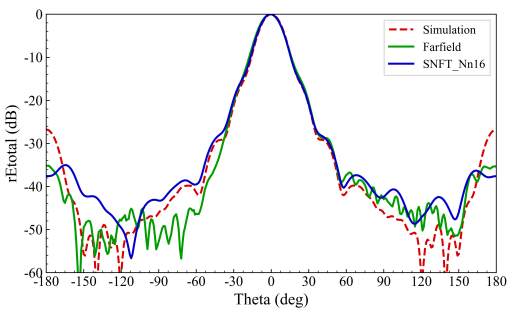


(a) $\phi = 0°$

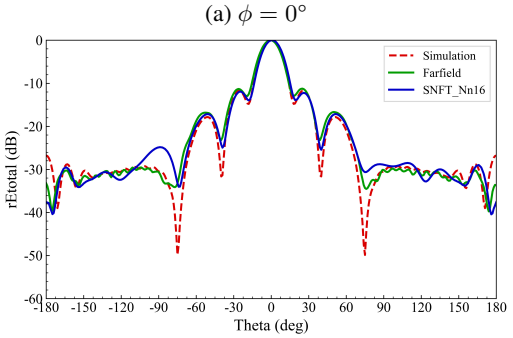


(b) $\phi = 90°$

FIGURE 7: Normalized near-field and far-field patterns of a standard gain horn antenna in FEKO software simulation for (a)$\phi = 0°$ (b)$\phi = 90°$



(a) $\phi = 0°$



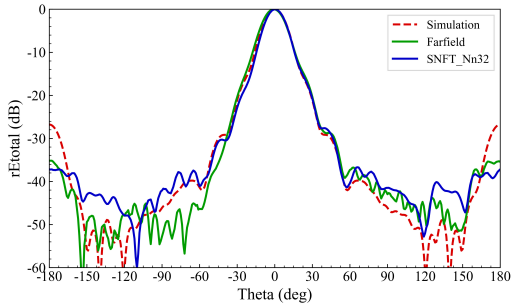
(b) $\phi = 90°$

FIGURE 8: Comparison of different measurement systems at 4.5 GHz

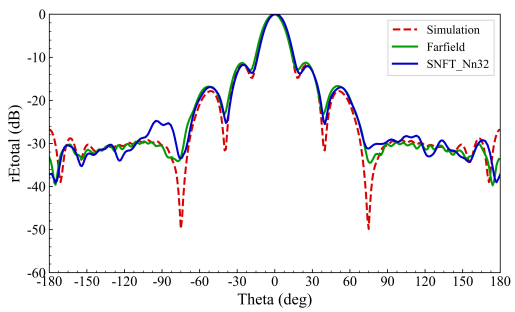
From the figures, it is found that when the spherical wave coefficient N_n is increased to 32, the spherical near-field

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3523182



(a) $\phi = 0°$



(b) $\phi = 90°$

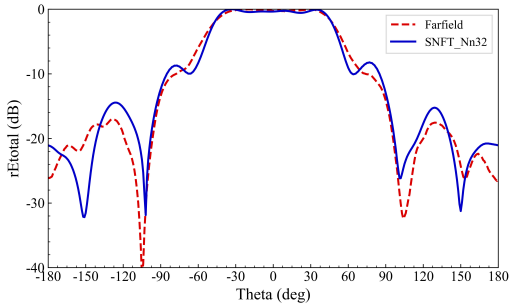
FIGURE 9: Comparison of different measurement systems at 5.5 GHz

test results are in good agreement with the far-field and the radiation pattern of the simulation data for both $\phi = 0°$ and $\phi = 90°$ which confirms the effectiveness of this measurement method. In this way, it is evident from the different figures that if the spherical wave coefficient $N_n = 32$ are used for such a measurement system, it can provide a simple, accurate, and cost-effective measurement system.

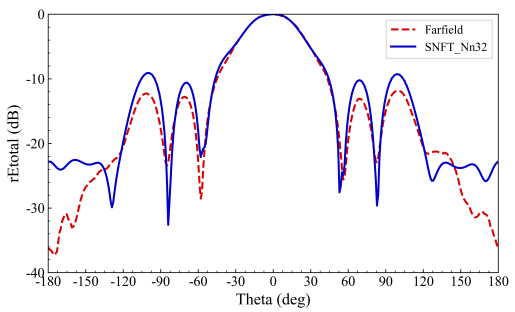
In addition, in order to more accurately verify the effectiveness of the proposed measurement method, a set of 0.4-8.5GHz dual-polarized broadband probe antennas were used as supplementary experiments, and far-field data were measured and compared. The comparison results were selected at the four frequency points of 3GHz, 4GHz, 5GHz, and 6GHz, as shown in Fig. 10. It is evident from the plots that the proposed method is accurate and robust.

These findings suggest that the robotic arm-based approach to replicating a multi-probe spherical near-field measurement system offers a versatile and cost-effective solution. By systematically varying probe positions, it is demonstrated how the optimal number of probes could be determined during the system design phase. This method provides significant flexibility in adapting to different measurement requirements, ultimately improving accuracy and efficiency.

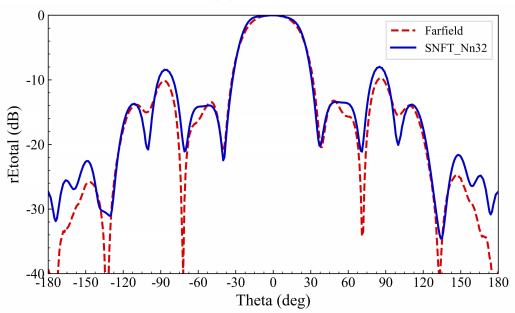
The experimental and simulation results indicate that the optimal number of probes is highly dependent on the specific measurement scenario, including the frequency, size of the AUT, and desired measurement accuracy. Future work could involve exploring the application of this approach to other types of near-field measurement systems and extending the analysis to higher frequencies and larger AUTs.



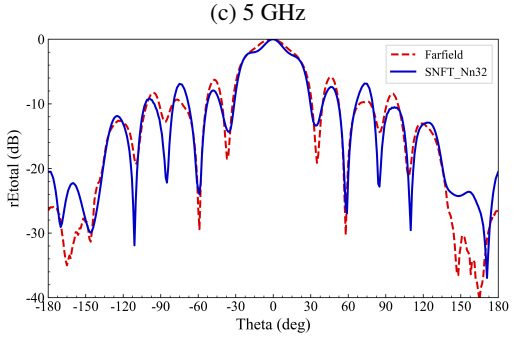
(a) 3 GHz



(b) 4 GHz



(c) 5 GHz



(d) 6 GHz

FIGURE 10: Comparison of SNFT and far-field measurements of dual polarization probe antennas at different frequencies

VI. CONCLUSION

A new method to estimate the number of probe counts for multi-probe spherical near-field measurement systems is proposed. This method helps to find the optimum number of

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3523182

probes needed for measurement accuracy, system complexity, and cost-effectiveness. In this work, it was found that when the spherical wave coefficient N_n = 32 (the number of probes is 35) would be sufficient for such a system. The experimental and simulation results validate the effectiveness and accuracy of this method, highlighting its potential as a versatile and cost-effective solution for antenna characterization. Future research will focus on further optimizing the probe configuration and extending the methodology to other types of near-field measurement systems.

VII. ACKNOWLEDGMENT

This work is funded by the Zhongshan Social Public Welfare Technology Research Project (2022B2004).

REFERENCES

[1] C. A. Balanis, *Modern Antenna Handbook*. Hoboken, NJ, USA: Wiley, 2008.
[2] J. E. Hansen, *Spherical Near-Field Antenna Measurements*. London, U.K.: Peter Peregrinus, 1988.
[3] J. Dong, "Computational electromagnetic modeling for accurate assessment of nearfield scanning systems," in *Proc. IEEE Int. Conf. Comput. Electromagn. (ICCEM)*, 2019, pp. 1–3.
[4] A. Yaghjian, "An overview of near-field antenna measurements," *IEEE Trans. Antennas Propag.*, vol. 34, no. 1, pp. 30–45, 1986.
[5] V. S. Kalashnikov et al., *Near-Field Antenna Measurements*. Springer, 2021.
[6] T. A. Laitinen, S. Pivnenko, and O. Breinbjerg, "Application of the iterative probe correction technique for a high-order probe in spherical near-field antenna measurements," *IEEE Antennas Propag. Mag.*, vol. 48, pp. 179–185, Aug 2006.
[7] J. Bolomey et al., "Rapid near-field antenna testing via arrays of modulated scattering probes," *IEEE Trans. Antennas Propag.*, vol. 36, no. 6, pp. 804–814, 1988.
[8] P. F. Wacker, "Non-planar near field measurements: Spherical scanning," *Final Report*, 1975.
[9] "IEEE recommended practice for near-field antenna measurements," *IEEE Std 1720-2012*, pp. 1–102, 2012.
[10] G. Li, H. Chen, and J. Chen, "SG128 multi-probe spherical near-field antenna test range and novel antenna measurements," in *Proc. 7th Int. Conf. Commun. Network. China*, 2012, pp. 689–693.
[11] S. Park et al., "High-speed antenna measurement system using multi-probe array technique for 5G applications," *Electronics*, vol. 11, no. 21, p. 3435, 2022.
[12] O. Cerny et al., "Optimization of far-field antenna range," *Radioengineering*, vol. 24, no. 4, 2015.
[13] R. Wittmann, "Probe correction in spherical near-field scanning, viewed as an ideal probe measuring an effective field," in *1984 Antennas and Propagation Society International Symposium*, vol. 22, 1984, pp. 674–677.
[14] T. B. Hansen, "Formulation of spherical near-field scanning for electromagnetic fields in the time domain," *IEEE Trans. Antennas Propag.*, vol. 45, no. 4, pp. 620–630, 1997.
[15] R. Cornelius, *Fast Spherical Near-Field Antenna Measurement Methods*. Aachen, Germany: Apprimus Verlag, 2019.
[16] F. H. Larsen, "Probe correction of spherical near-field measurements," *Electronics Letters*, vol. 14, no. 13, pp. 393–395, 1977.
[17] T. B. Hansen, "Spherical near-field scanning with higher-order probes," *IEEE Trans. Antennas Propag.*, vol. 59, no. 11, pp. 4049–4059, 2011.
[18] W. Kim et al., "Near-field to far-field rcs prediction on arbitrary scanning surfaces based on spherical wave expansion," *Sensors*, vol. 20, no. 24, p. 7199, 2020.

DONGMING WEI was born in Nanning City, Guangxi Province, China, in 1996. He received his B.S. degree in electronic information engineering from Shandong University of Science and Technology, Qingdao, China, in June 2019. He is currently working toward the master's degree in electronic and communication engineering at Changchun University of Science and Technology, Changchun, China. His current research interests focus on wireless communications and microwave technology.

JUNWEI DONG was born in Hebei, China, in 1983. He received B.S. degree in Measurement & Control and Instrumentation (Signal Processing) from Jilin University, Jilin, China, in 2006. He received his M.S. and Ph.D. in Electrical Engineering at Virginia Polytechnic Institute and State University (Virginia Tech), Virginia, USA, in 2009. He joined the Virginia Tech Antenna Group (VTAG) as a graduate student in 2006, where he conducted research on antennas, microwave lens designs and electromagnetic simulations. After graduation, he joined Microwave Engineering He joined the Virginia Tech Antenna Group (VTAG) as a graduate student in 2006, where he conducted research on antennas, microwave lens designs and electromagnetic simulations.

RAVI KUMAR ARYA is Director at Xiangshan Laboratory Wireless Group, Zhongshan Institute of Changchun University of Science and Technology (ZICUST), Zhongshan, Guangdong, China. He obtained his Ph.D. degree in Electrical Engineering from the Pennsylvania State University, USA. He received his master's degree in RF and Microwave Engineering from the Indian Institute of Technology, Kharagpur, India, and a bachelor's degree from Delhi Technological University, Delhi, India. Before joining ZICUST, he worked for several years in academia and industry. He has published over 80 technical papers in highly-ranked journals and refereed conference proceedings, including 7 book chapters. His research interests include Antenna Design, Computational Electromagnetics, Machine Learning Applications, and 3D modeling.

YIDA FAN was born in Meihekou City, Jilin Province, China, in 1999. He received his B.S. degree from Changchun University of Science and Technology, Jilin, China, in June 2021, He is currently working toward the doctor's degree at School of Optoelectronic Engineering, Changchun University of Science and Technology, Jilin, China. His current research interests focus on array antenna design and beamforming technology.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

Journal of Electronic Materials
https://doi.org/10.1007/s11664-024-11647-z

ORIGINAL RESEARCH ARTICLE

Electron Paramagnetic Resonance and Optical Absorption Studies of Cu^{2+} Ion-doped ZnAlBiB Glasses

K. Chandrika[1] · G. Snegitha[1] · M. Venkateswarlu[1] · M. V. V. K. Srinivas Prasad[1] · G. G. Dhar[2] · S. K. Mahamuda[1] · K. Swapna[1] · A. S. Rao[3]

Received: 18 February 2024 / Accepted: 26 November 2024
© The Minerals, Metals & Materials Society 2024

Abstract

Zinc alumino bismuth borate (ZnAlBiB) glasses enriched with various concentrations of Cu^{2+} ions were investigated using electron paramagnetic resonance (EPR) and optical absorption spectroscopy. The EPR spectra of all the glasses show a characteristic resonance, confirming the presence of Cu^{2+} ions within the glass matrices. The prepared glasses contained Cu^{2+} ions arranged in an octahedral coordination with tetragonal distortion, as evidenced by the spin-Hamiltonian parameters obtained from the EPR spectrum analysis. It was observed that the $g_{\perp}$ value increased as the concentration of copper ions increased. The optical absorption spectra displayed a single broad band characteristic of Cu^{2+} ions for all the prepared samples. The molecular orbital parameters were determined by comparing the EPR and optical absorption spectral data to understand the type of bonding present in these glasses. The optical bandgap energy (E_{opt}) and Urbach energy (E) were calculated for all ZnAlBiB glasses at the ultraviolet edges. The optical bandgap energy decreased as the concentration of copper ions increased, while the Urbach energy increased.

Keywords Electron paramagnetic resonance · spin-Hamiltonian · bandgap · Urbach energy · optical basicity

Introduction

Transition metal ion-doped glasses have sparked significant interest due to their memory and photoconductive properties. These materials hold potential for application in luminescent solar energy concentrating systems, optical fibers, solid-state lasers, and electrical memory switching devices.[1–4] An efficient method for obtaining comprehensive data on the local structural and dynamic features of paramagnetic centres in materials is electron paramagnetic resonance (EPR). EPR can be used to determine the site symmetry of surrounding transition metal ions in fabricated glasses, and it is a familiar technique for investigating how a paramagnetic material implanted in a material can absorb microwave radiation in the presence of a static magnetic field. This well-proven experimental method is used by researchers to gain insight into the site symmetry and glassy network around the transition metal ions doped into glasses through both theoretical and experimental approaches.[5,6] However, studies on the optical absorption of transition metal ions doped in glasses offer information on the energy levels related to the observed transitions. Correlating the optical absorption spectral data with the EPR data reveals the bonding nature of the glasses as well as the chemical and structural environment around the transition metal ions doped into the glasses.

Because of its special qualities and capacity to form glass, B_2O_3 is a useful ingredient in commercial glassware. It has good ion solubility, high transparency, chemical resistance, and thermal stability, and a low glass transition temperature. In addition, B_2O_3 is a good flux material due to its low fusion heat, small cation size, and excellent bond strength. Glasses with high B_2O_3 content have a lower expansion coefficient, higher viscosity, and better chemical resistance. The occurrence of the boron anomaly in glasses containing B_2O_3 has garnered significant scientific attention. These

✉ M. Venkateswarlu
mvrmphil@gmail.com

[1] Department of Physics, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh 522 302, India

[2] Department of Nanotechnology, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur, Andhra Pradesh 522510, India

[3] Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi 110 042, India

Published online: 23 December 2024

glasses provide good protection against infrared radiation, performing well as insulators and dielectrics. In light of their remarkable features, borate glasses are potential candidates for use in electro-optic switches, solid-state lasers, nonlinear optical parametric converters, and modulators.[7–12]

However, B_2O_3 glasses are of less importance due to their high phonon energy (~1300 cm^{-1}), which can affect their luminescence efficiency by promoting non-radiative transitions. Nevertheless, the higher phonon energy produced by borate glasses can be reduced by the addition of heavy metal oxides (HMO) such as Bi_2O_3 to achieve high quantum efficiency. The addition of Bi_2O_3 to borate glass not only decreases its phonon energy but also results in a high refractive index due to the generation of non-bridging oxygen in the glass network. It also increases the infrared transmission capacity of the borate glass by enhancing its nonlinear optical susceptibility and polarizability.[13,14]

Glasses containing Bi_2O_3 have gained technical importance in the field of glass ceramics, reflective windows, and mechanical and thermal sensors because of its low melting temperature (600–800°C). Bi_2O_3 alone cannot form glass because of its low field strength. However, Bi_2O_3 in the presence of B_2O_3 has good glass-forming ability. Bismuth borate glasses have attracted considerable research attention due to the nontoxic nature of bismuth, unlike other heavy elements such as Sb and Pb. Bi_2O_3 strongly absorbs gamma radiation and can act as a potential gamma ray shielding material due its high atomic number.[13,15]

On the other hand, ZnO is a direct-bandgap semiconductor which has a large bandgap (3.37 eV) and high excitation binding energy (60 meV). Glasses with the addition of ZnO have numerous applications including solar cells, optoelectronic devices, data storage materials, field-effect transistors, dielectric layers, transparent dielectrics, and barrier ribs in plasma display panels.[16,17] The addition of aluminates to borate glasses provides good optical properties, with applications in diverse products such as windshields, window curtains, and ophthalmic and space devices.[18,19] Transition metal ions are used as colouring agents in glasses. Copper in glasses can exist in two stable ionic states: divalent Cu^{2+} and monovalent Cu^{+}. Glasses containing monovalent Cu^{+} do not produce any colouring, while Cu^{2+}-doped glasses may appear bluish or green depending on the Cu^{2+} content and surrounding environment. EPR spectra of Cu^{2+} ions have been widely studied in a variety of glasses.[20,21]

In the present work, a zinc alumino bismuth borate (ZnAlBiB) glass system was developed and doped with various quantities of Cu^{2+} ions, taking into account the scientific implications of B_2O_3, Bi_2O_3, ZnO, and Al_2O_3 as mentioned above in various glasses for different applications. We carried out optical absorption and EPR studies to examine the structure of these new glasses. Our goal was to use EPR spectral analysis to examine the evolution of spin-Hamiltonian characteristics with varying Cu^{2+} concentrations. We aimed to understand the nature of bonding inside these glasses as well as the variations in site symmetry surrounding Cu^{2+} ions by establishing a correlation between EPR and optical absorption data.

Experimental

Glass Preparation

Copper-doped glasses were fabricated by the melt quenching technique. Table I presents the details of the glass sample composition used in the current study. The chemicals utilized to prepare the current glasses were of an analytical grade with 99.99% purity. An electrical balance with four-digit accuracy was used to precisely weigh the substances. Each batch of samples was carefully combined after being ground to a fine powder. The batches underwent 1 h of programmable electrical melting at 1200°C in a silica crucible. The melts were then immediately compressed with another plate after being quickly poured onto a heated polished brass plate. The produced glasses were annealed at 400°C for 1 h to remove thermal stresses. The resulting glasses have a bluish hue, are round in shape, are uniformly thick at 0.2 cm, and have a diameter of between 1 and 3 cm.

In order to obtain accurate optical measurements, the prepared glasses were polished. 1-Bromonaphthalene was used as a contact layer between the sample and the refractometer's prism in order to determine the glasses' refractive index ($\lambda = 5893$ Å). The Archimedes principle was used to calculate the densities of the glasses, with xylene serving as the immersion liquid. Other physical characteristics of the glasses, including density and refractive index, were also analysed using the conventional formulae[22] and are included in Table II. The powder x-ray diffraction (XRD) spectra of an un-doped glass were captured at ambient temperature under Cu-Kα radiation in the 2θ configuration using a Bruker D8 diffractometer. The XRD spectra contain broad spectra or absence of sharp peaks, confirming the amorphous nature of the prepared samples. In order to conduct the EPR studies, a JEOL FE 1X EPR spectrometer operating in the X-band at 9.205 GHz with a 100 kHz modulation frequency was

Table I Composition of glasses (mol.%) studied in the present work

Glass	B_2O_3	Bi_2O_3	Al_2O_3	ZnO	CuO
A	55	10	10	24.5	0.5
B	55	10	10	24	1
C	55	10	10	23.5	1.5
D	55	10	10	23	2
E	55	10	10	22.5	2.5

Table II Various physical properties of Cu^{2+} ion-doped ZnAlBiB glasses

Physical property	Glass A	Glass B	Glass C	Glass D	Glass E
Refractive index (n_d) at 589.3 nm	1.651	1.653	1.656	1.649	1.647
Density, d (g/cm^3)	4.0930	4.0966	4.1001	4.1037	4.1072
Average molecular weight $\overline{M}$ (g)	115.422	115.412	115.403	115.394	115.385
Cu^{3+}ion concentration N ($\times 10^{21}$ions/cm^3)	2.1358	4.2757	6.4196	8.5676	10.7196
Mean atomic volume (g/cm^3/atom)	6.6352	6.6289	6.6227	6.6163	6.6102
Optical dielectric constant (ε^{-1})	1.7258	1.7324	1.7423	1.7192	1.7126
Dielectric constant (ε)	2.7258	2.7324	2.7423	2.7192	2.7126
Reflection loss (R) (%)	6.0303	6.0585	6.1003	6.0024	5.9744
Molar refraction R_m(cm^{-3})	10.2982	10.3133	10.3410	10.2439	10.2094
Polaron radius r_p (A^0)	3.1293	2.4829	2.1683	1.9694	1.8277
Interionic distance r_i (A^0)	7.7650	6.1611	5.3805	4.8870	4.5353
Molecular electronic polarizability α ($\times 10^{-23}$ cm^3)	4.0818	2.0439	1.3662	1.0150	0.8093
Field strength F ($\times 10^{15}$cm^{-2})	2.0423	3.2441	4.2536	5.1562	5.9870

used. The magnetic field was scanned from 2200 to 4200 G at a scan rate of 250 G min^{-1}. For EPR measurements, a 100 mg sample of powdered glass was obtained in a quartz tube. The EPR spectrum of $CuSO_4{\cdot}5H_2O$ powder was also captured for use as a guide when determining the number of spins. The absorption spectrum of the Cu-doped glass was measured in the spectral range of 300–1100 nm using a PerkinElmer Lambda 950 double-beam spectrometer with a step width of 0.2 nm.

Results and Discussion

Electron Paramagnetic Resonance (EPR)

The absence of an EPR signal in an un-doped glass suggests that the raw materials employed in the glass fabrication process are devoid of contaminants originating from transition metal ions.

Figure 1 illustrates the EPR spectra of glasses containing Cu^{2+} dopants in the ZnAlBiB composition. The EPR spectra observed in this study exhibit similarities to the EPR spectra of copper-doped materials documented in previous studies.[19,23–25]

The spin of the electron (S) for the Cu^{2+} ion is determined to be 1/2, whereas the nuclear spin (I) is found to be 3/2 for both isotopes of copper, namely ^{63}Cu and ^{65}Cu. Therefore, it is anticipated that the EPR spectra of copper-doped materials will exhibit four ∥ and four ⊥ hyperfine (h_f) components. However, in the current study, we were only able to observe three somewhat faint ∥ components in the lower magnetic field zone. It appears that there was an overlap between the fourth ∥ component and the perpendicular component. Nevertheless, the ⊥ components that are found within the region of increased magnetic field exhibit clear resolution. The EPR spectra of Cu^{2+} ions in glasses can be effectively examined by employing an axial spin-Hamiltonian, as described by Abragam and Bleaney.[26]

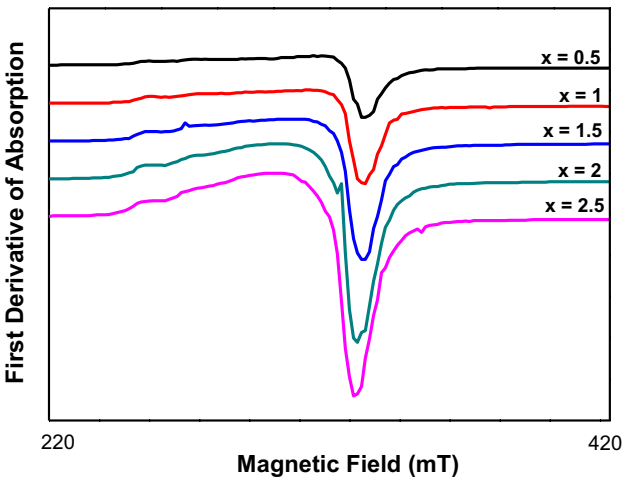


Fig. 1 EPR spectra of different mol.% of Cu^{2+} ion-doped ZnAlBiB glasses at room temperature.

$$\mathcal{H} = \beta\left[g_{\parallel}B_zS_z + g_{\perp}\left(B_xS_x + B_yS_y\right)\right] + A_{\parallel}S_zI_z + A_{\perp}\left(S_xI_x + S_yI_y\right),$$

where the variable z represents the symmetry axis; B_x, B_y, and B_z denote the static parameters of the magnetic field; S and I represent the operators for electron and nuclear spin, respectively; $g_{\parallel}$ and $g_{\perp}$ refer to the parallel and perpendicular parameters of the anisotropic g-tensor; $A_{\parallel}$ and $A_{\perp}$ represent the parallel and perpendicular parameters of the hyperfine tensor A; and β denotes the Bohr magneton. The nuclear quadrupole and nuclear Zeeman interaction factors are disregarded. The solution of the aforementioned spin-Hamiltonian yields the mathematical expression for the peak

Table III Observed optical band positions, optical bandgap, Urbach energy, optical basicity, number of spins, and susceptibility for Cu^{2+} ion-doped ZnAlBiB glasses

Glass sample	Band position (nm)	Optical bandgap (E_g) eV	Urbach energy (ΔE) eV	Optical basicity	Number of spins per kg N ($\times 10^{18}$)	Susceptibility χ ($\times 10^{-3}m^3/kg$)
A	777	2.5762	0.3346	0.4409	1.5879	5.0689
B	776	2.5695	0.3621	0.4408	2.0207	6.4751
C	776	2.5	0.3956	0.4407	2.9115	9.4892
D	777	2.4233	0.4211	0.4406	3.9383	12.8747
E	775	2.2772	0.4352	0.4405	4.4707	14.6151

position of the primary g and A tensors, as described by Bleaney et al.[27]

$$h\upsilon = g_{\parallel}\beta B + mA_{\parallel} + \left(15/4 - m^2\right)\frac{A_{\perp}^2}{2g_{\parallel}\beta B},$$

$$h\upsilon = g_{\perp}\beta B + mA_{\perp} + \left(15/4 - m^2\right)\frac{A_{\parallel}^2 + A_{\perp}^2}{4g_{\perp}\beta B},$$

where m is the magnetic quantum number of the copper nucleus ($m = \frac{3}{2}, \frac{1}{2}, \frac{-1}{2} and \frac{-3}{2}$) and is the microwave frequency, denoting parallel and perpendicular peaks, respectively. The solutions to the spin-Hamiltonian equation were used to determine the values of the spin-Hamiltonian parameters $A_{\parallel}$ (hyperfine splitting factor parallel to the principal axis), $g_{\parallel}$ (spectroscopic splitting factor parallel to the principal axis), and $g_{\perp}$ (spectroscopic splitting factor perpendicular to the principal axis), which are listed in Table IV. The evaluated $g_{\parallel}$ and $g_{\perp}$ values are the properties of Cu^{2+} ions in octahedral coordination elongated along the z-axis.[28] From the evaluated g value, it is observed that $g_{\parallel} > g_{\perp} > g_e$ (free electron $g_e = 2.0023$ value), indicating that the ground state for the paramagnetic ion is the $d_{x^2-y^2}$ orbital (i.e $^2B_{1g}$ state), the Cu^{2+} ion being located in distorted octahedral sites (D_{4h}) elongated along the z-axis.[6,19,29] This also suggests that Cu^{2+} ions are subjected to tetragonally elongated distortion.

The number of spins which are involved in the resonance can be evaluated with reference sample [$CuSO_4 \cdot 5H_2O$] using the following formula.[30]

$$N = \frac{A_x\left(\text{scan}_x\right)^2 G_{\text{std}}\left(B_m\right)_{\text{std}}\left(g_{\text{std}}\right)^2[s(s+1)]_{\text{std}}\left[P_{\text{std}}\right]^{1/2}}{A_{\text{std}}\left(\text{scan}_{\text{std}}\right)^2 G_x\left(B_m\right)_x\left(g_x\right)^2[s(s+1)]_x\left[P_x\right]^{1/2}}[\text{std}],$$

where A is the region under the absorption peak, scan_x is the magnetic field per unit length of the graph, G is the gain, B_m is the modulation field width, g is the g factor, S is the spin of the system in its ground state, and P is the microwave power. By performing a double integration of the first derivative of the absorption curve, we can obtain the area A under the curves. In the following equation, x and std stand for the measured and theoretically predicted amounts of Cu^{2+} ions in the glass sample and the standard, respectively. The number of spins in the standard deviation. The results of solving for N using the aforementioned equation are reported in Table III. The relationship between the concentration of Cu^{2+} ions and the number of spins in the ZnAlBiB glasses is depicted in Fig. 2.

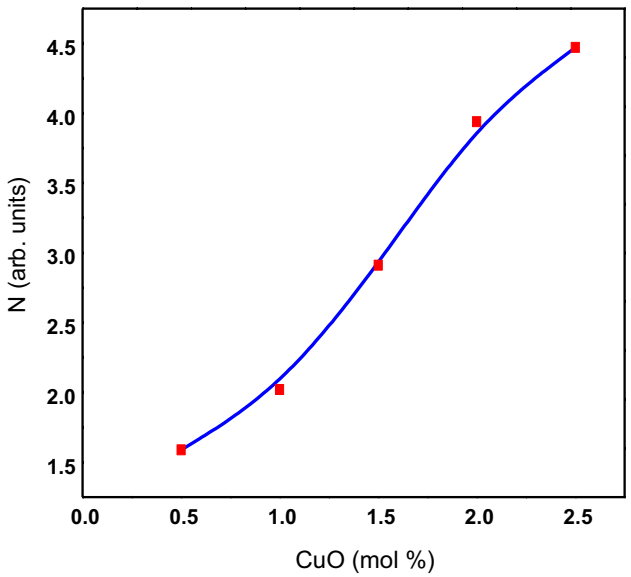


Fig. 2 Plot of the number of spins as a function of Cu^{2+} ion concentration in ZnAlBiB glasses.

Paramagnetic Susceptibility from EPR Data

The paramagnetic susceptibility can be calculated by

$$\chi = \frac{Ng^2\beta^2 J(J+1)}{3K_B T},$$

where N represents the number of spins/kg, J represents the total angular momentum, β is the Bohr magneton, K_B represents the Boltzmann constant, T is the absolute temperature, and g is the g-factor [$g = g_{\parallel}/3 + 2g_{\perp}/3$]. The parameters N

and g are derived from EPR measurements. Table III displays the resulting magnetic susceptibility values.

Optical Absorption Studies

Figure 3 presents the optical absorption spectra of the Cu^{2+} ion-doped ZnAlBiB glasses at room temperature, which are compared with the published scientific literature.[31–33] A single broad absorption band was identified in the near-infrared region for all the prepared glass samples. This band was attributed to the Cu^{2+} ions due to d–d transition, and it corresponded to the $^2B_{1g} \rightarrow {}^2B_{2g}$ transition. The observed positions of the absorption bands for the glasses in question are presented in Table III. The Cu^{2+} ions present in the glass do not demonstrate cubic symmetry, but rather display tetragonal distortion. The occurrence of tetragonal distortion results in the division of energy levels, leading to the formation of elongated or compressed structures. Additional calculations indicate that elongated structures exhibit greater energy favourability than compressed structures.[31,34,35] Typically, in glasses doped with copper, two wide absorption bands are found, which correspond to the $^2B_1g \rightarrow {}^2A_1g$ and $^2B_1g \rightarrow {}^2B_2g$ transitions. However, in the current study, we were only able to find a single absorption band ($^2B_1g \rightarrow {}^2A_1g$), while a band at a higher energy level ($^2B_1g \rightarrow {}^2B_2g$) was not detected.

Most earlier studies assigned this broad band to the $^2B_{1g} \rightarrow {}^2B_{2g}$ transition. The bonding coefficients of Cu^{2+} ions can be evaluated by relating the EPR data with the optical absorption spectral data. The bonding between the ligand and Cu^{2+} ion can be represented in terms of covalency components α^2 (in-plane σ bonding with Cu $d_{x^2-y^2}$ orbital), β^2 (= 1.00) (out-of-plane π bonding with d_{xz} and d_{yz} orbitals), and β_1^2(a measure of the in-plane π bonding with d_{xy} orbital). The in-plane σ bonding can be measured from the EPR data using the following relation.[28]

$$\alpha^2 = \frac{A_{\parallel}}{P} + \left(g_{\parallel} - g_e\right) + \frac{3}{7}\left(g_{\perp} - g_e\right) + 0.004.$$

Here, P (= 0.036 cm^{-1}) is the dipolar hyperfine coupling component. The in-plane π bonding (β_1^2) and out-of-plane π bonding (β^2) can be measured by correlating the optical absorption spectral data with EPR data using the following equations.[36]

$$g_{\parallel} = g_e\left[1 - \frac{4\lambda\alpha^2\beta_1^2}{E_1}\right],$$

$$g_{\perp} = g_e\left[1 - \frac{\lambda\alpha^2\beta^2}{E_2}\right].$$

Here, E_1 and E_2 are the energies relative to the $^2B_{1g} \rightarrow {}^2B_{2g}$ and $^2B_{1g} \rightarrow {}^2E_g$ transitions, respectively, and λ is the spin–orbit coupling constant (= −828 cm^{-1}). Since in the present work we have identified only one absorption band for all the glasses, the second band needed to calculate the bonding coefficients can be approximated using the following expression.[37]

$$E_2 = \frac{2k^2\lambda}{g_e - g_{\perp}}.$$

Here, $k2$ is the orbital reduction factor (= 0.77). A value of $\alpha 2$ equal to 1 indicates a purely ionic bond and a value equal to 0.5 indicates a purely covalent bond.[37] The values of β^2 and β_1^2 can be interpreted similarly. Here, $\alpha 2$ refers to the covalency of σ bonding between Cu (in-plane) and its ligands, and β_1^2 refers to the measure of covalency of π bonding between Cu (in-plane) and its ligands. For the prepared glasses, the values of α^2 and β_1^2 indicate a covalent nature for the in-plane σ bonding and an ionic nature for in-plane π bonding. The normalized covalency of Cu^{2+}-O in-plane bonding of σ and π symmetry[38] is computed using the parameters α^2 and β_1^2

$$\Gamma\sigma = \frac{200 * (1 - S)\left(1 - \alpha^2\right)}{1 - 2S}\%,$$

$$\Gamma\pi = 200\left(1 - \beta_1^2\right)\%.$$

Here, S = 0.076 represents the overlap integral for oxide ligands. The molecular orbital bonding coefficients and spin-Hamiltonian parameters are given in Table III.

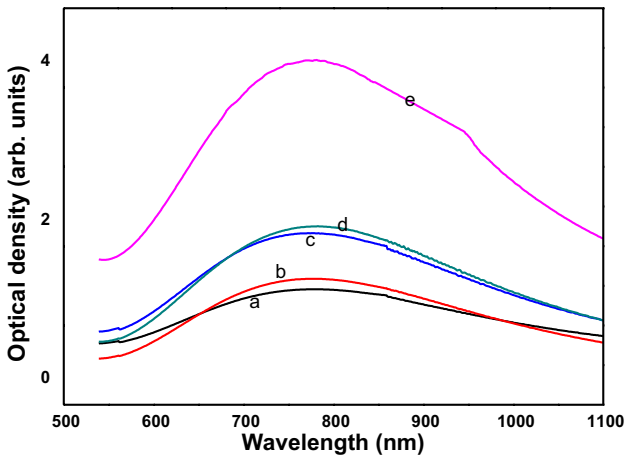


Fig. 3 Optical absorption spectra of Cu^{2+} ion-doped ZnAlBiB glasses at room temperature.

Optical Bandgap

The energy gap that exists between a semiconductor's conduction band and its valence band is comparable to the optical bandgap that exists in noncrystalline materials. The absorption edge study in the ultraviolet region is a useful method for better understanding the optical transitions and electronic band structure in glasses. This study is conducted in this region.

The indirect bandgap can be expressed as

$$E_{\mathrm{Opt}} = h\vartheta - \left(\frac{\alpha h\vartheta}{A}\right)^{1/2},$$

where α is a function of frequency ϑ, and A is a constant. This relation can be used to evaluate any oxide glasses.[39]

Figure 4 shows a plot between $h\vartheta$and$(\alpha h\vartheta)^{1/2}$ related to the UV absorption edges of Cu^{2+}-doped glasses. The optical bandgap (E_{opt}) can be determined by extending the linear section of the curve on the x-axis and identifying the point at which the extrapolation intersects the x-axis. An exponential increase in the absorption coefficient is the hallmark of the absorption edge in amorphous materials, especially at lower absorption coefficient values, $\alpha(\vartheta)$, with respect to photon energy, as described by the equation[32]

$$\alpha(\vartheta) = \mathrm{constexp}\left(\frac{h\vartheta}{\Delta E}\right).$$

The Urbach energy, denoted as ΔE, serves as a measure of the width of the band tails associated with localized states. The Urbach plots depicted in Fig. 5 are acquired by plotting the natural logarithm of $\alpha(\vartheta)$ versus the product of h and ϑ. The Urbach energy (ΔE) can be found by calculating the reciprocals of the slopes of the linear areas of the curves. Table IV provides the optical bandgap and Urbach energy values for the glasses.

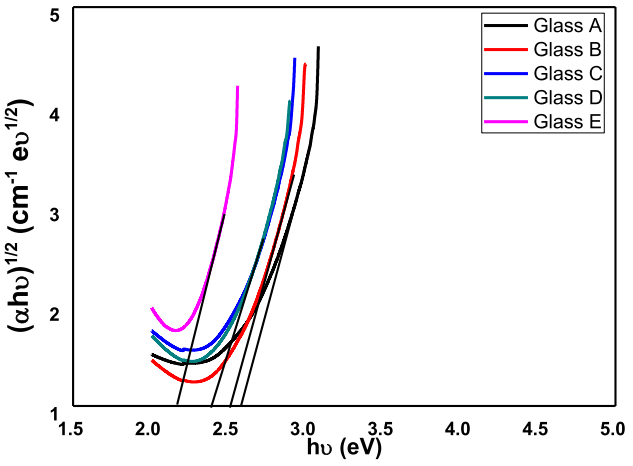


Fig. 4 Plot between $(\alpha h\nu)^{1/2}$ and hυ for Cu^{2+}-doped ZnAlBiB glasses.

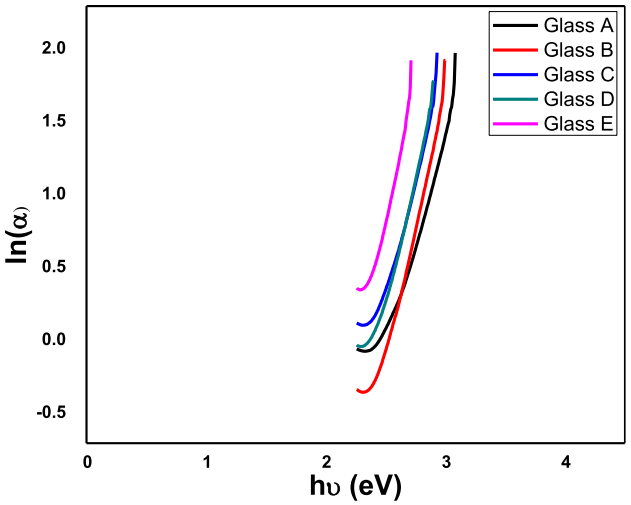


Fig. 5 Plot between ln(α) and hυ for Cu^{2+}-doped ZnAlBiB glasses.

The values found in this study align with those reported by previous researchers in the field of borate glasses.[23,29,40]

The data shown in Table IV demonstrate a clear correlation between the concentration of Cu^{2+} ions and the corresponding changes in the optical bandgap values and Urbach energy values.[40,41] Specifically, an increase in Cu^{2+} ion concentration leads to a decrease in the optical bandgap values and an increase in Urbach energy values. The presence of Cu^{2+} ions in the glass matrix can be linked to the introduction of additional defect states, such as colour centres.[34,41]

Optical Basicity

The optical basicity of the prepared glass refers to the capacity of oxygen atoms to donate an electron to a probe ion.[42] The optical basicity in glass can be determined by considering the glass composition and the moderating characteristics of different cations present within the glass composition, and then employing a mathematical function to determine its theoretical values.[43]

$$\Lambda_{\mathrm{th}} = \sum_{i=1}^{n} \frac{Z_i r_i}{2\gamma_i}.$$

Here, n represents the total number of cations present and Zi denotes the oxidation number of the ith cation, while γ_i represents the ratio of the number of ith cations to the number of oxides. Additionally, γ_i represents the basicity moderating parameter, which is determined by the equation[44]

Table IV The spin-Hamiltonian parameters and molecular orbital bonding coefficients for Cu^{2+} -doped ZnAlBiB glasses

Glass sample	$g_{\parallel}$	$g_{\perp}$	$A_{\parallel}(10^{-4})$ cm^{-1}	$A_{\perp}(10^{-4})$ cm^{-1}	α^2	β_1^2	$\Gamma\sigma$	$\Gamma\pi$
A	2.3174	2.0472	113	26	0.6890	0.9120	67.77	17.59
B	2.3029	2.0577	123	37	0.7072	0.8548	63.80	29.02
C	2.3287	2.0720	115	48	0.7180	0.9068	61.44	18.62
D	2.3470	2.0811	100	53	0.6980	0.9727	65.80	5.44
E	2.3656	2.0792	97	55	0.7068	0.9992	63.89	0.15

$$\gamma_i = 1.36\left[x_i - 0.26\right].$$

Here, x_i is the Pauling electronegativity[45] of the cation. The measured values of the optical basicity are tabulated in Table IV.

Conclusions

The following conclusions were derived based on the examination of copper-doped ZnAlBiB glasses using EPR and optical absorption measurements.

(1) The Cu^{2+} ions present in ZnAlBiB glasses are located in octahedral positions, exhibiting a tetragonal distortion. The octahedral sites reveals elongation across the *z*-axis, and the ground state of Cu^{2+} ions is characterized by the $d_(x^2-y^2)$ orbital.
(2) The glass samples exhibit a single broad absorption band in the near-infrared region, which can be attributed to the $^2B_1g \rightarrow {}^2B_2g$ transition.
(3) EPR and optical absorption data suggest that in-plane σ bonding is covalent, while in-plane π bonding is ionic.
(4) The Cu^{2+} ion content in ZnAlBiB glasses affects the optical bandgap energies and Urbach energies computed from ultraviolet absorption edges. The defect concentration increases with increasing copper ion concentration.

Acknowledgments The authors, Dr. K. Swapna (File Number: ECR/2015/000335), Dr. Sk. Mahamuda (File Number: ECR/2016/000376) and Prof. A. S. Rao (EMR/2016/007766) are thankful to the Department of Science and Technology (DST), Govt. of India, New Delhi, for the award of major research projects to them under DST-SERB. The authors also thank the Department of Science and Technology (DST), Govt. of India, for the award of DST-FIST Level-1 (SR/FST/PS-1/2018/35) scheme to the Department of Physics, KLEF.

Conflict of interest On behalf of all authors, the corresponding author declares that the authors have no financial interests or personal relationships with other people or organizations that could have appeared to inappropriately influence the work reported in this paper.

References

1. C.-C. Ding and B. Duan, Investigations on the role of doped 3d transition ions in $Li_2B_4O_7$ glasses: local structures, spectroscopic properties and optical basicities. *Mater. Res. Bull.* 156, 111995 (2022).
2. A. Bhogi, R.V. Kumar, and P. Kistaiah, Effect of alkaline earths on spectroscopic and structural properties of Cu^{2+} ions-doped lithium borate glasses. *J. Non-Cryst. Solids* 426, 47 (2015).
3. V. Aruna, S. Yusub, M. Venkateswarlu, A.R. Babu, and K. Anitha, Efficacy of copper ions on lithium ion conductivity, electron hopping, optical Bandgap, metallization criterion and morphology of Li_2O-B_2O_3-P_2O_5 glasses. *J. Non-Cryst. Solids* 536, 120015 (2020).
4. E.V. Zamyatina, V.V. Karzanov, and O.A. Zamyatin, Optical properties of the zinc-tellurite glasses doped with Cr^{3+} ions. *J. Non-Cryst. Solids* 528, 119759 (2020). https://doi.org/10.1016/j.jnoncrysol.2019.119759.
5. J. Lin, B. Zhang, Y. Yang, P. Xiao, and Y. Zhao, Theoretical studies of the local structure and the epr parameters for tetragonal Cu^{2+} center in mixed alkali cadmium phosphate glasses. *Rev. Mex. Física* 62, 325 (2016).
6. H.-M. Zhang, J.-H. Duan, W.-B. Xiao, and X. Wan, Theoretical studies of the local structure of Cu^{2+} center in aluminium lead borate glasses by their EPR and optical spectra. *J. Non-Cryst. Solids* 425, 173 (2015).
7. R.P. Sreekanth Chakradhar, G. Sivaramaiah, J.L. Rao, and N.O. Gopal, EPR and optical investigations of manganese ions in alkali lead tetraborate glasses. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* 62, 761 (2005). https://doi.org/10.1016/j.saa.2005.02.045.
8. H.A. ElBatal, A.M. Abdelghany, N.A. Ghoneim, and F.H. ElBatal, Effect of 3d-transition metal doping on the shielding behavior of barium borate glasses: a spectroscopic study. *Spectrochim. Acta. A. Mol. Biomol Spectrosc.* 133, 534 (2014). https://doi.org/10.1016/j.saa.2014.06.044.
9. A. Montreeuppathum, P. Kidkhunthod, S. Rujirawat, R. Yimnirun, S. Pinitsoontorn, and S. Maensiri, Effect of borate glass network to electrochemical properties: manganese-doped lithium borate glasses. *Radiat. Phys. Chem.* 170, 108677 (2020). https://doi.org/10.1016/j.radphyschem.2019.108677.
10. G. Sangeetha, A. Hameed, G. Ramadevudu, and M. Shareefuddin, Physical, EPR and optical studies of SrO-ZnO-$Na_2B_4O_7$-CuO glass system, in: AIP Conference Proceedings, AIP Publishing, (2023).
11. Y. Wu and Z. Chen, Investigation of the local structure of Cu^{2+} ions doped in alkali lead tetraborate glasses by their electron paramagnetic resonance and optical spectra. *Radiat. Eff. Defects Solids* 169, 485 (2014).
12. M.-Q. Kuang, S.-Y. Wu, G.-L. Li, and X.-F. Hu, Investigations on the EPR parameters and tetragonal distortions for Cu^{2+} in alkali lead tetraborate glasses. *Mol. Phys.* 113, 698 (2015).
13. B. Srinivas, A. Hameed, G. Ramadevudu, M.N. Chary, and M. Shareefuddin, Experimental and theoretical electron

paramagnetic resonance and optical studies of Cu^{2+} spin probe in BaO-TeO_2-Bi_2O_3-B_2O_3 glass system. *Optik* 156, 289 (2018).
14. S.P. Singh, R.P.S. Chakradhar, J.L. Rao, and B. Karmakar, Electron paramagnetic resonance, optical absorption and photoluminescence properties of Cu^{2+} ions in ZnO–Bi_2O_3–B_2O_3 glasses. *J. Magn. Magn. Mater.* 346, 21 (2013).
15. O. Sallam, A. Madbouly, N. Elalaily, and F. Ezz-Eldin, Physical properties and radiation shielding parameters of bismuth borate glasses doped transition metals. *J. Alloys Compd.* 843, 156056 (2020).
16. H.M. Gomaa, I. Yahia, B. Makram, A.H. El-Dosokey, and S.M. Elkatlawy, Optical and structural studies of some zinc calcium borate glasses for optoelectronic device applications. *J. Mater. Sci. Mater. Electron.* 32, 9392 (2021).
17. J. Ahlawat, S. Pawaria, N. Deopa, S. Dahiya, R. Punia, and A. Maan, Structural and optical characterization of IR transparent sodium-modified zinc borate oxide glasses. *Appl. Phys. A* 128, 923 (2022).
18. M.R. Ahmed and M. Shareefuddin, EPR, optical, physical and structural studies of strontium alumino-borate glasses containing Cu^{2+} ions. *SN Appl. Sci.* 1, 1 (2019).
19. B.R.V. Rao, K. Venkata Rao, S. Ravi Kumar, C. Rajyalakshmi, G.V.L. Kanth, G. RamaSwamy, and S. Cole, XRD, EPR and FT-IR Studies of Zinc aluminium lithium borate (ZALB) glases doped with Cu^{2+} Ions. *Phys. Chem. Res.* 10, 455 (2022).
20. B. Srinivas, A. Hameed, G. Ramadevudu, M.N. Chary, and M. Shareefuddin, Evaluation of EPR parameters for compressed and elongated local structures of VO^{2+} and Cu^{2+} spin probes in BaO-TeO_2-B_2O_3 glasses. *J. Phys. Chem. Solids* 129, 22 (2019).
21. D.F. Franco, E.E. Carvajal, J.P. Donoso, M.A. Silva, A.C. Sant'Ana, H. Fares, C.J. Magon, and M. Nalin, Structural and EPR studies of Cu^{2+} ions in $NaPO_3$–Sb_2O_3–CuO glasses. *J. Non-Cryst. Solids* 503, 169 (2019).
22. A.S. Rao, B.R.V. Rao, M. Prasad, J.S. Kumar, M. Jayasimhadri, J.L. Rao, and R.P.S. Chakradhar, Spectroscopic and optical properties of Nd^{3+} doped fluorine containing alkali and alkaline earth zinc-aluminophosphate optical glasses. *Phys. B Condens. Matter* 404, 3717 (2009).
23. G. Calas, N. Capobianco, and L. Galoisy, Spectroscopic properties of alkali borate glasses containing Cu^{2+}. *J. Non-Cryst. Solids* 591, 121711 (2022).
24. J.S. Krishna, D.S. Raju, V.V. Krishna, and C.L. Raju, Preparation and characterization of copper (Cu^{2+}) ions doped zincalumino-borophosphate (ZnAlBP) glasses, in Journal Physics Conference Seies, 1913:012013 (2021). https://doi.org/10.1088/1742-6596/1913/1/012013
25. B. Srinivas, A. Bhogi, P. Naresh, A. Hameed, M.N. Chary, and M. Shareefuddin, Effect of SrO and TeO_2 on the physical and spectral properties of strontium tellurite boro-titanate glasses doped with Cu^{2+} ions. *J. Non-Cryst. Solids* 575, 121218 (2022). https://doi.org/10.1016/j.jnoncrysol.2021.121218.
26. A. Abragam and B. Bleaney, Electron Paramagnetic Resonance of transition metal ions, Clarendon Oxford 175 (1970).
27. B. Bleaney, K.D. Bowers, and D.J.E. Ingram, Paramagnetic resonance in diluted copper salts I. Hyperfine structure in diluted copper Tutton salts, in Proceeding Royal Society London Series Mathematical Physics Science 228:147–157 (1955)
28. I. Ardelean, M. Peteanu, R. Ciceo-Lucacel, and I. Bratu, Structural investigation of CuO containing strontium-borate glasses by means of EPR and IR spectrometry. *J. Mater. Sci. Mater. Electron.* 11, 11 (2000).
29. X.-H. Xu and M.-Q. Kuang, Study on the local structure and EPR spectroscopy of copper ions in lead borate glass. *Optik* 152, 16 (2018).
30. J.A. Weil and J.R. Bolton, *Electron Paramagnetic Resonance: Elementary Theory and Practical Applications* (Hoboken: John Wiley & Sons, 2007).
31. B. Narsimha, K.C. Sekhar, M. Shareefuddin, M.N. Chary, and G. Ramadevudu, Influence of In_2O_3 on optical, EPR, and structural properties of barium borate glasses containing Cu^{2+} spin probe. *Braz. J. Phys.* 54, 40 (2024). https://doi.org/10.1007/s13538-023-01413-7.
32. G. Giridhar and M. Rangacharyulu, Electron paramagnetic resonance and optical absorption studies on copper ions in mixed alkali cadmium phosphate glasses. *J. Mater. Sci. Technol.* (2009). https://doi.org/10.1088/1757-899X/2/1/012058.
33. S. Prakash Singh, R.P.S. Chakradhar, J.L. Rao, and B. Karmakar, Electron paramagnetic resonance, optical absorption and photoluminescence properties of Cu^{2+} ions in ZnO–Bi_2O_3–B_2O_3 glasses. *J. Magn. Magn. Mater.* 346, 21 (2013). https://doi.org/10.1016/j.jmmm.2013.07.007.
34. Y.D. Li, B.J. Chen, H.M. Zhang, and C.D. Feng, Theoretical study of the local distortions and the EPR parameters for VO^{2+} and Cu^{2+} centers in oxyfluoroborate glasses. *J. Lumin.* 234, 117973 (2021). https://doi.org/10.1016/j.jlumin.2021.117973.
35. J. Su, S.-Y. Wu, X.-Y. Yu, T.-H. Guo, X.-X. Yang, and Q.-S. Zhu, An investigation of local structures and spin Hamiltonian parameters for Cu^{2+} in xBaO–(30–x)TeO_2–10TiO_2–58B_2O_3 glasses. *Russ. J. Phys. Chem. A* 98, 1860 (2024). https://doi.org/10.1134/S0036024424700948.
36. D. Kivelson and R. Neiman, ESR studies on the bonding in copper complexes. *J. Chem. Phys.* 35, 149 (1961).
37. A.K. Lonkowski, The structure of sodium aluminosilicate glass. *Phys. Chem. Glas.* 24, 166 (1983).
38. H. Kawazoe, H. Kokumai, H. Hosono, and T. Kanazawa, Microheterogeneities in R2O+ RO+ B_2O_3 glasses. *J. Non-Cryst. Solids* 38, 717 (1980).
39. M.A. Hassan and C.A. Hogarth, A study of the structural, electrical and optical properties of copper tellurium oxide glasses. *J. Mater. Sci.* 23, 2500 (1988).
40. S. Vedavyas, K.C. Sekhar, S. Ahammed, G. Ramadevudu, M.N. Chary, and M. Shareefuddin, Physical and structural studies of cadmium lead boro-tellurite glasses doped with Cu^{2+} ions. *J. Mater. Sci. Mater. Electron.* 32, 3083 (2021).
41. R. Stefan, E. Culea, and P. Pascuta, The effect of copper ions addition on structural and optical properties of zinc borate glasses. *J. Non-Cryst. Solids* 358, 839 (2012).
42. A.J. Easteal and A.T. Morcom, Variation of optical basicity with probe ion, for Cs_2O+ B_2O_3 and Li_2O+ B_2O_3 glasses. *J. Non-Cryst. Solids* 34, 29 (1979).
43. J.A. Duffy, A common optical basicity scale for oxide and fluoride glasses. *J. Non-Cryst. Solids* 109, 35 (1989). https://doi.org/10.1016/0022-3093(89)90438-9.
44. J.A. Duffy and M.D. Ingram, An interpretation of glass chemistry in terms of the optical basicity concept. *J. Non-Cryst. Solids* 21, 373 (1976).
45. C.A. Coulson, Pauling's Chemical Bond, (1961).

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

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

ISSN 0026-8933, Molecular Biology, 2024, Vol. 58, No. 6, pp. 1293–1311. © Pleiades Publishing, Ltd., 2024.

STRUCTURAL-FUNCTIONAL ANALYSIS OF BIOPOLYMERS AND THEIR COMPLEXES

UDC 577.21

Elucidating the Effects of Aromatic Mutations on the RNA Binding Efficacy of CELF2 Protein

N. Kumari[a, b], N. Sarovar Bhavesh[b, *], and Y. Hasija[a, *]

[a] *Department of Biotechnology, Delhi Technological University, Shahabad Daulatpur, Delhi, 110042 India*

[b] *Transcription Regulation, International Centre for Genetic Engineering and Biotechnology (ICGEB), Aruna Asaf Ali Marg, Vasant Kunj, New Delhi, 110067 India*

**e-mail: yashahasija@dtu.ac.in*

Received June 10, 2024; revised August 27, 2024; accepted September 4, 2024

Abstract—RNA Recognition Motifs (RRMs) are the most common type of protein domain found in eukaryotic RNA-binding proteins (RBPs). They regulate gene expression, RNA processing, and other important cellular processes. These functions of RRMs are dependent on their RNA recognition mode. However, the mechanism of binding between the RRM domain(s) of proteins and RNA is still an active area of investigation. Herein, we report the results of studies relating to the effect of aromatic amino acid mutations on the efficacy of RRM-RNA binding. Briefly, the RRM domain of CELF 2 protein and its six mutant complexes (F135Y, F152Y, F155Y, F176Y, F179Y, F209Y) were examined using Molecular Dynamics (MD). The mutations were introduced at aromatic amino acid phenylalanine (Phe) positions on the RRM2 Domain, substituting each Phe with tyrosine (Tyr). The binding interactions between the mutant and wild-type protein complexes were compared and studied. Comprehensive computational analysis was conducted to determine the effect of mutation on the stability of the protein structure and its interaction with RNA by doing RMSD/RMSF, BSA, PCA and MM/PBSA analysis. It was observed that induced mutations result in structure instability and flexibility, but the binding becomes stronger when Tyr replaced Phe in most mutation complexes. The mutation F135Y located on RRM showed a similar trend to the wild type in terms of binding interaction and stability. These results corelate with the finding that aromatic amino acids play a crucial role in RRM-RNA binding. This study offers insight into the RNA-binding mechanism and encourages further studies into the role of RRM primary sequence towards RNA binding.

Keywords: RNA-recognition motif, CELF-2, protein-RNA binding, mutational analysis, molecular dynamics simulation

DOI: 10.1134/S0026893324060062

INTRODUCTION

Proteins and RNA are the one of the primary building blocks of life and play a crucial role in the functioning of the human body. Proteins are an essential component of living cells and are involved in many biological processes, including enzyme catalysis, immune response, replication, metabolism, and the transport of materials and information inside and outside the cell [1]. RNA executes various necessary and significant tasks within the cell, including storing and transmitting genetic information and forming some parts of spliceosomes [2] and ribosomes [3] among others. The interaction between RNA and proteins is a recurring aspect among many of these biologically relevant functions. These interactions are highly dynamic and diverse; they involve binding proteins to specific RNA sequences. Over the course of its lifespan, a single RNA transcript can establish connections with multiple proteins [4].

The binding between proteins and RNA sequences regulates various cellular and molecular functions, such as translation, transcription, and stability, as well as the gene expression regulation [5]. Therefore, understanding the principles and mechanisms involved in protein-RNA recognition is a subject of great scientific interest and can have wide implications in the field of medicine. When RNA binds to proteins, ribonucleoprotein (RNP) complexes are formed. The proteins forming these complexes contain distinct RNA-binding domains (RBD), providing structural or sequence specificity for target RNA. There are various evolutionarily conserved known RBDs, like k-homology domain (KH), zinc finger domain (ZF), heterogenous ribonucleoprotein domain (hnRNP), and RNA recognition motif (RRM) [6, 7]. Of all these domains in eukaryotes, RRMs are the most widespread, abundant, and prominent protein domains [8], which bind to stem-looped single-stranded or short RNAs with high specificity [9]. The first-ever structure of a RRM was determined in 1990 by Nagai et al. [10]. Its structure is comprised of

approximately 80–90 amino acids with a conserved conformation of β1-α1-β2-β3-α2-β4, where four β-sheets are packed or placed against two α-helices [11]. Generally, RRM uses β-strands surface when binding to RNA [12], with certain exposed amino acids essential for the binding interactions. These exposed amino acids are generally one of the three aromatic amino acids—Phe, Tyr, and Trp [13]. Conserved residues in RRM are characterized by two stretches of sequences, RNP 1: Lys/Arg-Gly-Phe/Tyr-Gly/Ala-Phe/Tyr-Val/Ile/Leu-X-Phe/Tyr; and RNP2: Ile/Val/Leu-Phe/Tyr-Ile/Val/Leu-X-Asn-Leu. Any alteration or mutations introduced in these two sequences might change the binding ability of the RRM.

CUGBP Elav-Like Family Member 2 protein (also known as CELF2) is one of the RNA binding proteins that has a significant role in the regulation of several post-transcriptional events, such as splicing, editing, stability, and translation. It comprises of two N-terminals and one C-terminal RRMs designated as, RRM1, RRM2, and RRM3, respectively, with a segmental difference of about 160–230 amino acid residues between RRM2 and RRM3. These RRMs can identify and bind to multiple RNA targets to effectively perform their functions [14]. This protein has also been implicated in cancer and myotonic dystrophy (DM) pathophysiology (https://www.uniprot.org/uniprotkb/O95319/entry). DM is the most prevalent neuromuscular disease inherited by adults, with a prevalence of 1 in 8000 individuals worldwide. Myotonia and increasing muscle weakness are the primary characteristics of DM, a multisystemic, autosomal dominant disease [15]. CELF2 plays a crucial role in DM due to its interaction with abnormal RNA sequences, particularly the CUG repeat expansions characteristic of the disease. The binding of CELF2 to these repeats disrupts its normal function in RNA splicing, leading to altered splicing patterns. This mis-splicing affects gene expression and protein function, contributing to muscle weakness and other symptoms of DM. Additionally, dysregulation in the expression levels of CELF2 further exacerbates these effects [16]. Other than DM, CELF2 is also related to diseases like acute myeloid leukemia (AML), loss of CELF2 accelerates myeloid cell transformation and AML development. Researchers have discovered that CELF2 deficiency leads to the stabilization of FAT10 mRNA. Stabilized FAT10 mRNA, in turn, activates the AKT phosphorylation and mTORC1 signaling pathway, contributing to AML development [17]. CELF2 expression is downregulated in ovarian cancer. Altered CELF2 expression affects the proliferation, migration, and invasion of ovarian cancer cells.

One of the PDB structures of CELF2 is the 4LMZ crystal structure available in the PDB database containing both RRM1 and RRM2 domain. Canonically, an RRM binds to two to three nucleotides through conserved amino acids (mostly aromatic) on β-sheet of RNPs [14]. In CELF2, conserved Phe on the β-strand of RRMs stacks with the uracil residues, and the protein backbone and sidechains form specific hydrogen bonds to recognize all four nucleotides in the RNA molecule. Aromatic residues on RRM3 stack with four out of six nucleotides, and specific recognition is facilitated by a dense network of hydrogen bonds between RNA molecules and the protein backbone and polar sidechains [14]. Mutations that impact the interaction of RNA with the RRM domain can modify the binding affinity. For example, mutations that disrupt key interactions between the RRM domain and RNA can reduce the binding affinity of the protein. Conversely, mutations that strengthen these interactions can also increase the binding affinity. By exploring the effects of certain mutations on RNA binding, it is possible to obtain a deeper understanding of the mechanisms underlying protein-RNA interactions.

This work is aimed to study the effects of mutations induced in the RRM2 domain of CELF2 on the CELF2-RNA interactions via computational methods. RRM2 is positioned in the central region of CELF2 and has a higher affinity for RNA compared to RRM1 and RRM3. Mutations in RRM2 can disrupt CELF2's RNA-binding ability, affecting downstream processes such as alternative splicing and mRNA stability. In this work, the conserved aromatic amino acid Phe on the RRM2 domain of CELF2 was replaced with another aromatic amino acid residue Tyr. These mutated complexes were then simulated, and the effects were studied. The rationale behind replacement of Phe with Tyr was to study the effects of change in a single side chain while maintaining the aromatic nature of the replaced amino acid. Additionally, the substitution of Phe with Tyr in RBPs is significant because they both have aromatic phenyl rings, only difference is the alcohol group at the end of the ring in Tyr which allows for hydrogen bonding [18]. These consensus sequences are known to be crucial for RNA binding. Both Phe and Tyr participate in stacking interactions with RNA nucleotide bases [19]. Therefore, the effects of this substitution on RNA binding can be investigated to gain insights into the specific interactions that occur at the interface between the RRM domain and RNA. The dynamical properties of RRM and RNA were explored using MM/PBSA, PCA (principal component analysis), Buried Surface Area (BSA) and pyContact. This study provides new insights into the molecular mechanisms of RRM-RNA interactions and highlights the importance of amino acid mutations on binding affinities and overall RRM-protein interactions. This study paves the way for studying the RRM-RNA interactions at the submolecular and atomic levels, which can bring new insights into the pathogenesis of several diseases including DM and cancer.

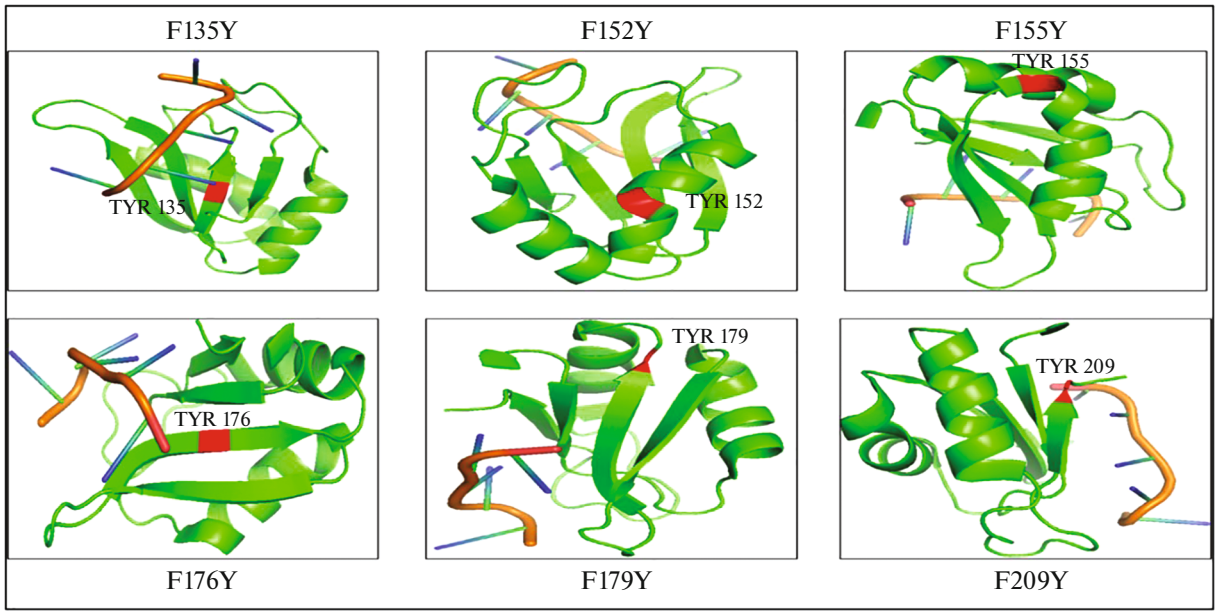


Fig. 1. Placements of induced mutations of Tyr in CELF 2 Protein RRM2 Domain.

METHODS

Molecular dynamics simulation of 4LMZ and its mutants. MD simulations were performed to understand the dynamics and implications of structural changes in the mutant protein with respect to the original structure and its binding with RNA. The structure of wild-type complex was obtained from RCSB PDB as the "crystal structure of RRM 1 + 2 domain of human ETR-3 protein" (PDB ID: 4LMZ, Uniprot ID-O95319) (https://www.rcsb.org/structure/4LMZ) with a resolution of 2.78 Å. Based on the complex structure of 4LMZ, six mutant systems were designed by introducing mutations into the structure of wild type using the PyMOL software (http://www.pymol.org/pymol). The CHARMM-GUI interface [20] (https://www.charmm-gui.org/) was used to prepare structures of a wild-type protein-RNA complex and six mutants—F135Y, F152Y, F155Y, F176Y, F179Y, F209Y (Fig. 1). For the experiment, the solution builder option found in the input generator menu of CHARMM-GUI was used to build the systems for every simulation. The rectangular box of the TIP3 water model was used to solvate the systems, and every structure was positioned at minimum distance of 10 Å from the box edge to the nearest solute atom.

The coordinates and topology for each system were created using the all-atom additive force field CHARMM36 [21]. Na^+ and Cl^- ions were added in every system for the neutralization process. The GROMACS version 20.3 was used to simulate every system [22]. To reduce the steric repulsion in every system, steepest descent energy minimizations of 50000 steps were run. To stabilize the temperature of the system, 500 picoseconds (ps) of NVT (number of particles (N), system volume (V), and temperature (T)) equilibration were operated, and for the stabilization of pressure system, 500 ps of short position restraint NPT was operated. Thereafter, the system was subjected to 200 ns of constraints-free simulations. GROMACS utilities were used to analyze root mean square deviation (RMSD) and root mean square fluctuations (RMSF) trajectories files. Values of RMSF and RMSD were measured using the GROMACS command line "gmxrmsf" and "gmxrmsd" over the carbon (Cα) atoms of the protein backbone of the RRM domain.

Binding free energy calculation. Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) is a tool used to approximate the free energy of binding between a protein and a ligand complex using samples from MD trajectory [23]. This tool is reliable and efficient as it can integrate the factor of structural configuration in calculating binding energy, especially when MD simulation is integrated [24]. The MM-PB/GBSA analysis was performed using the gmx_MMPBSA tool [25] to analyze the MD trajectories of the frames acquired from protein-RNA complex simulation. The binding energy was calculated over the last 50 ns (150–200 ns) from 200 ns of MD trajectory simulation. A total of 501 frames were composed within these 50 ns [26]. The net binding free energy of the protein-RNA complex (ΔG_{bind}) is given by the following equation:

$$\Delta G_{bind} = \Delta G_{Complex(protein + RNA)} - (\Delta G_{protein} + \Delta G_{RNA}).$$

The ΔG values for the protein and RNA were calculated from the sum of the energy of different interactions, including Van der Waals and electrostatic forces,

represented as E_{vdW} and E_{el}, respectively, along with contributions of the solvation-free energy by both non-polar (E_{surf}/E_{nPolar}) and polar ($E_{PB/GB}$) interactions [26].

Principal component analysis. Principal component analysis (PCA), a technique for reducing dimensions, is a popular approach in MD simulations for elucidating the overall functional movements of biomolecules that occur slowly over time [27]. It is a technique to determine prevalent molecular motion patterns from an MD simulations trajectory [28] and helps in the collective motion and conformational exploration of protein atoms. It has been regarded as an essential tool for studying the protein dynamics with various ligands. The present study employed PCA to measure the variations in the trajectories of proteins-RNA complexes, both native and mutant. PCA incorporates calculating and diagonalizing the covariance matrix for Cα atoms. These steps aid in identifying the patterns of accumulated structural variations in proteins [29]. To identify variations in wild-type and mutant proteins based on their simulation trajectories of the last 50 ns, a package of Python called "mdtraj" was used [30] to run on MD trajectory files. Cluster groups were formed by extracting eigenvectors or principal components (PCs) from wild-type and mutant protein RNA complex structures.

Contact analysis. Hydrogen, hydrophobic, ionic bonds, and intermolecular stacking interactions predominate in the interfacial contacts between RRM domains and RNA [31]. These interactions play a significant role in molecule identification and binding. The PDB structures for all the MD trajectories after the simulations were extracted. A web tool, "Protein—Ligand interaction Profiler" (PLIP), was used for visualizing the RRM-RNA interactions [32]. It is highly configurable and user-friendly software with GUI. It can facilitate rapid determination of noncovalent interactions from PDB coordinates files of protein and ligand. PyContact was used for additional analysis of residues surrounding native and mutated protein systems for the last 50 ns of the simulation trajectory [33]. It is a GUI-based, user-friendly application for analyzing biomolecular interactions in MD trajectory. It enables the detection of significant noncovalent interactions by providing rapid data analysis and visual data representation [33].

Solvent Accessible Surface Area (SASA) quantifies the area of biomolecules, like proteins, accessible to solvent molecules, measured in square nanometers (nm^2). It indicates the ability of solvent atoms to contact the surface atoms of a molecular structure [34]. SASA is crucial for analyzing protein structure and function, particularly in understanding how hydrophobic residue placement affects protein folding and stability [35]. The lesser exposure of these hydrophobic residues to the solvent, the greater the stability of the protein. Area of protein-RNA surface were measured using the GROMACS command line "gmx-sasa" over the atoms of the protein backbone of the RRM domain [36]. Using SASA, the buried surface area was calculated.

The Buried Surface Area (BSA) gives an indication of how much surface area is hidden from solvent upon complex formation. This is an important metric for understanding the stability and affinity of molecular interactions. First, the SASA for the entire complex (e.g., a protein-ligand complex) will be calculated. This gives you the surface area accessible to the solvent when the molecules are interacting. Next, SASA for each component of the complex in isolation will be calculated. This might involve creating separate trajectory files where the individual molecules are alone, or simply using a frame where the molecules are sufficiently separated. The BSA is then calculated by subtracting the sum of the SASAs of the individual components from the SASA of the complex.

$$\mathrm{BSA} = (\mathrm{SASA\ complex}) - \left(\mathrm{SASA}_{\mathrm{Protein}} + \mathrm{SASA}_{\mathrm{RNA\ (ligand)}}\right).$$

RESULTS

Molecular Dynamics Stability

MD trajectories of 4LMZ Protein-RNA complex and its mutants (F135Y, F152Y, F155Y, F176Y, F179Y, F209Y) were used for the experiment. 4LMZ (uniport sequence 36–211) contains two RRM domains, RRM1 (40–123) and RRM2 (132–212); simulations were conducted on RRM1 and RRM2 domain as 4LMZ comprise both the domains. To assess the stability of simulated protein-RNA complexes, the RMSD trajectory of the protein Cα atoms backbone was measured and compared to the original structure in nanometers along the time trajectory. This was achieved by comparing the RMSD values obtained from all frames generated during the simulation to the native complex (Fig. 2). All the complexes were stable after 50 ns except for the F152Y (blue) and F209Y (pink) complexes, depicting maximum deviation from the wild-type complex. It is also seen that F152Y is placed on the alpha helix chain, and F209Y is on the beta-sheet chain of the complex structure. F135Y (green) and F155Y (yellow) complexes showed similar behavior as the wild type and appeared stable throughout the simulation. F176Y (purple) and F179Y (red) complexes showed deviation at the beginning. However, they stabilized after that, and their trajectories overlapped with that of the wild type. These results indicate that while the complexes' structure stability was initially affected by the induced mutations, they gained stability after some time.

To gain more insights into the deviation of the complexes' structures, RMSD values of the RNA were also calculated. Figure 3 shows the RMSD of the RNA in different complexes. The trajectories of RNA in

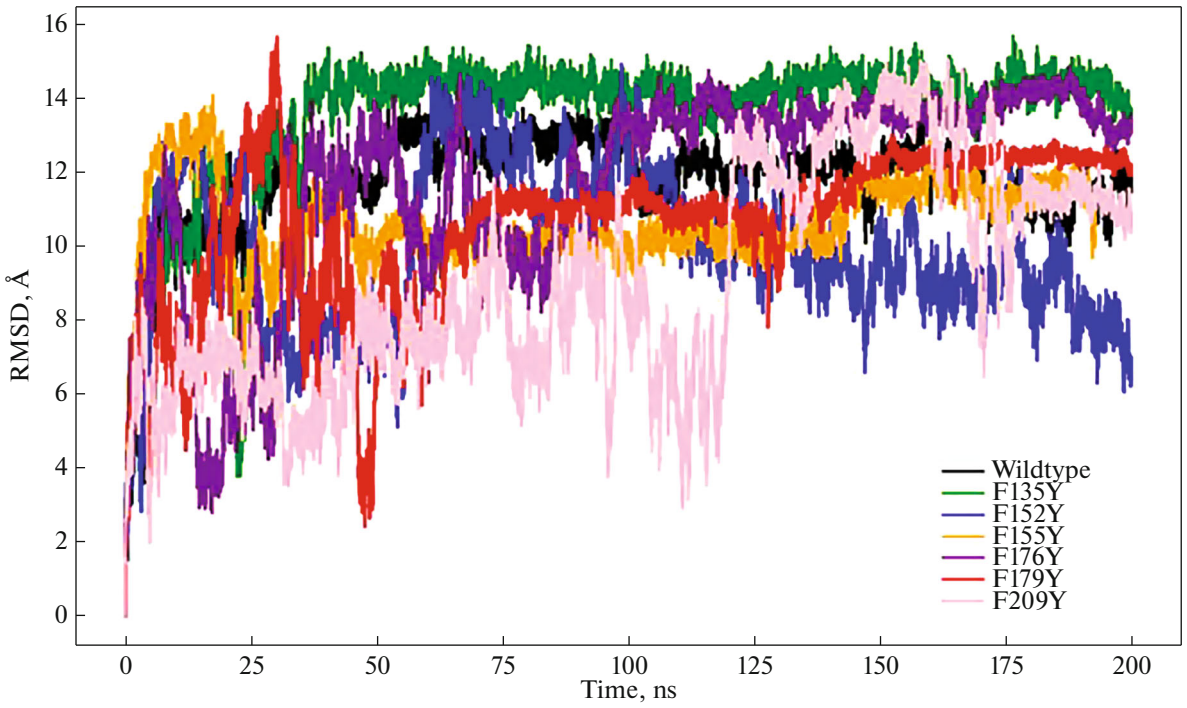


Fig. 2. RMSD plot for 200 ns of C-alpha backbone of all the mutated and wild type protein-RNA complexes for 200 ns, showing deviation between the mutant's complexes compared to wild type. F135Y (green), F152Y (blue), F155Y (yellow), F176Y (purple), F179Y (red), F209Y (pink), Wildtype (black).

F135Y (green), F155Y (yellow), and F176Y (purple) complexes are almost similar, while deviations were observed in F152Y and F209Y complexes.

Further, RMSF values were determined for both native and mutant protein-RNA complexes to examine the effect of a mutation on the dynamic behavior of protein residues. The values were estimated for the simulation trajectories' last 50 ns (150–200 ns). The RMSF patterns of all 7 complexes, including native, have been shown in Fig. 4. The highest peak of fluctuation was observed for F209Y (pink) with a rise to 0.9 nm, whereas the flexibility of the wild protein-RNA complex (black) structure was in the range of 0.3–0.4 nm. The mutant F209Y (pink) complex had the highest flexibility compared to other protein-RNA complexes.

The region of residues number from 70–85 and 120–135 showed the maximum flexibility during the simulation in every complex. F135Y (green) mutant complex showed a similar trend of residue flexibility as the wild type over the 50 ns trajectory. F152Y (blue) complex depicted wild-type-like rigidity till residue number 165 however, the residues after 165 to 211 showed a higher flexibility, which included the induced mutation. The degree of fluctuations displayed by the F155Y (yellow) complex were higher, followed F209Y. Further, it displayed an overlap of trajectories with F152Y and wild type at some residues but showed higher flexibility for the end residues. F176Y (purple) and F179Y (red) displayed overlapping trajectories throughout indicating fluctuations at similar residue positions, thus leading to similar flexibility values. These results indicate that the introduction of mutations at the Phe positions led to increased flexibility in F209Y, F155Y, and F152Y (in decreasing order) mutated complexes, while higher rigidity was observed in F179Y and F176Y complex compared to wild type. It should be noted that the F135Y complex depicted skewed overlapping with the wild-type trajectory. The RMSF graph showed that all mutated complexes observed increased flexibility than wild type except F179Y and F176Y. However, that does not seem to damage the structure, which is evident from the RMSD plots.

MM-PBSA/MM-GBSA Analysis

MM-PB/GBSA analysis is a dynamic and credible tool for free energy calculation. It has been used extensively to calculate protein-RNA binding interactions [23]. MM-PB(GB)SA on MD simulations trajectory allows exploring different conformations and capturing the binding energy changes associated with ligand binding. By decomposing the overall binding free energy into several components, such as Van der Waals, electrostatic, and solvation energies, MM-PB(GB)SA provides a detailed understanding of the factors that contribute to protein-ligand binding affin-

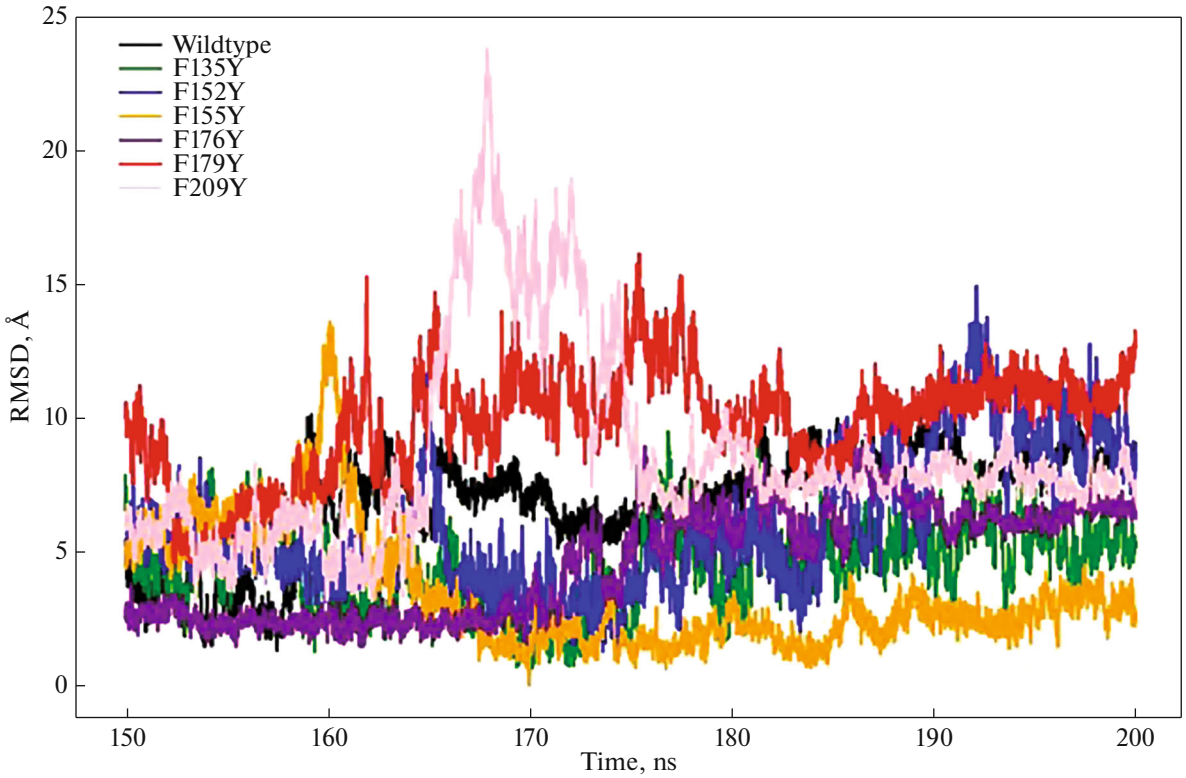


Fig. 3. RMSD plot of RNA in the protein-RNA complex for last 50 ns (150–200 ns) F135Y (green), F152Y (blue), F155Y (yellow), F176Y (purple), F179Y (red), F209Y (pink), Wildtype (black).

ity. It can also be used to identify key residues and interactions that are important for ligand binding [26].

The contributions of different bonds in total binding energies of all 7 complex systems were calculated using Gmx_mmpbsa and are shown in Table 1. The calculation of ΔG was performed on 500 frames, corresponding to the MD trajectory of all the complexes from 150–200 ns. The mutant complex F176Y (−63.87 ± 7.07 kcal/mol) has the lowest total free binding energy, followed by F135Y (−56.71 ± 3.80 kcal/mol), suggesting a strong binding interaction between protein and RNA. While the wild type of complex (−23.96 ± 7.89 kcal/mol) has the highest total binding energy, followed by F179Y mutant complex (−33.05 ± 8.22 kcal/mol), indicating relatively weaker binding ability. Electrostatic interactions (ΔE_{ele}), Van der Waals forces (ΔE_{vdW}), and solvation energy (ΔE_{SURF}) contributed to the binding nature identification of the protein-RNA complexes. F176Y had the highest E_{surf}, E_{vdW}, and E_{ele} values compared to the wild type. Van der Waals forces were found to be the highest contributor to the total GBSA value for F176Y. All the mutated complexes displayed a lower binding energy than the wild type, implying that they possess stronger binding interactions at the interface between RRM domain and RNA.

MM-PBSA results indicated that instead of F176Y, F135Y (−37.91 ± 7.19 kcal/mol) had the lowest binding energy. F155Y (−11.77 ± 5.82 kcal/mol) had the highest PBSA values followed by the wild type (−15.38 ± 6.51 kcal/mol). F152Y (−33.90 ± 8.90 kcal/mol), F176Y (−33.26 ± 8.94 kcal/mol) and F209Y (−33.76 ± 8.00 kcal/mol) displayed similar PBSA values. The ΔE_{vdW} values were found to be similar in the MMGB/PBSA analyses, however E_{ele} was found to have a more prominent contribution in the PBSA analysis. The mutated complexes follow the trend of having lower total binding energies and stronger binding interactions.

The differences in the results obtained by MMPBSA and MMGBSA can be understood by the differences in the way they operate. MMPBSA uses the Poisson–Boltzmann equation to compute the electrostatic contribution to free energy. In contrast, MMGBSA uses the Generalized Born approximation, an approximate and faster treatment of the Poisson–Boltzmann equation. MMGBSA provides faster results, however MMPBSA often yields more precise results even though it is slow [23].

Buried Surface Area

The Buried Surface Area (BSA) is crucial because it directly correlates with the stability and specificity of

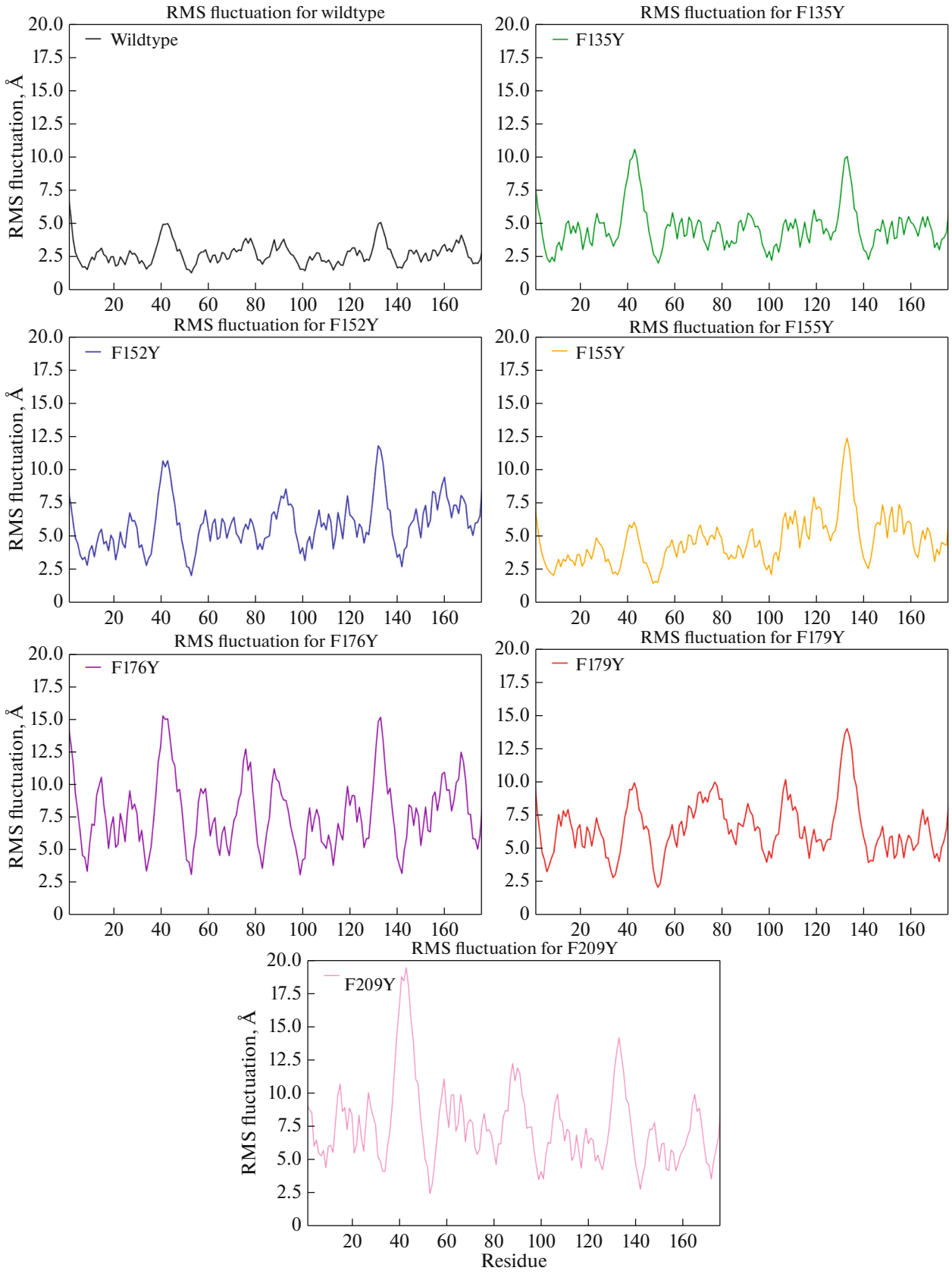


Fig. 4. RMSF plot of all the mutated and wild-type protein-RNA complexes showing variability in mutated complexes compared to wild type. F135Y (green), F152Y (blue), F155Y (yellow), F176Y (purple), F179Y (red), F209Y (pink), Wildtype (black).

Table 1. MM-GBSA & MM-PBSA analysis results for free binding energy of wild type and mutant proteins-RNA complexes

MM-GBSA analysis					
Energy, kcal/mol	ΔEEL	ΔVDW	ΔESURF	ΔEGB	ΔGBSA
Wild type	−23.98 ± 4.07	−21.99 ± 7.09	−2.97 ± 1.00	24.98 ± 3.75	**−23.96 ± 7.89**
F135Y	−34.73 ± 3.38	−49.67 ± 4.00	−7.20 ± 0.31	34.90 ± 2.87	−56.71 ± 3.80
F152Y	−38.68 ± 4.74	−41.24 ± 5.96	−6.47 ± 0.80	38.25 ± 4.01	−48.15 ± 6.54
F155Y	−26.43 ± 3.32	−39.32 ± 4.26	−5.12 ± 0.53	28.55 ± 2.90	−42.33 ± 4.61
F176Y	−43.39 ± 4.45	−56.13 ± 6.31	−7.80 ± 0.95	43.45 ± 3.65	**−63.87 ± 7.07**
F179Y	−31.88 ± 4.78	−29.15 ± 7.59	−4.34 ± 1.01	32.32 ± 4.20	−33.05 ± 8.22
F209Y	−44.17 ± 3.77	−33.37 ± 5.50	−5.21 ± 0.70	42.80 ± 3.22	−39.94 ± 5.85
M-PBSA analysis					
Energy, kcal/mol	ΔEEL	ΔVDW	ΔENPOLAR	ΔEPB	ΔPBSA
Wild type	−239.78 ± 40.70	−21.99 ± 7.09	−2.81 ± 0.78	249.21 ± 39.94	**−15.38 ± 6.51**
F135Y	−347.34 ± 33.82	−49.67 ± 4.00	−5.27 ± 0.20	364.38 ± 31.00	**−37.91 ± 7.19**
F152Y	−386.79 ± 47.43	−41.24 ± 5.96	−5.13 ± 0.51	399.25 ± 44.59	−33.90 ± 8.90
F155Y	−264.34 ± 33.25	−39.32 ± 4.26	−4.17 ± 0.29	296.05 ± 32.45	−11.77 ± 5.82
F176Y	−433.93 ± 44.46	−56.13 ± 6.31	−6.24 ± 0.58	463.04 ± 41.52	−33.26 ± 8.94
F179Y	−318.82 ± 47.76	−29.15 ± 7.59	−3.76 ± 0.71	332.95 ± 46.85	−18.78 ± 7.41
F209Y	-441.65 ± 37.68	-33.37 ± 5.50	-4.29 ± 0.48	445.55 ± 34.79	-33.76 ± 8.00

molecular interactions, such as those in protein-protein and protein-ligand binding. A larger BSA typically indicates stronger and more specific interactions, as more surface area is involved in binding, leading to greater interaction energy. To find the buried surface area of protein RNA complexes, first the SASA was measured of all the three groups (protein, ligand and complex) of all the mutations and wild type. Mutants BSA were then compared to the wild type of structure BSA in nanometer square along the time trajectory (in picoseconds). Figure 5 shows the values obtained from the last 50 ns frames of the total 200 ns simulation.

The wild-type complex demonstrated the highest buried surface area relative to the other mutants, indicating strongest and more specific interactions between the protein and RNA. Following the wild type, both the F135Y and F209Y mutants displayed (Fig. 5) overlapping trajectories over time for BSA. These mutants exhibited higher buried surface area values compared to other mutants, resulting in strong interactions albeit slightly lower than those of the wild type. The F179Y mutant showcased the lowest BSA among the studied variants, indicating weak interactions between the molecules. This finding implies a reduction in surface contact and potentially a decrease in the stability of the complex. The BSA values for mutants F176Y and F155Y lay (Fig. 5) on the lower side of all mutants and wild type complexes, yet higher than the F179Y mutant.

Principal Component Analysis

PCA helps to identify the most significant modes of motion and explore protein conformations at atomic levels. It attempts to determine the cumulative fluctuation trends in the trajectory of protein complexes by calculating and diagonalizing the Cα atoms covariance matrix. The eigenvectors generated having the highest values are called principal components (PCs). In this study, PCA was conducted to determine fluctuations in the wild-type and mutated protein complexes. MD trajectory files of the last 50 ns were used to extract protein interactions and atomic gaps between wild-type and mutated proteins. The projections of Cα atoms of proteins and two prominent PCs were plotted opposite each other with respect to the time of simulation (Fig. 6). It was observed that mutant proteins exhibited large dynamic movements and evident fluctuations in terms of atomic vibrations for the last 50 ns simulation period. The depiction of yellow to purple color indicates the shifting of conformations for the mutant and wild-type proteins throughout the interval of the observed simulation. It represents the significant periodic bounces amid the different conformational postures of the structures of wild-type and mutant protein-RNA complexes.

Contact Analysis

Recognition between biomolecules considerably relies on non-covalent interactions. Similarly, the RRM-RNA interfacial interactions are dominated by

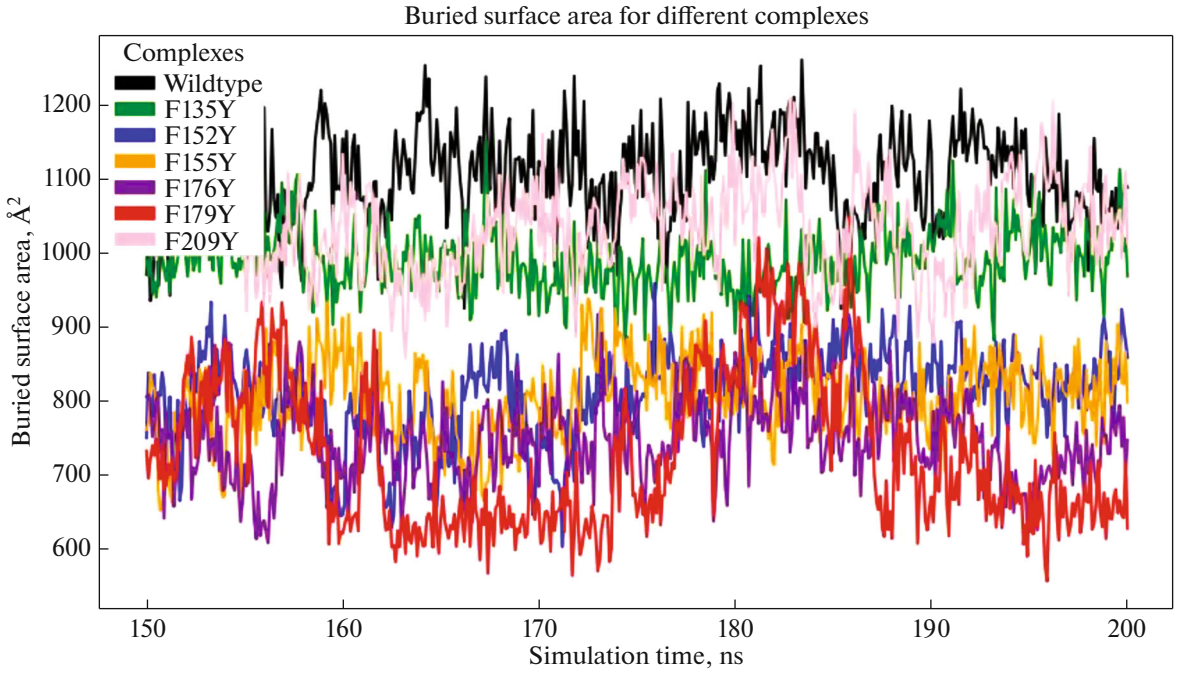


Fig. 5. BSA (Buried Surface Area) plot of last 50 ns (150–200 ns) for each mutant and wild type complexes residue. *Y* axis depicting area in nano-meter square and *x* axis depicting time in picoseconds. F135Y (green), F152Y (blue), F155Y (yellow), F176Y (purple), F179Y (red), F209Y (pink), Wildtype (black).

intermolecular stacking interactions and hydrogen bonds [31]. Hydrogen bonds stabilize the secondary and tertiary structures of proteins, including alpha helices and beta sheets. In addition, they contribute to the determination of the specificity of protein-ligand interactions and the stability of protein-protein interactions. Hence, finding the H-bond percentage in the desired interactions would provide insights into the binding between RRM and proteins. The last 50 ns MD trajectories of wild-type and mutant complexes were analyzed using a Python package to extract intermolecular bonds. Table 2 shows that in most of the mutation complexes, Phe135 was involved in π–π interactions, while Phe179 contributed to the wild-type and F152Y pi interactions, respectively. Additionally, it was observed that only Tyr135 in mutant complex F135Y made the π–π interactions among all mutant complexes. Amino acids like Arg, Lys, Asp, and Ser were the major contributors to hydrogen bond formation between protein and RNA. Moreover, Phe135 and Phe176, along with π–π interactions, also depicted hydrophobic interactions with the nucleotides in case of wild-type complex.

PyContact was used for the identification of hydrogen bond interactions. The bond distance range for hydrogen bond interaction was set to be from 2.7 to 3.3 Å, with the cut-off angle set to 120°. The H-bond interaction percentage for all complexes was derived and is shown in Fig. 7. The input files for PyContact were MD topology (.psf) and trajectory files(.xtc). For the wildtype protein complex in the last 50 ns, the maximum number of hydrogen bond occupancy of 70% was shown by GLN 46 residue with URA 4, followed by GLN 117, MET 118, and ARG172 within the range of 30–40% hydrogen bond occupancy. Interestingly, the ARG172, GLY178, and ASP 211 residues in F135Y, and the ARG172, ASP211residues in F152Yand F209Y—mutant complexes showed the highest H-bond occupancy of 100%, which was higher than the wild type. In the F135Y mutant, the mutated residue Tyr 135 also showed 70% hydrogen bond occupancy with GUA3. In F155Y, ASP 211 (75%) showed the highest occupancy percentage. For the F176Y mutant complex, residue LYS141 has 85% occupancy and HSD (Hydroxysteroid dehydrogenases) was seen with 88% bond occupancy. In F179Y, maximum occupancy was shown by ASP 211 with 59%, and F209Y has ARG 172, showing the highest occupancy of 100% with URA1 and 4. Overall, the residues involved significantly in H-bond interactions were found to be ARG and ASP.

An overall analysis of various key attributes (MMGBSA/MMPBSA, No. of hydrogen bonds and buried Surface area) were done to find rank average. It is a statistical measure used to combine rankings from multiple attributes into a single, aggregate ranking. These factors collectively contribute to our understanding of the binding affinity, interaction stability, and interaction surface between protein and RNA. The F135Y mutation is ranked highest overall, excelling particularly in MMGBSA values and hydrogen bonding, suggesting a significantly enhanced binding

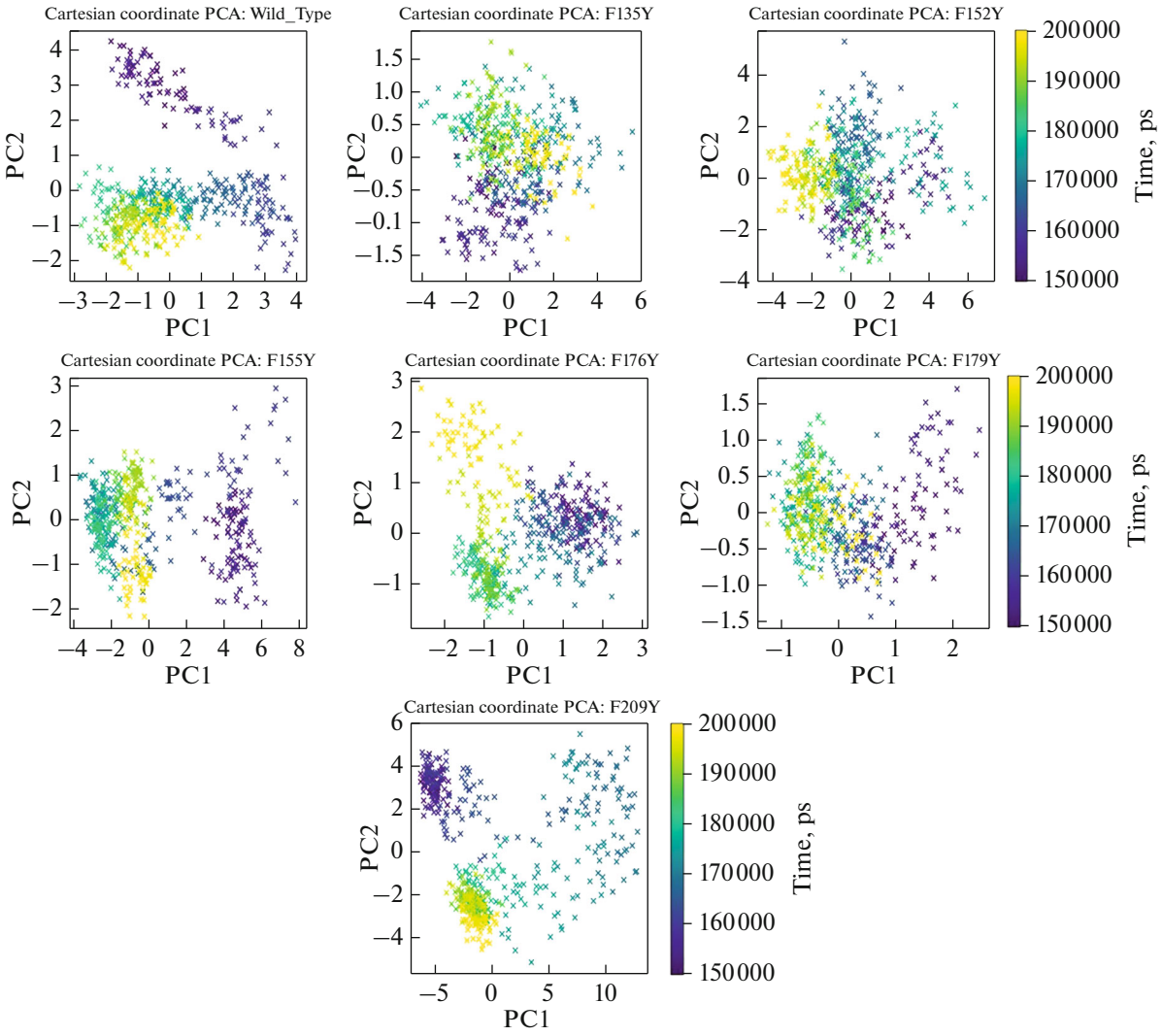


Fig. 6. PCA graphs for native and mutant proteins exhibiting large dynamic movements.

affinity and interaction stability compared to the wild type. F209Y is ranked second, showing moderate decrease in binding affinity and a slight increase in buried surface area. The F152Y and F176Y mutations share the same average rank but for different reasons; F152Y shows balanced attributes with moderate binding affinity and surface area, while F176Y stands out for its exceptionally high MMGBSA value, despite its average hydrogen bonding and smaller BSA. The F179Y mutation has higher binding energies (weaker binding interactions) and a smaller BSA than most other mutants. The lower rank reflects compromised interaction capabilities, likely due to reduced hydrogen bonding and hydrophobic interactions.

DISCUSSION

A protein-RNA interaction is like a molecular dance between two partners, wherein the protein and RNA molecules gracefully weave their unique sequences together, creating a beautiful symphony of biological function.

RRM binding with nucleotides plays a crucial role in various intermolecular interactions, serving as a key mechanism for regulating the recognition and binding of DNA and RNA molecules [37]. Hence, any changes in its sequence may affect different aspects of the biological processes involving RRM, such as transcription, translation, and regulation of gene expression. Despite the prevalence and biological relevance of RNA-protein interactions, little definitive information is available about them at the sub-molecular level [38]. A prominent characteristic observed in binding single-stranded DNA (ssDNA), and RNA is the direct interaction between the protein and the nucleotide bases. Various studies have highlighted the significance of Van der Waals interactions, particularly aromatic stacking, in stabilizing protein-base interactions

Table 2. Types of π–π stacking interactions of protein residues with nucleotides of RNA

Residue no.	AA	RNA Nuc	Distance	Angle	Stacking type
135A (Wild)	**PHE**	**3G**	**4.92**	**89.51**	**T (pi-stacking, perpendicular)**
135A (F135Y)	TYR	3G	5.32	78.52	T (pi-stacking perpendicular)
135A (F152Y)	PHE	3G	5.27 5.11	81.86 88.27	T (pi-stacking perpendicular)
135A (F155Y)	PHE	3G	5.06	85.28	T (pi-stacking perpendicular)
135A (F176Y)	PHE	3G	5.22	87.67	T (pi-stacking perpendicular)
135A (F179Y)	PHE	3G	5.08	84.81	T (pi-stacking perpendicular)
135A (F209Y)	PHE	3G	5.13	85.46	T (pi-stacking perpendicular)
135A (Wild)	**PHE**	**4U**	**3.77**	**4.88**	**P (pi-stacking parallel)**
135A (F135Y)	TYR	4U	3.82	13.71	P (pi-stacking parallel)
135A (F152Y)					**No pi-pi interaction**
135A (F155 Y)	PHE	4U	3.59	21.40	P (pi-stacking, parallel)
135A (F176Y)					**No pi-pi interaction**
135A (F179Y)	PHE	4U	3.54	10.96	P (pi-stacking, parallel)
135A (F209Y)	PHE	4U	3.54	14.32	P (pi-stacking, parallel)

[39]. Conserved amino acids in RRMs are key factors for binding with RNA. The three most highly conserved residues in RRMs are aromatic, substantiating the importance of aromatic stacking interactions in RRM-RNA binding [39, 40]. The upregulation of CELF2 seems to play a significant role in the pathogenesis of DM type 1, which is a debilitating disease affecting human musculature. Thus, a detailed understanding of the molecular (and sub-molecular) mechanisms of binding between CELF2 and RNA can shed light into newer treatment options for DM type 1. The present study aimed to investigate the impact of point mutations in the RRM2 domain of CELF 2 protein on its binding interaction with RNA. This is done by replacing Phe with Tyr, thus maintaining the aromatic nature of the position. Tyr and Phe are two structurally similar amino acids, differing only by the presence or absence of a hydroxyl (-OH) group. The analysis of the change in trajectory between wild-type and mutant systems revealed that the overall structural stability of the protein-RNA complex varied based on the position of mutation induced in the sequence and the residues surrounding it.

In 200 ns MD simulation analysis, the single amino acid mutations F135Y, F155Y, and F176Y displayed significantly similar molecular characteristics to the wild type in RMSD (Fig. 2). Additionally, the most deviated mutant complex was F209Y, which might be because of the position at the end loop and ends seem to be looser with increased flexibility. At the same time, the F152Y mutant showed deviation till 100 ns and mutant F179Y showed deviation till 50 ns, getting stabilized after that. The RMSF analysis (Fig. 4) revealed that the F155Y mutant had increased fluctuations concerning the other two (F135Y and F176Y) and wild type at the residue level. The amino acid residues of F209Y depicted maximum fluctuations followed by F152Y. The F179Y mutant exhibited minimum fluctuations (increased rigidity) compared to other mutant and wild types.

The MMPBSA results (Table 1) showed that all the mutated complexes demonstrated lower binding energies than the wild type. Mutant F176Y (-63.87 ± 7.07 kcal/mol) and F135Y (-37.91 ± 7.19 kcal/mol) had the lowest GB and PB values, respectively, depicting stronger binding than the wild type. The contribution of various bonds in the total binding energies of all 7 complex systems was also calculated, and the patterns showed that Van der Waals interactions were the highest energy contributors in the binding nature identification of protein-RNA complexes. These results reiterate previous findings that Van der Waals interactions play a significant role in RRM-RNA binding. Jones et al. have reported that Van der Waals contacts are prevalent, with an amino acid preference of Phe (F) and Tyr (Y) [41].

While both Phe and Tyr can participate in hydrophobic interactions, the hydroxyl group in Tyr allows for additional interactions with polar groups. Tyrosine –OH groups generally make a favorable contribution to protein stability, which is substantially more significant when the –OH group is hydrogen-bonded [42]. This explains the similar RMSD and RMSF results of F135Y mutant to the wild type, that even after inducing the mutation, the mutant complex does not deviate much from the wild type. In mutant F135Y, Tyr mutation positioned at 135 made hydrogen and π-stacking interaction, leading to stronger bonds and lower energies (Figs. 8 and 9a, Table 2). Positively charged amino acids like Arg (172) and Lys (141, 208), which have a

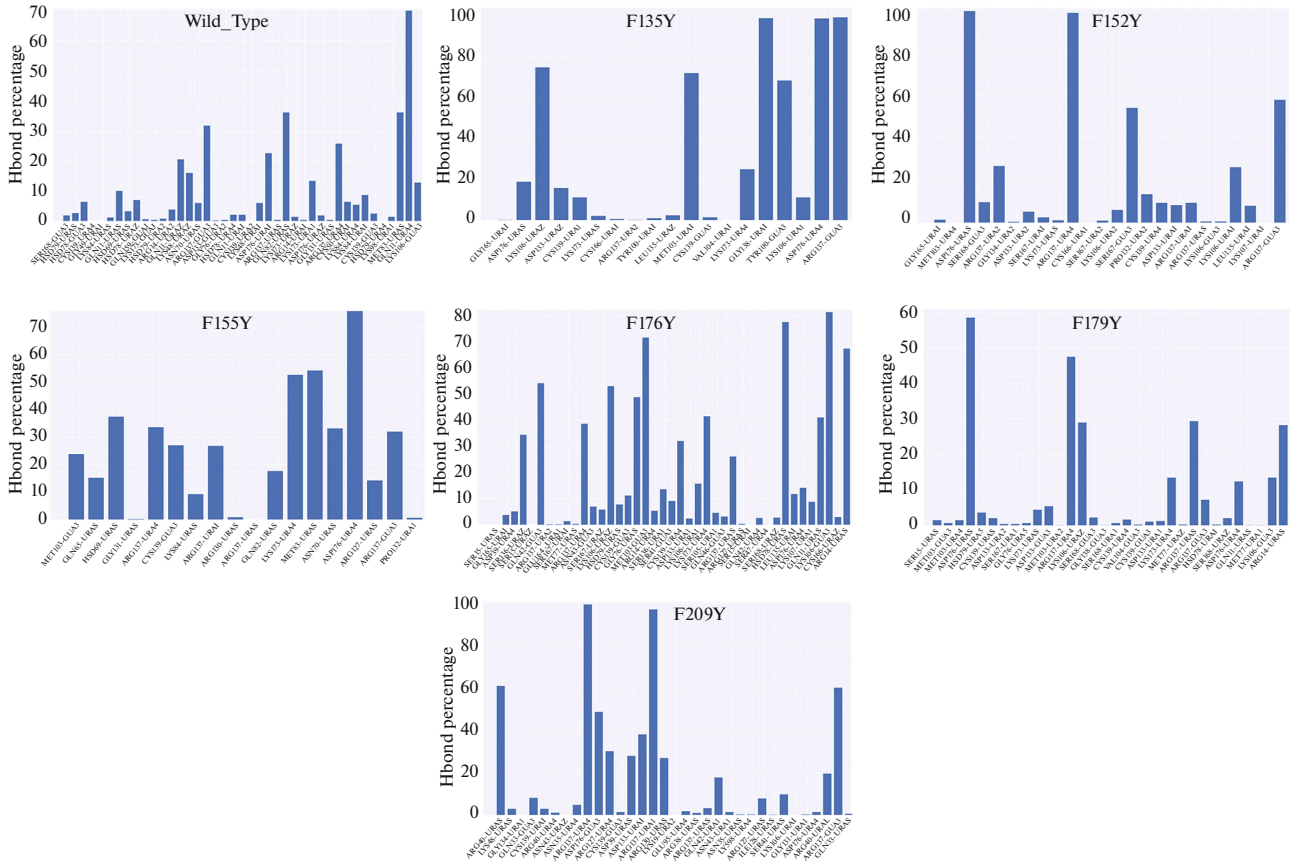


Fig. 7. PyContact analysis graphs for wild and mutant protein-RNA complexes' MD trajectories for potential hydrogen bond occupancy.

high propensity to bind with RNA, form strong ionic hydrogen bonds with RNA, [43] favorably contributing to the stability of the mutant F135Y. However, the combination of these hydrogen bond interactions and Tyr –OH binding in F135Y is not present in the wild type or any other mutant, making the mutant complex comparatively more stable and stronger.

The F152Y mutant exhibited a similar number of hydrophobic and hydrogen bond interactions (Fig. 9b) as the wild type. The lower binding energy (Table 1) of the mutant complex may be attributed to the number of these interactions around the RNA molecule. The amino acids (ASP 168, ASP 211, SER 208, ARG 172) facilitate the primary hydrogen bond interactions within the protein structure's loop region. In addition to hydrophobic interactions, the presence of these amino acids also contributed to the overall binding energy.

Still, their positioning within loops of the protein structure might result in a significantly increased flexibility and deviation when compared to wild type and other mutants. Loops connected to β-strands or α-helices are involved in RRM recognition and responsible for the recognition of RNA shape [44]. The flexibility of loops enables the correct positioning of catalytic residues. We believe that upon the introduction of mutation in F152Y, the catalytic amino acids within the loop repositioned to achieve stability. This subsequently might have resulted in more fluctuations and flexibility [45]. Furthermore, the tyrosine mutation is not in direct proximity to the RNA molecule, as it is positioned further back in the structural arrangement, hence not being able to contribute to the stability of the mutant complex. Based on these observations, it can be inferred that the spatial arrangement of an amino acid within the protein-RNA complex structure exerts a notable influence on the binding affinity of both biomolecules.

In contrast to F152Y, F155Y did not exhibit elevated binding energies despite their proximity of only three amino acids. The F155Y variant exhibits reduced hydrophobic interactions that contribute to its low binding energy. It is observed that the amino acids aspartic acid (ASP) and lysine (LYS) form hydrogen bonds with uracil 4 of the RNA simultaneously (Fig. 9b). ASP is an amino acid that carries a negative charge due to the presence of a carboxylate side chain. On the contrary, LYS has an amino group as its side chain and bears a positive charge. When both amino acids bind to uracil simultaneously, the proximity of

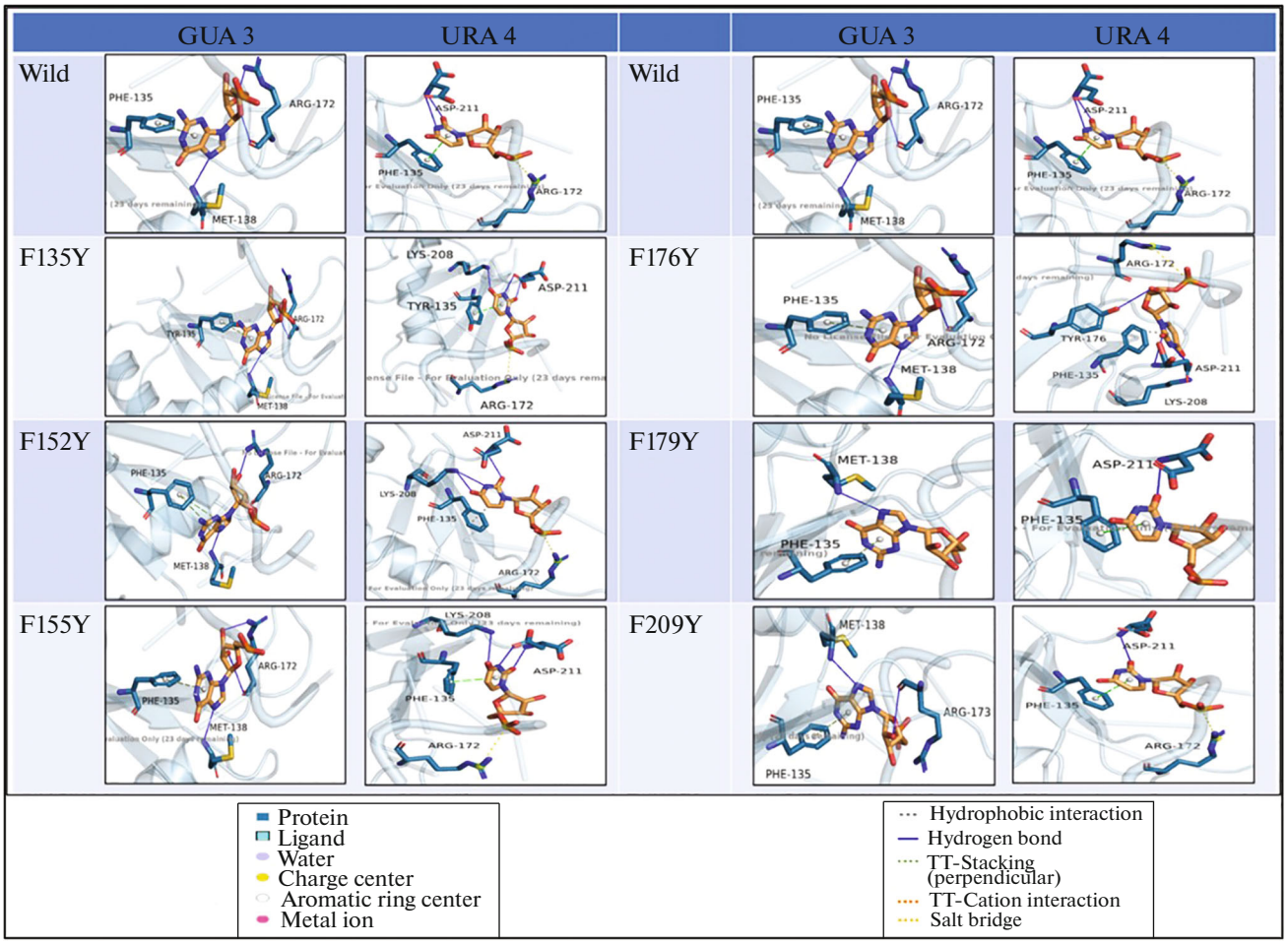


Fig. 8. Interaction and proximity of residues around RNA in 200 ns average structure of wild type and mutant complexes with GUA 3 and URA 4.

their charged and polar side chains may lead to attractive forces, stabilizing the mutant complex. Nucleic acids possess surface charges, which implies that the binding of proteins to these targets is anticipated to be significantly affected by electrostatic interactions. Residues can have either positive or negative charge, where opposite charges attract and like charges repel each other; interactions between charged residues can confer specificity [46].

Figure 9b shows the binding interactions of HSD104, GLN117, and ARG162 with the O2 atom of URA5 residue in the RNA. Due to the presence of only two lone pairs of electrons, oxygen exhibits a maximum bonding capacity of two, limiting its ability to form bonds with more than two amino acids at a particular instant. The generated diagram exhibits dynamic behavior over 50 ns, illustrating the competitive interaction among three amino acids as they gradually form bonds with URA5. This competition might induce steric hindrance among the three amino acids, resulting in weakened bond and binding energy, as evidenced by our RMSD and MMPBSA findings (Fig. 9b, Table 1).

The F176Y variant exhibits lower binding energy (−33.26 ± 8.94 kcal) than the wild type and corresponds to F152Y (−33.90 ± 8.90 kcal) and F209Y (−33.76 ± 8.00 kcal). However, the F176Y variant displays the least fluctuation compared to these mutants. Hydrogen bonds are formed by residues LYS 141, SER 171, LYS 208, and MET 138, among which only lysine carries a positive charge, and others are neutral. Therefore, there is no steric hindrance or repelling between the amino acids. Thus, making the mutant has similar binding energy and reduced flexibility. Similarly, another study by Nolan et al. (1999) showed that introducing Tyr mutation led to lower binding affinity and slight destabilization of the RRM-RNA complex structure [47].

Compared to other mutants, F179Y exhibited the highest binding energy (−18.78 ± 7.41 kcal), nearly equivalent to the wild type. The RMSD and RMSF (Figs. 2, 4) exhibited a similar pattern to that of the wild type. This observation indicates that the mutation

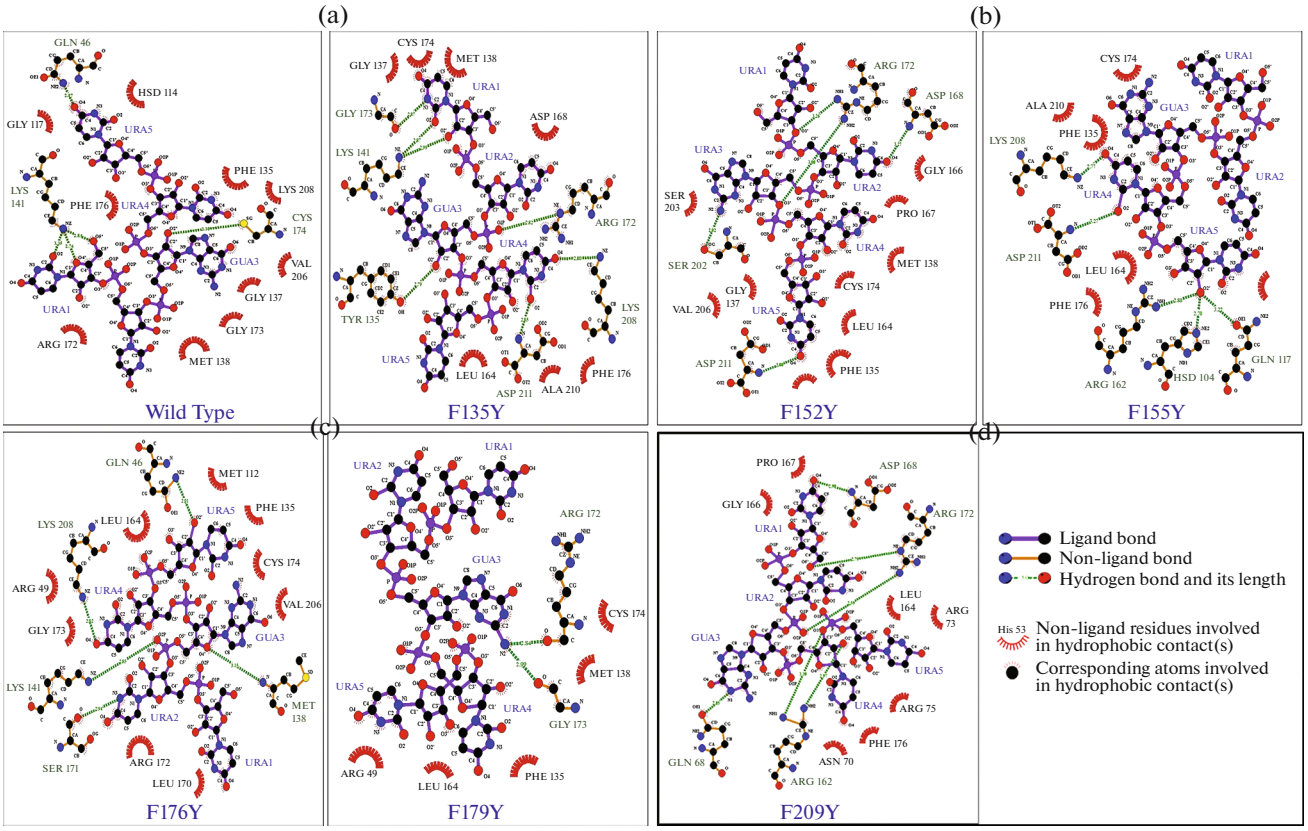


Fig. 9. Interaction of atoms and residues around RNA nucleotide of wild type and F135Y (a). Interaction of atoms and residues around RNA nucleotide of F152Y and F155Y (b). Interaction of atoms and residues around RNA nucleotide of F176Y and F179Y (c). Interaction of atoms and residues around RNA nucleotide of F209Y (d).

occurring at position 179 had minimal impact on the original protein-RNA complex. Additionally, there is a limited presence of binding residues that form hydrogen bonds (Fig. 7) and a reduced contribution of hydrophobic interactions (Fig. 9c) in comparison to other mutants. These results suggest that the presence of Tyr in Phe's position is only mildly disruptive. Similar results were shown in a study by Law et al. [48] when Phe was mutated by Tyr.

F209Y mutant exhibited the maximum flexibility while having a similar binding affinity to F152Y and F176Y. Residues like ASP, ARG, and GLN contribute to hydrogen bond formation. However, even after stronger hydrogen and hydrophobic interaction, F209Y was the most flexible mutant. The flexible behavior of the mutant might be explained by quoting the first lines of the paper [49]. The terminal regions of proteins differ in several ways from more internal segments. The termini are often surface-exposed and flexible. As F209Y is located at the end terminal of the protein chain, this could be the possible reason for it. Also, mutating Tyr at the Phe in the wild type, the interaction site becomes rigid, locking the molecules together. In contrast, adjacent elements in the two molecules loosen to balance the decrease in entropy. In this way, nucleotides or residues that do not directly interact can still be instrumental in binding if they direct the necessary compensatory changes for binding [50].

The "stacking" of aromatic moieties is a characteristic feature of complexes of proteins with ssRNA. In general, the planes of the aromatic rings stack in a parallel orientation, with an average distance of ~ 3.3 Å between the planes [51]. Intermolecular stacking is also observed at protein-RNA binding interfaces, consisting of rings of nucleic acid bases with aromatic side chains such as phenylalanine, tyrosine, tryptophan, and/or histidine. Two potential orientations are identified in these protein-RNA intermolecular stacking-parallel (P) and perpendicular (T) configurations, as these orientations represent the energy minima. In this study, we found that all the mutants displayed both P and T orientations except for F152Y and F176Y. These two mutants displayed only T- pi-stacking orientation with 4U (Table 2), while the rest showed both T- and P-type stacking with 3G. This could be explained by the observation that purines are better stacking partners than pyrimidines, as demonstrated by experiments on ss-polynucleotides and isolated nucleosides [52]. Generally, guanine (G) forms stronger and more

Table 3. Rank Average analysis of mutants compared with Wild type over 50 ns, for various attributes

Rank	Analysis (50 ns)	MMGBSA, kcal/mol	MMPBSA, kcal/mol	No. of Hydrogen bond (Avg)	BSA, nm^2 (Avg)	Rank Average
1	F135Y	−56.71	**−37.91**	6.55	9.902	1.75
2	F209Y	−39.94	−33.76	6.12	10.35	3
3	F152Y	−48.15	−33.9	3.69	8.12	3.5
4	F176Y	**−63.87**	−33.26	4.53	7.437	3.5
5	Wild Type	**−23.96**	**−15.38**	3.78	11.038	4.5
6	F155Y	−42.33	−11.77	1.56	8.051	5.75
7	F179Y	−33.05	−18.78	3.59	7.291	6

favorable stacking interactions than uracil (U). Guanine has a more planar structure compared to uracil. The planar conformation of guanine allows for better alignment and overlap of the aromatic rings, which promotes stronger stacking interactions.

On the other hand, Uracil exhibits some degree of puckering, which can introduce steric hindrance and disrupt stacking interactions. Guanine also possesses a more extensive π-electron system compared to uracil. The extended π-electron system in guanine provides more opportunities for π–π stacking interactions, resulting in stronger and more stable stacking. However, according to a general statistical study of protein-RNA interactions, all four nucleotide bases are involved in stacking interactions, most often with phenylalanine [53]. For mutant F152Y, the observations mentioned above can be explained.

Furthermore, PCA analysis was performed to investigate the fluctuations between wild-type and mutated protein complexes. The outcomes demonstrated that the mutations affected the conformations of the protein atoms throughout the duration of the observed simulation and were individual entities after mutation.

Our results indicate that only the F135Y mutation is the most stable and stronger mutation (Table 3) supported by π–π stacking interactions and increased hydrogen bond occupancy. Additionally, in the RNP1 motif (**F**/T-I-G-M-V) from 135–139 residues, where F was at conserved position at 135, did not lead to significant disruptions. Whereas F179Y mutation did not result in stronger binding affinity and decreased hydrogen bond occupancy. In RNP1 (R-G-C-A-**F**/T-V-T-**F**/T) from 172–179 residues, for mutations F176Y and F179Y, the interactions were less stable, leading to a less adaptable interaction with RNA. This resulted in making RNP1 more unstable with the introduced mutations compared to F135Y in RNP2. In case F179Y, the presence of two adjacent tyrosine residues introduces steric hindrance which could potentially make the binding site less accessible or alter the orientation of the tyrosine's, affecting the overall binding affinity and specificity.

The study further observed that Arg acts as another key residue in the association of RRM with RNA bases. These mutations can affect adjacent residues that interact with the RNA molecule in 3D space. In protein–RNA interactions, the Van der Waals contacts significantly outnumber the hydrogen bonding. The proteins in these complexes prefer contacting the G (purine) and U (pyrimidine), using Van der Waals contacts and H-bonds. There seems to be a preference for the presence of Arg, Tyr, and Phe at the RNA binding site [43]. Free contribution energy calculations show that the binding energy of mutants was higher than that of the wild type. This could be possible because Tyr increases the hydrogen bonding possibilities in its vicinity. The additional hydroxyl group can further potentially expand the repertoire of RNA sequences that the RRM can recognize and bind to and subsequently strengthen the interaction between the RRM domain and its RNA targets. While Tyr mutation is suggested to make stronger bonds with higher binding energy, nature has favored Phe, which forms slightly weaker bonds with the RNA than Tyr. These bonds are much easier to make and break in the case of phenylalanine (K_D = 45 μM) as compared to those formed by tyrosine (K_D = 23 μM) [54]. We believe this allows the CELF2 protein to engage in dynamic protein-protein interactions and form complexes with partner proteins, which play an important role in modulating RNA processing and functional regulation. This protein is also involved in post-transcriptional regulation, including alternative splicing. Weaker bonds (in part of Phe) in CELF2 may contribute to its ability to remodel RNA structures or induce conformational changes in target RNAs. This remodeling activity influences alternative splicing patterns and other RNA processing events. Future studies, including functional studies and characterization of the mutated RRM domains, should focus on assessing the potential therapeutic implications of this mutation in elucidating the role of CELF2 in normal cellular functions and the pathogenesis of related abnormalities such as MD.

CONCLUSIONS

RRMs are an important part of RBPs and contribute to various cellular functions. Hence, evaluating and understanding the effects of mutations in the RRM sequence is necessary. RNA-binding proteins such as CELF 2 (CUG-BP2) contain multiple RRM copies and are involved in several regulatory processes. The comparative analysis of MD trajectories indicates that the residue mutations of the RRM domain induce fluctuations of complex systems but increase the binding energy with RNA. The calculated binding energies based on MM/PBSA agree well with the binding affinities. Depending on the position, the mutations caused significant structural fluctuations in some of the complexes and affected the dynamic behavior of the protein residues. The mutated complexes showed stronger binding interactions with RNA than the wild type. Van der Waals interactions were identified as the highest energy contributors in the binding nature identification of protein-RNA complexes. The findings of this study could help find how RRM recognizes and interacts with RNA. By investigating the impact of a specific amino acid mutation on protein structure, binding affinity, and flexibility, this study enhances our understanding of how amino acid substitutions can affect protein properties and binding with RNA.

The information provided by MD simulations and further derivative studies can be applied to various fields, including protein engineering, drug design, and understanding the functional consequences of genetic variations. This analysis provides valuable insights into the specific interactions and dynamics of protein-RNA complexes, aiding in the characterization of protein structure-function relationships with RNA. While aromatic amino acids often play a crucial role, their impact is not uniform and can be influenced by the surrounding molecular environment. Further experimental and in-depth studies are required to uncover the unknown features of the RRM-RNA interaction world. Additionally, it is crucial to recognize the limitations of the point mutation approach. The substitution of even a single atomic group can induce alterations in a sphere with a radius of up to 5 Å, and the effects of such changes cannot be excluded. This study paves the way towards exploration of other amino acid substitutions in RRMs, assess the effects of mutations in different contexts, and elucidate the role of sub molecular protein-RNA interactions on specific biological functions and diseases.

There are certain restrictions to the point mutation approach, used in this research of aromatic amino acid mutational analysis. The point mutation approach, though valuable, has inherent limitations that must be acknowledged. Even the substitution of a single atomic group can cause structural changes within a 5 Å radius, potentially affecting both nearby and distant regions of the protein. These alterations might disrupt the protein's overall structure and function. Additionally, point mutations can alter RNA Recognition Motifs (RRMs), influencing their ability to bind RNA with the correct affinity and specificity. This can unintentionally affect the protein's interactions with RNA, possibly disrupting processes like splicing, stability, and translation. Given the complex and far-reaching impacts of these mutations, interpreting their results requires careful consideration of potential off-target effects and structural changes. It is essential to validate findings through further experiments to confirm their significance. Therefore, conclusions drawn from point mutation studies should be made with caution, avoiding overgeneralization.

ABBREVIATIONS AND NOTATION

RRM	RNA Recognition Motif
RBP	RNA-binding Proteins
RNA	Ribonucleic Acid
CELF2	CUGBP Elav-like family member 2
MD	Molecular Dynamics
Phe	Phenylalanine
Tyr	Tyrosine
RMSD	Root mean Square Deviation
RMSF	Root mean Square Fluctuations
PCA	Principal Component Analysis
MMPBSA	Molecular Mechanics Poisson–Boltzmann Surface Area
MMGBSA	Molecular mechanics with Generalized Born and Surface Area Solvation
RNP	Ribonucleoprotein
RBD	RNA Binding Domain
KH	K-Homology Domain
ZF	Zinc Finger Domain
hnRNP	Heterogenous Ribonucleoprotein
DM	Myotonic Dystrophy
AML	Acute Myeloid Leukemia
BSA	Buried Surface Area
NVT	number of particles (N), system volume (V), and temperature (T)
NPT	Constant temperature, constant pressure
PLIP	Protein-Ligand interaction Profiler
SASA	Solvent Accessible Surface Area
Arg	Arginine
Lys	Lysine
Asp	Aspartic Acid
Ser	Serine
GLN	Glutamine
MET	Methionine
GLY	Glycine
HSD	Hydroxysteroid dehydrogenases

ACKNOWLEDGMENTS

We express our gratitude to Delhi Technological University, Delhi, for its invaluable help and resources that enabled us to successfully write this research study.

AUTHOR CONTRIBUTION

Neha Kumari: Methodology, Software, Writing-Original draft preparation, Visualization, Investigation. **Yasha Hasija & Neel Sarovar Bhavesh:** Conceptualization, Supervision, Reviewing and Editing.

FUNDING

This work was supported by ongoing institutional funding. No additional grants to carry out or direct this particular research were obtained.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This work does not contain any studies involving human and animal subjects.

CONFLICT OF INTEREST

The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. Alberts B., Johnson A., Lewis J., Raff M., Roberts K., Walter P. 2002. Protein Function. Chapter 3. In *Molecular Biology of the Cell*. 4th ed. New York: Garland Science. https://www.ncbi.nlm.nih.gov/books/NBK26911/.
2. Lührmann R., Kastner B., Bach M. 1990. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochim. Biophys. Acta*. **1087** (3), 265–292. https://doi.org/10.1016/0167-4781(90)90001-I
3. Moore P.B. 1998. The three-dimensional structure of the ribosome and its components. *Annu. Rev. Biophys. Biomol. Struct*. **27**, 35–58. https://doi.org/10.1146/annurev.biophys.27.1.35
4. Stefl R., Skrisovska L., Allain F.H.T. 2005. RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO Rep*. **6** (1), 33–38. https://doi.org/10.1038/sj.embor.7400325
5. Riley K.J., Steitz J.A. 2013. The "Observer Effect" in genome-wide surveys of protein-RNA interactions. *Mol. Cell*. **49** (4), 601–604. https://doi.org/10.1016/j.molcel.2013.01.030
6. Dominguez D., Freese P., Alexis M.S., Su A., Hochman M., Palden T., Bazile C., Lambert N.J., Van Nostrand E.L., Pratt G.A., Yeo G.W., Graveley B.R., Burge C.B. 2018. Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell*. **70** (5), 854–867.e9. https://doi.org/10.1016/j.molcel.2018.05.001
7. Darai N., Mahalapbutr P., Wolschann P., Lee V.S., Wolfinger M.T., Rungrotmongkol T. 2022. Theoretical studies on RNA recognition by Musashi 1 RNA-binding protein. *Sci. Rep.* **12**, 12137. https://doi.org/10.1038/s41598-022-16252-w
8. Cléry A., Blatter M., Allain F.H.T. 2008. RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol*. **18** (3), 290–298. https://doi.org/10.1016/j.sbi.2008.04.002
9. Krepl M., Cléry A., Blatter M., Allain F.H.T., Sponer J. 2016. Synergy between NMR measurements and MD simulations of protein/RNA complexes: Application to the RRMs, the most common RNA recognition motifs. *Nucleic Acids Res*. **44** (13), 6452–6470. https://doi.org/10.1093/nar/gkw438
10. Nagai K., Oubridge C., Jessen T.H., Li J., Evans P.R. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*. **348** (6301), 515–520. https://doi.org/10.1038/348515A0
11. Muto Y., Yokoyama S. 2012. Structural insight into RNA recognition motifs: versatile molecular Lego building blocks for biological systems. *Wiley Interdiscip. Rev. RNA*. **3** (2), 229–246. https://doi.org/10.1002/wrna.1107
12. Allain F.H.T., Bouvet P., Dieckmann T., Feigon J. 2000. Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J*. **19** (24), 6870–6881. https://doi.org/10.1093/emboj/19.24.6870
13. Afroz T., Cienikova Z., Cléry A., Allain F.H.T. 2015. One, two, three, four! How multiple RRMs read the genome sequence. *Methods Enzymol*. **558** (1), 235–278. https://doi.org/10.1016/bs.mie.2015.01.015
14. Diarra Dit Konté N., Krepl M., Damberger F.F., Ripin N., Duss O., Šponer J., Allain F.H. 2017. Aromatic side-chain conformational switch on the surface of the RNA recognition motif enables RNA discrimination. *Nat. Commun.* **8** (1), 654. https://doi.org/10.1038/s41467-017-00631-3
15. Suominen T., Bachinski L.L., Auvinen S., Hackman P., Baggerly K.A., Angelini C., Peltonen L., Krahe R., Udd B. 2011. Population frequency of myotonic dystrophy: higher than expected frequency of myotonic dystrophy type 2 (DM2) mutation in Finland. *Eur. J. Hum. Genet.* **19**, 776–782. https://doi.org/10.1038/ejhg.2011.23
16. Timchenko L.T., Miller J.W., Timchenko N.A., DeVore D.R., Datar K.V., Lin L., Roberts R., Caskey C.T., Swanson M.S. 1996. Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic Acids Res*. **24** (22), 4407–4417. https://doi.org/10.1093/nar/24.22.4407
17. Guo T., T, Wang Y., Sun X., Hou S., Lan Y., Yuan S., Yang S., Zhao F., Chu Y., Ma Y., Cheng T., Yu J., Liu B., Yuan W., Wang X. 2024. Loss of RNA-binding protein CELF2 promotes acute leukemia development via FAT10-mTORC1. *Oncogene*. **43** (19), 1476–1487. https://doi.org/10.1038/s41388-024-03006-3
18. Bandziulis R.J., Swanson M.S., Dreyfuss G. 1989. RNA-binding proteins as developmental regulators.

Genes Dev. **3** (4), 431–437.
https://doi.org/10.1101/gad.3.4.431

19. Moras D., Poterszman A. 1995. RNA-protein interactions: diverse modes of recognition. *Curr. Biol*. **5** (3), 249–251.
https://doi.org/10.1016/S0960-9822(95)00051-0
20. Lee J., Cheng X., Swails J.M., Yeom M.S., Eastman P.K., Lemkul J.A., Wei S., Buckner J., Jeong J.C., Qi Y., Jo S., Pande V.S., Case D.A., Brooks C.L. 3rd, MacKerell A.D. Jr, Klauda J.B., Im W. 2016. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput*. **12** (1), 405–413.
https://doi.org/10.1021/acs.jctc.5b00935
21. Huang J., Mackerell A.D. 2013. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem*. **34** (25), 2135–2145.
https://doi.org/10.1002/jcc.23354
22. Blokzijl F., de Ligt J., Jager M., Sasselli V., Roerink S., Sasaki N., Huch M., Boymans S., Kuijk E., Prins P., Nijman I.J., Martincorena I., Mokry M., Wiegerinck C.L., Middendorp S., Sato T., Schwank G., Nieuwenhuis E.E., Verstegen M.M., van der Laan L.J., de Jonge J., IJzermans J.N., Vries R.G., van de Wetering M., Stratton M.R., Clevers H., Cuppen E., van Boxtel R. 2016. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature*. **538** (7624), 260–264.
https://doi.org/10.1038/nature19768
23. Hou T., Wang J., Li Y., Wang W. 2011. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model*. **51** (1), 69–82.
https://doi.org/10.1021/ci100275a
24. Homeyer N., Gohlke H. 2012. Free energy calculations by the molecular mechanics poisson-boltzmann surface area method. *Mol. Inform*. **31** (2), 114–122.
https://doi.org/10.1002/minf.201100135
25. Valdés-Tresanco M.S., Valdés-Tresanco M.E., Valiente P.A., Moreno E. 2021. Gmx_MMPBSA: A new tool to perform end-state free energy calculations with GROMACS. *J. Chem. Theory Comput*. **17** (10), 6281–6291.
https://doi.org/10.1021/acs.jctc.1c00645
26. Genheden S., Ryde U. 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug. Discovery* **10** (5), 449–461.
https://doi.org/10.1517/17460441.2015.1032936
27. Wan H., Hu J.P., Tian X.H., Chang S. 2012. Molecular dynamics simulations of wild type and mutants of human complement receptor 2 complexed with C3d. *Phys. Chem. Chem. Phys*. **15** (4), 1241–1251.
https://doi.org/10.1039/c2cp41388d
28. Amadei A., Linssen A.B.M., Berendsen H.J.C. 1993. Essential dynamics of proteins. *Proteins*. **17** (4), 412–425.
https://doi.org/10.1002/prot.340170408
29. David C.C., Jacobs D.J. 2014. Principal component analysis: A method for determining the essential dynamics of proteins. *Methods Mol. Biol*. **1084**, 193–226.
https://doi.org/10.1007/978-1-62703-658-0_11
30. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay É. 2011. Scikit-learn: Machine learning in python. *J. Machine Learning Res*. **12** (85), 2825–2830.
31. Teplova M., Farazi T.A., Tuschl T., Patel D.J. 2016. Structural basis underlying CAC RNA recognition by the RRM domain of dimeric RNA-binding protein RBPMS. *Q. Rev. Biophys*. **49**, e1.
https://doi.org/10.1017/S0033583515000207
32. Salentin S., Schreiber S., Haupt V.J., Adasme M.F., Schroeder M. 2015. PLIP: Fully automated protein–ligand interaction profiler. *Nucleic Acids Res*. **43** (W1), W443–W447.
https://doi.org/10.1093/nar/gkv315
33. Scheurer M., Rodenkirch P., Siggel M., Bernardi R.C., Schulten K., Tajkhorshid E., Rudack T. 2018. PyContact: Rapid, customizable, and visual analysis of noncovalent interactions in MD simulations. *Biophys. J*. **114** (3), 577–583.
https://doi.org/10.1016/j.bpj.2017.12.003
34. Lee B., Richards F.M. 1971. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol*. **55** (3), 379–400.
https://doi.org/10.1016/0022-2836(71)90324-x
35. Eisenberg D., Mclachlan A.D. 1986. Solvation energy in protein folding and binding, *Nature*. **319** (6050), 199–203.
https://doi.org/10.1038/319199a0
36. Eisenhaber F., Lijnzaad P., Argos P., Sander C., Scharf M. 1995. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem*. **16** (3), 273–284.
https://doi.org/10.1002/jcc.540160303
37. Guallar V., Borrelli K.W. 2005. A binding mechanism in protein-nucleotide interactions: implication for U1A RNA binding. *Proc. Natl. Acad. Sci. U. S. A*. **102** (11), 3954–3959.
https://doi.org/10.1073/pnas.0500888102
38. Wilson K.A., Holland D.J., Wetmore S.D. 2016. Topology of RNA–protein nucleobase–amino acid π–π interactions and comparison to analogous DNA–protein π–π contacts. *RNA*. **22** (5), 696.
https://doi.org/10.1261/rna.054924.115
39. Jones S., Daley D.T.A., Luscombe N.M., Berman H.M., Thornton J.M. 2001. Protein–RNA interactions: A structural analysis. *Nucleic Acids Res*. **29** (4), 943–954.
https://doi.org/10.1093/nar/29.4.943
40. Rahman M., Muhseen Z., Junaid M., Zhang H. 2015. The aromatic stacking interactions between proteins and their macromolecular ligands. *Curr. Protein Pept. Sci*. **16** (6), 502–512.
https://doi.org/10.2174/1389203716061507021315l6
41. Ellis J.J., Broom M., Jones S. 2007. Protein-RNA interactions: Structural analysis and functional classes. *Proteins*. **66** (4), 903–911.
https://doi.org/10.1002/prot.21211

42. Pace C.N., Horn G., Hebert E.J., Bechert J., Shaw K., Urbanikova L., Scholtz J.M., Sevcik J. 2001. Tyrosine hydrogen bonds make a large contribution to protein stability. *J. Mol. Biol.* **312** (2), 393–404. https://doi.org/10.1006/jmbi.2001.4956

43. Gupta A., Gribskov M. 2011. The role of RNA sequence and structure in RNA–protein interactions. *J. Mol. Biol.* **409** (4), 574–587. https://doi.org/10.1016/j.jmb.2011.04.007

44. Cléry A. Allain F.H.-T. 2011. From structure to function of RNA binding domains. In *Madame Curie Bioscience Database Internet.* https://www.ncbi.nlm.nih.gov/books/nbk63528/.

45. Gunasekaran K., Nussinov R. 2007. How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. *J. Mol. Biol.* **365** (1), 257–273. https://doi.org/10.1016/j.jmb.2006.09.062

46. Zhou H.X., Pang X. 2018. Electrostatic interactions in protein structure, folding, binding, and condensation. *Chem. Rev.* **118** (4), 1691–1741. https://doi.org/10.1021/acs.chemrev.7b00305

47. Nolan S.J., Shiels J.C., Tuite J.B., Cecere K.L., Baranger A.M. 1999. Recognition of an Essential adenine at a protein–RNA interface: Comparison of the contributions of hydrogen bonds and a stacking interaction. *J. Am. Chem. Soc.* **121** (38), 8951–8952. https://doi.org/10.1021/ja991617n

48. Law M.J., Chambers E.J., Katsamba P.S., Haworth I.S., Laird-Offringa I.A. 2005. Kinetic analysis of the role of the tyrosine 13, phenylalanine 56 and glutamine 54 network in the U1A/U1 hairpin II interaction. *Nucleic Acids Res.* **33** (9), 2917–2928. https://doi.org/10.1093/nar/gki602

49. Jacob E., Unger R. 2007. A tale of two tails: why are terminal residues of proteins exposed? *Bioinformatics.* **23** (2), e225–e230. https://doi.org/10.1093/bioinformatics/btl318

50. Corley M., Burns M.C., Yeo G.W. 2020. How RNA-binding proteins interact with RNA: Molecules and mechanisms. *Mol. Cell.* **78** (1), 9–29. https://doi.org/10.1016/j.molcel.2020.03.011

51. Šponer J., Hobza P. 2003. Molecular interactions of nucleic acid bases. A review of quantum-chemical studies. *Collect. Czechoslov. Chem. Commun.* **68** (12), 2231–2282. https://doi.org/10.1135/cccc20032231

52. Kerwin S.M. 2000. Nucleic Acids: Structures, Properties, and Functions by Victor A. Bloomfield, Donald M. Crothers, and Ignacio Tinoco, Jr., with contributions from John E. Hearst, David E. Wemmer, Peter A. Kollman, and Douglas H. Turner. Sausalito, CA: University Science Books. 2000 (Book Review). *J. Med. Chem.* **43** (24), 4721–4722. https://doi.org/10.1021/jm000396p

53. Allers J., Shamoo Y. 2001. Structure-based analysis of protein-RNA interactions using the program ENTANGLE. *J. Mol. Biol.* **311** (1), 75–86. https://doi.org/10.1006/jmbi.2001.4857

54. Ae M.Y., Joseph J., Ae W., Knight R. 2009. RNA-amino acid binding: A stereochemical era for the genetic code. *J. Mol. Evol.* **69** (5), 406–429. https://doi.org/10.1007/s00239-009-9270-1

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

AI tools may have been used in the translation or editing of this article.

Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

EmoMBTI-Net: Introducing and Leveraging a Novel Emoji Dataset for Personality Profiling with Large Language Models

Akshi Kumar

Akshi.Kumar@gold.ac.uk

Goldsmiths University of London

Dipika Jain

Delhi Technological University

Research Article

Keywords: sentiment analysis, personality, MBTI, emojis, LLM, natural language understanding

Posted Date: December 4th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4768237/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Emojis, integral to digital communication, often encapsulate complex emotional layers that enhance text beyond mere words. This research leverages the expressive power of emojis to predict Myers-Briggs Type Indicator (MBTI) personalities, diverging from conventional text-based approaches. We developed a unique dataset, EmoMBTI, by mapping emojis to specific MBTI traits using diverse posts scraped from Reddit. This dataset enabled the integration of Natural Language Processing (NLP) techniques tailored for emoji analysis. Large Language Models (LLMs) such as FlanT5, BART, and Pegasus were trained to generate contextual linkages between text and emojis, further correlating these emojis with MBTI personalities. Following the creation of this dataset, these LLMs were applied to understand the context conveyed by emojis and were subsequently fine-tuned. Additionally, transformer models like Roberta, DeBERTa, and BART were specifically fine-tuned to predict MBTI personalities based on emoji mappings from MBTI dataset posts. Our methodology significantly enhances the capability of personality assessments, with the fine-tuned BART model achieving an impressive accuracy of 0.875 in predicting MBTI types, which notably exceeds the performances of Roberta and DeBERTa, at 0.82 and 0.84 respectively. By leveraging the nuanced communication potential of emojis, this approach not only advances personality profiling techniques but also deepens insights into digital behaviour, highlighting the substantial impact of emotive icons in online interactions.

1. Introduction

Automatic personality detection utilizes machine learning algorithms and natural language processing (NLP) techniques to analyze digital communication, such as texts, social media posts, or emails, inferring individuals' personality traits based on established models like the Myers-Briggs Type Indicator (MBTI)[1–8]. By examining linguistic patterns, word choices, and communication styles, algorithms can categorize individuals into different personality types, such as Introversion (I) vs. Extraversion (E), Sensing (S) vs. Intuition (N), Thinking (T) vs. Feeling (F), and Judging (J) vs. Perceiving (P). This process offers valuable insights into behaviour, enables personalized digital experiences, and contributes to psychological research by studying large-scale datasets. Leveraging the MBTI framework, automatic personality detection algorithms can decode the nuances of digital behaviour, providing actionable insights for personal development, relationship building, targeted marketing, and a deeper understanding of human psychology in the digital age.

User-generated content on social media platforms has become a valuable resource for personality detection, as it encompasses more than just text. Beyond written posts and comments, user-generated content includes images, videos, likes, shares, and other interactions, providing a rich source of data for inferring individuals' personality traits. Thus, a multi-modal approach to personality detection, incorporating textual, visual, and emotive elements, allows a more comprehensive understanding of users' personalities and preferences, facilitating personalized experiences and targeted interventions in various digital contexts. More specifically, emojis, beyond their colourful appearance, are integral to digital communication, offering visual cues that enhance the expression of emotions and attitudes

alongside text. This symbiosis provides a deeper understanding of individuals' digital self-expression. For example, in a text message conversation, a person might use a series of heart emojis □□□ to convey affection and excitement about a forthcoming event. These emojis complement the text, adding layers of emotional nuance and conveying the sender's enthusiasm more vividly than words alone. Similarly, a laughing emoji □ might accompany a joke or humorous comment, amplifying the sense of amusement and creating a more engaging interaction.

As emojis play a crucial role in enriching digital communication by conveying emotions and attitudes in a concise and visually appealing manner, enhancing the overall expressive capacity of written text. analyzing emoji usage to infer personality types based on MBTI principles merges the realms of digital communication and psychological typology. This integration offers a novel approach to understanding the psychological underpinnings of individuals through their digital expressions.

Building on this foundation, this research aims to refine predictive models that adeptly correlate specific emojis with MBTI traits to discern personality types from textual conversations. For instance, frequent use of emotive emojis like hearts or smileys may suggest a preference for Feeling over Thinking, while the use of more ordered or structured symbols like checkmarks or clocks might indicate a Judging rather than Perceiving inclination. This methodology leverages the inherent expressiveness of emojis to tap into subtle psychological cues that may be less apparent in plain text. The utilization of emoji-based personality detection has practical applications across various domains. In social media, understanding user personality can enhance content personalization and ad targeting, improving user engagement. In educational technologies, it can help tailor learning experiences to suit different personality-driven learning styles, potentially increasing effectiveness and satisfaction. Furthermore, in professional settings, this approach could assist in team formation and leadership strategies by providing insights into team members' preferred communication styles and decision-making processes.

Transformers models and large language models (LLMs) have revolutionized natural language processing (NLP) with their capacity for complex tasks like personality detection from textual data. These models are designed to be fine-tuned on specialized datasets, which enhances their ability to discern and predict personality traits effectively. The inherent flexibility and robustness of transformer architectures enable them to process and generate text that closely mimics human communication, grounded in the nuances learned from extensive training data. The adaptability of transformers and LLMs is a key advantage, particularly for dynamic applications such as personality profiling. They are capable of ongoing learning, which allows them to refine and adapt their predictions as they are exposed to new data and evolving contexts. This capability is essential in environments where communication styles and interactions are constantly changing, ensuring that the models remain relevant and accurate over time.

Building upon the theoretical and technical groundwork, this research specifically aims to leverage emojis as a novel data source for inferring personality traits based on the MBTI model, which will serve

to advance our understanding of digital behaviour and its psychological implications. The research objectives are as follows:

- *Develop the Emo-MBTI Dataset*: Curate and refine a novel dataset by integrating emoji usage with textual data from the MBTI dataset and additional posts scraped from the r/mbti subreddit on Reddit, aiming to map emojis to specific MBTI personality types for enhanced personality analysis.
- *Implement and Evaluate Advanced LLMs for Emoji-Text Contextual Mapping*: Employ and assess the effectiveness of advanced language models like FlanT5[9], BART[10], and Pegasus in understanding and generating the contextual relationships between text and emojis, preparing the groundwork for accurate personality type predictions.
- *Fine-Tune Transformers for MBTI Personality Prediction*: Utilize fine-tuned transformer models such as Roberta[11], DeBERTa[12], and BART to predict MBTI personality types from emoji usage, leveraging the nuanced comprehension of emoji contexts to enhance personality profiling accuracy.

This paper is structured to methodically explore the interface between emoji usage and personality profiling within the digital communication realm. We begin with

a Literature Review that scrutinizes prior research on personality detection. Following this, the section on MBTI Personalities and Emojis discusses the theoretical underpinnings of our study, explaining how emojis can serve as practical indicators of MBTI personality traits. The Emo-MBTI Dataset section describes the meticulous process of dataset assembly—detailing the collection, integration, and preparation of data that combines emoji usage with textual content from diverse sources. In the Emoji-Based MBTI Personality Prediction Model section, we detail the methodologies used for training and fine-tuning Large Language Models (LLMs) to predict MBTI personality types from emojis. The Results and Discussion segment presents the outcomes of our experiments, showcasing how different models perform in the context of emoji-based personality prediction. This section highlights key findings, discusses model efficiencies, and examines the implications of the results in the broader context of digital communication and psychological analysis. Finally, the Conclusion and Future Work section summarizes the contributions of this research and outlines directions for future inquiry. It suggests potential expansions of the current work, including the exploration of additional models, further diversification of data sources, and the practical application of our findings in various domains such as marketing, mental health, and user experience design.

2. Formal Problem Definition and Mathematical Formulation

In this research, we address the challenge of predicting Myers-Briggs Type Indicator (MBTI) personality types by analyzing the use of emojis in textual communications. This problem involves interpreting the emotional and contextual nuances conveyed through emojis within text messages, a critical aspect of understanding digital human interactions.

2.1. Notation and Definitions:

- **Emojis Set (E)**: Let $E=\{e_1,e_2,\ldots,e_n\}$ represent the set of all possible emojis that may appear in text messages.
- **Text Messages (X)**: Define $X=\{x_1,x_2,\ldots,x_m\}$ as the collection of input text messages, where each x_i consists of words interspersed with emojis from E.
- **MBTI Types (Y)**: Let $Y=\{y_1,y_2,\ldots,y_k\}$ denote the set of MBTI personality types, with each y_j corresponding to one of the 16 distinct personality types recognized by the MBTI framework.

2.2. Problem Modelling: The objective is to construct a predictive function $f:X\rightarrow Y$ that assigns an MBTI type to a given text message based on its content and the emojis used. This function f is derived from a predictive model trained on a dataset $D\subset X\times Y$, where each pair (x_i,y_j) indicates that message x_i is associated with personality type y_j.

To optimize the performance of our model, we minimize a loss function $L(f(x_i),y_i)$ across the dataset DD. The loss function typically employed is the cross-entropy loss, which is particularly suited for classification tasks

3. Literature Review

The study of personality through textual analysis, particularly English texts, has seen a growing interest over the years. A notable dataset in this field is the publicly available Kaggle_MBTI dataset, which has been instrumental in analyzing personalities across 16 different dimensions. Throughout the years, various researchers have utilized this dataset to deepen our understanding of personality traits. For instance, in 2017, Hernandez and Scott [4] applied a Recurrent Neural Network (RNN) model to this dataset, while Ismail et al.[5] undertook a questionnaire-based approach to analyze MBTI personality traits. In subsequent years, more advanced models were employed; Ren et al. [6] used a BERT model in 2020 to achieve 54% accuracy, and Jain et al. [7] improved upon this in 2022 with a BERT-base model, reaching 69% accuracy. The following year, Kumar et al. [8] tested the effectiveness of a kernel-based soft voting ensemble model on these personality dimensions.

Research in personality detection has also extended beyond textual data to include audio-visual and multimodal data. The First Impression dataset from Chalearn, which focuses on Big-Five personality traits, is a prominent example. In 2016, Gurpinar et al. [13][14] explored this dataset using a pre-trained DCC model to analyze facial expressions and employed the OpenSMILE tool to extract facial emotions, achieving an accuracy of 91.3%. Following this, Nishant Rai [15] emphasized leveraging visual information through deep networks. Zhang et al. [16] developed a deep bimodal regression model using the same dataset, achieving an identical accuracy of 91.3%. Over the next few years, various approaches were tested on this dataset, including Kaya et al. [17] use of a CNN with a KELM model in 2017 which reached an accuracy of 91.7%, and Gucluturk et al.'s[18] application of a ResNet model in 2018, achieving 91.18% accuracy.

In recent years, there has been an increasing interest in integrating non-verbal cues such as facial expressions and body language into personality analysis. This trend is reflected in the work of Junior et al. [19] and Aslan et al. [20] in 2019, who employed spatial-temporal modelling with CNN and RNN models along with pre-trained deep learning models such as ResNet and VGG, achieving an accuracy of 91.16%. Artha Agastya et al. [21] continued this exploration by utilizing a range of deep learning models like DNN, CNN, RNN, and ResNet.

In 2022, Kosan et al. outlined a study aimed at predicting personality traits using social media data, particularly from Twitter. The research emphasized the importance of understanding individuals' personal characteristics for various sectors such as law enforcement, human resources, and advertising. It highlighted the need for labelled datasets and structural analysis of social media platforms to develop accurate prediction models. The methodology involved creating a personality dataset from Twitter, transforming unstructured data into meaningful formats, and utilizing LSTM-based prediction models. Evaluation was conducted on both the created dataset and an existing benchmark dataset (PAN-2015-EN).

More recently in 2024, Kennison et al. examined the relationship between personality traits, language use, and emoji usage on X (formerly Twitter). Through linguistic analysis and surveys across two samples, they found that higher emoji usage correlated with lower levels of openness to experience, while also revealing associations with specific language patterns in social media posts. These findings underscore the nuanced nature of personality expression and linguistic cues in online communication. Liao et al. [22] also proposed a comprehensive framework that incorporates audio, video, and non-verbal audio-visual elements for personality analysis using the same benchmark multimodal dataset.

Emojis, like text, images, videos, and audio, also serve as an important means of communication, particularly in conveying emotions, thoughts, and feelings. This aspect of digital communication has not been overlooked in personality research. A recent innovative study by Saedi et al.[23] involved analyzing screenshots from WhatsApp conversations that included emojis, providing new insights into personality dimensions through this unique form of expression. This research highlights how even seemingly trivial elements like emojis can offer valuable information about an individual's personality traits. We demonstrate how sophisticated algorithms can leverage emojis and language models to infer personality traits, offering insights into digital behaviour and its psychological underpinnings. Through such advancements, the intricate connection between emojis, personality, and digital communication continues to be explored, unlocking new avenues for understanding human behaviour in the digital age.

4. MBTI Personalities and Emojis

The use of emojis in digital communication can significantly illuminate aspects of an individual's personality, as evidenced by their choices of specific emojis in various contexts. This is particularly relevant when considering the Myers-Briggs Type Indicator (MBTI), which categorizes personality types

based on how individuals perceive the world and make decisions. Emojis, as non-verbal cues, offer a unique window into how personality traits manifest in modern communication methods.

For example, consider a conversation where Person A texts, "I had a challenging day at work 😩," choosing the weary face emoji to express feelings of stress and exhaustion. The emoji here does more than simply emphasize the sentiment; it also gives clues to Person A's emotional state and possibly their personality type. If Person A frequently uses such emojis to express discomfort or distress, they might have a personality that leans towards sensitivity or introversion, where personal and professional struggles are internalized deeply. In response, Person B replies, "I totally get it, let's plan something fun to unwind! 🎉," employing the party popper emoji to communicate enthusiasm and a proactive attitude. This response, marked by a contrasting emoji, suggests a personality that might be more extroverted or feeling-oriented, focusing on emotional support and active engagement in social activities as coping mechanisms. The choice of the party popper emoji, which is vibrant and celebratory, indicates a disposition towards positivity and a preference for interpersonal interaction as a way to alleviate stress.

Analyzing such emoji usage provides profound insights into personality traits, particularly in how different MBTI types might choose to express emotions and interact. Introverts and extroverts, for instance, might consistently use different sets of emojis that align with their typical behaviors—introverts opting for more subdued, reflective emojis (like 🤔 or 📚) and extroverts choosing more expressive, outgoing emojis (like 🎉 or 🗣️). Moreover, this kind of analysis can also uncover how individuals with a 'thinking' versus a 'feeling' orientation might use emojis. Thinkers might prefer emojis that are less about emotional expression and more about depicting thoughts or activities (like 💡 or 🔍), while feelers are likely to use emojis that convey empathy, affection, or emotional states directly (like 🤗 or 💖).

Thus, the use of emojis in digital texts is not just a casual or aesthetic choice but is deeply tied to an individual's personality and communication style. By examining patterns in emoji use, we can gain deeper insights into how people of different MBTI types express themselves and interact with others in a digital context. This understanding not only enhances interpersonal communication but also has potential applications in areas like personalized marketing, mental health assessments, and the development of AI-driven communication tools that are sensitive to individual personality differences.

5. Methodology

As we embark on detailing the methodology of our study, we adopt the OODA (Observe, Orient, Decide, Act) Loop framework, a systematic approach guiding our data collection, preparation, model training, evaluation, decision-making, and iterative refinement, ensuring a comprehensive and rigorous analysis. This process is integral in ensuring the robustness and accuracy of our analyses.

5.1. Data Collection and Preparation

- *Observation (Observe)*: Data was collected from various sources, including social media platforms and online forums, to observe patterns in textual expressions and emoji usage related to personality

traits.

- *Orientation (Orient)*: Relevant features were extracted from the observed data, such as text content and emoji sequences, to create inputs for the AI models.

5.2. Model Training and Evaluation

- *Observation (Observe)*: Patterns and trends within the data were observed to inform the training of AI models for personality analysis.
- *Orientation (Orient)*: The performance of AI models was evaluated using metrics such as accuracy, F1 scores, and Rouge scores to orient decisions regarding model selection and optimization.

5.3. Model Selection and Optimization

- *Decision (Decide)*: The most suitable AI models for personality analysis were selected based on their performance metrics, computational efficiency, and interpretability.
- *Action (Act)*: Hyperparameters of the selected models were optimized through techniques such as hyperparameter tuning to improve their performance on personality classification tasks.

5.4. Model Deployment and Iterative Improvement

- *Decision (Decide)*: Deployed models were monitored for performance, and feedback was collected to inform iterative improvements.
- *Action (Act)*: Iterative refinement of models was conducted based on new data or insights gained from their usage to enhance their accuracy and applicability.

Figure 1 illustrates the OODA Loop framework, used for this study.

The following sections probe the specific details, shedding light on each phase of our methodology and explicating the systematic approach we employ to conduct our study.

6. Emo-MBTI Dataset

The Emo-MBTI Dataset is a specialized resource designed to bridge the gap between digital communication, represented by emoji usage, and psychological profiling through the Myers-Briggs Type Indicator (MBTI). This dataset innovatively combines natural language processing and psychological analysis to explore the intersection of emoji usage and personality traits.

6.1. Core Composition of the Emo-MBTI Dataset

The primary foundation of the Emo-MBTI Dataset is derived from the MBTI Dataset, which includes data from over 8600 individuals. These data snippets, taken from the last 50 posts of each individual on Reddit, are categorized into one of the 16 distinct MBTI personality types. Each post is segmented using

"|||" (three pipe characters) to demarcate individual entries, facilitating a structured analysis of textual content related to distinct personality profiles as shown in Fig. 2.

6.2. Data Enrichment through Advanced Scraping

To enrich the foundational MBTI dataset, we employed advanced web scraping techniques targeting the r/mbti subreddit on Reddit. Our focus was on users who frequently use emojis in their posts, a criterion that suggests heightened emotional expression. Specifically, we collected the last 50 posts from users who utilized more than three emojis per post, using these emojis as contextual labels for the text. Using the Python Reddit API Wrapper (PRAW), we meticulously extracted data from users on the r/mbti subreddit. The initial scraping yielded 6280 data points, which were then filtered focus on users demonstrating consistent engagement:

- 6029 data points included users with more than 10 comments.
- 5527 data points were further refined to include only those with more than 50 comments, ensuring a robust representation of active community members.

The snapshot of this enriched dataset is shown in Fig. 3

This process not only expanded our dataset but also ensured the relevance and depth of the data collected, focusing on users with significant subreddit participation.

6.3. Integration of Advanced NLP Models for Emoji Mapping

Post-collection, the data underwent processing where a sophisticated model was trained to map text to emojis effectively. This model leverages the nuances of sequence-to-sequence generation, using generative AI models, specifically Large Language Models (LLMs), known for their capability in handling complex language tasks, For the task of mapping text to emojis, three sophisticated models were utilized: FlanT5, Pegasus, and BART, each chosen for their unique strengths in summarization and contextual understanding.

- **FlanT5**: This transformer-based model excels at generating concise, contextually accurate summaries by deeply understanding the intricacies within extensive texts. Its proficiency in handling complex narratives makes it highly effective for detailed summarization tasks.
- **Pegasus**: Renowned for its abstractive summarization capabilities, Pegasus employs a unique pre-training strategy tailored specifically for summarization. This approach enables it to distill extensive text into essence-focused summaries, adeptly capturing the core messages.
- **BART**: Utilizing a denoising autoencoder architecture, BART is designed to reconstruct and simplify text, making it particularly suitable for summarizing complex content into more digestible forms, including the translation of text sentiments and themes into corresponding emojis.

These models were chosen for their ability to efficiently process and condense textual information, facilitating accurate and expressive emoji mappings that reflect the nuanced emotional and thematic undertones of the original content. The dataset after mapping is as shown in Fig. 4:

6.4. Dataset Analysis and Bias Mitigation

We conducted a thorough statistical analysis of the emoji usage across different MBTI personality types within our dataset. This analysis revealed significant variations in emoji usage patterns. For example, we found that extroverted personality types tend to use a wider range of emojis more frequently compared to introverted types. These insights are critical as they help identify inherent dataset imbalances that could influence the model's performance.

To address the identified biases, we implemented synthetic minority oversampling techniques (SMOTE) to improve the representation of underrepresented personality types in our dataset. This method enhances the dataset's balance by artificially generating new instances for less prevalent classes, thereby equalizing the training conditions across all personality types. Such an approach significantly diminishes the model’s inclination to favour classes that are more commonly represented, ensuring a fairer and more accurate analysis across the spectrum of personality traits.

6.5. Final Dataset and Utility: The final Emo-MBTI Dataset, as visualized in the snapshots, features posts labelled with emojis that correlate with the inferred MBTI types based on the textual analysis provided by the models. This dataset not only facilitates the exploration of how different personality types might prefer certain types of emojis but also serves as a valuable tool for psychological research, digital behaviour analysis, and the development of AI-driven tools that can understand and predict human emotional and psychological states through digital footprints.

7. EmoMBTI-Net: The Emoji-Based MBTI Personality Prediction Model

In this study, our goal is to predict MBTI personality types based on the usage of emojis in textual communications. As introduced earlier, the integration of emojis within natural language processing frameworks offers a novel approach to understanding complex human behaviours and personality traits. The rationale behind this methodology lies in the richness of emojis as expressive tools that convey emotional and contextual nuances beyond the capacity of plain text. Emojis encapsulate a spectrum of emotions and sentiments, making them valuable in the psychological profiling inherent to personality assessment.

The principal challenge we faced was selecting appropriate models equipped to process emojis as fundamental components of language. Traditional models like BERT [24], DistilBert[25], and ALBERT[26], while robust in general language tasks, do not inherently support emoji interpretation within their tokenization processes. This limitation necessitates the exploration of alternative models that natively incorporate emoji understanding to ensure the accuracy of personality predictions. Consequently, our approach has focused on identifying and leveraging transformer-based models that are inherently

capable of processing and interpreting complex, emoji-rich text. These models, such as NLI-RoBERTa, BART, and NLI-DeBERTa, offer sophisticated mechanisms for embedding contextual and emotional nuances, which are critical in capturing the subtleties required for accurate personality profiling from textual communications. Embracing these advanced technologies allows us to explore the depths of emoji semantics, translating seemingly simplistic symbols into profound insights into human psychology and behaviour, thereby enhancing the predictive capabilities of our system. This shift towards models adapted for emoji interpretation not only addresses the shortcomings of conventional NLP tools but also sets a new standard in the nuanced analysis of digital communication.

7.1 NLI-RoBERTa

NLI-RoBERTa, an extension of the RoBERTa[3] (Robustly optimized BERT approach) model, is designed for Natural Language Inference (NLI) tasks. The model employs a transformer architecture known for its efficiency in grasping contextual relationships within text. NLI-RoBERTa is pre-trained on extensive text datasets using a self-supervised learning paradigm, enabling it to capture complex semantic subtleties essential for discerning between entailment, contradiction, and neutrality in texts. Such capabilities make NLI-RoBERTa highly effective for sequence classification, including MBTI personality prediction, where understanding the layered meanings conveyed by emojis is crucial. The strength of NLI-RoBERTa lies in its transformer-based design, incorporating multiple layers of self-attention mechanisms and feedforward neural networks. This architecture is adept at learning contextual embeddings from vast amounts of data, crucial for emoji-based text interpretation. By eliminating the next sentence prediction task and adopting dynamic masking, the model enhances its focus on relevant textual segments, crucial for accurate personality assessment.

7.2. BART

BART is versatile in handling both sequence-to-sequence and sequence classification challenges. Its architecture combines a bidirectional transformer encoder with an auto-regressive decoder, forming a denoising autoencoder that excels in reconstructing noisy inputs to predict original sequences. This innovative training approach enables BART to identify and retain critical textual features, including emojis, making it suitable for detailed personality analysis. The bidirectional encoder ensures comprehensive understanding of context, capturing dependencies in text from both directions, which is vital for interpreting the emotional and contextual layers conveyed by emojis. The auto-regressive decoder enhances the model's ability to generate coherent outputs that are contextually aligned with the input, thereby supporting sophisticated sequence classification tasks like MBTI personality prediction.

7.3 NLI-DeBERTa

NLI-DeBERTa refines the DeBERTa[4] model with a focus on Natural Language Inference. Its standout feature is the disentangled attention mechanism, which optimizes focus on different segments of input text independently, crucial for parsing complex emoji-laden communications. This ability to focus selectively on relevant parts of the text facilitates a deeper understanding of long-range dependencies

and intricate semantic relationships, essential for accurate personality profiling. The specialized architecture of NLI-DeBERTa, with its nuanced attention to detail in context modelling, makes it an exceptional choice for tasks requiring an advanced grasp of subtle textual interactions. Its efficacy in MBTI classification is enhanced by its capacity to disentangle attention weights during pre-training, thus improving the model's overall ability to interpret and analyze emoji-based communication in the context of personality assessment.

The following Fig. 5 outlines a sophisticated model for processing and predicting personality traits based on textual input, particularly focusing on the Hindi language.

7.4. Workflow Summary

The Emo-MBTI Dataset comprises posts labelled with emojis that reflect inferred MBTI types derived from textual analysis by the transformer models. The dataset is processed independently by each model: NLI-RoBERTa leverages its natural language inference capabilities to deduce personality traits from text and emojis; BART re-encodes or transforms text to emphasize emotional or stylistic features for personality prediction; and NLI-DeBERTa uses its decoding-enhanced attention mechanisms for a deeper analysis of emoji-text interaction. Each model undergoes training, tuning, and testing on separate data splits to ensure comparable performance metrics. The outputs from each model are not only the predicted MBTI types but also insights into which emojis correspond to specific personality traits. These results are thoroughly compared to evaluate accuracy, precision, recall, and F1-scores, providing a basis for understanding which model best handles the nuances of emoji usage in textual communications and why.

8. Results and Discussion

This section presents the results from our comprehensive evaluation, illuminating the performance of diverse language models across both emoji mapping and MBTI personality classification tasks. As the first research initiative using this dataset, we lack baseline comparisons, making this study foundational in setting initial benchmarks for future explorations. The application of meticulous analysis using Rouge scores and F1 metrics has revealed clear patterns of strengths and weaknesses, offering valuable insights into the nuanced interpretation of textual expressions and the efficacy of different models in capturing the intricacies of human communication.

8.1. Contextual Connection

As we set out to explore the predictive capabilities of large language models in interpreting emojis within digital communications, the results presented here underscore our objective to enhance personality profiling through textual and emoji analysis. These findings not only affirm our hypotheses but also pave the way for innovative applications in computational linguistics.

8.2. Performance Across Models

The Rouge scores, often used as a measure of similarity between the expected and the model-generated outcomes, were utilized here to gauge the efficacy of each model in emoji mapping, as shown in Table 1.

Table 1
Emoji mapping Rouge scores using LLMs

Model	Final Training Loss (Rouge Score)
Flan-T5	44.1280
Pegasus	11.5288
BART	0.0250

The rouge scores reveal several key points about the performance of these models:

- **Higher Efficiency in Learning Representations**: BART's exceptionally low rouge score of 0.0250 suggests a superior ability to capture and reproduce the nuances required for emoji mapping. This implies that BART's training process effectively minimized the loss, leading to more accurate emoji predictions.
- **Comparison with Other Models**: Flan-T5, despite being a powerful and versatile model, showed a much higher final training loss in this task. This indicates that while Flan-T5 is generally effective across various tasks, it may not be as optimized as BART for the specific nuances of emoji mapping. Pegasus, known for its summarization capabilities, performed better than Flan-T5 but still fell short compared to BART. Its middle-range score suggests it has moderate capabilities in emoji mapping, potentially due to its underlying architecture which may prioritize certain aspects of language understanding differently.
- **Choice for MBTI Classification**: Given BART's outstanding performance in minimizing training loss, it was the logical choice for implementing the final emoji mappings for MBTI personality classification. The low rouge score indicates a high level of precision in BART’s emoji mappings, which is crucial for accurately capturing personality traits through emojis.

This graph in Fig. 6 illustrates the final training losses (Rouge Scores) for Flan-T5, Pegasus, and BART models. BART demonstrates a significantly lower Rouge Score, indicating a superior ability to capture the nuances required for emoji mapping due to minimized loss during training.

Overall, BART consistently outperformed NLI-Roberta and NLI-DeBERTa, achieving the highest overall accuracy of 86.160% as shown in table

Table 2
Accuracy of different models

Accuracy	NLI-Roberta	NLI-DeBERTa	BART
Overall	0.78571	0.80654	0.8616
Introversion	0.84824	0.85504	0.88497
Extroversion	0.63736	0.63636	0.76331
Intuition	0.92096	0.92359	0.95000
Sensing	0.14035	0.18181	0.19318
Thinking	0.73437	0.76948	0.83742
Feeling	0.84308	0.83631	0.87172
Judging	0.73818	0.76351	0.81111
Perceiving	0.82897	0.83870	0.88356

Notably, BART displayed notable strengths in discerning traits related to introversion, intuition, and perceiving. However, the models demonstrated varying degrees of success in understanding traits such as extroversion and sensing.

The F1 scores presented in Table 3 for the models NLI-Roberta, NLI-DeBERTa, and BART provide an in-depth look at their performance across various MBTI personality traits. These scores are crucial for understanding how well each model balances precision and recall in their classifications, which is particularly important in personality analysis where both false positives and false negatives can lead to significant misinterpretations.

Table 3
F1- scores of different models

F-1	NLI-Roberta	NLI-DeBERTa	BART
Overall	0.786	0.807	0.862
Introversion	0.848	0.855	0.885
Extroversion	0.637	0.636	0.763
Intuition	0.921	0.924	0.950
Sensing	0.140	0.182	0.193
Thinking	0.734	0.769	0.837
Feeling	0.843	0.836	0.872
Judging	0.738	0.764	0.811
Perceiving	0.829	0.839	0.884

Here's a detailed discussion of the F1 scores:

Overall Performance

- **BART** shows the highest overall F1 score at **0.862**, indicating its superior general performance in classifying MBTI personalities.
- **NLI-DeBERTa** follows with **0.807**, and **NLI-Roberta** has an F1 of **0.786**. This suggests that while all models are relatively effective, BART excels in balancing precision and recall most effectively.

The graph in Fig. 7 showcases the accuracy and F1 scores for MBTI personality classification using EmoMBTI-Net, focusing on how well each model performs in this specific context.

Figure 8 presents the Radar Chart for model performance on MBTI Traits. Each spoke of the radar chart represents a different personality trait, and each model's performance on these traits is plotted as a separate line on the chart. BART (in red) shows superior performance in most traits compared to NLI-Roberta (in blue) and NLI-DeBERTa (in green), particularly noticeable in Extroversion and Intuition.

Trait-Specific Performance

- **Introversion and Extroversion**:
    - BART outperforms in both traits, with particularly strong gains in Extroversion (**0.763**), which is notably higher than the scores for NLI-Roberta (**0.637**) and NLI-DeBERTa (**0.636**). This indicates BART's better capability in detecting features related to extroverted behaviour through text.
- **Intuition**:

    - All models perform well with Intuition, with BART leading at **0.950**. This suggests that the models are effective at picking up on the abstract and theoretical content often associated with intuitive personalities.
- **Sensing**:
    - All models struggle with Sensing, with F1 scores remaining below **0.200**. This is likely due to the difficulty of capturing concrete and detail-oriented expressions through text, which are less prevalent or explicitly marked compared to intuitive expressions.
- **Thinking and Feeling**:
    - BART again leads in these traits, with Thinking at **0.837** and Feeling at **0.872**. These traits involve logical versus empathetic content, indicating BART's stronger ability to distinguish between structured argumentation and emotional expressions.
- **Judging and Perceiving**:
    - BART performs best in both Judging (**0.811**) and Perceiving (**0.884**), showing its effectiveness in identifying structured versus spontaneous expressions in text.

The results indicate BART's robustness in nuanced text interpretation, making it particularly suited for tasks requiring deep semantic understanding, such as personality assessment through text analysis. The superior F1 scores for BART across most traits suggest that its training and underlying model architecture may be more attuned to the subtleties of personality-expressive language.

The lower scores in Sensing across all models highlight a common challenge in natural language processing—capturing concrete and sensory-related information, which tends to be less explicitly expressed. This suggests a potential area for further model refinement and training, possibly by incorporating more sensory-specific training data or employing techniques that better capture these aspects.

In conclusion, the F1 scores reveal the strengths and weaknesses of each model in handling different aspects of personality classification. BART's leading performance across multiple traits recommends it as the preferable model for applications in personality analysis, offering more reliable and nuanced insights into individual personality types based on their textual expressions.

The choice of pre-trained language model and hyperparameters significantly influenced the performance of both emoji mapping and MBTI classification tasks. For emoji mapping, the hyperparameters such as learning rate and epochs were tuned for each LLM, optimizing their performance. Similarly, in MBTI classification, different learning rates and epochs were applied to NLI-Roberta, NLI-DeBERTa, and BART as shown in Table 4. An iterative tuning process was crucial in selecting appropriate learning rates and epochs, which were dynamically adjusted based on model performance during validation to prevent overfitting while maximizing accuracy.

Table 4
Hyperparameters used

Model	Learning Rate	Epochs
Emoji Mapping		
Flan-T5	5e-5	3
Pegasus	3e-4	3
BART	1e-5	3
MBTI Classification		
NLI-Roberta	3e-5	10
NLI-DeBERTa	3e-6	10
BART	1e-5	10

Overall, the study highlights the effectiveness of large language models in leveraging emojis for personality detection tasks, with BART demonstrating superior performance in both emoji mapping and MBTI classification. The findings underscore the importance of model selection and hyperparameter tuning in optimizing the performance of such tasks and provide valuable insights into the nuances of interpreting personality traits conveyed through emojis.

8.3. Limitations and Ethical Considerations

While our study provides valuable insights into the performance of language models in emoji mapping and MBTI personality classification tasks, several limitations should be acknowledged. Firstly, the size and diversity of the dataset used for training and evaluation may impact the generalizability of our findings. Future research could focus on collecting larger and more diverse datasets, encompassing a wider range of linguistic expressions and cultural backgrounds. Additionally, domain-specific biases inherent in the training data could influence the models' predictions and performance. To address this limitation, researchers could explore techniques for mitigating bias in language models, such as adversarial training, data augmentation, or fairness-aware learning algorithms. Moreover, fine-tuning models on domain-specific tasks, such as personality analysis, could enhance their effectiveness in capturing nuanced textual expressions relevant to the target domain. Furthermore, computational resources, including hardware limitations and training time constraints, may restrict the scalability and efficiency of language model training and evaluation. Future research could investigate strategies for optimizing model architectures, hyperparameters, and training procedures to maximize performance while minimizing computational overhead. In terms of evaluation metrics, while Rouge scores and F1 metrics provide valuable insights into the performance of language models, exploring novel evaluation metrics tailored to personality classification tasks could offer a more comprehensive assessment of model performance. For instance, sentiment analysis metrics, semantic similarity measures, or

psychological validity assessments could provide additional perspectives on the accuracy and effectiveness of personality classification algorithms.

As with any application of artificial intelligence, there are ethical considerations associated with using language models for personality analysis. Privacy concerns arise from the potential collection and analysis of sensitive textual data, raising questions about data ownership, consent, and confidentiality. Moreover, there is a risk of biases in model predictions, stemming from biases inherent in the training data or algorithmic decision-making processes. Biased predictions could have adverse consequences, particularly in sensitive domains like mental health or hiring, where decisions based on personality assessments may impact individuals' opportunities and well-being. To address these ethical considerations, it is essential to prioritize responsible deployment of AI technologies, ensuring transparency, fairness, and accountability throughout the development and deployment lifecycle. This includes implementing robust data governance practices, conducting bias audits, and providing mechanisms for recourse and redress in case of algorithmic harm. Furthermore, ethical guidelines and regulatory frameworks should be established to govern the use of language models for personality analysis, protecting individuals' privacy rights, promoting algorithmic transparency, and mitigating the risk of discriminatory outcomes. By adhering to ethical principles and best practices, developers can harness the potential of language models for personality analysis while minimizing potential risks and safeguarding societal well-being.

8.4. Practical Implications

The findings of our study have significant practical implications for real-world applications, particularly in the development of personalized recommendation systems, chatbots, and mental health assessment tools. By leveraging insights gained from analyzing personality traits through text, developers can enhance user experiences, tailor content recommendations, and provide more personalized interactions. For personalized recommendation systems, understanding users' personality traits can facilitate the delivery of content and product recommendations tailored to their individual preferences and characteristics. By integrating personality-based profiling into recommendation algorithms, platforms can enhance user engagement, satisfaction, and retention rates.

Similarly, in the context of chatbots and virtual assistants, incorporating personality-aware dialogue generation techniques can enhance the conversational experience and foster more meaningful interactions. Chatbots capable of adapting their language and tone to match users' personality traits can improve communication effectiveness and user satisfaction. Furthermore, in the field of mental health assessment and therapy, analyzing personality traits through text can provide valuable insights for psychological research and clinical practice. By identifying linguistic markers associated with various personality dimensions, clinicians can gain a deeper understanding of patients' psychological profiles and tailor interventions accordingly.

9. Conclusion and Future Work

The study unveils the distinct performances of DeBERTa and BART models in emoji-based MBTI personality trait classification. BART consistently emerges as the most accurate model overall, showcasing particular proficiency in discerning traits associated with introversion, intuition, and perceiving. However, capturing traits such as extroversion and sensing remains a challenge, with NLI-DeBERTa showing relative improvement over NLI-Roberta. These findings underscore the importance of selecting models based on their strengths and weaknesses for specific applications. The emoji-based approach to personality classification is promising and offers a novel method for understanding human characteristics. Nevertheless, further research is necessary to enhance accuracy and applicability across diverse datasets. Addressing challenges such as potential biases, data cleaning, and privacy considerations is crucial for maintaining methodological rigor and ethical research practices.

Firstly, systematic ablation studies will be essential to dissect the influence of specific model components on performance. This will allow us to refine our understanding of feature importance and optimize model configurations to prevent overfitting. Additionally, expanding our research to incorporate additional datasets, possibly through synthetic augmentation or adapting existing datasets from related fields, will test the generalizability of our models across diverse data environments.

We also plan to explore multi-modal learning techniques that integrate textual content with emojis to better capture the nuanced interplay of communication elements. This could significantly improve the accuracy of personality predictions by providing a more comprehensive dataset for model training.

Further, a temporal analysis of emoji usage patterns over time could provide insights into dynamic personality expressions, while collaborations with psychologists to validate predictions against established psychological assessments would enhance the credibility of our results. Finally, extending the application of our models to other digital platforms could uncover broader insights into how emojis reflect personality traits across different communication styles and user demographics. This expansion would not only validate our findings in varied contexts but also broaden the applicability of our research, contributing to the fields of digital communication and psychological profiling.

Declarations

Author Contribution

"A.K. conceptualized the study, designed the methodology, led the data curation process, and wrote the original draft of the manuscript. A.K. also supervised the project and handled the project administration." "D.J. was responsible for the development and implementation of the computational models, conducted the data analysis, and prepared all figures and tables for the manuscript. D.J. also contributed to the literature review and assisted in writing the manuscript." Both authors A.K. and D.J. contributed to the final version of the manuscript.

References

1. Keh SS, Cheng I (2019) Myers-Briggs Personality Classification and Personality Specific Language Generation Using Pre-trained Language Models. ArXiv, abs/1907.06333
2. Hernández Y, Peña CA, Martínez A (2018) Model for Personality Detection Based on Text Analysis. In: Batyrshin I, Martínez-Villaseñor M, Ponce Espinosa H (eds) Advances in Computational Intelligence. MICAI 2018. Lecture Notes in Computer Science, vol 11289. Springer, Cham. https://doi.org/10.1007/978-3-030-04497-8_17
3. Cui B, Qi C (2017) Survey Analysis of Machine Learning Methods for Natural Language Processing for MBTI Personality Type Prediction
4. Hernandez RK, Scott I (2017) Predicting Myers-Briggs type indicator with text
5. Ismail S, Bashari Rad B, Ismail S (2017) Significant of MBTI personality model on decision making in university program selection. 10.1109/ICITISEE.2017.8285560
6. Ren Z, Shen Q, Diao X, Xu H (2021) A sentiment-aware deep learning approach for personality detection from text. Inf Process Manag 58(3):102532
7. Personality BERT A Transformer-Based Model for Personality Detection from Textual Data. In: Bashir, A.K., Fortino, G., Khanna, A., Gupta, D. (eds) Proceedings of International Conference on Computing and Communication Networks. Lecture Notes in Networks and Systems, vol 394. Springer, Singapore. https://doi.org/10.1007/978-981-19-0604-6_48
8. Kumar A, Beniwal R, Jain D (2023) Personality Detection using Kernel-based Ensemble Model for leveraging Social Psychology in Online Networks ACM Transactions on Asian and Low-. https://doi.org/10.1145/3571584. Resource Language Information Processing (ACM TALLIP)-
9. Liusie A, Manakul P, Gales M (2024), March LLM comparative assessment: Zero-shot NLG evaluation through pairwise comparisons using large language models. In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 139–151)
10. Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O, Zettlemoyer L (2019) Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461
11. Liao W, Zeng B, Yin X, Wei P (2021) An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa. Appl Intell 51:3522–3533
12. He P, Liu X, Gao J, Chen W (2020) Deberta: Decoding-enhanced bert with disentangled attention. arXiv preprint arXiv:2006.03654
13. Gürpınar F, Kaya H, Salah AA (2016) Combining deep facial and ambient features for first impression estimation. In Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8–10 and 15–16, 2016, Proceedings, Part III 14 (pp. 372–385). Springer International Publishing
14. Gürpinar F, Kaya H, Salah AA (2016), December Multimodal fusion of audio, scene, and face features for first impression estimation. In 2016 23rd International conference on pattern recognition (ICPR) (pp. 43–48). IEEE

15. Rai N (2016) Bi-modal regression for Apparent Personality trait Recognition. 2016 23rd International Conference on Pattern Recognition (ICPR). 10.1109/icpr.2016.7899607
16. Zhang CL, Zhang H, Wei XS, Wu J (2016), October Deep bimodal regression for apparent personality analysis. In European conference on computer vision (pp. 311–324). Cham: Springer International Publishing
17. Kaya H, Gurpinar F, Ali Salah A (2017) Multi-modal score fusion and decision trees for explainable automatic job candidate screening from video cvs. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 1–9)
18. Gucluturk Y, Guclu U, Perez M, Escalante J, Baro H, Guyon X, Van Lier I (2017) R. Visualizing apparent personality analysis with deep residual networks. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 3101–3109)
19. Junior JCJ, Güçlütürk Y, Pérez M, Güçlü U, Andujar C, Baró X, Escalera S (2019) First impressions: A survey on vision-based apparent personality trait analysis. IEEE Trans Affect Comput 13(1):75–95
20. Aslan S (2019) Multimodal video-based personality recognition using Long Short-Term Memory and convolutional neural networks. Computer Vision and Pattern Recognition, arXiv
21. Artha Agastya IM, Dwi Handayani DO, Mantoro T (2019) A Systematic Literature Review of Deep Learning Algorithms for Personality Trait Recognition, 2019 5th International Conference on Computing Engineering and Design (ICCED), pp. 1–6, 10.1109/ICCED46541.2019.9161107
22. Liao R, Song S, Gunes H An Open-source Benchmark of Deep Learning Models for Audio-visual Apparent and Self-reported Personality Recognition, in IEEE Transactions on Affective Computing, 10.1109/TAFFC.2024.3363710
23. Saeidi S (2024) Identifying personality traits of WhatsApp users based on frequently used emojis using deep learning. Multimedia Tools Appl 83(5):13873–13886
24. Lin J, Liu Y, Zeng Q, Jiang M, Cleland-Huang J (2021), May Traceability transformed: Generating more accurate links with pre-trained bert models. In 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE) (pp. 324–335). IEEE
25. Tamburini F, Cimiano P, Preite S Deep Question Answering: A New Teacher For DistilBERT
26. Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, Soricut R (2019) Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942

Figures

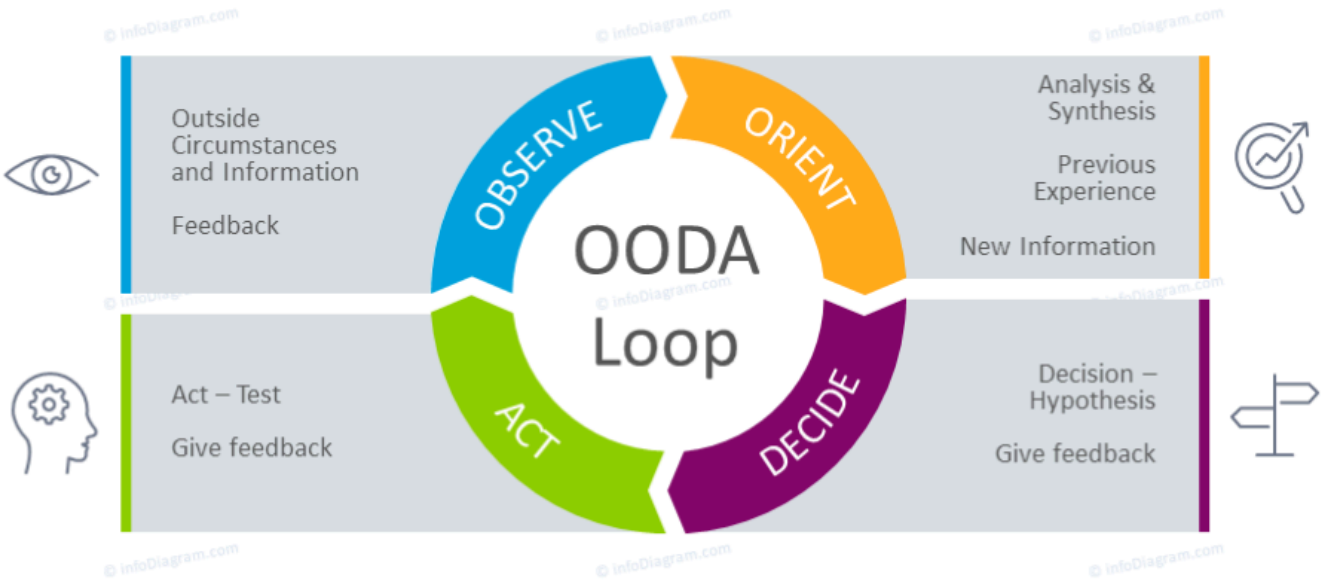


Figure 1

OODA loop framework

	type	posts
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw|||...
1	ENTP	'I'm finding the lack of me in these posts ver...
2	INTP	'Good one _____ https://www.youtube.com/wat...
3	INTJ	'Dear INTP, I enjoyed our conversation the o...
4	ENTJ	'You're fired.|||That's another silly misconce...
5	INTJ	'18/37 @.@|||Science is not perfect. No scien...
6	INFJ	'No, I can't draw on my own nails (haha). Thos...
7	INTJ	'I tend to build up a collection of things on ...
8	INFJ	I'm not sure, that's a good question. The dist...
9	INTP	'https://www.youtube.com/watch?v=w8-egj0y8Qs||...

Figure 2

Snapshot of MBTI Dataset

index	label	text
0	🎭😳👹	While currently on trial for fraud ||| The Sau...
1	🚗🦾🎩	I love how all these analysts and fans contin...
2	🪶📸😭	I immediately burst into tears ||| Look into ...
3	💩🦃👶	No ||| sorry ||| I had b rob too it feels b...
4	🥚🍊👀	Thanks man ||| I may hit you up I'm in SEPA ...
5	🎆🌄💍	Rock vs Austin with Mankind as special ref |||...
6	🧀🥤🍟	That makes total sense ||| Lolz ||| the ide...
7	🧍🔪🖌️	Heidi Gardner is *FORTY* ||| Revoke Stefan Kra...
8	💔🕺🎛️	That still seems really low but I don't know w...
9	😈📙🐖	Some books I have used and like are* Suomen m...

Figure 3

MBTI posts relabelled with Emojis

Unnamed: 0	type	posts	emoji
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw|||...	👥💻💻💻🎉🎭👥
1	ENTP	'I'm finding the lack of me in these posts ver...	💏😄😂
2	INTP	'Good one _____ https://www.youtube.com/wat...	🙌❤️👥💔❤️😱
3	INTJ	'Dear INTP, I enjoyed our conversation the o...	🧠🌎🤔💀💀💀
4	ENTJ	'You're fired.|||That's another silly misconce...	😤💥❓👯✈️🌴
...	...	...	...
8670	ISFP	'https://www.youtube.com/watch?v=t8edHB_h908||...	😾❤️💻 Dutch 🇳🇱😄
8671	ENFP	'So...if this thread already exists someplace ...	💻🎛️💔👥🍦
8672	INTP	'So many questions when i do these things. I ...	🤔💭💰🐴🐴
8673	INFP	'I am very conflicted right now when it comes ...	😱💔👶🇳🇱👫🤦
8674	INFP	'It has been too long since I have been on per...	💻🤷🌎🌈😱❤️

Figure 4

Snapshot of the Emoji Labelled MBTI Dataset

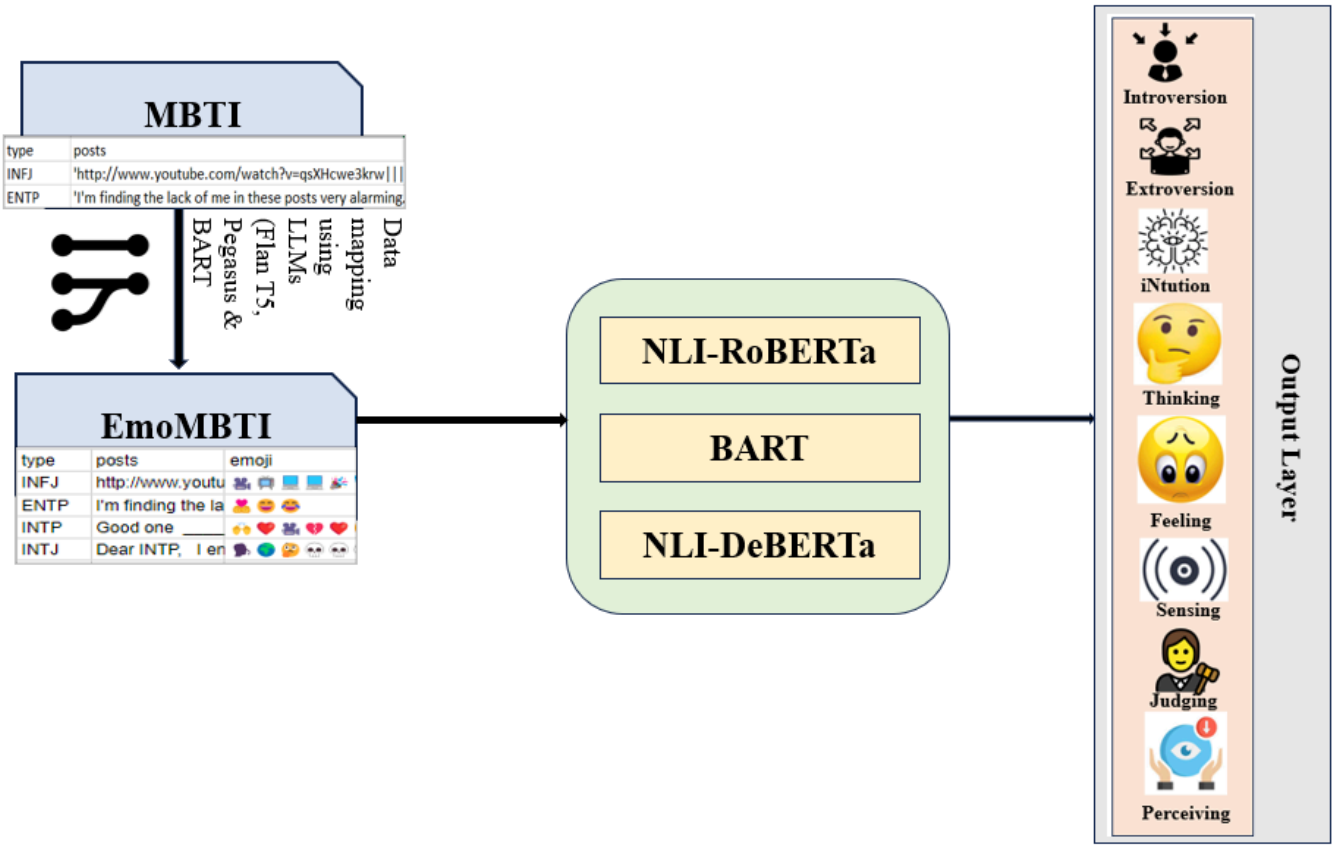


Figure 5

The MBTIEmoNet Model

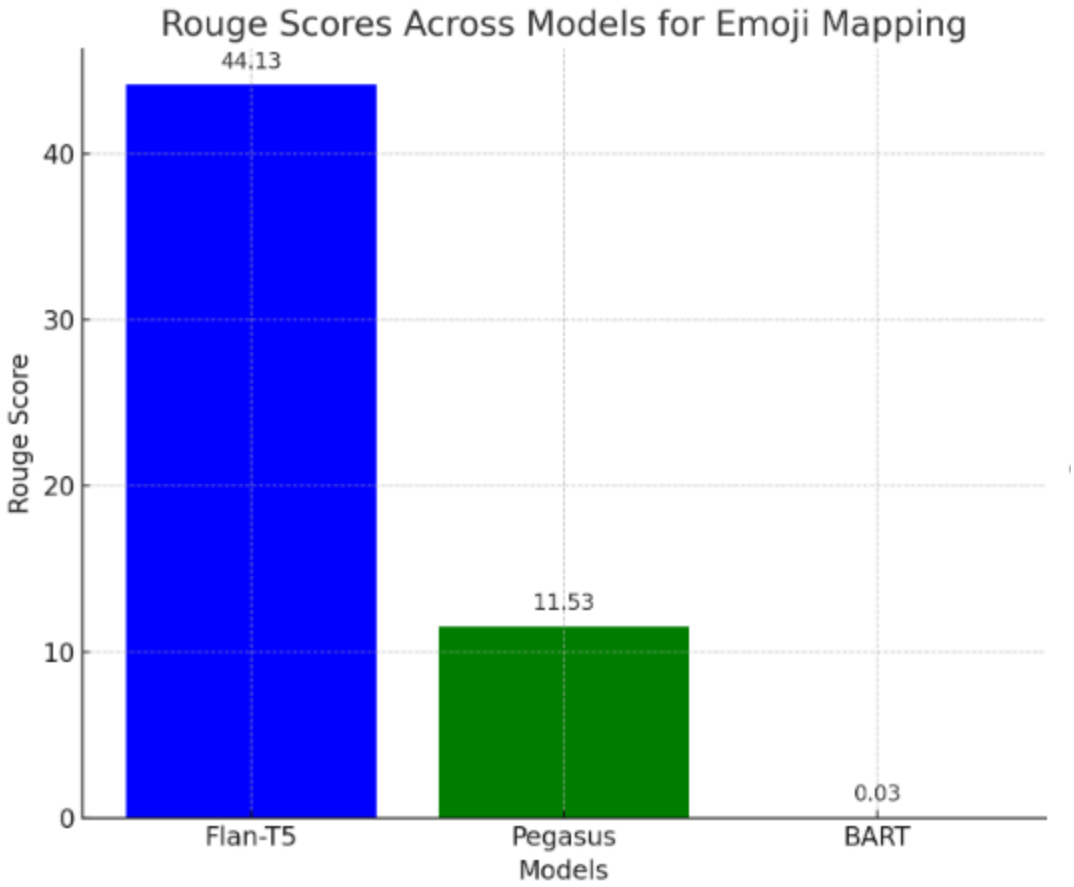


Figure 6

Rouge Scores Across Models for Emoji Mapping

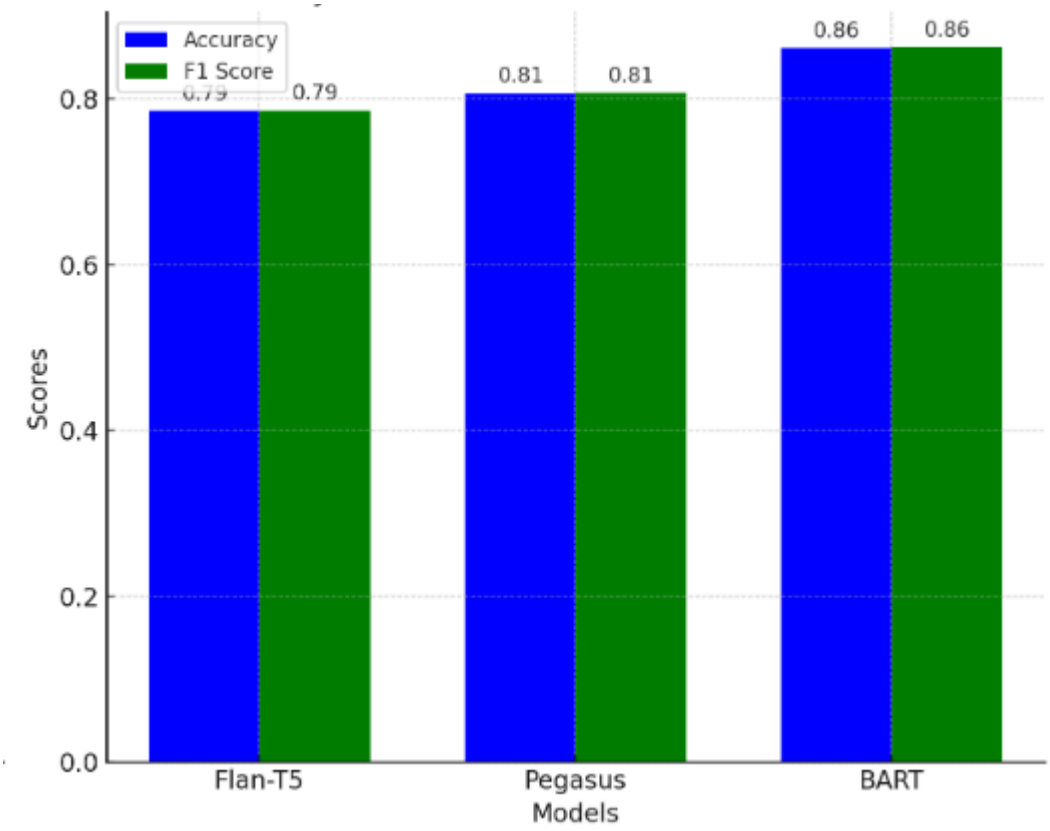


Figure 7

Overall Accuracy and F1 Scores Across Models in EmoMBTI-Net

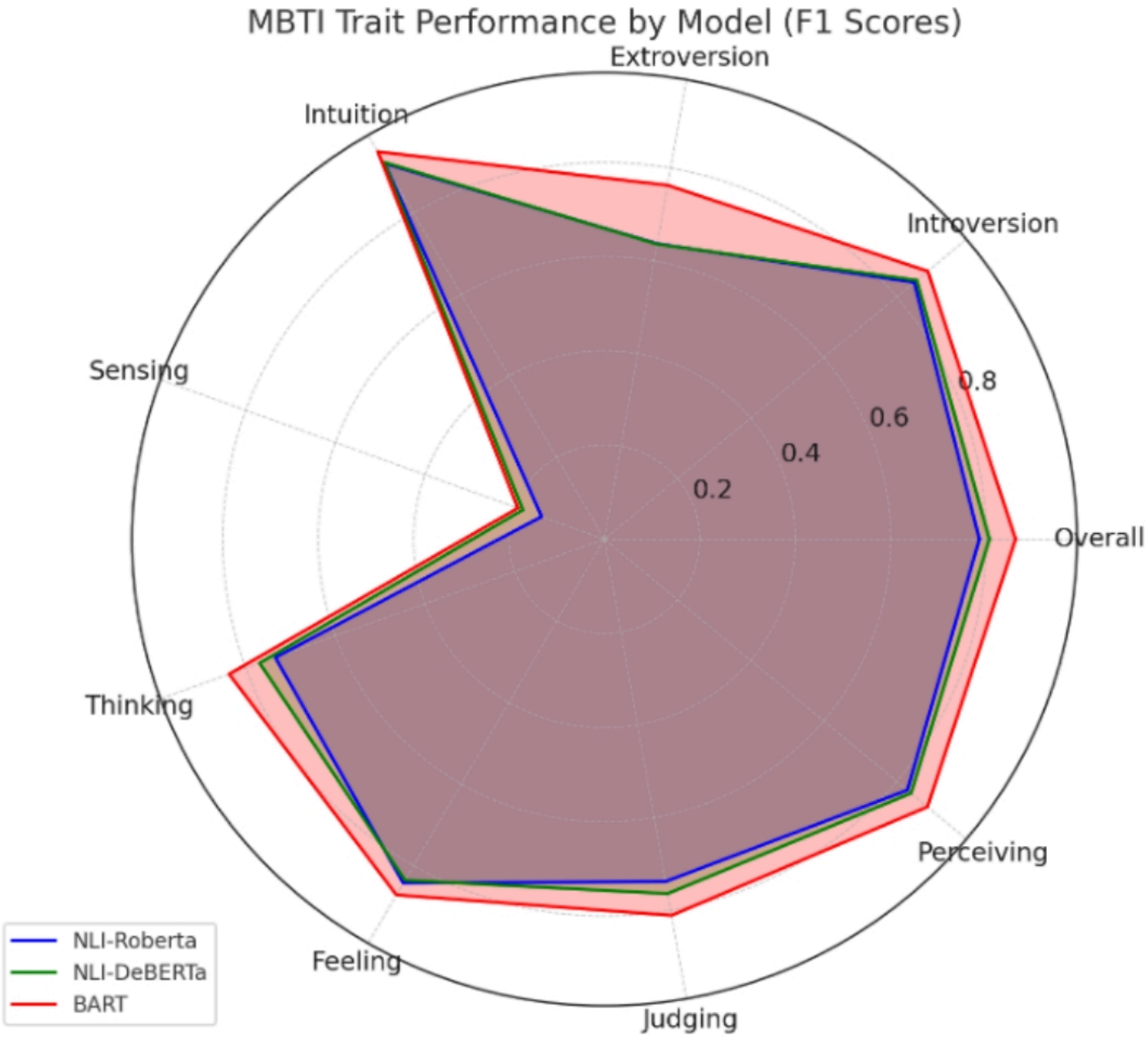


Figure 8

Radar Chart for Model Performance on MBTI Traits

Journal of Fluorescence
https://doi.org/10.1007/s10895-024-04066-0

RESEARCH

Energy Transfer Studies on Red Emitting $CaGdSbWO_8$: Sm^{3+}/Eu^{3+} Phosphor for w-LEDs and Indoor Plant Growth Lighting System

Jatin Parashar[1] · Gaurank Yadav[1] · Seema[1,2] · Anu[1] · Aman Prasad[3] · A. S. Rao[1]

Received: 24 September 2024 / Accepted: 4 December 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

In the present research work, a solid-state reaction method was employed to synthesize a series of $CaGd_{(0.95-x)}SbWO_8:Sm_{0.05}Eu_x$ (x = 1, 1.5, 2, 3 and 4 mol%) phosphors. The phase purity, crystallinity, morphological and compositional studies were analysed via X-ray diffraction (XRD), scanning electron microscopy (SEM) imaging, and energy dispersive spectroscopy (EDS) analysis. The Fourier transform infrared (FT-IR) spectroscopy was utilized to measure OH content in the prepared phosphor material. The diffused reflectance spectroscopy (DRS) analysis was applied to measure the optical energy band gap. The photoluminescence (PL) spectroscopy at excitation wavelengths 275 and 405 nm was carried out to analyze the luminescence properties of the prepared phosphors. The intensity of emission peaks of Sm^{3+} ions decreased as the concentration of Eu^{3+} ions increased showing the energy transfer mechanism from Sm^{3+} to Eu^{3+} ions. The dipole-dipole interaction mechanism is responsible for the energy transfer between Sm^{3+} to Eu^{3+} ions as assessed using the Reisfeld approximation. The CIE coordinates of the prepared samples lie in the warm red region represented by the relatively low (3200 K) value of correlated color temperature. These results show the suitability of the prepared phosphor for white light emitting diode (w-LEDs) and also in indoor plant growth lighting systems.

Keywords Phosphor · Spectroscopy · Photoluminescence (PL) emission · Energy transfer · CIE coordinates

Introduction

Reducing global CO_2 emissions is one of the main goals of the research that is going on all over the world for a greener future. About 6% of the global CO_2 emission is caused by the lighting technologies [1]. Energy-efficient alternatives for conventional lighting systems are required. LEDs are the best alternatives of traditional incandescent and fluorescence lamps because of their low cost and longer lifetime. An LED chip has a semiconductor layer that gives output according to their energy band gap when electrical energy is provided and phosphors are the materials that produce luminescence when irradiated by that energy output inside the LEDs. Nowadays, white light LEDs (w-LEDs) have found tremendous usage in various fields of lasers, displays, fiber optics, lighting, and other optoelectronic device applications due to their various advantages like low-cost effectiveness, high energy efficiency, durability and eco-friendliness. Commercially, w-LEDs are created using a blue LED chip and yellow phosphor. But, because of the absence of a red-light component, their correlated color temperature (CCT) is high (> 3200 K) and their color rendering index (CRI) is low [2, 3]. It is reported that cool light (CCT > 3200 K) reduces positive mood as compared to warm light [4]. Sometimes a blue/UV or n-UV LED chip can be used to excite the red, blue and green phosphor to generate white light [5]. As UV light has energy much higher than blue light, this method has higher luminous efficiency and CRI. Thus, an efficient UV/n-UV excited red phosphor is needed to be developed.

Light is an essential part of the plant life cycle. Plants absorb sunlight and convert light energy to chemical energy by the process called photosynthesis. This happens because plant pigments absorb different wavelengths of light. Not all

✉ A. S. Rao
drsrallam@gmail.com

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

[2] Department of Physics, Government College Nagina, Nuh, Haryana 122108, India

[3] Department of Physis and Computer Science, Dayalbagh Educational Institute (DEI), Deemed University, Dayalbagh, Agra 282005, India

Published online: 18 December 2024

wavelengths are absorbed by plant pigments, Chlorophyll A absorbs blue light (450 nm), chlorophyll B absorbs red light (620 nm), Phytochromes Pr and Pfr absorb deep red (660 nm) and far-red light (730 nm), respectively [6]. Red and deep red light affects plant growth throughout the plant life cycle [7]. Thus, a red-emitting phosphor-based lighting system is suitable for indoor plant lighting systems to enhance yield and energy conservation [8].

Rare Earth (RE) ions are commonly used for obtaining colors in the field of solid-state illuminations because of their unique optical properties, such as narrow band emission peaks due to the presence of many 4f energy levels. The host environment does not affect these energy levels because of the shielding effect [9]. Among the RE^{3+} ions, Sm^{3+} and Eu^{3+} ions are used to generate red light-emitting phosphor materials. Sm^{3+}ions activated phosphors emit reddish-orange light owning to its $^4G_{5/2}\rightarrow{}^6H_{k/2}$ (k = 5, 7, 9, 11) transitions while Eu^{3+}ions activated phosphors have bright emission in the red region owing to its $^5D_0\rightarrow{}^7F_J$ (J = 1, 2, 3, 4) transitions [10]. However, the complete n-UV region cannot be used for excitation in the case of Eu^{3+} ions because the absorption in this range is limited due to the forbidden transitions. This limitation can be overcome by adding Sm^{3+} ions which show absorption in the whole n-UV region, act as good sensitizers for Eu^{3+} ions and can have a very efficient energy transfer mechanism [11].

Tungstate, as a host material, is self-activating, has a high refractive index, chemical and thermal stability, and low phonon energy, which allows it to efficiently transfer energy from the host to luminescent centers [12]. It is reported that luminescent intensity improved with the addition of Lanthanides (Ln) in the Scheelite tungstate host [13].

The present study aims to develop a red-emitting phosphor by Sm^{3+} and Eu^{3+} doping in a tungstate host, and study the energy transfer process from host to sensitizer Sm^{3+} ions and then from Sm^{3+} to activator Eu^{3+} ions. In previous works, Y. Yang et al. synthesized $CaGdSbWO_8$:Sm^{3+}phosphor samples and optimized them at Sm0.05 concentration [14]. They studied its luminescent properties under the exciting wavelength of 405 nm. In the current research work, however, $CaGdSbWO_8$: Sm^{3+}/Eu^{3+} phosphors were prepared using a conventional solid-state synthesis route. Their structural and optical properties with energy transfer phenomena were studied for their potential utility in w-LEDs and indoor plant growth lighting systems applications.

Experimental

Synthesis

A series of $CaGd_{(0.95-x)}SbWO_8$:$Sm_{0.05}Eu_x$ (x = 1, 1.5, 2, 3, 4 mol%) (CGSW: S5Ex) was synthesized by solid-state reaction method in air. All the precursors $CaCO_3$, Gd_2O_3, Sb_2O_3, WO_3, Sm_2O_3, and Eu_2O_3 were weighed in stoichiometric ratio and thoroughly mixed. The mixture was manually ground with mortar and pestle to make it homogeneous and then, calcined at 1300˚C in an alumina crucible for 5 h. The obtained sample was then cooled naturally to room temperature and then reground uniformly to obtain the final product.

Characterization

The XRD patterns of the samples were measured using Bruker D-8 Model advance X-Ray diffractometer with x-rays of wavelength 1.546Å (Cu-Kα) between 2θ range of 20˚ to 60˚ to study the phase and crystalline nature of samples. The morphological and chemical compositional analysis was done by using a JEOL JSM-6490 scanning electron microscope (SEM) at 10 kV accelerating voltage. The IR spectrum was measured by Perkin Elmer frontier FT-IR spectrometer in the wave number range 400–4000 cm^{-1}. The Jasco V-770 spectrophotometer was used to measure the UV-Vis spectrum of the prepared sample. The excitation and emission data were recorded under the Xenon lamp with the help of a Jasco 8300 FP spectrofluorophotometer. The fluorescence decay curves were measured via Horiba PTI QM-8450-11using a xenon lamp. All measurements were performed at room temperature.

Results and Discussion

Structural and Morphological Study

The XRD patterns of the prepared un-doped CGSW and CGSW: S5Ex (x = 1, 1.5, 2, 3, 4 mol%) phosphors are shown in Fig. 1. The sharp peaks suggest that the prepared samples are crystalline. The diffraction peaks are closely matched with the standard card JCPDS no. 00-064-0231 which shows that the samples have a single-phase tetragonal structure in the space group $I_{41/a}$ (Space group no. 88) with lattice constants a = b = 5.2408Å and c = 11.2692Å. There are two additional small peaks at 30˚ and 50˚ which may be due to any impurity that might have crept in during preparation of the samples. However, the luminescent properties of the phosphors are not affected by such a small amount of impurity [15]. Despite the doping of Sm^{3+} and Eu^{3+} ions, no extra diffraction peaks or peak shifting is observed which shows that after doping, crystal structure of the host remains unchanged.

The average crystallite size is determined by considering major peaks using Debye-Scherrer formula [16]:

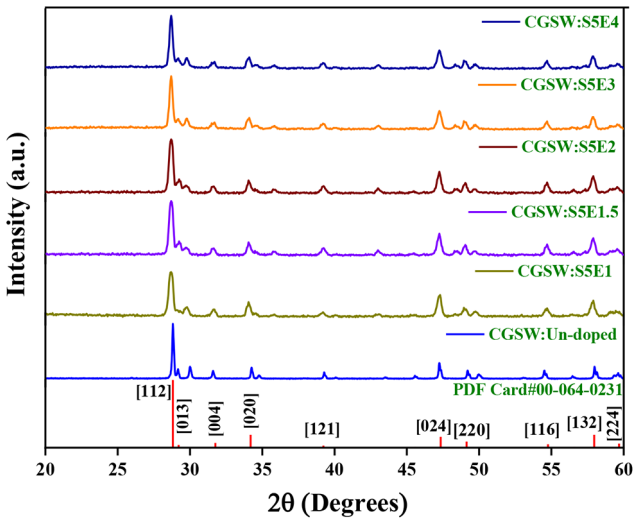


Fig. 1 XRD pattern of synthesized un-doped CGSW and CGSW: S5Ex (x = 1, 1.5, 2, 3, 4 mol%) phosphors along with the standard JCPDS data

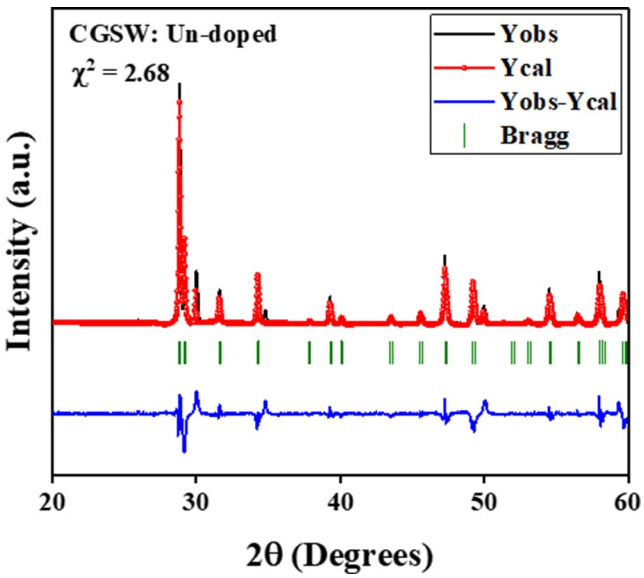


Fig. 2 The Rietveld refinement map for an un-doped CGSW sample

$$\beta \cos\theta = \frac{K\lambda}{D} \quad (1)$$

where β represents full width at half maxima of XRD peaks, θ represents the peak positions, λ is the wavelength of the X-ray used for the recording of XRD pattern (Cu K_α=1.54 Å), D here is the average crystallite size and K = 0.94 is the Scherrer constant. The average crystallite size of the optimized CGSW: S5E3 phosphor is 24.13 nm.

The XRD data of un-doped CGSW host matrix was analyzed further via Rietveld refinement using the FullProf Suite programme for a thorough structural investigation.

Table 1 Refinement parameters for an un-doped CGSW host lattice

Parameters	JCPDS	CGSW: Un-doped
a = b (Å)	5.2408	5.2340
c (Å)	11.2692	11.3286
α = β = γ	90°	90°
V $(Å)^3$	309.52	310.34
χ^2	-	3.8

The backgrounds were improved using fifteen cycles of linear interpolation, and for each composition, the pseudo-Voigt function was employed to fit the experimental data. The atomic positions of O, Ca, Gd, Sb and W ions did not change significantly throughout the refining process. Figure 2 shows the Rietveld refinement map for an un-doped CGSW sample and Table 1 lists all the related crystal structure parameters. The black line in the refined map for each composition represents the observed intensity data, the red line with a circle represents the refined calculated intensity, the blue line illustrates the difference between the calculated and the observed intensities, and the green perpendicular line symbolizes the positions of Bragg's peaks.

Figure 3 shows the SEM images of the CGSW un-doped and CGSW: S5E3 phosphor. SEM image shows that the shape of the particles is irregular, and doping does not change the surface morphology of the prepared phosphor.

The particles were agglomerated because of high synthesis temperature and the size was found to be in micrometer range.

The EDS spectrum of CGSW: S5E3 phosphor is shown in the Fig. 4, which shows that elements Ca, Gd, Sb, W, O, Sm and Eu are present in the sample, with an additional element Pt, which may be due to the preparation of sample to perform EDS.

FT-IR

To gain deep insights into the nature of bonding, Infrared (IR) spectroscopy analysis was performed. Figure 5 shows the FT-IR spectrum of un-doped CGSW and CGSW: S5E3 phosphor observed in the range 400–4000 cm^{-1}. The figure shows several bands near 451, 662, 905, 1064, 1247, 1395, 2900, 2988, and 3672 cm^{-1}. The bands between 450 and 950 cm^{-1} correspond to the $[WO_4]^{2-}$ group [17]. The band at 2900 cm^{-1} shows the existence of -OH radicals [18]. The existence of a band at 2988 cm^{-1} represents the antisymmetric stretching of the H_2O molecule [17] and at 3672 cm^{-1}, the stretching of the -OH group [17]. The spectrum confirms the existence of $[WO_4]^{2-}$ and hence that of the formation of CGSW. All the band positions alongwith their type of vibrations are reperesented in Table 2.

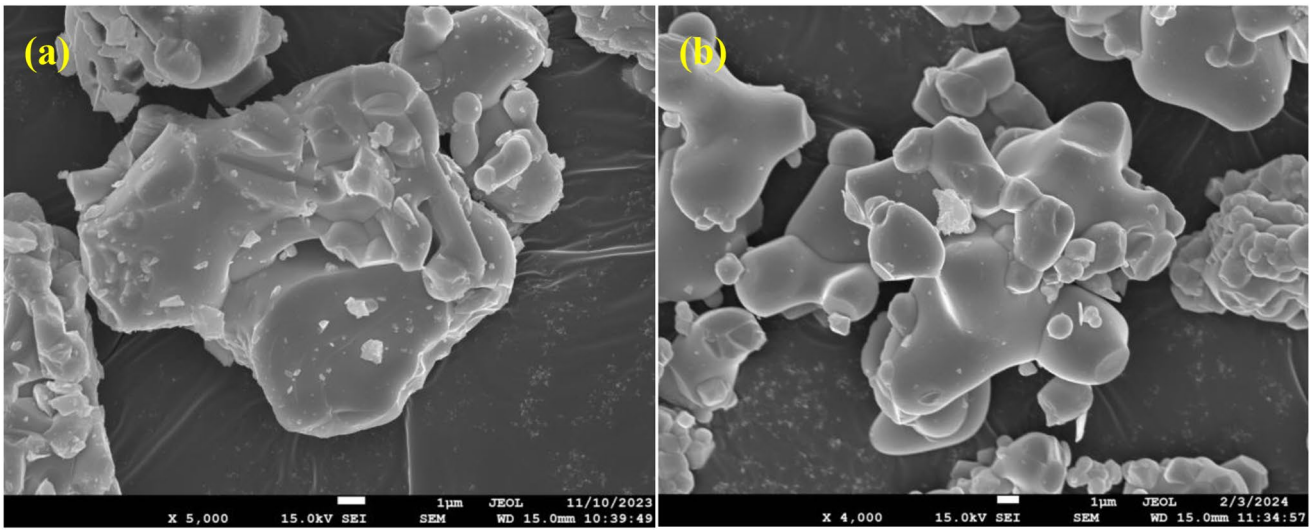


Fig. 3 SEM images of (**a**) CGSW un-doped, (**b**) CGSW: S5E3phosphor

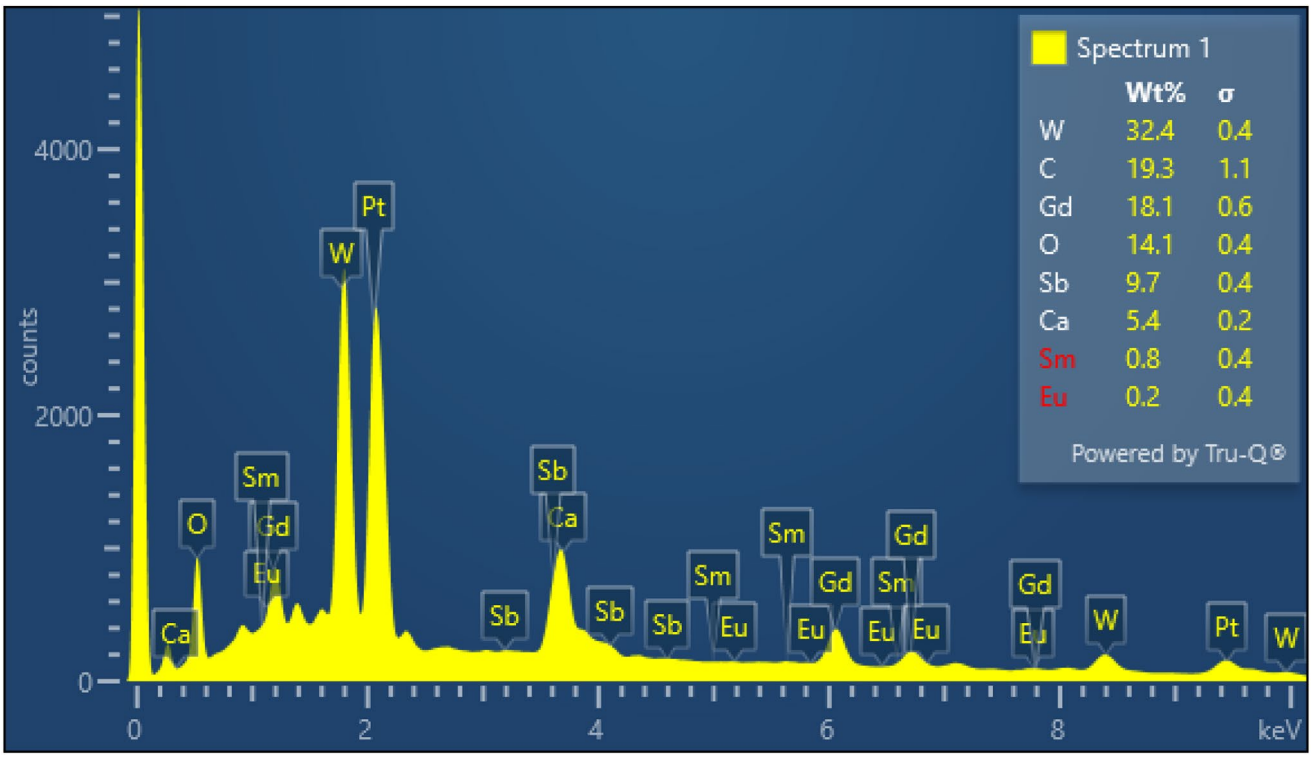


Fig. 4 EDS spectrum of CGSW: S5E3 phosphor

UV-Vis Spectroscopy

Figure 6 shows the diffuse reflectance spectra (DRS) of the prepared samples monitored in the wavelength range 200–1600 nm. The absorbance band at 250–300 nm common to all the samples corresponds to band-to-band transition in host lattice [22] while the absorbance bands between 400 and 1600 nm are attributed to transitions of f-f orbitals in Sm^{3+} and Eu^{3+} ions. The spectra show absorbance at 406 nm which corresponds to both Eu^{3+} ($^7F_0 \rightarrow {}^5L_6$) and Sm^{3+} ($^6H_{5/2} \rightarrow {}^4F_{7/2}$) ions transitions. Other peaks at 472, 955, 1094, 1249, 1399, 1510 and 1564 nm correspond to the absorption of electrons of the lower energy state $^6H_{5/2}$ of Sm^{3+} to higher energy states $^4I_{11/2}$, $^6F_{11/2}$, $^6F_{9/2}$, $^6F_{7/2}$, $^6F_{5/2}$, $^6H_{13/2}$ and $^6H_{11/2}$, respectively [23].

The optical energy band gap is calculated from the DRS curve using the formula [24]:

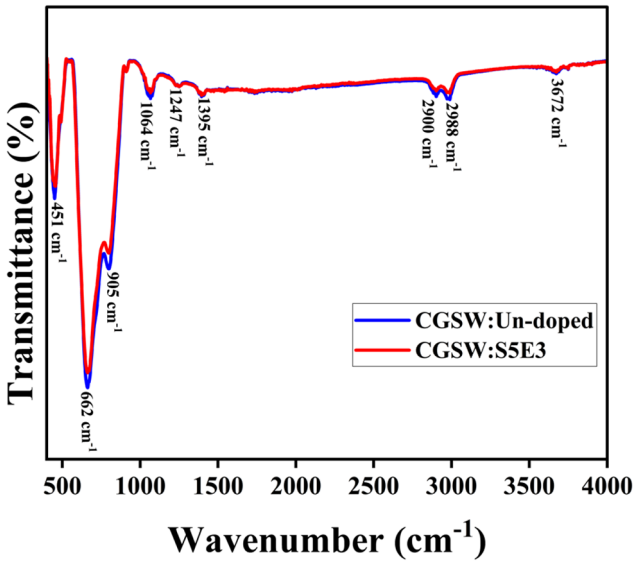


Fig. 5 FT-IR spectra of un-doped CGSW and CGSW: S5E3 phosphor

Table 2 FT-IR bands and their corresponding type of vibrations for un-doped CGSW and CGSW: S5E3 phosphor

Bands Position (cm^{-1})	Types of vibrations	References
451	Bending of W-O bond	[17, 19]
662, 905	Stretching of W-O bond	[19]
1064, 1247	Asymmetric stretching of W-O bond	[20]
1395	Symmetric stretching of –COO-bond	[21]
2900	The existence of –OH radicals	[18]
2988	Antisymmetric stretching of H_2O molecule	[18]
3672	Stretching of –OH group	[18]

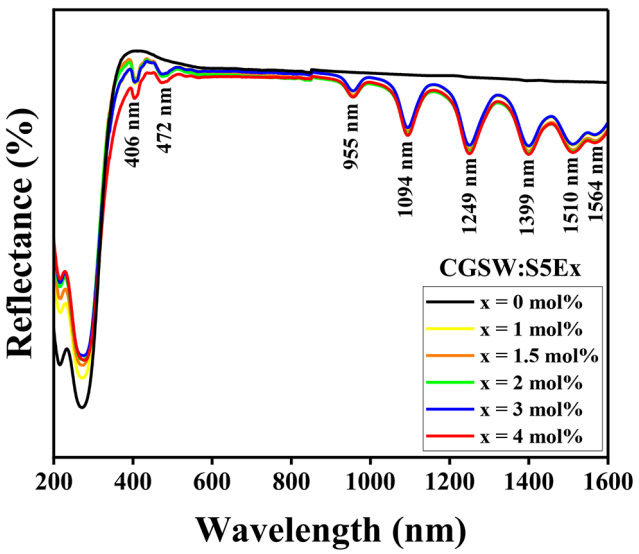


Fig. 6 UV-Vis DRS of CGSW: un-doped and CGSW: S5Ex (x = 1, 1.5, 2, 3, 4 mol%) phosphors

$$[F(R_\infty)h\nu]^n = C(h\nu - E_g) \quad (2)$$

Here, hν represents the energy of a photon of frequency ν, $E_g$is the optical energy band gap, and $F(R_\infty)$ the Kubelka-Munk function (R_∞=R_{sample}/$R_{standard}$). Here, C is the proportionality constant and n = ½ or 2 for indirect or direct band gaps, respectively. The Kubelka-Munk (K-M) function can be obtained using the following equation [25]:

$$F(R) = \frac{(1-R)^2}{2R} = \frac{k}{s} \quad (3)$$

In the above equation, the sample's reflectance, molar absorption coefficient, and scattering coefficient are represented by the variables R, k, and s, respectively. On putting the value of $n=2$ in Eq. 2, we get:

$$[F(R_\infty)h\nu]^2 = C(h\nu - E_g) \quad (4)$$

The plot between the K-M function and hν is shown in Fig. 7. The linear part of the plot is extrapolated to find E_g. The calculated direct energy band gap of CGSW: S5Ex is 4.0, 3.97, 3.98, 3.96, 3.93, and 3.93 eV for x = 0, 1, 1.5, 2, 3, and 4 mol%, respectively. The slight change in the E_g shows not much effect on the structure of the synthesized phosphor, The bandgap decreases with an increase in the concentration of dopant ion which may be due to energy levels of incorporated rare earth ions which makes electron transitions more favorable from valence band to conduction band [26].

Photoluminescence Properties

The PL excitation (λ_{em} = 612 nm) and PL emission spectra (λ_{ex} = 275 nm) of CGSW: S5E3 are compared with that of CGSW: Sm in Fig. 8. The excitation spectra consist of a broad band at 200–340 nm with maximum intensity at 275 nm, and several sharp peaks between 340 and 540 nm, as shown in Fig. 8(a). The reason behind the existence of broadband is the charge transfer from $O^{2-}\rightarrow W^{6+}$ and $O^{2-}\rightarrow Sm^{3+}$. So, it is called the charge transfer band (CTB) where electrons in 2p states of O^{2-} are excited to 5d and 4f states of W^{6+}, and Sm^{3+}, respectively [27]. The sharp peaks are caused by the transitions from the 4f orbital to the other 4f orbital of Sm^{3+} and Eu^{3+} ions. The peaks at 346, 363, 377, 405, 418, 440, 465, 480 and 527 nm are attributed to the transitions from $^6H_{5/2}$ level of Sm^{3+} ions to $^4D_{5/2}$, $^4D_{3/2}$, $^6P_{7/2}$, $^4F_{7/2}$, $^4M_{19/2}$, $^4I_{15/2}$, $^4I_{13/2}$, $^4M_{15/2}$, and $^4F_{3/2}$ levels respectively [28, 29]. The additional peaks arising in co-doped samples at 383, 395, and 535 nm can be attributed to $^7F_0\rightarrow{}^5L_7$, $^7F_0\rightarrow{}^5L_6$ and $^7F_0\rightarrow{}^5D_1$ transitions of Eu^{3+} ions respectively [30]. The highly intense CTB suggests energy migration from the host

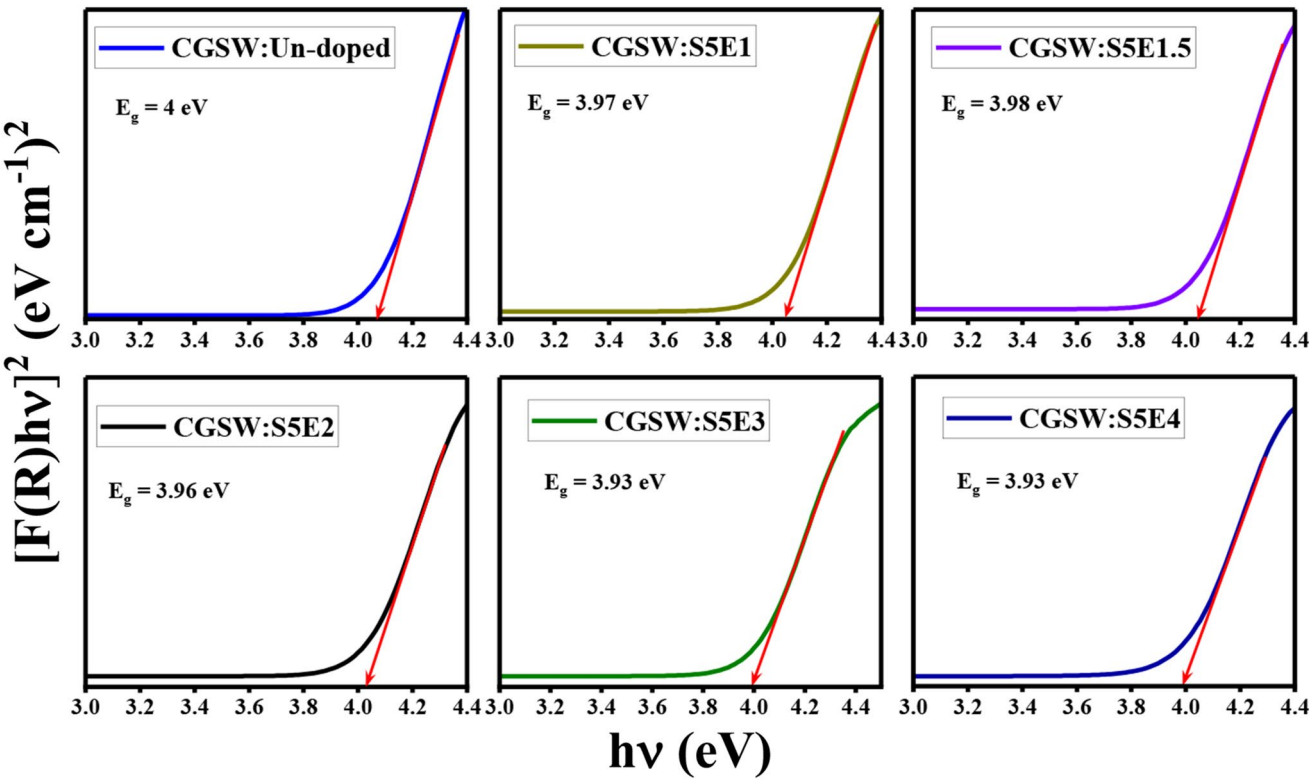


Fig. 7 The plot between K-M function and hν for CGSW: S5Ex (x = 0, 1, 1.5, 2, 3, 4 mol%) phosphors

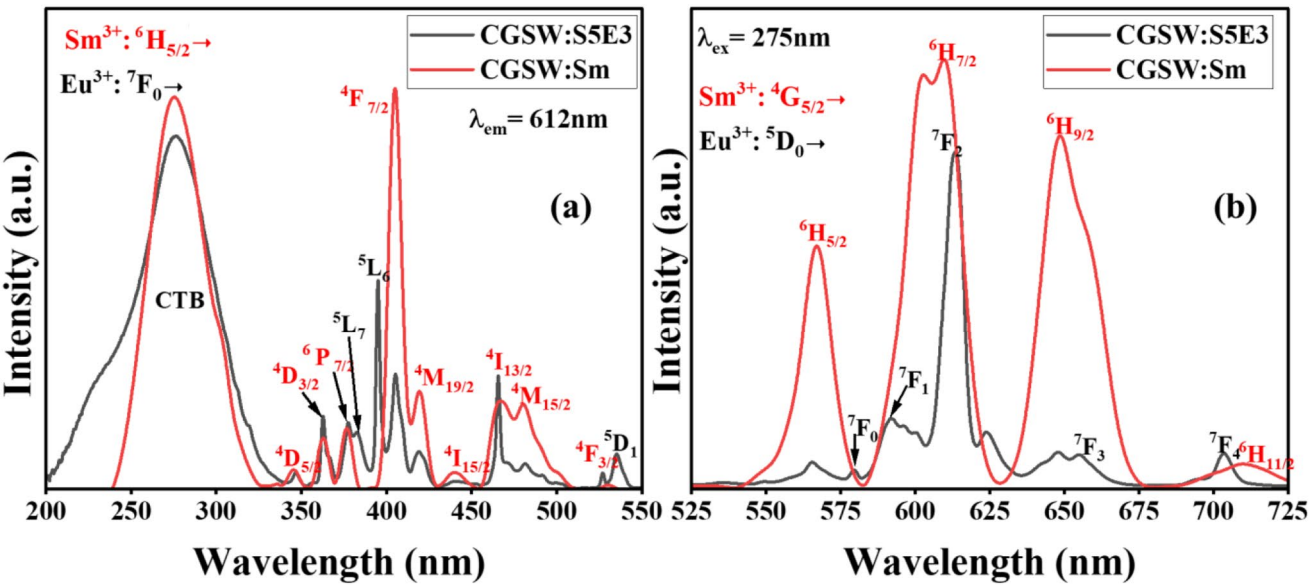


Fig. 8 PL (**a**) excitation and (**b**) emission spectra of CGSW: S5E3 and CGSW: Sm phosphors

matrix to the Sm^{3+} ions, and host excitation wavelength can give enhanced emission. The excitation peaks of Sm^{3+} that can be seen in the excitation spectra of the co-doped sample recorded at the emission wavelength of 612 nm ($^5D_0 \rightarrow {}^7F_2$ level of Eu^{3+}) suggest transfer of energy from Sm^{3+} to Eu^{3+}. The emission spectra, as shown in Fig. 8(b), show peaks at 565, 600, 649, and 710 nm corresponding to $^4G_{5/2} \rightarrow {}^6H_{5/2}$, $^6H_{7/2}$, $^6H_{9/2}$, and $^6H_{11/2}$ transitions of Sm^{3+} ions respectively whereas additional peaks at 579, 592, 612, 654, and 704 nm in co-doped sample correspond to $^5D_0 \rightarrow {}^7F_J$ (J = 0, 1, 2, 3, and 4) transitions of Eu^{3+} ions [31, 32].

Now, the excitation spectra of CGSW: S5E3 phosphor monitored at 600 nm and 612 nm emission wavelengths are shown in Fig. 9 (a). Figure 9(b) shows the comparison of emission spectra of CGSW: S5E3 at host excitation wavelength $\lambda_{ex}=275$ nm and Sm^{3+} ion excitation wavelength $\lambda_{ex}=405$ nm resulting a relatively intense PL emission under 275 nm.

The emission spectra of CGSW: S5Ex(x = 1, 1.5, 2, 3, 4 mol%) under excitation wavelength of (a) 275 nm, and (b) 405 nm are shown in Fig. 10 (a) and (b), respectively. The presence of Eu^{3+} characteristics peaks in the emission spectra of the sample at $\lambda_{ex}=405$ nm (related to Sm^{3+}) also indicates the Sm^{3+} to Eu^{3+} energy transfer. The photographs of the as-prepared CGSW: S5Ex(x = 1, 1.5, 2, 3, 4 mol%) phosphors under UV lamp has been presented in Fig. 10 (c).

The phenomenon of energy migration from one ion to another ion is very common in rare earth ions as they have many energy levels, especially in crystalline materials. These levels split due to the crystal field because of which the chances of level matching increase, and energy transfer occurs [28].

Figure 11, shows the intensity variation of the characteristic emission peaks of Sm^{3+} ions and Eu^{3+} ions with variation in the Eu^{3+} ions concentration in the samples. The intensity of Sm^{3+} ions peaks decreases whereas the intensity of Eu^{3+} ions characteristic peak increases as the concentration of Eu^{3+} increases which indicates the energy migration from Sm^{3+} to Eu^{3+} [32]. The emission intensity of peak at 612 nm ($^5D_0 \rightarrow {}^7F_2$ of Eu^{3+}) is highly intense and it shows that Sm^{3+} to Eu^{3+} energy transfer is efficient. The enhanced red emission of Eu^{3+} due to co-doping with Sm^{3+} makes this phosphor suitable for proposed w-LED and indoor plant lighting systems.

The intensity variation of $^5D_0 \rightarrow {}^7F_2$ (612 nm) transition of Eu^{3+} ions is shown in the inset of Fig. 10(a) and (b). The intensity of the $^5D_0 \rightarrow {}^7F_2$ transition of Eu^{3+} ions initially increases with the concentration of Eu^{3+}, reaches a maximum at 3 mol% and then decreases. This phenomenon is called concentration quenching [35]. At higher concentrations, the distance between the Eu^{3+} ions decreases which promotes non-radiative transfer and as a result, concentration quenching occurs [36]. The concentration at which concentration quenching occurs is called critical concentration (X_c) and the distance between the ions at this concentration is called critical distance (R_c). The relation between R_c and X_c is given by the Blasse's equation [33]:

$$R_c = 2\left[\frac{3V}{4\pi Z X_c}\right]^{\frac{1}{3}} \tag{5}$$

where V, Z, and X_c stand for the unit cell volume, the number of available dopant sites in the unit cell, and the total concentration of dopant ions, respectively. For as-prepared CGSW: S5E3 phosphor, Z is 4, V is 309.52Å^3 and X_c is 0.08 (5 mol% of Sm^{3+} and 3 mol% of Eu^{3+}). From Eq. (5), R_c comes out to be 12.27Å. As the R_c is greater than 5Å, it suggests that energy transfer occurs through electric multipolar interaction. The determination of the kind of interaction was done by plotting $I_0/I = C^{\frac{n}{3}}$ using Dexter's theory along with Reisfeld's approximation [34, 35] where I and I_0 refer to the PL intensity of $^4G_{5/2} \rightarrow {}^6H_{7/2}$ transition of Sm^{3+} ion in the presence and absence of Eu^{3+} ion respectively, C is the total concentration of Sm^{3+} and Eu^{3+} ion in the sample and $n=6$ for dipole-dipole (d-d), 8 for dipole-quadrupole(d-q), and 10 for quadrupole-quadrupole (q-q) interactions. The $I_0/I\ vs.\ C^{\frac{n}{3}}$ ($n=6$, 8, 10) the plot is shown in Fig. 12. The

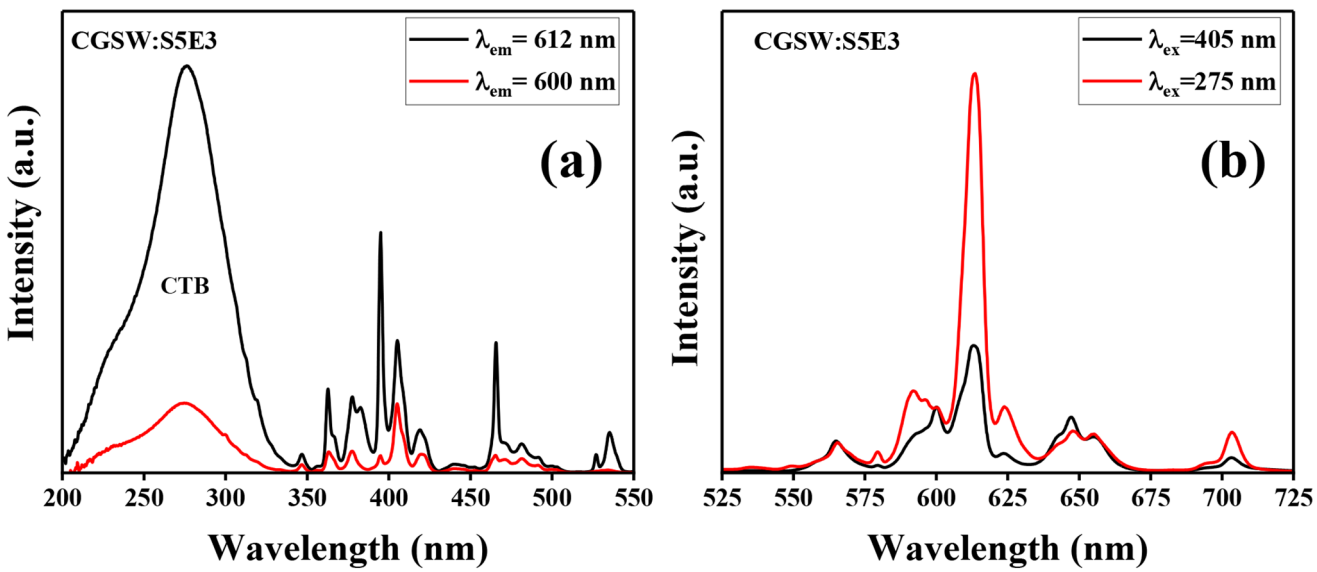


Fig. 9 (**a**) Excitation spectra at $\lambda_{em}=600$, 612 nm (**b**) Emission spectra at $\lambda_{ex}=405$, 275 nm of CGSW: S5E3 phosphor

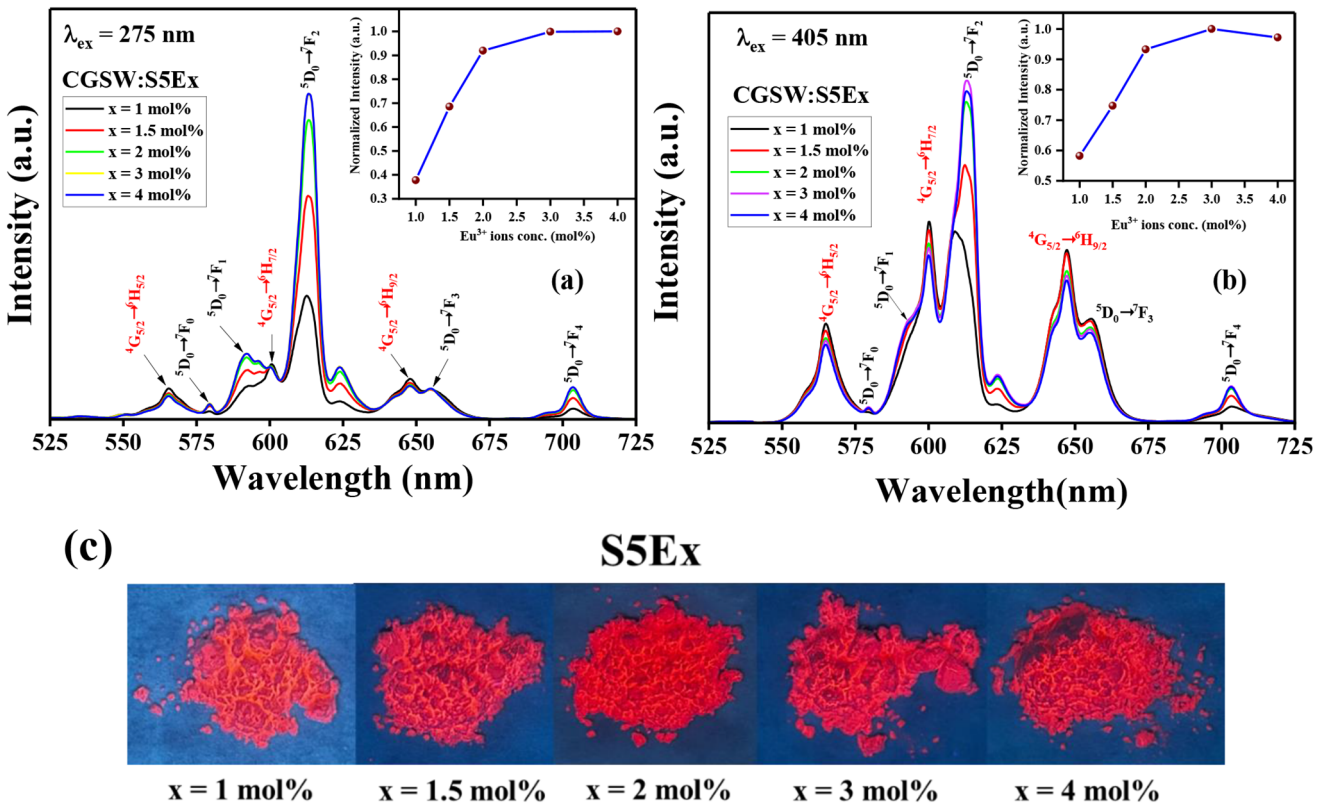


Fig. 10 The PL emission spectra of CGSW: S5Ex (x = 1, 1.5, 2, 3, 4 mol%) phosphors under λ_{ex} = (**a**) 275 nm, (**b**) 405 nm and (**c**) photographs of CGSW: S5Ex (x = 1, 1.5, 2, 3. 4 mol%) phosphors under UV light

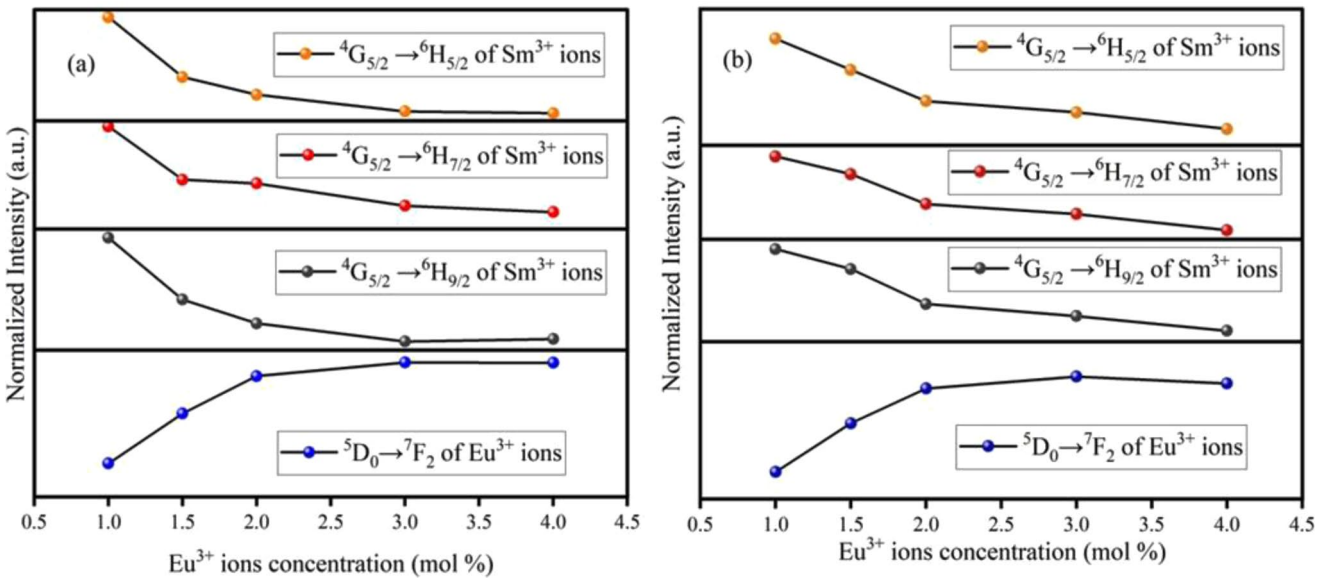


Fig. 11 The intensity variation of Sm^{3+} ions and Eu^{3+} ions emission peaks with the variation of Eu^{3+} ions concentration in the as-prepared phosphor sample at (**a**) λ_{ex} = 275 nm, and (**b**) λ_{ex} = 405 nm

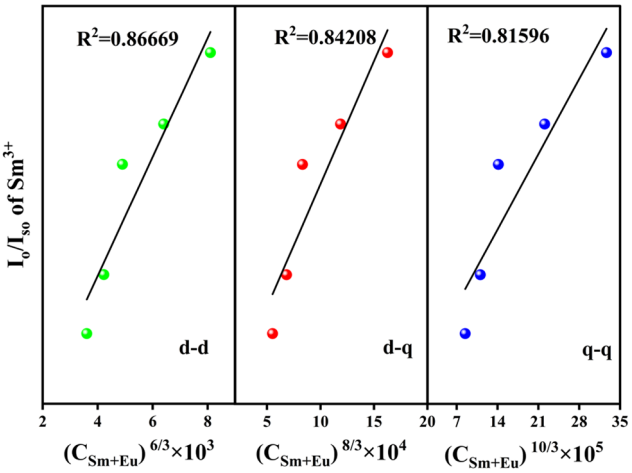


Fig. 12 The I_o/I_{so} versus $C^{\frac{n}{3}}$ ($n=6, 8, 10$) plot for CGSW: S5Ex (x = 1, 1.5, 2, 3, 4 mol%) phosphors

kind of interaction involved in the energy transfer between Sm^{3+}-Eu^{3+} is dipole-dipole interaction since the best fit is for $n=6$.

The fluorescence decay curves of 600 nm emission peak ($^4G_{5/2} \rightarrow {}^6H_{7/2}$ of Sm^{3+}) under 275 nm excitation are shown in Fig. 13. The lifetime curves are fitted using bi-exponential function given by the formula [36]:

$$I = I_o + A_1 e^{\frac{-t}{\tau_1}} + A_2 e^{\frac{-t}{\tau_2}} \tag{6}$$

Where I_0 and I represent the intensity at any time initial time t = 0 and at any time "t", respectively, A_1 and A_2 are constants of the fitting. τ_1 and τ_2 represent fast and slow lifetime components of the equation. The average lifetime values of prepared samples were calculated using the below Eq. (7):

$$\tau_{avg} = \frac{A_1\tau_1^2 + A_2\tau_2^2}{A_1\tau_1 + A_2\tau_2} \tag{7}$$

Table 3 The fluorescence lifetime values for CGSW: S5Ex (x = 1, 1.5, 2, 3, 4 mol%) phosphors

CGSW: S5Ex	Lifetime τ (μs)
x = 1 mol%	44.25
x = 1.5 mol%	35.50
x = 2 mol%	32.51
x = 3 mol%	31.57
x = 4 mol%	28.99

The average lifetimes calculated for CGSW: S5Ex (x = 1.0, 1.5, 2.0, 3.0, 4.0 mol%) phosphors are shown in the Table 3. As the Eu^{3+} ions concentrations increase, the lifetime values decrease which shows that the energy is being transferred to Eu^{3+} from Sm^{3+} ions [37].

Energy Transfer Mechanism

Figure 14 shows the energy transitions from host $[WO_4]^{2-}$ to Sm^{3+} ions and from Sm^{3+} ions to Eu^{3+} ions. Firstly, at an excitation wavelength of 275 nm, electrons are excited to the higher energy states of $[WO_4]^{2-}$ and then come back to

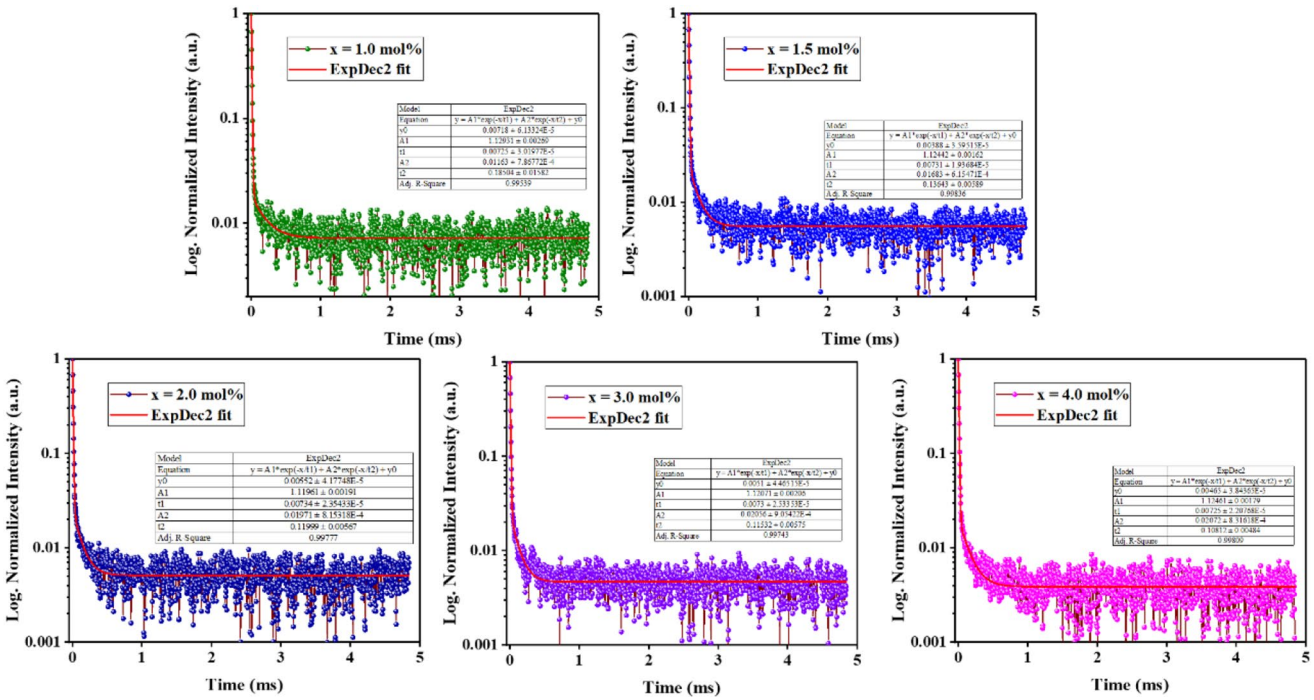


Fig. 13 The fluorescence decay curves of CGSW: S5Ex (x = 1, 1.5, 2, 3, 4 mol%) phosphors

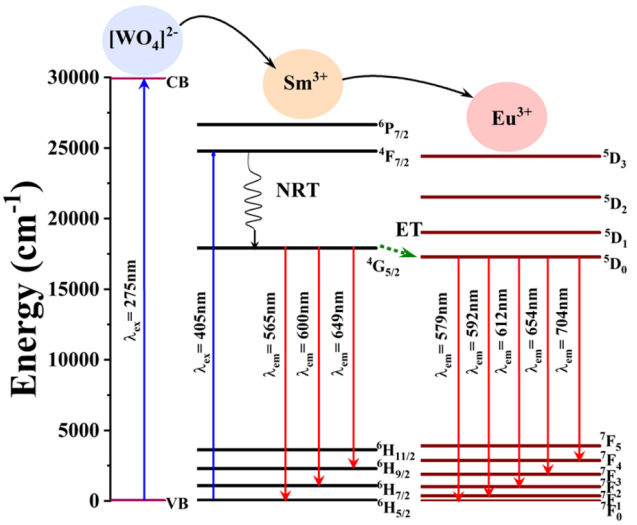


Fig. 14 Energy level diagram showing energy transfer processes from host to Sm^{3+} and then from Sm^{3+} to Eu^{3+} ions

lower energy states by non-radiative transfer as no emission peaks of $[WO_4]^{2-}$ are observed. Then, as the host transfers energy to Sm^{3+} ions, the excitation of electrons of Sm^{3+} ions take place to the higher energy state $^4G_{5/2}$. Subsequently, electrons relax to the lower energy states $^6H_{k/2}$ (k = 5, 7, 9) by radiative processes exhibiting emission at wavelengths 565, 600 and 649 nm, respectively. As the $^4G_{5/2}$ state of Sm^{3+} is very close to the $^5D_0$ state of the Eu^{3+} ion, some energy is transferred to the Eu^{3+} ion from Sm^{3+} making it unstable. Then, the electrons de-excite to the lower energy states ^7F_j (j = 0, 1, 2, 3, and 4) of Eu^{3+} showing emissions at 579, 592, 612, 654 and 704 nm respectively.

Under 405 nm excitation, the Sm^{3+} ions electrons go to higher energy level $^4F_{7/2}$ and then come back to the lower level $^4G_{5/2}$ level via non-radiative transfer and then further relaxe to the lower levels transferring energy to Eu^{3+} ions as discussed above. First, the host acts as a sensitizer and transfers energy to Sm^{3+} and then, Sm^{3+} transfers energy to the Eu^{3+} ions.

Colorimetry Analysis

CIE chromaticity coordinates and CCT values are the important parameters to determine the luminescence properties of the prepared phosphor. The CIE chromaticity coordinates were evaluated using the CIE 1931 software. Correlated color temperature (CCT) is the absolute temperature of the blackbody whose chromaticity is almost the same as the light source [38]. The warm light (red, orange, and yellow) has low CCT (< 3200 K) values whereas cool light (blue light) has high CCT value (> 4000 K) [39]. CIE color coordinates were determined by the emission spectra of CGSW: S5Ex at an excitation wavelength of 275 nm using the formulae given in the literature. CCT values were evaluated using the McCamy relation [40]. The color purity (CP), another important factor in characterizing a phosphor, is the relative distance of the light source under the test from the center of the chromaticity diagram. It determines the purity or monochromatic quality of color at a specific level of lightness [41]. The CP is evaluated by using the following relation [42]:

$$CP = \frac{\sqrt{(x_s - x_i)^2 - (y_s - y_i)^2}}{\sqrt{(x_d - x_i)^2 - (y_d - y_i)^2}} \times 100\% \quad (8)$$

Here (x_s, y_s); (x_d, y_d) and (x_i, y_i) are the CIE coordinates of the samples, dominant wavelength and pure white light, respectively. Also, (x_i, y_i) are valued as (0.33, 0.33). CIE chromaticity coordinates of all samples are plotted in Fig. 15. CIE coordinates, CCT, and CP values of prepared samples are shown in the Table 4 as well. The CIE coordinates of the samples lie in red region and low CCT values suggest that the prepared phosphor can have potential application in making warm w-LEDs.

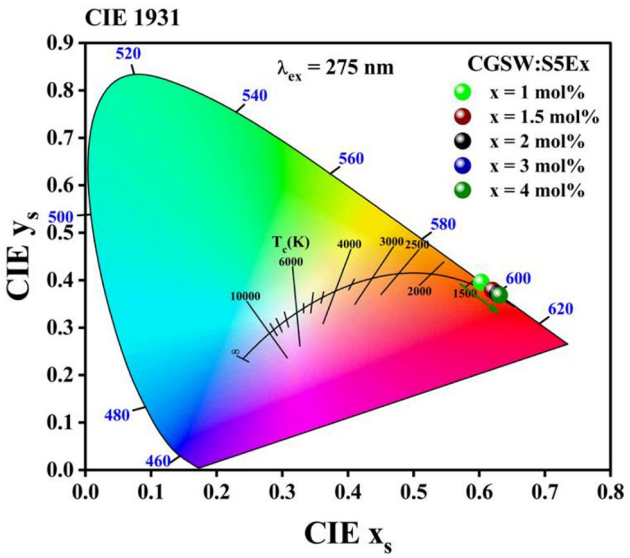


Fig. 15 CIE chromaticity coordinates for CGSW: S5Ex (x = 1, 1.5, 2, 3, 4 mol%) phosphors

Table 4 CIE coordinates, CCT, and color purity (CP) of CGSW: S5Ex (x = 1, 1.5, 2, 3, 4 mol%) phosphors

CGSW: S5Ex	CIE coordinates (x_s,y_s)	CCT(K)	CP (%)
x = 1 mol%	(0.60841, 0.39068)	1742	82.12
x = 1.5 mol%	(0.62414, 0.37513)	1899	85.92
x = 2 mol%	(0.63104, 0.36831)	2020	87.68
x = 3 mol%	(0.63346, 0.36593)	2072	87.67
x = 4 mol%	(0.63397, 0.36543)	2084	88.46

Conclusions

In summary, a red-light emitting phosphor CGSW: S5Ex was successfully synthesized using a conventional solid-state synthesis route. To study its structural and compositional properties, XRD, SEM and FT-IR analyses were carried out. The average crystallite size of the optimized sample CGSW: S5E3 was calculated to be 24.13 nm using the Debye-Scheerer equation. The FT-IR study shows that OH content was low in the prepared phosphor. The optical properties, luminescent properties, and decay kinetics were studied using UV-Vis spectroscopy, PL spectroscopy, and time-resolved fluorescence spectroscopy respectively. The DRS spectra were used to calculate the optical energy band gap values of the prepared phosphors which lie in the range 3.93 eV to 4.0 eV. The PL emission spectra suggest an intense emission intensity while using the host excitation wavelength rather than that of the dopant ion excitation. The PL emission spectra of samples under host excitation and under Sm^{3+} ion excitation wavelength suggest energy transfer from Sm^{3+} ions to Eu^{3+} ions. The energy transfer process has been studied and it was that observed the energy transfer process occurs through dipole-dipole interaction. The emission intensity was found to be maximum for CGSW: S5E3 phosphor sample. The prepared phosphor shows emission in the red region with low CCT and high CP values. The above results make the prepared phosphor sample a potential candidate for w-LEDs and indoor plant growth lighting systems.

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). The authors are also thankful to the authorities of Delhi Technological University, New Delhi, INDIA for providing the necessary infrastructure facilities to carry out the present research work. The authors would also like to thank Dr. Manju Srivastava and Dr. Shyama Prasad, Department of Chemistry, Dayalbagh Educational Institute (DEI), Deemed University, Agra for providing access to the FE-SEM facility.

Author Contributions J.P.: Investigation, Methodology, Formal analysis, Software, Validation, Visualization, Writing – original draft, Writing, review & editing. G.Y.: Validation, Visualization, Formal analysis, Writing – review & editing. S.: Validation, Visualization, Formal analysis, Writing – review & editing. A.: Resources, Visualization, Supervision, Project administration, Writing – review & editing. A.P.: Resources, Visualization, Supervision, Project administration, Writing – review & editing. A.S.R.: Conceptualization, Methodology, Resources, Visualization, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Funding Not applicable.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Ethical Approval The authors declare that no animal or animal data was used in any study or experiment mentioned in this manuscript. We approve all the ethics.

Competing Interests The authors declare no competing interests.

References

1. Hong WY, Rahmat BNNN (2022) Energy consumption, CO_2 emissions and electricity costs of lighting for commercial buildings in Southeast Asia. Sci Rep 12:13805. https://doi.org/10.1038/s41598-022-18003-3
2. Anu, Rao AS (2023) Luminescence and optical thermometry strategy based on emission spectra of $Li_2Ba_5W_3O_{15}:Pr^{3+}$ phosphors. Opt Mater (Amst) 145:114476. https://doi.org/10.1016/j.optmat.2023.114476
3. Kaur S, Rao AS, Jayasimhadri M, Jaiswal VV, Haranath D (2020) Tb^{3+} ion induced colour tunability in calcium aluminozincate phosphor for lighting and display devices. J Alloys Compd 826:154212. https://doi.org/10.1016/j.jallcom.2020.154212
4. Smolders KCHJ, de Kort YAW (2017) Investigating daytime effects of correlated colour temperature on experiences, performance, and arousal. J Environ Psychol 50:80–93. https://doi.org/10.1016/j.jenvp.2017.02.001
5. Anu, Kumar A, Ravina, Deopa N, Rao AS (2023) Efficient tunable photoluminescence of Dy^{3+}/Eu^{3+} co-doped OFSZBS glasses for warm white LEDs. J Non Cryst Solids 616:122421. https://doi.org/10.1016/j.jnoncrysol.2023.122421
6. Yang J, Lai X, Luo L, Li W, Du P (2024) Regulating the concentration quenching in Eu^{3+}-activated $Lu_2W_3O_{12}$ red-emitting phosphors through phase transition for white light-emitting diode and plant growth applications. Ceram Int 50:34403–34411. https://doi.org/10.1016/j.ceramint.2024.06.259
7. Zhou X, Yang C, Liu S, Wang Y, Hu R, You Q, Liu W, Liu M, Yang S, Jiang X (2023) Enhanced photoluminescence of $K_7CaY_2(B_5O_{10})_3:Eu^{3+}$, Sm^{3+}/Gd^{3+} phosphors through energy transfer and their applications in LED lights for plant growth. J Alloys Compd 969:172471. https://doi.org/10.1016/j.jallcom.2023.172471
8. Lai X, Ge X, Li J, Zhu J, Du P (2024) Regulating the luminescence properties of $Eu_2W_3O_{12}$ red-emitting phosphor via rare-earth ions doping for optical thermometry, white light-emitting diode and plant growth applications. J Lumin 275:120815. https://doi.org/10.1016/j.jlumin.2024.120815
9. Mehare MD, Mehare CM, Swart HC, Dhoble SJ (2023) Recent development in color tunable phosphors: a review. Prog Mater Sci 133:101067. https://doi.org/10.1016/j.pmatsci.2022.101067
10. Wei Q, Chen D (2009) Luminescence properties of Eu^{3+} and Sm^{3+} coactivated gd(III) tungstate phosphor for light-emitting diodes. Opt Laser Technol 41:783–787. https://doi.org/10.1016/j.optlastec.2008.12.003
11. Fu M, Wu Q, Gu C, Hu S, Lu S, Wang B, Hong Y, Wang Y (2023) Energy transfer in Sm^{3+}/Eu^{3+} doped $Li_6Gd(BO_3)_3$ orange-red phosphor. Vacuum 214:112147. https://doi.org/10.1016/j.vacuum.2023.112147
12. Xie F, Gu J, Xu D, Lin H, Shi Y, Zhang S, Wang D, Zhu R, Yang J (2024) Tunable emission color, energy transfer mechanisms and high thermal stability of Tm^{3+}/Dy^{3+} co-doped single-phase white luminescent tungstate phosphors. J Lumin 269:120519. https://doi.org/10.1016/j.jlumin.2024.120519

13. Zhang W, Long J, Fan A, Li J (2012) Effect of replacement of ca by ln (ln = Y, Gd) on the structural and luminescence properties of $CaWO_4$:Eu^{3+} red phosphors prepared via co-precipitation. Mater Res Bull 47:3479–3483. https://doi.org/10.1016/j.materresbull.2012.06.075
14. Yang Y, Li F, Lu Y, Du Y, Wang L, Chen S, Ouyang X, Li Y, Zhao L, Zhao J, Deng B, Yu R (2022) $CaGdSbWO_8$:Sm^{3+}: a deep-red tungstate phosphor with excellent thermal stability for horticultural and white lighting applications. J Lumin 251:119234. https://doi.org/10.1016/j.jlumin.2022.119234
15. Mondal K, Manam J (2018) Investigation of photoluminescence properties, thermal stability, energy transfer mechanisms and quantum efficiency of $Ca_2ZnSi_2O_7$: Dy^{3+}, Eu^{3+} phosphors. J Lumin 195:259–270. https://doi.org/10.1016/j.jlumin.2017.11.028
16. Anu, Rao AS (2023) Novel orange-red-emitting $Li_2Ba_5(WO_5)_3$:Eu^{3+} phosphor for w-LEDs and non-contact thermometry applications. J Mater Sci Mater Electron 34:2191. https://doi.org/10.1007/s10854-023-11602-4
17. Xie J, Cheng L, Tang H, Yu X, Wang Y, Wang C, Mi X, Liu Q, Zhang X (2020) Synthesis and photoluminescence properties of $NaBi(WO_4)_2$:Eu^{3+} red-emitting phosphor for NUV-based WLEDs. J Lumin 219:116841. https://doi.org/10.1016/j.jlumin.2019.116841
18. Anu, Seema, Kumar A, Deopa N, Sahu MK, Prasad A, Rao AS (2024) A single phase $Li_2Ba_5W_3O_{15}$:Dy^{3+}/Eu^{3+} phosphor for color tunable devices and non-contact optical thermometry. J Lumin 269:120444. https://doi.org/10.1016/j.jlumin.2024.120444
19. Baby B, Thomas S, T K, Jose J, Biju PR, Joseph C (2023) Optimized microwave assisted synthesis of $La_2(WO_4)_3$:Tb^{3+} phosphors and analysis of photoluminescence behavior. J Lumin 254:119510. https://doi.org/10.1016/j.jlumin.2022.119510
20. Sreeja E, Vidyadharan V, Jose SK, George A, Joseph C, Unnikrishnan NV, Biju PR (2018) A single-phase white light emitting Pr^{3+} doped Ba_2CaWO_6 phosphor: synthesis, photoluminescence and optical properties. Opt Mater (Amst) 78:52–62. https://doi.org/10.1016/j.optmat.2018.02.003
21. Kumari P, Manam J (2016) Influence of Dy^{3+} ions doping on structural and luminescent properties of $GdVO_4$. J Mater Sci Mater Electron 27:9437–9447. https://doi.org/10.1007/s10854-016-4990-7
22. Rathnakumari EA, Kennedy SMM (2021) Photoluminescence properties of certain trivalent RE ions (Dy/Sm/Eu) activated $NaSrBi_2(PO_4)_3$ phosphors and cool to warm white color tuning in Dy^{3+}-Sm^{3+} and Dy^{3+}-Eu^{3+} codoped $NaSrBi_2(PO_4)_3$ phosphors for indoor lighting applications. J Lumin 235:118018. https://doi.org/10.1016/j.jlumin.2021.118018
23. Carnall WT, Fields PR, Rajnak K (1968) Electronic energy levels of the Trivalent Lanthanide Aquo ions. IV. Eu^{3+}. J Chem Phys 49:4450–4455. https://doi.org/10.1063/1.1669896
24. Escobedo Morales A, Sánchez Mora E, Pal U (2007) Use of diffuse reflectance spectroscopy for Optical characterization of Unsupported nanostructures. 67 Bf 53
25. Guo Y, Moon BK, Choi BC, Jeong JH, Kim JH (2017) Color–tunable luminescence and energy transfer behaviors of Dy^{3+}/Eu^{3+} co–doped $SrLaMgTaO_6$ phosphors for solid state lighting applications. Mater Res Bull 88:166–173. https://doi.org/10.1016/j.materresbull.2016.12.035
26. Zhao R, Guo X, Zhang J, Zhang C, Deng C, Cui R (2023) Thermally-stable novel NaBaBi2(PO4)3: Sm^{3+}/Dy^{3+} white phosphors with tunable photoluminescence. Ceram Int 49:25795–25805. https://doi.org/10.1016/j.ceramint.2023.05.126
27. Kang F, Hu Y, Wu H, Mu Z, Ju G, Fu C, Li N (2012) Luminescence and red long afterglow investigation of Eu^{3+}–Sm^{3+} co-doped $CaWO_4$ phosphor. J Lumin 132:887–894. https://doi.org/10.1016/j.jlumin.2011.11.022
28. Jiang HX, Lü SC (2019) Intense red emission and two-way energy transfer in Sm^{3+}, Eu^{3+} co-doped $NaLa(WO_4)_2$ phosphors. Mater Res Bull 111:140–145. https://doi.org/10.1016/j.materresbull.2018.11.004
29. Seema, Rao AS (2024) Photoluminescence and energy transfer studies in the Sm^{3+} and Eu^{3+} co-doped $Sr_2ZnSi_2O_7$ red-emitting phosphors. J Lumin 275:120742. https://doi.org/10.1016/j.jlumin.2024.120742
30. Li L, Chang W, Chen W, Feng Z, Zhao C, Jiang P, Wang Y, Zhou X, Suchocki A (2017) Double perovskite $LiLaMgWO_6$:Eu^{3+} novel red-emitting phosphors for solid sate lighting: synthesis, structure and photoluminescent properties. Ceram Int 43:2720–2729. https://doi.org/10.1016/j.ceramint.2016.11.093
31. Han B, Ren J, Teng P, Zhu J, Shen Y, Li Z, Zhu X, Yang X (2022) Synthesis and photoluminescence properties of a novel double perovskite Ca_2LaSbO_6:Sm^{3+} phosphor for w-LEDs. Ceram Int 48:971–980. https://doi.org/10.1016/j.ceramint.2021.09.182
32. Sahu MK, Jayasimhadri M, Jha K, Sivaiah B, Rao AS, Haranath D (2018) 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. https://doi.org/10.1016/j.jlumin.2018.06.002
33. Anu, Kumari S, Deopa N, Rao AS (2024) Spectral studies of thermally stable Dy^{3+}/Sm^{3+} co-doped $Li_2Ba_5W_3O_{15}$ phosphors for warm white LEDs. J Phys D Appl Phys 57:315107. https://doi.org/10.1088/1361-6463/ad4566
34. Liu N, Mei L, Liao L, Fu J, Yang D (2019) High Thermal Stability Apatite Phosphors $Ca_2La_8(SiO_4)_6O_2$:Dy^{3+}/Sm^{3+} for White Light Emission: synthesis, structure, Luminescence properties and Energy transfer. Sci Rep 9:15509. https://doi.org/10.1038/s41598-019-51915-1
35. Kaur S, Rao AS, Jayasimhadri M (2019) 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. https://doi.org/10.1016/j.materresbull.2019.04.022
36. Anu, AS Rao (2024) Temperature sensing materials based on the fluorescence intensity ratio in $Li_2Ba_5W_3O_{15}$:Dy^{3+} phosphors. Sens Actuators Phys 372:115336. https://doi.org/10.1016/j.sna.2024.115336
37. Rohilla P, Rao AS (2023) Energy transfer induced colour tunable photoluminescence performance of thermally stable Sm^{3+}/Eu^{3+} co-doped Ba_3MoTiO_8 phosphors for white LED applications. J Mater Sci Mater Electron 34:1662. https://doi.org/10.1007/s10854-023-11025-1
38. Shang M, Li C, Lin J (2014) How to produce white light in a single-phase host? Chem Soc Rev 43:1372–1386. https://doi.org/10.1039/C3CS60314H
39. Anu, Deopa N, Rao AS (2022) Structural and luminescence characteristics of thermally stable Dy^{3+} doped oxyfluoride strontium zinc borosilicate glasses for photonic device applications. Opt Laser Technol 154:108328. https://doi.org/10.1016/j.optlastec.2022.108328
40. McCamy CS (1992) Correlated color temperature as an explicit function of chromaticity coordinates. Color Res Appl 17:142–144. https://doi.org/10.1002/col.5080170211
41. Kumari S, Anu A, Prasad P, Rohilla AS, Rao (2023) Prospective applications of thermally stable Dy^{3+} doped potassium zinc

strontium borate (KZSB) glasses in w-LEDs. J Mater Sci Mater Electron 34:907. https://doi.org/10.1007/s10854-023-10272-6
42. Anu, Rao AS (2023) Synthesis, structural and fluorescence investigations of Novel $Li_2Ba_5W_3O_{15}$: Sm^{3+} phosphors for Photonic device applications. J Fluoresc. https://doi.org/10.1007/s10895-023-03449-z

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

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

bioengineering

MDPI

Article

GAN-Based Novel Approach for Generating Synthetic Medical Tabular Data

Rashid Nasimov [1], Nigorakhon Nasimova [2], Sanjar Mirzakhalilov [2], Gul Tokdemir [3], Mohammad Rizwan [4], Akmalbek Abdusalomov [2,5,*] and Young-Im Cho [5,*]

1 Artificial Intelligence, Tashkent State University of Economics, Tashkent 100066, Uzbekistan; rashid.nasimov@tsue.uz
2 Department of Software Information Technologies, Tashkent University of Information Technologies Named After Muhammad Al-Khwarizmi, Tashkent 100200, Uzbekistan; nigorakhon.nasimova@tuit.uz (N.N.); mirzaxalilov86@tuit.uz (S.M.)
3 Department of Computer Engineering, Faculty of Engineering, Cankaya University, 06790 Ankara, Turkey; gtokdemir@cankaya.edu.tr
4 Centre of Excellence for Electric Vehicle and Related Technologies, Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India; rizwan@dce.ac.in
5 Department of Computer Engineering, Gachon University, Sujeong-gu, Seongnam-si 461-701, Gyeonggi-do, Republic of Korea
* Correspondence: akmaljon@gachon.ac.kr (A.A.); yicho@gachon.ac.kr (Y.-I.C.)

Abstract: The generation of synthetic medical data has become a focal point for researchers, driven by the increasing demand for privacy-preserving solutions. While existing generative methods heavily rely on real datasets for training, access to such data is often restricted. In contrast, statistical information about these datasets is more readily available, yet current methods struggle to generate tabular data solely from statistical inputs. This study addresses the gaps by introducing a novel approach that converts statistical data into tabular datasets using a modified Generative Adversarial Network (GAN) architecture. A custom loss function was incorporated into the training process to enhance the quality of the generated data. The proposed method is evaluated using fidelity and utility metrics, achieving "Good" similarity and "Excellent" utility scores. While the generated data may not fully replace real databases, it demonstrates satisfactory performance for training machine-learning algorithms. This work provides a promising solution for synthetic data generation when real datasets are inaccessible, with potential applications in medical data privacy and beyond.

Keywords: GAN; synthetic medical tabular data; statistical data; custom loss function

check for updates

Citation: Nasimov, R.; Nasimova, N.; Mirzakhalilov, S.; Tokdemir, G.; Rizwan, M.; Abdusalomov, A.; Cho, Y.-I. GAN-Based Novel Approach for Generating Synthetic Medical Tabular Data. *Bioengineering* **2024**, *11*, 1288. https://doi.org/10.3390/bioengineering11121288

Academic Editors: Ujjwal Baid and Bhakti Baheti

Received: 30 October 2024
Revised: 7 December 2024
Accepted: 11 December 2024
Published: 18 December 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent years, the generation of synthetic data has emerged as a prominent trend in the medical domain. This is because accessing real medical databases is very challenging, simply because medical records include private or sensitive data with special protections from the Health Insurance Portability and Accountability Act (HIPAA) [1] and General Data Protection Regulation (GDPR) [2]. In developed countries, hospitals have adopted Electronic Health Records (EHRs) to store patients' medical histories, reduce the need for physical records and manual paperwork, streamline processes like billing, scheduling and lab results and give the opportunity to medical professionals to easily access and share patient information across different healthcare facilities. Considering that the majority of hospital systems are integrated, every developed country currently possesses a massive database containing hundreds of thousands, if not millions, of patient records. Nevertheless, due to privacy policy, this database is not publicly available. Traditional methods of data anonymization, such as data masking, pose a serious threat to data utility while also risking privacy [3]. To solve this problem, synthetic data is the best choice, which can help to anonymize real hospital data and protect sensitive information while maintaining

the statistical features of actual data. Despite its several disadvantages such as limited realism—synthetic data may fail to capture the true variability and complex patterns of real-world medical data, reduced generalizability—synthetic data might not include rare or unforeseen conditions present in real-world datasets [4]; synthetic data can be an important factor for conducting research, especially in the field of healthcare since synthetic data give some benefits such as improving data accessibility and high-quality data augmentation. For this reason, in the last several years, the number of research papers in the field of generating synthetic medical data has increased significantly, and would be more useful in the future. However, real databases are used as the primary tool in all practical methods, including statistical and probabilistic methods that employ real databases to obtain the necessary distributions. However, as previously mentioned, an unauthorized individual cannot obtain access, even if their purpose is to convert real databases into a synthetic one. Such access is only available to individuals who have obtained special permission under special agreements. Because of this, extremely few studies use a database that is not openly accessible to generate synthetic data [5–7]. Almost all other studies use open databases. The question arises what is the purpose of creating a synthetic database from an already open database. There are two goals: the first is to augment the existing small database, and the second is to prove the effectiveness of the proposed method.

However, no matter how effective the proposed method is, it cannot create any synthetic data without the author having direct access to the real database. Therefore, despite the huge amount of research on synthetic data generation in recent years, very few databases are being made available for open use. That is, the problem of data shortage here is still an urgent problem to solve. As a solution to this problem, the use of statistical information about real tabular databases, rather than the database itself, is proposed to create synthetic data. This is because today there is statistical information about a huge number of important databases that are not publicly available. In particular, if there are databases collected within the framework of a specific project or research, then full statistical information about them is presented in the articles. For example, in [8–11], a lot of statistical information is given about a very large database while the database is not openly available. This study introduces a novel method for generating fully synthetic databases using only statistical data, leveraging modified GANs. To the best of our knowledge, this approach represents the first attempt to generate synthetic tabular data directly from statistical data using deep learning models. Despite certain limitations, the proposed method makes a significant contribution by pioneering a new direction for converting statistical data into tabular formats, laying the foundation for future advancements in this area.

The article is structured as follows. Section 2 is devoted to related works. Section 3 explains the proposed algorithm. The architecture of the modified GAN, which is considered an important part of the algorithm, the additional loss function used in it, and the formulas for calculating it are presented. Section 4 briefly provides information about the evaluation criteria and their calculation formulas. Section 5 presents the results of the evaluation based on the selected evaluation criteria to assess the effectiveness of the proposed method and compares them with the Gaussian Multivariate (GM), Synthetic Data Vault (SDV) and Conditional Tabular Generative Adversarial Network (CTGAN) methods. The reason for comparing with different methods, even though the proposed method has no analogues, is also explained at the end of Section 5. This chapter also provides a discussion of the results. Section 6 provides the conclusion of the article and outlines future work.

2. Related Works

The number of studies on synthetic tabular data generation has increased drastically in recent years [12]. Among them, methods such as CTGAN [13], Tabular Variational Auto encoders (TVAE) [14], Large Language Model-based [15] and Transformer-based hybrid methods [16] have been found to be the most effective. However, since these works are beyond the scope of this research work, the information in detail may be found in the literature [17–20].

The closest works to the proposed work are statistical and probabilistic methods. Therefore, the works conducted in this direction will be considered. Based on the statistical distributions and correlations of the variables in the real data, the statistical and probabilistic-based approaches create synthetic data. This type of approach is used by algorithms such as bootstrapping [21], the Bayesian models [22] and the probabilistic Bayesian networks [23]. These methods rely on their specific distribution functions to create synthetic databases. As a result, they are unable to generate tabular data directly from arbitrary statistical summaries of real datasets.

To our knowledge, only two methods that are close to ours have been developed so far [24–26]. Authors of [24] named their algorithm Synthea (Synthetic Health Data Generation). This is an open-source project designed to simulate realistic, synthetic patient health records for research, software testing and training purposes. Synthea uses a rule-based approach to model patient behavior, disease progression and healthcare encounters. It begins by generating a population of synthetic individuals. Demographic attributes such as age, gender and location are assigned based on real-world statistics to ensure realism. Environmental factors, including geography, access to healthcare and regional epidemiology, are configured based on external datasets. The algorithm uses disease modules that define the natural history of illnesses based on clinical guidelines and epidemiological data. Each disease is represented as a finite state machine with nodes and transitions. Nodes represent states (e.g., "Healthy", "Mild Disease", "Severe Disease"), and transitions are triggered by probabilities, time progression, or external factors. Disease progression and treatment are guided by evidence-based clinical guidelines, ensuring that simulated scenarios align with real-world healthcare practices. However, the quality and accuracy of the data generated by this method are highly dependent on the clinical guidelines, epidemiological data and expert knowledge used to build the state-transition models. Collecting and organizing these rules and knowledge requires the intervention of medical professionals. Also, knowledge preprocessing processes require a lot of time and effort. And this is the disadvantage of the method. Refs. [25,26] present a synthetic data-generation approach that uses statistical information of data and deep learning algorithms. This method involves modifying the primary database, which is initially developed based on statistical data, through the application of a specialized shuffle algorithm. The shuffle algorithm is designed to rearrange or manipulate the data in a way that ensures the resulting dataset meets predefined criteria for quality and realism, ultimately achieving a satisfactory outcome. Following the transformation process, the modified dataset is evaluated using a neural network. The neural network is utilized to assess the dataset's quality, ensuring that it retains meaningful patterns, adheres to expected distributions, and maintains coherence with the original statistical characteristics. The evaluation result of the neural network is used as a loss function to adjust the parameters in the shuffling algorithm. However, one disadvantage of this strategy is the difficulty in reaching convergence, as the training loss function of this deep neural network depends on a shuffling technique, leading to long training times.

Thus, a new approach that uses statistical information and deep learning models was proposed to fill this research gap.

3. Proposed Methods

The original GAN is composed of two parts: a generator neural network and a discriminator neural network. The generator tries to trick the discriminator by producing samples from random noise, while the discriminator attempts to accurately differentiate generated and real data. In this way, these two networks compete in a zero-sum game, in which one part's gain is proportional to another's loss. The generator takes random noise as an input, and the discriminator receives generated and real data, respectively. The basic loss function is an adversarial or Minimax loss function. The terms min and max will stand for generator loss minimization and discriminator loss maximization, respectively. Expanding upon the basic model, we suggested three significant enhancements: (1) a modified generator network architecture with two inputs was designed, (2) a new loss

function was used to train a generator neural network. Furthermore, to produce fully synthetic data, a unique training network method was proposed, which will be explained in Section 3.1.

3.1. Proposed Algorithm

Before presenting the algorithm of the suggested method, the main idea of the work should be described in greater depth. First of all, the goal of the proposed work, unlike previous methods, is to generate synthetic data based solely on given statistical data without directly accessing the database. In contrast to conventional statistical or probabilistic methods, tabular data are generated using a Deep Neural Network (DNN) rather than complex mathematical/statistical formulas. In this method, DNN is trained on a dataset which consisted of statistical data and tabular data to convert statistical data into tabular data. In other words, this approach is similar to text-to-image (DALL-E, OpenAI) or voice-to-text techniques. The GAN first learns to generate a table based on statistical data, then based on the given table, it has the same statistical properties, but produces tabular data with slightly different row values. Therefore, the database needed to train this network consists of two parts, a database of tables and the corresponding statistical properties of these tables. Each step of this algorithm is described in Figure 1.

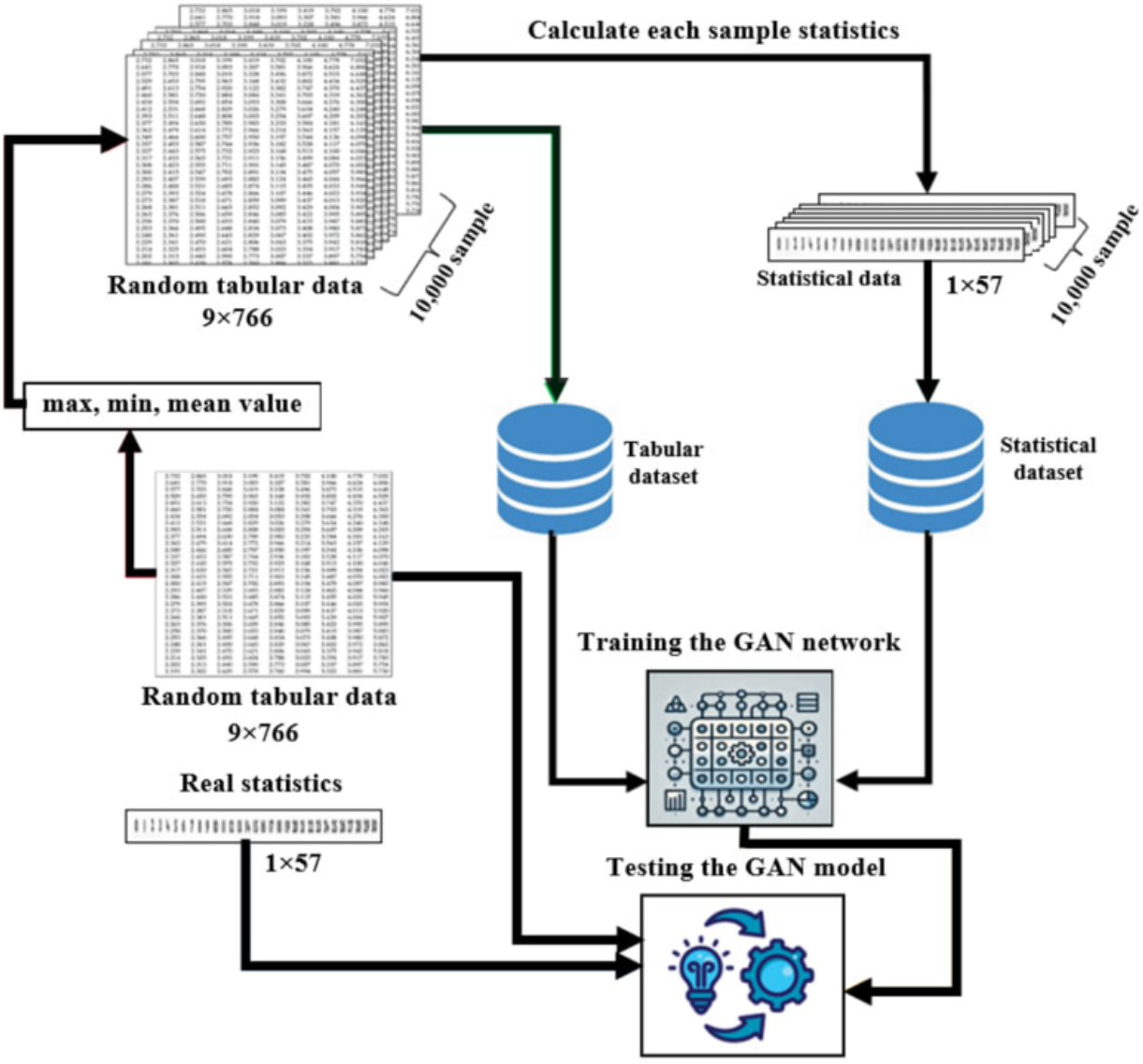


Figure 1. Steps of generating synthetic tabular data.

The first step is generating a dataset of statistical information. Since the tabular data are not given, first of all, random tabular data are generated. However, in order for the network to learn faster, the randomness of the data must be limited. This indicates that statistics that are similar or at least close to those contained in real datasets can be utilized to generate these raw data. For example, maximum, minimum and mean values of each column, percentage of elements with values greater than or less than a threshold value, etc., can be used to create raw tabular data. However, during this process, it should be noted that there will be at least one datum for each class.

More precisely, if a real dataset has y target classes, the outcome/target column of synthetic data must contain at least one element for each class. After generating random tabular data, all tabular data will be stored in separate files to compose the tabular database. Subsequently, the statistics of each table in the generated tabular dataset will be generated. All the values of the generated statistics will be combined into a single row and this row will be stored in a .csv file. As a proof of concept, the real dataset and its corresponding statistical information were taken. These real data and their statistics are taken away for testing the network performance, or as a ground truth.

The second step is to train the GAN with two datasets, i.e., tabular and statistical dataset. This network is a modified version of the GAN, which has two inputs. Furthermore, for training this GAN, an additional loss function for the generator network will be used. The details of the modified GAN and additional loss functions are given in Sections 3.3 and 3.4, respectively.

The third step is the final step or evaluation step. For evaluation, different methods are used and their descriptions are provided in Section 4.

3.2. Modified GAN Architecture

A Generative Adversarial Network is made up of two parts, the generator and the discriminator, that are trained together via adversarial training. The discriminator is used to improve its ability to discern between actual and fake data, and the generator takes random noise as input and generates synthetic data close to the real data. Moreover, it attempts to generate data during training that the discriminator is unable to distinguish from real data. In the simplest terms, the two networks are playing a game against one another: the generator wants to generate believable fake data, while the discriminator is trying to differentiate between real and fake data. Over time, the generator produces ever-better data as a result of this adversarial process.

Random noise will constitute the network's input, and by adding noise and sampling from various locations within the target distribution, GAN can produce a variety of data. However, our goal is to produce tabular data featuring all the internal relationships that are precisely like the provided statistical data; in other words, the generator should produce such tabular data that, even if its value differs from real data, the statistical properties should be the same as the given statistical data. In order to achieve this, statistical data were added to the generator's input in addition to random noise as illustrated in Figure 2. It can be seen from the Figure 2 that statistical data are first fed into the generator's input, and after several layers statistical data are fed again. The purpose of inputting twice is to focus the network on the information being entered. This should be done as far as possible after the first layers, because in the subsequent layers the network begins to recover the properties of the table. However, for input after the first layer, the output of the first layer and the statistical data form must be suitable for concatenation. However, this requires the use of very large filters in the first layer, which dramatically increases the number of parameters of the network, so it is advisable to input after two layers instead of one. Concatenated data are then used to produce fake tabular data. Generated and corresponding real data are sequentially entered into the discriminator. Then, the discriminator begins to identify features of both data. Appropriate statistics are entered into the discriminator when reaching a certain nth layer. In this instance, the statistical data and feature map of the last layer are concatenated and transmitted to the next layer. At the

end of the training, the discriminator comes to a conclusion about how different the fake data are from the real ones. This conclusion is presented in the form of a loss function (2). The corresponding loss function (1) is sent to the discriminator and generator, respectively, to back-propagate.

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[logD\left(x^{(i)}\right) + log\left(1 - D\left(G(z^i)\right)\right)\right] \quad (1)$$

Moreover, the loss function of the generator is presented below (2):

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \log(1 - D\left(G(z^i)\right)) + L_{TS} \quad (2)$$

Here, L_{TS} is the custom loss function; we named it the Table similarity (TS) loss function. A comprehensive description of the custom loss function is provided in Section 3.3.

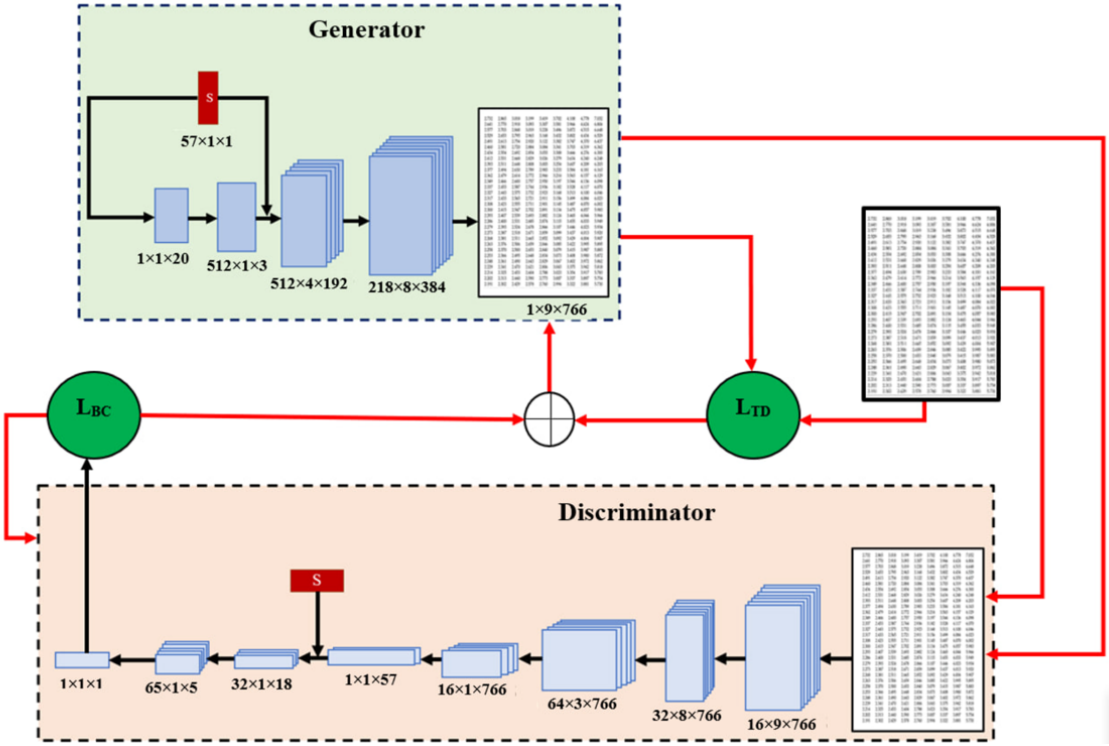


Figure 2. Architecture of modified GAN.

3.3. Custom Loss Function

In conventional GAN Binary Cross Entropy (BCE), a loss function is used. The purpose of using the BCE loss function for the generator is that the discriminator should evaluate the probability of real-likeness of the images generated by the generator. However, in this particular case, not only the probability of real-likeness but also resemblance to the real table should be estimated. It is not required to generate a completely new table that combines the attributes of multiple tables, but rather a table that has precisely the same statistical properties and correlational relationship as the appropriate table. In other word, the purpose of the model is converting statistical data into tabular data.

The output shape of the generator is 9 × 766 in our case. In contrast to previous table-generating GAN approaches, the generator in this network produces a sample table that includes data of multiple patients. In our example, 766 patient data are generated all at once.

Although it looks like a 9 × 766 pixel image, mean squared error (MSE) and mean absolute error (MAE) losses could not be used to evaluate resemblance of real and synthetic samples. The images are compared pixel by pixel; the MSE or MAE loss function is typically employed to assess the similarity of the images. Nevertheless, unlike images, tables should be similar row-wise, not pixel-wise. This is due to the fact that the statistical information of the table remains unchanged even if the rows' positions in the table alter. For this reason, the most optimal way to evaluate the similarity of tables is to compare them row by row.

Tabular data are typically compared using the Kullberg–Leibler (KL) loss function, the Jaccard Index and the cosine embedding loss function. However, employing these functions comes with a variety of issues. Specifically, because the Jaccard index relies on sets, it cannot be utilized directly as a loss function. The loss function for cosine embedding is designed for binary data and is not applicable to non-binary data. The KL loss function is a relatively appropriate loss function. However, the network has trouble converging and is challenging to train when KL is utilized as a loss function. Due to this, we proposed a new loss function to assess the similarity of tables, and we named the new loss function TS loss function. The new loss function is easy to calculate and allows you to compare table rows even when they change their positions. The sequence of its calculation is as follows:

First step—computing the sum for each row: Let the real table be represented as a matrix R and the synthetic table be represented as a matrix S. In addition, where x_{nm} is the element in the *n-th* row and *m-th* column of R, y_{nm} is the element in the *i-th* row and *j-th* column of S. In our case, the matrices have 766 rows and 9 columns.

The sum of R matrix elements in the *n-th* row is given by:

$$r_n = \sum_{m}^{9} x_{nm} \tag{3}$$

The sum of S matrix elements in the *i-th* row is given by:

$$s_i = \sum_{j}^{9} y_{ij} \tag{4}$$

The above summation for each row n and i will be performed, where n = 1, 2, . . ., 9 and i = 1, 2, . . ., 9. Resulting in two rows of sums, r and s.

Second step—computing the sum for each column: The sum of R matrix elements in the *m-th* column is given by:

$$R_m = \sum_{n}^{766} x_{nm} \tag{5}$$

For synthetic data, the sum of elements in the *j-th* column is given by:

$$S_j = \sum_{i}^{766} y_{ij} \tag{6}$$

The above summation for each column m and j will be performed, where m = 1, 2, . . ., 766 and we will receive two rows of sums, R and S.

Third step—defining the mean of vertical subtractions: After computing the sum of columns of each datum, element-wise subtraction will be performed. As a result, the set of subtractions is created, such that $C = \{\left| R_m - S_j \right|\}$. Next, the mean value μ of the subtractions is calculated:

$$\mu = \frac{1}{9} \sum_{k=1}^{9} C_k \tag{7}$$

where C_k represents the *k-th* value in the set C.

Fourth step—calculating minimum subtraction for each row: The module of subtraction of each r_n element from each s_i is calculated: $|r_n - s_i|$. Then for each *n-th* element of r, the minimum value of the subtractions module will be found:

$$\delta_n = \min_i(|r_n - s_i|) \tag{8}$$

Thus, a set M which consists of minimum values δ_n will be formed: $M = \{\delta_1, \delta_2, \ldots, \delta_n\}$, where, n = 1, 2, . . ., 766.

Fifth step—finding percent of repeated minimum value: Now, 766 minimum values are obtained, of which N has equal value. Then, the repeated minimum value percent is equal to $P = N/766$.

Sixth step—defining the mean of minimum values: From the M set, its mean value will be defined:

$$\nu = \frac{1}{766}\sum_{n=1}^{766}\delta_n \tag{9}$$

Seventh step—calculating horizontal loss: Finally, horizontal loss will be found:

$$D = \frac{\nu}{P} \tag{10}$$

Overall loss for one sample will be defined as:

$$L_{onesample} = D + \mu \tag{11}$$

During training $L_{1,onesample}$, $L_{2,onesample}$ and $L_{k,onesample}$ loss values are obtained for each instance in a batch, where k is the number of instances in the batch. The final step is to take the square root of the mean value of the loss function for the batch:

$$L_{TS} = \sqrt{\frac{1}{k}\sum_{l=1}^{k}L_l} \tag{12}$$

3.4. Dataset

The Pima Diabetes Database (PDD) was chosen to use as a dataset in our experiment. The dataset is named after the Pima Indians, a Native American group from the Pima Country region in Arizona, USA. It was collected as part of a study conducted by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) to investigate diabetes prevalence and related factors among this population [27]. The dataset contains 766 samples, with each instance corresponding to a patient record. Each record is described by 8 features, including pregnancies, glucose, blood pressure, skin thickness, insulin, BMI, diabetes pedigree function, age and the binary outcome variable.

The proposed method does not necessitate the use of an actual database; instead, it relies on statistical information about the database. However, for the purpose of evaluating the generated database, a real database is used. More precisely, its statistical information was extracted and used to train the network. A tabular database is employed to test the network and for comparative evaluation. The primary reason for selecting this database was that the model is designed to generate a small-scale database.

From this database, the maximum and minimum values, percentage of risky values and OR value for each column were determined. A total of 57 values were extracted. A dataset of 10,000 statistics that differed by 1% from that statistics was then generated. For this, the random generator from the Python library was used.

As a result of this, the dataset consisting of 10,000 data, each with 766 rows and 9 columns, was generated. These dataset was almost, similar to the PIMA dataset, or rather, the statistics differ by 1% and this dataset is then called the primary dataset.

4. Evaluation Criteria

Once the synthetic data are developed, assessing their quality is the most crucial stage. Even though a universal and standardized assessment methodology for synthetic data quality evaluation has not yet been established, several evaluation metrics and methodologies have been proposed and used [28,29]. Based on their intended purpose, the evaluation techniques used to evaluate the quality of the data can be categorized into three classes:

1. Fidelity-assessment metrics.
2. Utility-assessment metrics.
3. Privacy-assessment metrics.

However, two metrics have been utilized in this work, and the reason of selection is provided in the "Results and Discussion" section.

4.1. Fidelity-Assessment Metrics

There are numerous metrics proposed and used to evaluate fidelity [29], and all of them are categorized into three groups: statistical test for numerical attributes, statistical test for categorical attributes and distance calculation. Every category includes many metrics. One of the most significant and commonly used method is considered in this work. More precisely,

1. For statistical test for numerical attributes: Student's *t*-test.
2. For statistical test for categorical attributes: Chi-square (χ^2) test.
3. For distance calculation: Jensen–Shannon divergence.

4.1.1. Student's *t*-Test for the Comparison of Means

The formula for Student's *t*-test is defined as (13):

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}} \tag{13}$$

where t = Student's *t*-test, x_1 = mean of first group, x_2 = mean of second group, σ_1 = standard deviation of group 1, σ_2 = standard deviation of group 1, n_1 = number of observations in real dataset, n_2 = number of observations in synthetic dataset.

In cases of multiple-column tabular data, a separate *t*-test should be carried out for every column. In ref. [30], it is mentioned that the critical value can be chosen as 0.05. This means that if the test's *p*-value is greater than this amount, it can be considered as a mean of real data and the synthetic features are similar. Otherwise, the attributes of the two datasets will be different from each other.

4.1.2. Chi-Square Test

When comparing the independence of categorical features, Chi-square analysis is employed. If the synthetic data attributes maintain the characteristics of the real data attributes, the null hypothesis is accepted and for this the *p*-value was set as 0.05. The formula of the Chi-square test is given below:

$$\aleph^2 = \sum \frac{(S-R)^2}{R} \tag{14}$$

where S is statistical data and R is real data.

4.1.3. Wasserstein Distance

The Wasserstein distance estimates the distance, in a given metric space, between two probability distributions. The *p-th* Wasserstein distance is given by:

$$Wass_dist(r,s) = \int_{-\infty}^{+\infty} |R - S| \tag{15}$$

R and S are the cumulative distribution functions of the real and synthetic datasets, respectively.

4.1.4. Univariate Resemblance Assessment (URA)

In ref. [29], the authors introduced a verbal evaluation method of the generative model performance. Firstly, statistical tests and distances are calculated for every feature of the dataset, and then the threshold value is used. It is determined whether each character meets the h0 hypothesis or not. After that, the number of features that fulfill the requirements is counted. The generative model is labelled "Excellent" if this number is higher than 50% percent of the number of all the features. If less than 50% percent, it is labelled "Good"; if the number is zero, then the model is labelled "Poor". After labelling the model based on each metric, a URA score is calculated. For this purpose, labels are converted to numerical values as follows: "Excellent" = 3, "Good" = 2, "Poor" = 1. And next, every number is multiplied by a proportionality coefficient.

The proportionality coefficient is defined according to this formula:

$$M = \frac{1}{n} \tag{16}$$

where n is the number of metrics used to evaluate fidelity.

In our case n = 3, so the converted numbers will be multiplied by 0.33 and thesum of all multiplications will be calculated. The outcomes are first rounded, and the resulting numerical value is subsequently determined. This numerical value is then translated into descriptive categories as follows: "Poor" for a value of 1, "Good" for a value of 2 and "Excellent" for a value of 3.

4.2. Utility-Assessment Metrics

Usually, the synthetic dataset is mainly used to train machine learning or deep learning algorithms. For this reason, Train on Synthetic, Test on Real (TSTR), Train on Real, Test on Real (TRTR) and Train on Real, Test on Synthetic (TRTS) methods are used to evaluate the usefulness of synthetic data. These methods are described in detail in the following:

The TSTR metric involves training a machine-learning model on synthetic data and then testing its performance on real data [31]. TSTR is a more important metric than TRTS because the purpose of generating synthetic data is using them for training AI algorithms. For this reason, the TRTS method was not used to evaluate utility.

Obviously, the accuracy of the model depends not only the dataset, but also on the Machine Learning (ML) model. This means that valuating synthetic data utility only with one ML model is insufficient. For this reason, several ML models were used to assess the utility metric, namely, Random Forest (RF) classifier (1), k-Nearest Neighbors (KNN) (2), Decision Tree (DT) classifier (3) and Multilayer Perceptron (MLP) classifier (4). The parameters for these models were configured as follows:

1. RF: Number of estimators = 88, Maximum depth = 11, Random state = 28 and Criterion = "Gini". All other parameters were set to their defaults.
2. KNN: Number of neighbors = 4, Metric = cosine value. All other parameters were set to their defaults.
3. DT: Maximum depth = 7, Random state = 27, Maximum features = 2. All other parameters were set to their defaults.

4. MLP: Hidden layer sizes = (32,128,256), Maximum iteration = 30, Random state = 9, Verbose = 1. All other parameters were set to their defaults. The training process stopped if training loss did not improve more than 0.0001 for 10 consecutive epochs.

After training ML models, in order to evaluate model-classification performance traditional metrics, accuracy, precision, recall and F1-score were used.

4.3. Comparison with Other Methods

To assess the quality of the dataset generated by the proposed method, a comparative analysis was conducted. For this purpose, three synthetic database-generation methods are selected on the basis of their effectiveness. These methods include:

GM—a technique based on modeling the joint probability distribution of the data using multivariate Gaussian distributions.

SDV—a framework that leverages probabilistic modeling and statistical techniques to generate synthetic datasets while preserving the original data's structural and statistical properties.

CTGAN—a generative model that employs GANs tailored for tabular data, enabling the generation of realistic datasets even in the presence of imbalanced and categorical data.

5. Results and Discussion

The training of the GAN is performed using the Adam optimizer. To mitigate the risk of mode collapse, the learning rate for the discriminator was set to 0.00005, while the learning rate for the generator was set to 0.0001. The batch size was fixed at 64. As outlined previously, two distinct datasets—statistical and tabular databases—were employed for training, each comprising 10,000 samples.

Throughout the training process, the GAN's performance was continuously monitored to evaluate the network's progress, and scatterplot graphs were generated to illustrate the correlation between two specific features (the fifth and seventh) from the tables produced in each batch. These scatterplots were then visually compared against those generated from the real datasets to assess the similarity between the real and synthetic data.

The network demonstrated improvements in performance up until the 70th epoch (10,920th iteration) (Figure 3). Beyond this point, the quality of the generated data began to decline; even Loss values remained the same. This prompted the decision to terminate the training at the 70th epoch to prevent overfitting and ensure optimal model performance.

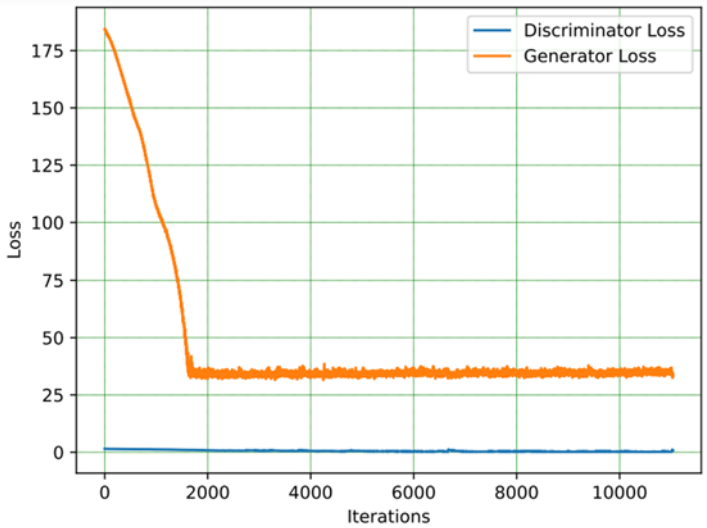


Figure 3. Results of generator and discriminator loss during training.

After the training process, the next and final step was testing the performance of the generator. As a ground truth, the real PIMA diabetes dataset is used. Three criteria were

used to evaluate the performance—fidelity, utility and privacy. The outcomes of these assessment techniques were as follows.

For fidelity evaluation of synthetic data, Student's *t*-test, Chi-square and Jensen–Shannon distance estimation methods were utilized for every attribute of data.

In Table 1, the *p*-values derived from the statistical *t*-tests for each column are given. From the table, it can be seen that all features generated by SDV have *p*-values of Student's *t*-test greater than 0.05, indicating similarities in mean values as real dataset features. *p*-values for 7 out of 8 features are higher than the threshold value; that is, GM also generated data with mean values very close to the real dataset; seven of the eight features generated by GM had *p*-values of *t*-test higher than the threshold value, which shows that they also have mean values close to the real data features' means. In the data generated by the proposed method and CTGAN, only two features satisfied the required condition. The *t*-test showed that only a few features' mean values were simulated closely using the proposed method and CTGAN. However, it is indicated in Section 4 that if less than half of the features' *p*-values match the condition, the result of the test can be classified as "Good"; therefore, we rate the performance of the proposed and CTGAN methods as "Good" and the performance of the GM and SDV as "Excellent".

Table 1. Student's *t*-test results.

Attributes\Model Names	Proposed Model	GM	SDV	CTGAN
Pregnancies	0.0000	0.0271	**0.4539**	0.0000
Glucose	0.0000	**0.9834**	**0.6115**	0.0000
Blood Pressure	0.0000	**0.3075**	**0.8030**	**0.1696**
Skin Thickness	**0.1127**	**0.8747**	**0.9245**	**0.6030**
Insulin	0.0000	**0.8664**	**0.6294**	0.0000
BMI	0.0019	**0.7159**	**0.8157**	0.0000
Diabetes Pedigree Function	0.0009	**0.8929**	**0.1311**	0.0000
Age	**0.3811**	**0.3766**	**0.6342**	0.0010

The *p*-values acquired for every category attribute are displayed in the Chi-square test results. It can be seen from Table 2 that for all methods *p*-values were above the established critical level (>0.05), showing the acceptance of the alternative hypothesis ($h1$) and a statistical relationship between the two features. The performance was labelled "Excellent", since the alternative hypothesis was accepted for every categorical attribute produced by every technique. The last point that should be noted is that our proposed method earned the highest score.

Table 2. Chi-square test results.

Model Names	Proposed Model	GM	SDV	CTGAN
Outcome	**0.977744**	0.9486	0.8987	0.901

Table 3 shows the proposed and CTGAN techniques which produced data with Wasserstein distances less than 0.1 for every feature. Only one feature of the GM generated dataset and two features of the SDV generated dataset had a Wasserstein distance bigger than the threshold value. As previously stated, even though not all features fulfill the conditions set forth concerning closeness, it is enough that more than half of the features meet the criteria in order to classify the data as "Excellent". For this reason, all approaches were categorized as "Excellent" for Wasserstein distances.

Table 3. Wasserstein distance values.

Attributes\Model Names	Proposed Model	GM	SDV	CTGAN
Pregnancies	**0.077049**	**0.050503**	0.210039	**0.090343**
Glucose	0.139552	**0.029401**	**0.022376**	0.211049
Blood Pressure	0.119209	**0.077486**	**0.054027**	0.135793
Skin Thickness	0.166620	**0.082580**	0.192674	**0.063013**
Insulin	**0.047145**	0.381628	0.444365	0.161786
BMI	0.155688	**0.026149**	**0.053898**	0.111652
Diabetes Pedigree Function	0.176867	**0.051625**	0.125081	**0.057591**
Age	**0.086909**	**0.022935**	0.385533	**0.031314**

In Table 4, labels are provided for each fidelity metric, along with the overall URA performance. With the exception of CTGAN and the proposed method, every technique preserved the univariate characteristics of real data accurately. Even the proposed approach exhibits a worse performance, with Numerical Statistical Test total URA values similar to GM and SDV models. This means that using the proposed method, it is possible to create data whose statistical properties are very similar and close to those of real data.

Table 4. Fidelity assessment's overall results.

Approach	Numerical Statistical Tests	Categorical Statistical Tests	Distance Calculation	Total URA
Proposed	Good	Excellent	Good	Good
GM	Excellent	Excellent	Excellent	Excellent
SDV	Excellent	Excellent	Excellent	Excellent
CTGAN	Good	Excellent	Excellent	Excellent

For the utility-evaluation TSTR test, RF, KNN, DT and MLP models are used. For every model, accuracy, precision, recall and F1 scores were calculated. Tables 5 and 6 present the information about the TSTR and TRTR test results.

Table 5. Utility assessment with KNN and DT.

Method Name	KNN				DT			
	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)
Real	0.6623	0.6388	0.6623	0.6256	0.7078	0.709	0.7078	0.7084
GM	0.6234	0.5839	0.6234	0.5824	0.6818	0.6739	0.6818	0.6764
SDV	0.5065	0.5511	0.5065	0.5167	0.5649	0.6167	0.5649	0.5735
CTGAN	0.5584	**0.6384**	0.5584	0.5627	0.4545	0.4954	0.4545	0.4659
Proposed model	**0.6494**	0.6304	**0.6494**	**0.6326**	**0.6883**	**0.6859**	**0.6883**	**0.6873**

Table 6. Utility assessment with RF and MLP.

Method Name	RF				MLP			
	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)
Real	0.7532	0.7554	0.7532	0.7542	0.6688	0.648	0.6688	0.6307
GM	**0.7468**	0.7426	**0.7468**	**0.7438**	0.5649	0.6122	0.5649	0.5738
SDV	0.6494	**0.7648**	0.6494	0.6494	0.3636	0.7712	0.3636	0.2017
CTGAN	0.3312	0.388	0.3312	0.2812	0.4545	0.5942	0.4545	0.4246
Proposed	0.7078	0.6966	0.7078	0.6926	**0.6558**	**0.6311**	**0.6558**	**0.5831**

The first row of the table shows TRTR results for every ML algorithm. From Tables 5 and 6, it can be seen that, when trained using the dataset produced by the proposed method, three algorithms, namely KNN, DT and MLP, achieved the maximum performance. Thus, the proposed method gained the highest TSTR value in 3 out of 4 tests. Only in one case, i.e., RF, GM took first place in term of accuracy, recall and F1-score. However, the proposed method also performed well when TSTR was tested with the RF model, placing second among all models behind GM. The CTGAN model, whose architecture is closest to the proposed method, showed the lowest result in all cases. To analyze results visually, all results are plotted in the form of a bar graph (Figure 4). Concerning the variances in accuracy, precision, recall and F1-score values between the TRTR, it can be seen from the pictures that almost all values of the dataset created by the proposed method were the lowest when checked by DT, KNN and MLP methods. However, only when testing with the KNN model, the difference in precision values between TRTR and TST were marginally smaller than that of the proposed technique for CTGAN. Furthermore, the GM approach produced the smallest values for all parameters only in cases of the RF method test. Nevertheless, even when the dataset generated by the proposed method was tested by the RF method, the difference of all metrics of TSTR and TRTR was less than 0.1, which is a very acceptable outcome.

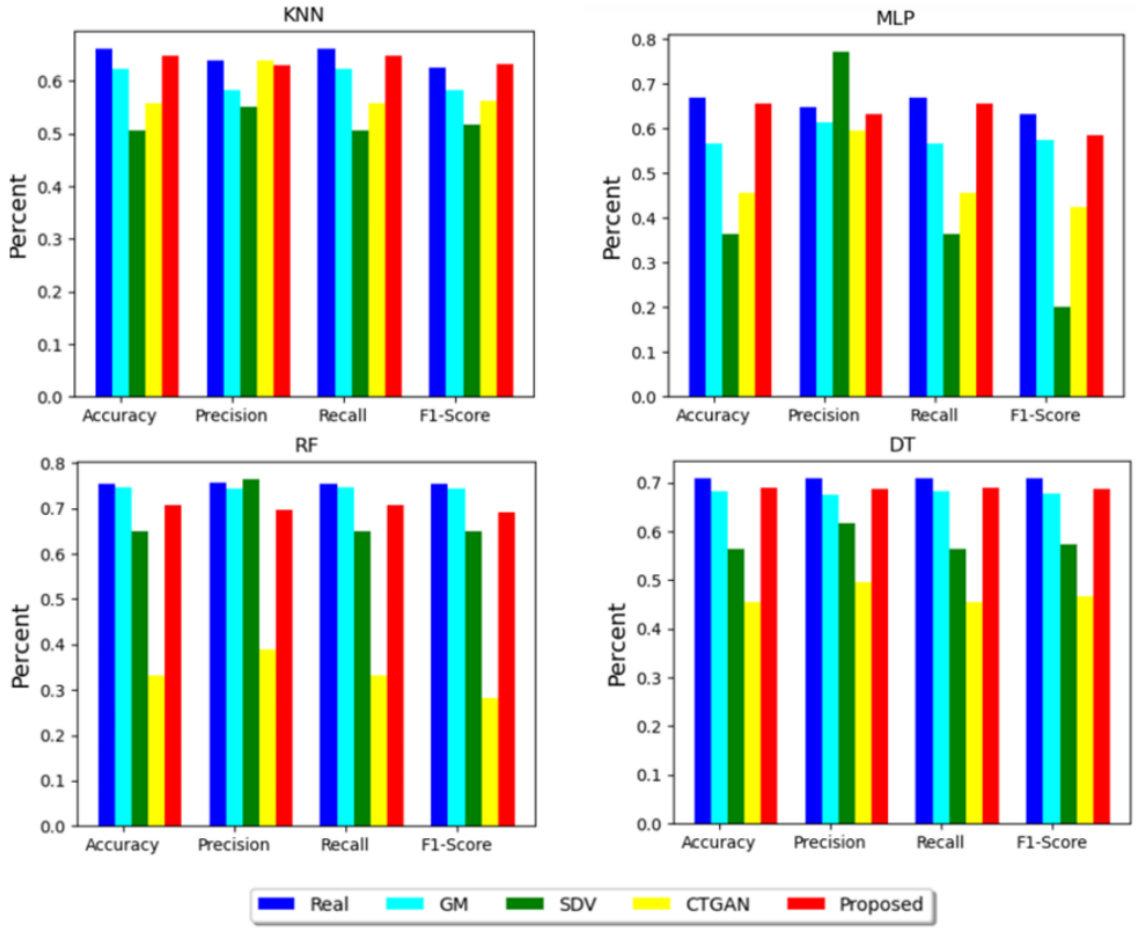


Figure 4. Results of TSTR test carried out with different ML algorithms.

The Figure 4 shows that the approach is categorized as the "Excellent" if the difference in the test performance between TSTR and TRTR is less than 0.4. In our instance, the differences between all models (CTGAN excluded) on all metrics were less than 0.4 across all tests (KNN, DT, RF and MLP). CTGAN also performed well in 3 out of 4 cases, but obtained poor results when tested by the RF method. Therefore, the utility metrics of all methods were classified as "Excellent" for all approaches.

From the above results, the proposed method demonstrates promising results in terms of both similarity and usefulness. While it performs on par with the CTGAN algorithm in terms of similarity and generates a synthetic table satisfactorily similar to the real one under the URA criterion, it falls slightly behind the GM and SDV methods. However, its ability to outperform other methods in three out of four usefulness tests highlights its strength in accurately representing relationships between variables. Unlike GM, which can produce negative values that disrupt variable relationships, the proposed method maintains consistency, contributing to its superior usefulness scores. GM also showed results close to the proposed method, but usually there are negative values in the table generated using GM, and if one filters them using the thresholding method, they can still give values close to real, but they affect the relationships between variables. Therefore, the GM method showed a slightly lower result.

Since the proposed method does not directly use a real table to generate the synthetic table, this method is not evaluated according to the security criteria. Because it does not rely on real tables or personal data, ensuring that no actual patient information is stored or processed. So, from a security perspective, the method has significant advantages, this makes it a robust choice for privacy-preserving synthetic data generation.

This proposed method was the first step towards opening up an area with great potential for the future. By enabling the generation of realistic tabular datasets from statistical data, these algorithms bridge a crucial gap, providing accessible, privacy-compliant and cost-effective alternatives to real-world datasets across diverse applications. The availability of extensive medical, economic and financial databases for training machine learning algorithms has the potential to significantly accelerate advancements in research within these fields, fostering unprecedented progress and innovation. However, much work needs to be conducted in the future for this area to develop. The proposed method showed the better results in terms of fidelity and usefulness as compared to other existing methods. In particular, the network is trained only to recognize specific statistical data and convert them into a table. Each time different types of statistical data are used, one has to train the network again. Furthermore, as the number of variables in the database increases, training the network also becomes more difficult. The network architecture needs to be changed to make it relatively easier to train. The method proposed in this article is designed only for creating a small database; the same method can be adapted for a large database, and in our future work we aim to consider the issues of creating a large database.

6. Conclusions and Future Work

This article introduced a novel approach for generating synthetic tabular datasets derived from statistical data. The proposed method is evaluated based on fidelity and utility characteristics. While not all attributes of the synthetic table met the fidelity requirements for replicating the real table, the overall similarity is rated as good. In terms of utility, the method excelled, achieving good results in the KNN, DT and MLP tests. The TSTR accuracy differed minimally from that of the real database, with deviations of only 0.12%, 0.19% and 0.13% for KNN, DT and MLP, respectively. These utility scores surpassed those of all three compared models, highlighting the proposed method as a superior choice for training machine learning models.

Despite the aforementioned advantages, several improvements can be implemented in future work: enhancement of network architecture, adaptation of the algorithm to generate large-scale databases and optimization of network parameters to facilitate efficient training, even as the number of variables in the table increases.

Author Contributions: This manuscript was written and designed by N.N., R.N., A.A. and M.R. who analyzed, validated the proposed model. Y.-I.C., S.M. and G.T. supervised the study and contributed to the analysis and discussion of the algorithm and experimental results. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is supported by Korea Institute of Marine Science & Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (G22202202102401), and by Korean Agency for Technology and Standard under Ministry of Trade, Industry and Energy in 2023, project number is 1415181638 ("Establishment of standardization basis for BCI and AI Interoperability").

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Winter, J.S. AI in healthcare: Data governance challenges. *J. Hosp. Manag. Health Policy* **2020**, *5*. [CrossRef]
2. Appenzeller, A.; Leitner, M.; Philipp, P.; Krempel, E.; Beyerer, J. Privacy and utility of private synthetic data for medical data analyses. *Appl. Sci.* **2022**, *12*, 12320. [CrossRef]
3. Giuffrè, M.; Shung, D.L. Harnessing the power of synthetic data in healthcare: Innovation, application, and privacy. *NPJ Digit. Med.* **2023**, *6*, 16. [CrossRef]
4. Goyal, M.; Mahmoud, Q.H. A Systematic Review of Synthetic Data Generation Techniques Using Generative AI. *Electronics* **2024**, *13*, 3509. [CrossRef]
5. Clinical Practice Research Datalink. CPRD Cardiovascular Disease Synthetic Dataset (Version 2020.06.001) [Data Set]. Clinical Practice Research Datalink. 2020. Available online: https://www.cprd.com/cprd-cardiovascular-disease-synthetic-dataset (accessed on 12 October 2024).
6. Clinical Practice Research Datalink. CPRD COVID-19 Symptoms and Risk Factors Synthetic Dataset December 2021 (Version 2021.12.001) [Data Set]. Clinical Practice Research Datalink. 2021. Available online: https://www.cprd.com/cprd-covid-19-symptoms-and-risk-factors-synthetic-dataset-december-2021 (accessed on 12 October 2024).
7. Guyet, T.; Allard, T.; Bakalara, J.; Dameron, O. An open generator of synthetic administrative healthcare databases. In Proceedings of the IAS 2021—Atelier Intelligence Artificielle et Santé, Bordeaux, France, 28 June 2021; pp. 1–8, ffhal-03326618f. Available online: https://hal.science/hal-03326618/document (accessed on 12 October 2024).
8. Wang, T.; Zhao, Z.; Wang, G.; Li, Q.; Xu, Y.; Li, M.; Hu, R.; Chen, G.; Su, Q.; Mu, Y.; et al. Age-related disparities in diabetes risk attributable to modifiable risk factor profiles in Chinese adults: A nationwide, population-based, cohort study. *Lancet Healthy Longev.* **2021**, *2*, e618–e628. [CrossRef] [PubMed]
9. Lee, Y.-T.; Lin, R.S.; Sung, F.C.; Yang, C.-Y.; Chien, K.-L.; Chen, W.-J.; Su, T.-C.; Hsu, H.-C.; Huang, Y.-C. Chin-Shan Community Cardiovascular Cohort in Taiwan–baseline data and five-year follow-up morbidity and mortality. *J. Clin. Epidemiol.* **2000**, *53*, 838–846. [CrossRef] [PubMed]
10. Vashist, P.; Senjam, S.S.; Gupta, V.; Manna, S.; Gupta, N.; Shamanna, B.R.; Bhardwaj, A.; Kumar, A.; Gupta, P. Prevalence of diabetic retinopahty in India: Results from the National Survey 2015-19. *Indian J. Ophthalmol.* **2021**, *69*, 3087–3094. [CrossRef] [PubMed] [PubMed Central]
11. Han, K.; Kim, B.; Lee, S.H.; Kim, M.K. A nationwide cohort study on diabetes severity and risk of Parkinson disease. *NPJ Park. Dis.* **2023**, *9*, 11. [CrossRef] [PubMed] [PubMed Central]
12. Wang, A.X.; Chukova, S.S.; Simpson, C.R.; Nguyen, B.P. Challenges and opportunities of generative models on tabular data. *Appl. Soft Comput.* **2024**, *166*, 112223. [CrossRef]
13. Xu, L.; Skoularidou, M.; Cuesta-Infante, A.; Veeramachaneni, K. Modeling tabular data using conditional GAN. In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Curran Associates Inc.: Red Hook, NY, USA, 2019. Article 659. pp. 7335–7345.
14. Kiran, A.; Kumar, S.S. A Comparative Analysis of GAN and VAE based Synthetic Data Generators for High Dimensional, Imbalanced Tabular data. In Proceedings of the 2023 2nd International Conference for Innovation in Technology (INOCON), Bangalore, India, 3–5 March 2023; pp. 1–6. [CrossRef]
15. Ling, Y.; Jiang, X.; Kim, Y. MALLM-GAN: Multi-Agent Large Language Model as Generative Adversarial Network for Synthesizing Tabular Data. *arXiv* **2024**, arXiv:2406.10521.
16. Li, X.; Metsis, V.; Wang, H.; Ngu, A.H.H. Tts-gan: A transformer-based time-series generative adversarial network. In *International Conference on Artificial Intelligence in Medicine*; Springer International Publishing: Cham, Switzerland, 2022.
17. Fonseca, J.; Bacao, F. Tabular and latent space synthetic data generation: A literature review. *J. Big Data* **2023**, *10*, 115. [CrossRef]
18. Panagiotou, E.; Roy, A.; Ntoutsi, E. Synthetic Tabular Data Generation for Class Imbalance and Fairness: A Comparative Study. *arXiv* **2024**, arXiv:2409.05215.
19. Pezoulas, V.C.; Zaridis, D.I.; Mylona, E.; Androutsos, C.; Apostolidis, K.; Tachos, N.S.; Fotiadis, D.I. Synthetic data generation methods in healthcare: A review on open-source tools and methods. *Comput. Struct. Biotechnol. J.* **2024**, *23*, 2892–2910. [CrossRef] [PubMed] [PubMed Central]
20. Gonzales, A.; Guruswamy, G.; Smith, S.R. Synthetic data in health care: A narrative review. *PLoS Digit. Health* **2023**, *2*, e0000082. [CrossRef] [PubMed] [PubMed Central]

21. Smania, G.; Jonsson, E.N. Conditional distribution modeling as an alternative method for covariates simulation: Comparison with joint multivariate normal and bootstrap techniques. *CPT Pharmacomet. Syst. Pharmacol.* **2021**, *10*, 330–339. [CrossRef]
22. Kharya, S.; Soni, S.; Swarnkar, T. Generation of synthetic datasets using weighted bayesian association rules in clinical world. *Int. J. Inf. Technol.* **2022**, *14*, 3245–3251. [CrossRef]
23. Kiagias, D.; Russo, G.; Sgroi, G.; Pappalardo, F.; Juárez, M.A. Bayesian augmented clinical trials in TB therapeutic vaccination. *Front. Med. Technol.* **2021**, *3*, 719380. [CrossRef]
24. Walonoski, J.; Kramer, M.; Nichols, J.; Quina, A.; Moesel, C.; Hall, D.; Duffett, C.; Dube, K.; Gallagher, T.; McLachlan, S. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *J. Am. Med. Inform. Assoc.* **2017**, *25*, 230–238. Erratum in *J. Am. Med. Inform. Assoc.* **2018**, *25*, 921. [CrossRef] [PubMed] [PubMed Central]
25. Tagmatova, Z.; Abdusalomov, A.; Nasimov, R.; Nasimova, N.; Dogru, A.H.; Cho, Y.-I. New Approach for Generating Synthetic Medical Data to Predict Type 2 Diabetes. *Bioengineering* **2023**, *10*, 1031. [CrossRef]
26. Nasimov, R.; Nasimova, N.; Mumimov, B.; Usmanxodjayeva, A.; Sobirova, G.; Abdusalomov, A. Development of Fully Synthetic Medical Database Shuffling Method. In *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*; NEW2AN ruSMART 2023; Lecture Notes in Computer Science; Koucheryavy, Y., Aziz, A., Eds.; Springer: Cham, Switzerland, 2024; Volume 14543. [CrossRef]
27. Schulz, L.O.; Bennett, P.H.; Ravussin, E.; Kidd, J.R.; Kidd, K.K.; Esparza, J.; Valencia, M.E. Effects of traditional and western environments on prevalence of type 2 diabetes in Pima Indians in Mexico and the US. *Diabetes Care* **2006**, *29*, 1866–1871. [CrossRef]
28. Abdusalomov, A.B.; Nasimov, R.; Nasimova, N.; Muminov, B.; Whangbo, T.K. Evaluating Synthetic Medical Images Using Artificial Intelligence with the GAN Algorithm. *Sensors* **2023**, *23*, 3440. [CrossRef] [PubMed]
29. Hernadez, M.; Epelde, G.; Alberdi, A.; Cilla, R.; Rankin, D. Synthetic Tabular Data Evaluation in the Health Domain Covering Resemblance, Utility, and Privacy Dimensions. *Methods Inf. Med.* **2023**, *62*, e19–e38. [CrossRef] [PubMed] [PubMed Central]
30. Yoon, J.; Drumright, L.N.; van der Schaar, M. Anonymization through data synthesis using generative adversarial networks (ADS-GAN). *IEEE J. Biomed. Health Inform.* **2020**, *24*, 2378–2388. [CrossRef] [PubMed]
31. Esteban, C.; Hyland, S.L.; Rätsch, G. Real-valued (medical) time series generation with recurrent conditional gans. *arXiv* **2017**, arXiv:1706.02633.

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.

Volume 44, 2024

Electromagnetics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uemg20

High isolation dual band circularly polarized CPW based MIMO antenna with improved diversity performance for WLAN and C band wireless applications

Rakesh Kumar, Asok De & Priyanka Jain

To cite this article: Rakesh Kumar, Asok De & Priyanka Jain (07 Dec 2024): High isolation dual band circularly polarized CPW based MIMO antenna with improved diversity performance for WLAN and C band wireless applications, Electromagnetics, DOI: 10.1080/02726343.2024.2428324

To link to this article: https://doi.org/10.1080/02726343.2024.2428324

Published online: 07 Dec 2024.

Submit your article to this journal

Article views: 2

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=uemg20

ELECTROMAGNETICS
https://doi.org/10.1080/02726343.2024.2428324

Taylor & Francis
Taylor & Francis Group

Check for updates

High isolation dual band circularly polarized CPW based MIMO antenna with improved diversity performance for WLAN and C band wireless applications

Rakesh Kumar, Asok De, and Priyanka Jain

Department of Electronics and Communication Engineering, Delhi Technological University (DTU), New Delhi, India

ABSTRACT

This paper presents a 2-port MIMO antenna with a CPW feed, designed for dual-band operation with high isolation. We resolved bandwidth and polarization issues in the antenna's initial CPW-fed design by modifying the radiating patch and incorporating an asymmetrical ground plane. These adjustments enabled dual-band and dual-polarization functionality. The first band, with linear polarization, operates between 2.75 GHz and 2.86 GHz for sub-6 GHz wireless applications, while another band exhibits circular polarization with wide-band resonating at two frequencies, in the (4.7–5.6) GHz and (7.1–8.5) GHz ranges, respectively, covering the entire C-band and suitable for wide-band applications. The antenna, built on a Rogers-5880 substrate (1.57 mm thickness, $\varepsilon_r = 2.2$, tan $\delta = 0.0009$), offers low loss and excellent thermal stability. The antenna, measuring $45.5 \times 27 \times 1.57\ mm^3$, attains port isolations of 27 dB and 22.4 dB, with peak realized gains of 2.8 dB and 5.4 dB in the two bands. Furthermore, we analyze the key performance of the MIMO antenna, including the envelope correlation coefficient (ECC), diversity gain (DG), total active reflection coefficient (TARC), and channel capacity loss (CCL), and confirm that they fall within an acceptable range, making it an excellent choice for many portable wireless applications.

ARTICLE HISTORY
Received 24 August 2024
Revised 29 October 2024
Accepted 5 November 2024

KEYWORDS
5G; circularly polarized; CPW; dual band; high isolation; MIMO; wide band

Introduction

In the rapidly evolving landscape of wireless communication, the demand for high data rates, seamless connectivity, and reliable transmission has driven significant advancements in antenna technology. MIMO systems have emerged as a critical component of modern communication networks due to their ability to enhance channel capacity and signal quality (Ishteyaq and Muzaffar 2022). Among the various MIMO configurations, the 2×1MIMO antenna stands out for its balance between performance and complexity, making it particularly suitable for compact devices and applications requiring efficient spatial diversity (Jamshed et al. 2021). One of the primary challenges in contemporary wireless communication is to design a wideband antenna with reasonably adequate gain. Wideband MIMO antennas make it possible

CONTACT Priyanka Jain priyankajain@dtu.ac.in Department of Electronics and Communication Engineering, Delhi Technological University (DTU), New Delhi, India

© 2024 Taylor & Francis Group, LLC

to send large amounts of data over a wide range of frequencies. This is important for uses that need to send a lot of data quickly and reliably across multiple frequency bands (Yuan et al. 2020). Additionally, dual-band MIMO antennas offer enhanced versatility by operating in two distinct frequency bands, supporting multiple-communication standards simultaneously, such as Wi-Fi and 5 G, and thus optimizing spectrum resource utilization (Huang et al. 2021). Circular polarization (CP) and linear polarization (LP) play a crucial role in antenna design to improve performance. Circularly polarized antennas exhibit reduced vulnerability to orientation mismatch and multipath fading, making them highly suitable for dynamic situations characterized by the varying orientations of transmitting and receiving antennas. This function is especially pertinent in mobile devices and satellite communications, as the positioning of the antenna can have a major impact on performance (Ullah, Mabrouk, and Koziel 2020). Conversely, linearly polarized antennas offer simple designs and are easier to include, especially in situations where the orientation remains constant. The integration of circular polarization (CP) and linear polarization (LP) in a dual-polarized MIMO antenna system capitalizes on the advantages of both techniques, ensuring resilience against multipath interference while maintaining a simple structure. Various decoupling techniques are available to improve the isolation between the elements of the MIMO antenna. One such technique involves placing an electronic bandgap (EBG) structure between the antennas, which creates a stopband to block electromagnetic waves, thereby improving the isolation (Parvathi and Gupta 2021). We extensively use metamaterial absorbers (MA) to enhance the characteristics of various microwave components (Puri and Singh 2022). In addition, we can introduce a Neutralization Line (NL) to create an antiphase current between the array's elements (Li, Jiang, and Yeung 2020). We can also use Defected Ground Structure (DGS) to enhance isolation, as it influences the antenna's backward radiation pattern (Xing et al. 2020).

Antenna technology for WLAN and C-band applications has significantly advanced in recent years. For instance, researchers have optimized cavity-backed slot antennas to enhance isolation in full-duplex systems (Benkhadda et al. 2023). Additionally, researchers have increasingly employed methods such as the Substrate Integrated Waveguide approach, slotted ground technique, and meander line designs to develop dual-band miniaturized printed monopole antennas for WiFi and wireless local area networks, owing to their improved gain and efficiency (Behera, Midya, and Mishra 2024; Gad and Sallam 2024; Turan and Odabasi 2023). Nevertheless, planar CPW feed antennas remain widely favored because of their compactness.

This paper introduces, designs, and validates a compact CPW feed 2×1MIMO antenna. The suggested antenna is capable of operating in the sub-6 GHz and C band frequency range, which is ideal for Wi-Fi, radar systems, LTE applications, and satellite communication. We organize the paper as follows. Section II outlines the design approach for a single element of the MIMO system along with its simulated

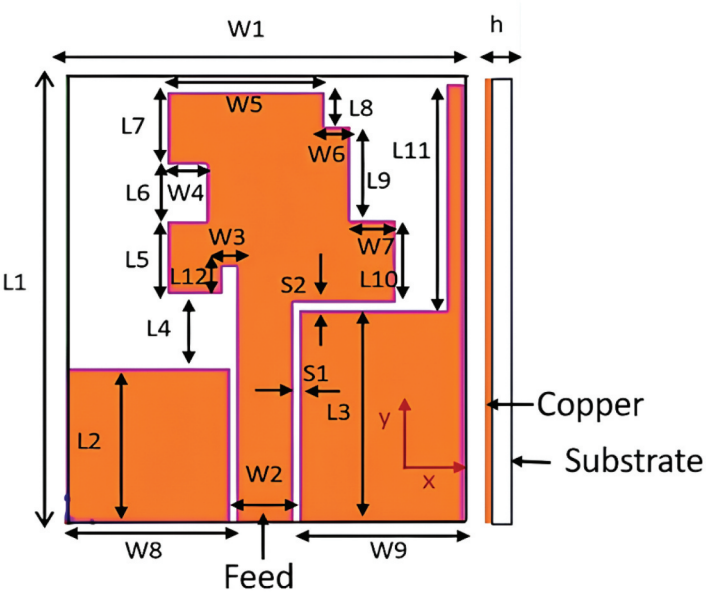
results. Section III discusses the suggested MIMO antenna's performance along with the simulated and experimental results of the return loss and radiation pattern plots. Section IV presented the diversity performance of the recommended MIMO antenna. Section V compares the performance of the proposed MIMO antenna with recent research. Section VI presents the conclusion.

Procedure for designing a single element

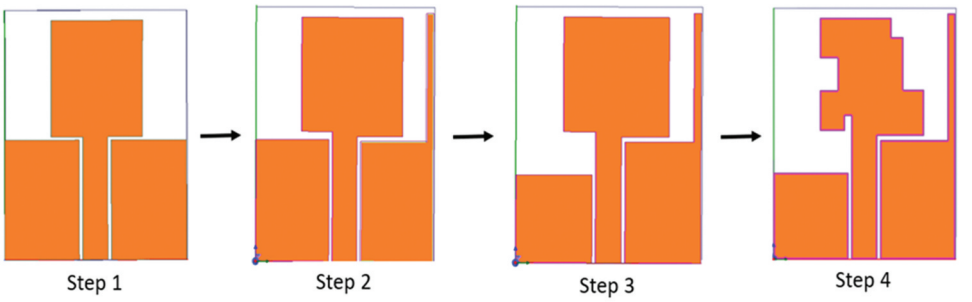
To design the single element, we use a modified square patch with an unsymmetrical ground plane and CPW feed. We design the proposed single element on the Rogers-5880 substrate, which has a relative permittivity of 2.2. We conduct all the simulations using ANSYS HFSS, a 3D electromagnetic (EM) simulation software. Figure 1(a) displays the schematic of the single element, while Table 1 provides its design parameters. Figure 1(b) presents the evolution steps of the proposed single-element antenna. Step 1 represents the original rectangular patch antenna with a CPW feed and a ground plane. In step 2, we adjusted the feed line and modified the ground plane to achieve the dual band characteristic as it changes the current distribution and creates additional resonance modes. In step 3, we introduce a defected ground structure into the ground plane, which results in the wide band characteristic due to an increase in coupling between elements and ground (Jenath Sathikbasha and Nagarajan 2020). In the final design, step 4, we modify the patch to enhance the axial ratio of the entire structure. Due to the introduction of an asymmetric radiating element, the two orthogonal modes are excited with a 90° phase difference, leading to circular polarization. The lower frequency band is from 2.4 GHz to 3.1 GHz, and the upper frequency band is from 4.8 GHz to 8 GHz, as shown by Figure 1(c). The axial ratio for each evolution step is shown in Figure 1(d), which shows that the proposed single-element antenna has circular polarization from 5.1 GHz to 10 GHz.

Table 1. Proposed single-element-antenna design parameters.

Parameters	Value (mm)	Parameters	Value (mm)
L1	25	W1	20.5
L2	8.5	W2	2.84
L3	11.8	W3	0.768
L4	4.21	W4	1.98
L5	5	W5	7.98
L6	3.2	W6	1.34
L7	4	W7	2.3
L8	1.95	W8	8.4
L9	5.2	W9	8.3
L10	4.52	S1	0.39
L11	12.6	S2	0.49
L12	1.5	H	1.57

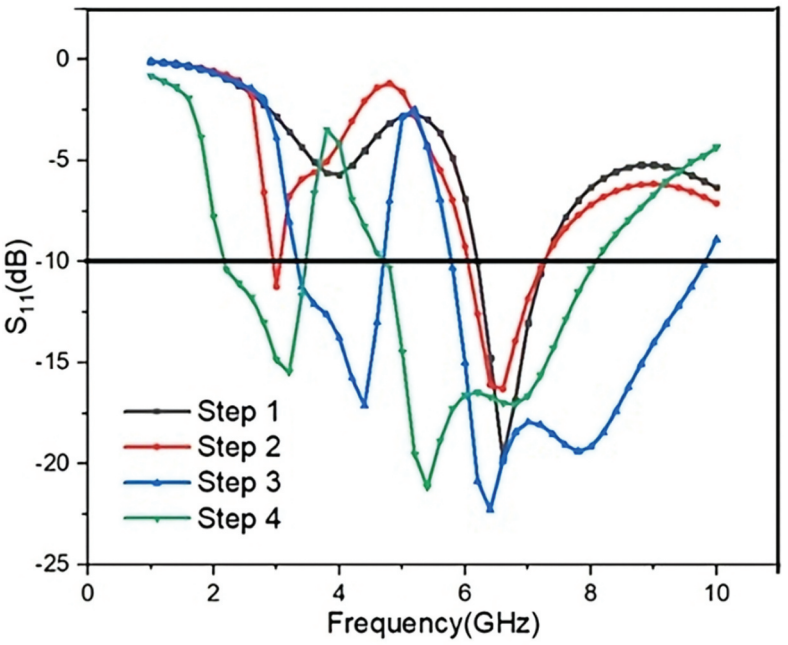

W1
h
W5
L8
L7
W6
L11
L9
L6
W4
W3
W7
L5
L12
S2
L10
L1
L4
S1
L3
y
Copper
L2
Substrate
W2
x
W8
W9
Feed

(a)

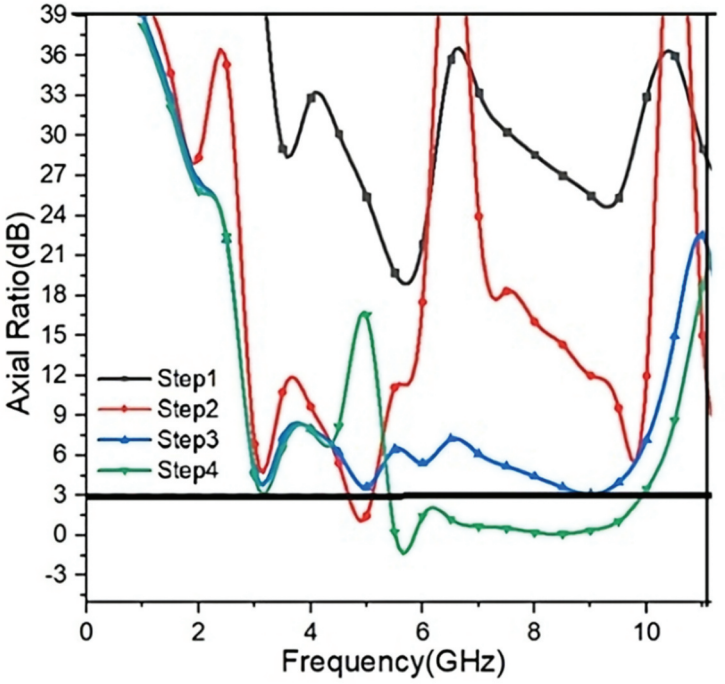

Step 1
Step 2
Step 3
Step 4

(b)

Figure 1.



(c)



(d)

Figure 1. (a) Proposed single element antenna (b) evolution steps (c) return loss steps (d) axial ratio steps.

Parametric analysis

In order to improve the axial ratio and frequency response of the proposed single-element antenna, the rectangular-shaped patch structure (RP) attached to the radiator element is moved from the bottom of the radiator to the top of it. Figure 2(a) illustrates the reflection coefficient at different positions of the RP, revealing a minimum reflection coefficient at

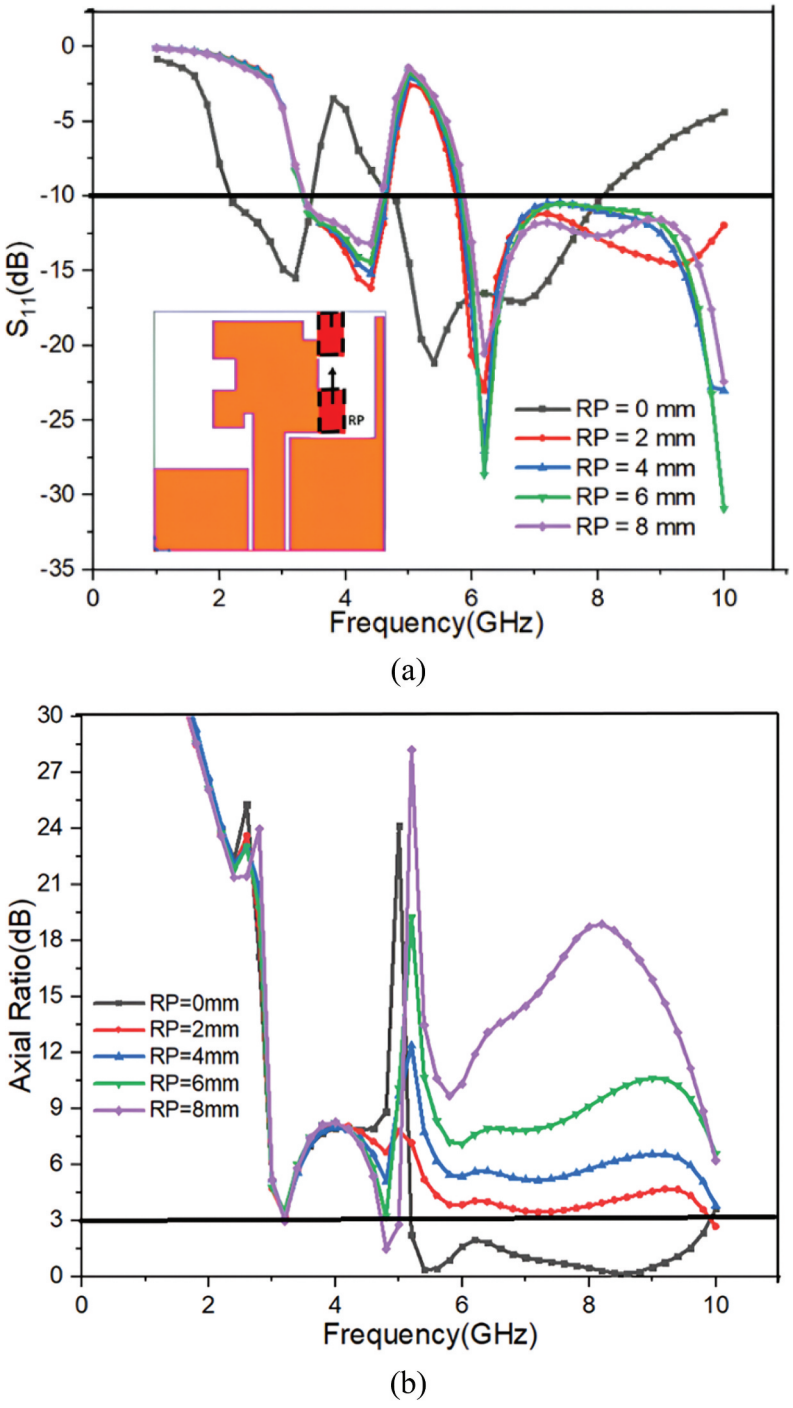


Figure 2. Proposed single element-antenna RP variations in terms of (a) return loss (b) axial ratio.

RP = 8 mm. Figure 2(b) illustrates the variation of RP w.r.t. axial ratio and finds axial ratio <3 dB for RP = 0 mm, so while considering the design for further calculation, we select RP = 0 mm.

To achieve a characteristic impedance close to 50 Ω, we initiate the initial design of a CPW feed antenna using the equations provided below (Hu, Pan, and Di Yang 2021).

$$Z0 = \frac{30\pi}{\sqrt{\varepsilon_{eff}}} \frac{K(K')}{K(K)} \tag{1}$$

Where Z_0, $K(K) and K(K')$ represents the characteristic impedance of the feed line and complete elliptic integrals of the first kind, respectively.

While

$$K' = \sqrt{1 - K^2} \tag{2}$$

and

$$\varepsilon_{eff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1/\sqrt{1 + 12(h/w)}\right) \tag{3}$$

$$K = w/(w + 2s) \tag{4}$$

In the preceding equations ϵ_r,h,w, ϵ_{eff}, S are the relative permittivity of the substrate, the thickness of the dielectric substrate, the width of the CPW fed line, effective dielectric constant of the substrate, and the gap between CPW- feed line and the ground.

We can use the above formula for calculating the characteristic impedance of the line. In our calculation, as mentioned in Table 1, we take the following values:

s = S1 = 0.39 mm and w = W1 = 2.84 mm.

Using equation (4), we get K = 2.84/{2.84 + 2(0.39)} = 0.784

Therefore $K' = \sqrt{1 - (0.784)^2} = 0.621$

$K(K) and K(K')$ can be computed using numerical methods (Simons 2004) where we get $K(K) \simeq 1.6858$ for K = 0.784 and $K(K') \simeq 1.3183$ for K′ = 0.621

Also, $\varepsilon_{eff} = \frac{2.1+1}{2} + \frac{2.1-1}{2}(1/\sqrt{1 + 12(1.57/2.84)} = 1.817$

Hence, $Z0 = \frac{30\pi}{\sqrt{1.817}} \frac{(1.3183)}{(1.6858)} \simeq 54.66\Omega$

This value of characteristic impedance is near to 50Ω.

MIMO antenna design configuration

The dual-band MIMO antenna configuration is designed with the help of the proposed single-element configuration. The two-element MIMO antenna configuration is shown in Figure 3, in which both antennas are placed opposite each other with a neutralization line (NL) in between them acting as a stop band filter. The two radiators are separated by a minimum distance of 0.11 λ, where λ is the free space wavelength at 2. 8 GHz. This shows that the proposed MIMO antenna is compact, as the minimum distance is less than 0.25λ (Salamin et al. 2023).

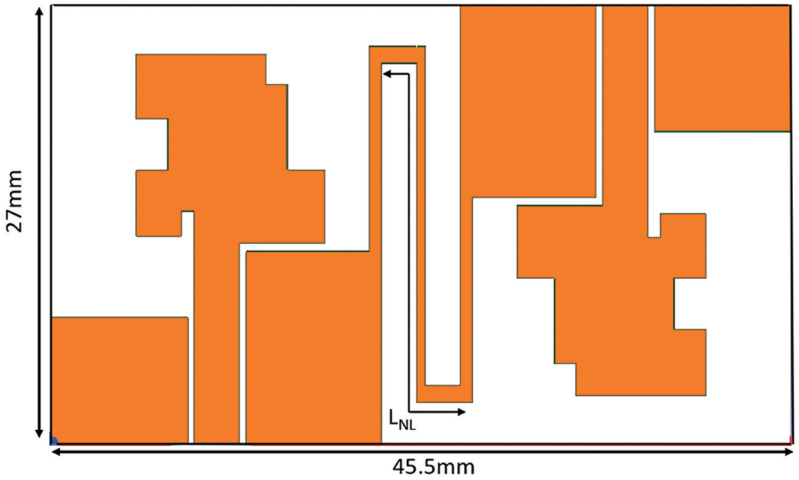


Figure 3. Proposed 2x 1 MIMO antenna.

Decoupling structure

The proposed 2 × 1MIMO antenna has a common NL in between. They serve as a stop-band filter and enhance isolation (Birwal et al. 2021). In order to fully analyze the effect of the NL between elements, we conduct an analysis on the parametric study. Initially, we set the length of the NL, L_{NL} at 4.78 mm, but this length resulted in poor isolation. With just a little shift in bandwidth, the isolation at 2.8 GHz became better with longer NL. Following the parametric variation of the length of NL, we achieved the best isolation with improved bandwidth while maintaining the circular polarization characteristic at L_{NL} = 26 mm. Figure 4(a) shows that the best impedance matching happens at L_{NL} = 26 mm. Figure 4(b) shows that the whole frequency range achieved a minimum isolation level of 22.5 dB, which is good.

The simulated bandwidth for the first band is from 2.75 GHz to 2.86 GHz, and for the second band it is from 4.8 GHz to 8.5 GHz. The minimum simulated isolation for the first band is 27 dB, while for the second band, it is 22.5 dB, which shows excellent isolation between the radiators of the antenna, as shown in Figure 5(a). Three bandwidths from (3.1–3.3) GHz, (4.7–5.6) GHz, and (7.1–8.5) GHz, respectively, show the 3 dB axial ratio bandwidth in Figure 5(b). However, in our case, we have only considered these two bands (4.7–5.6) GHz and (7.1–8.5) GHz, as they fall within our resonating frequency range. The simulated peak realized gain for Band 1 is 2.8 dB, whereas for Band 2 it is 5.4 dB, as shown by Figure 5(c).

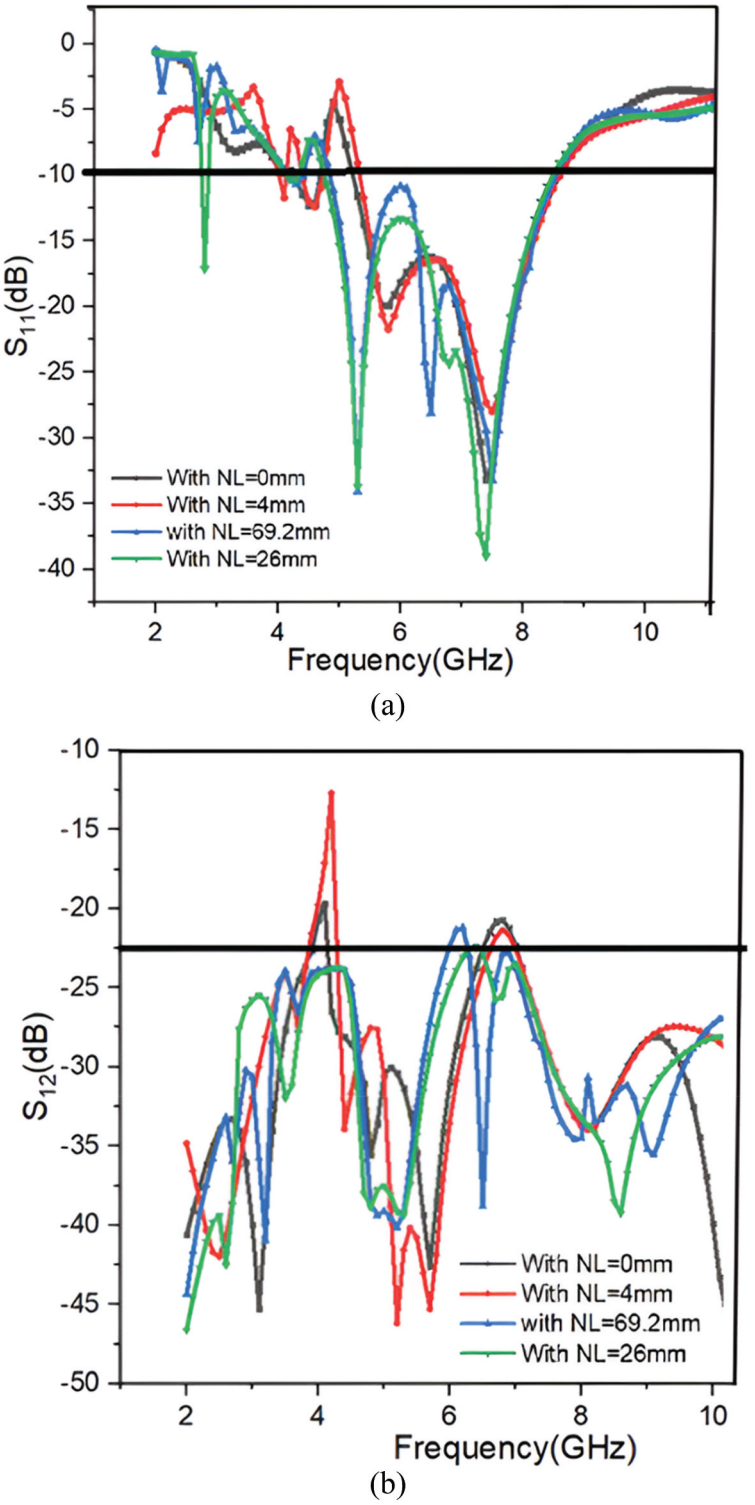
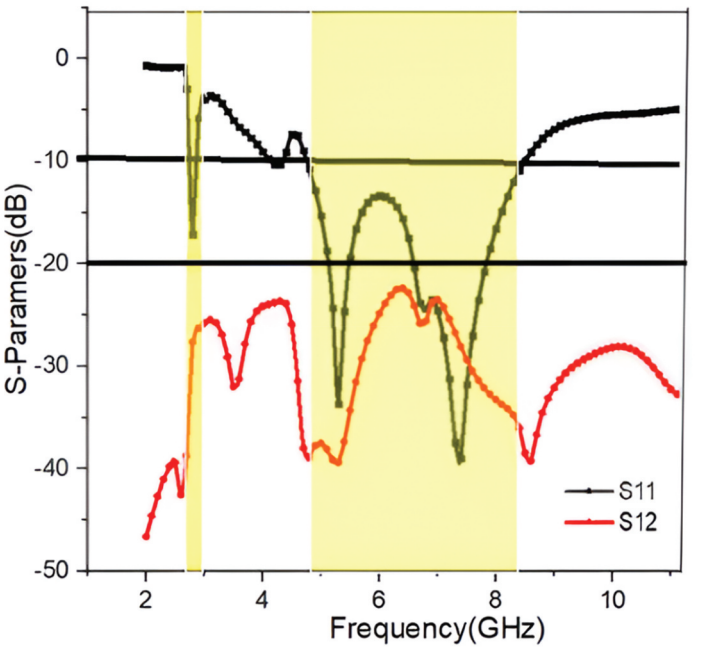
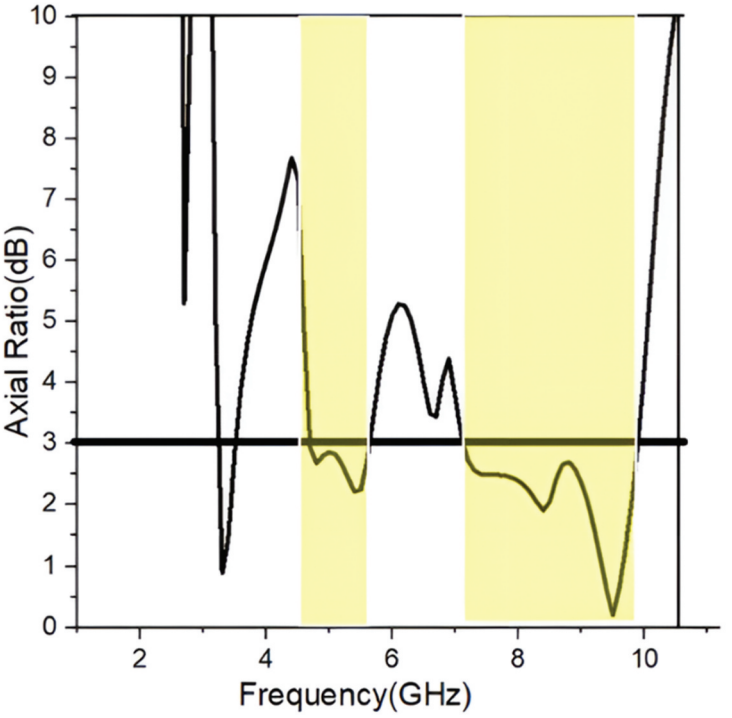


Figure 4. Simulation result of proposed antenna for different values of NL in terms of (a) reflection coefficient (S_{11}) (b) transmission coefficient (S_{12}).

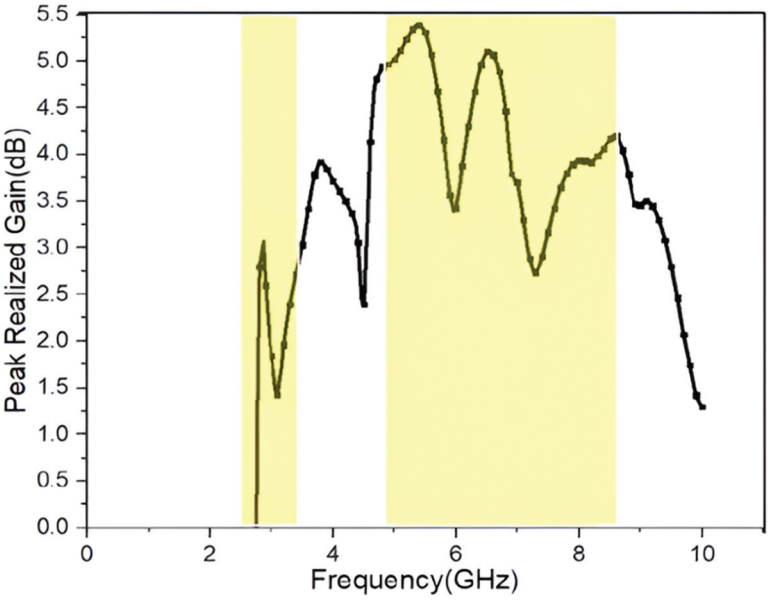

S-Paramers(dB)
0
-10
-20
-30
-40
-50
2
4
6
8
10
Frequency(GHz)
S11
S12

(a)


Axial Ratio(dB)
10
9
8
7
6
5
4
3
2
1
0
2
4
6
8
10
Frequency(GHz)

(b)

Figure 5.



(c)

Figure 5. Simulated results of the proposed antenna (a) $|S_{11}|$ & $|S_{12}|$ (b) axial ratio (c) peak realized gain.

Experimental results

In order to assess the performance of the suggested antenna, it is manufactured and subsequently tested using the Agilent PNA-X N5247A Network Analyzer. The measurement procedure involves connecting Port 1 to the excitation source and Port 2 to a 50-Ω load. Figure 6(a) shows a photograph of the fabricated antenna. Figure 6(b) displays the set-up for testing the proposed antenna, showing that the S_{11} of the fabricated antenna resonates at two frequencies in the second band, similar to the simulated results. Meanwhile, Figure 6(c) shows the set-up for radiation pattern measurement using an anechoic chamber. Figure 6(d) displays the measured and simulated S_{11} and S_{12} characteristics of the proposed antenna, which show almost the same behavior. Figure 6(e) shows the measured and simulated axial ratio of the proposed antenna. The frequency ranges of (4.7–5.6) GHz and (7.1–8.5) GHz are significantly below the 3 dB range, exhibiting characteristics of circular polarization. For a variety of reasons, fabricated antennas exhibit variation due to various factors, including the complexities of chemical etching and soldering connections, as well as the impact of laboratory testing conditions such as temperature, humidity, and electromagnetic interference.

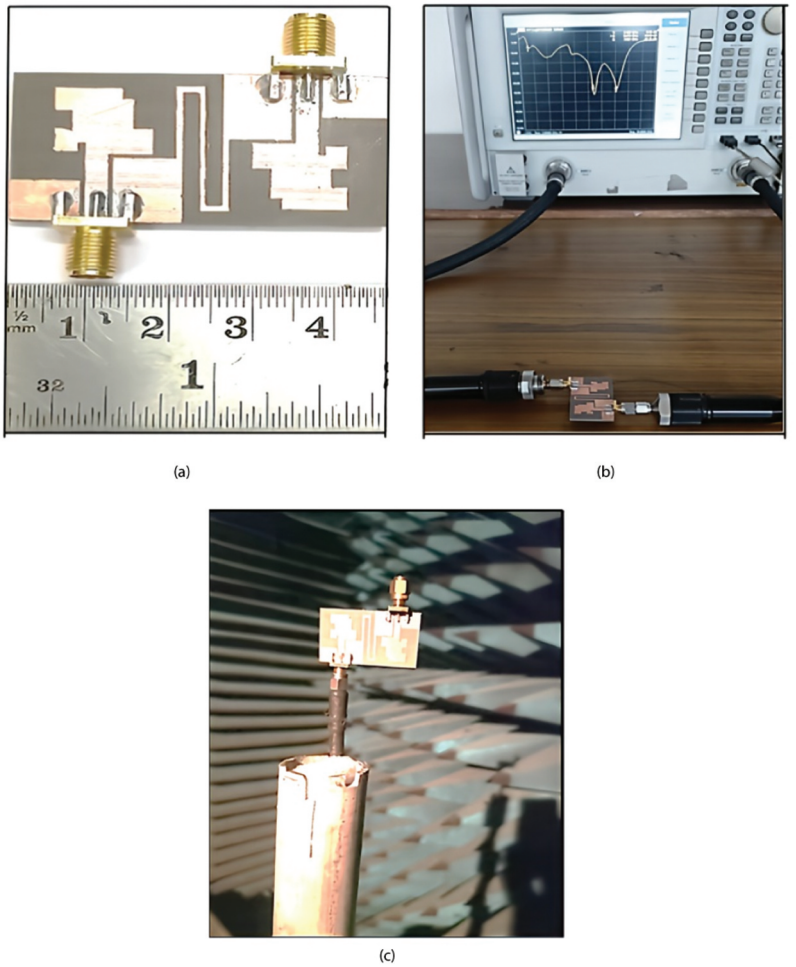
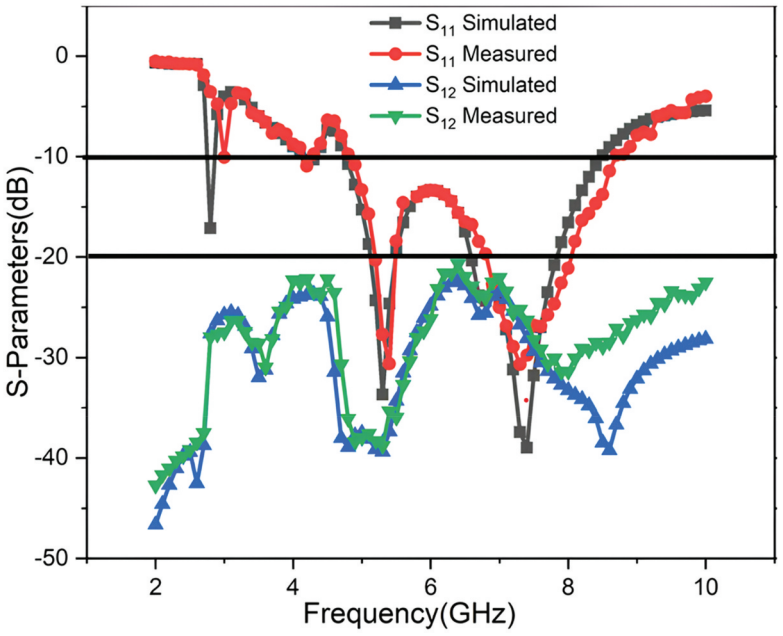
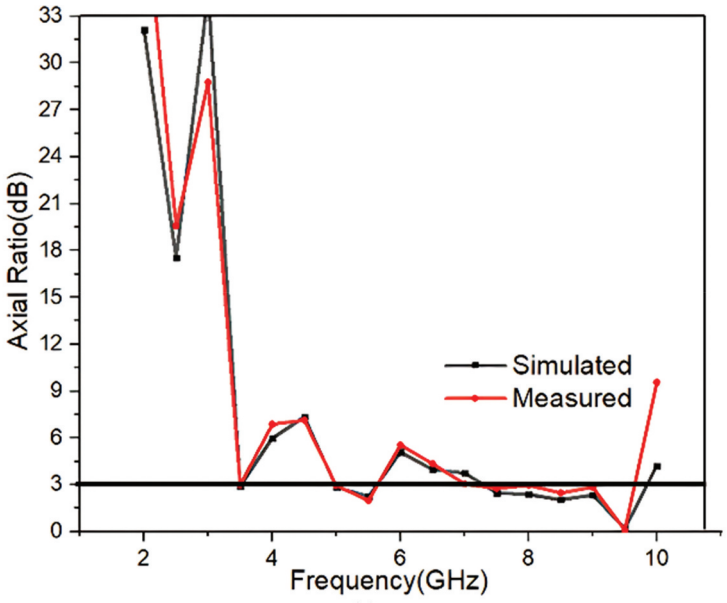


Figure 6.



(d)



(e)

Figure 6. Proposed antenna (a)prototype photograph (b)set up using network analyzer (c) radiation pattern measurement using anechoic chamber (d)simulated and measured $|S_{11}|$ & $|S_{12}|$ (e) simulated and measured axial ratio.

Radiation pattern

The suggested MIMO antenna design's 2-D radiation patterns in the E-plane (Φ = 0°) and H-plane (Φ = 90°) are depicted in Figure 7(a, b). The simulated and measured results are presented. These figures illustrate that the proposed antenna demonstrates a right-handed

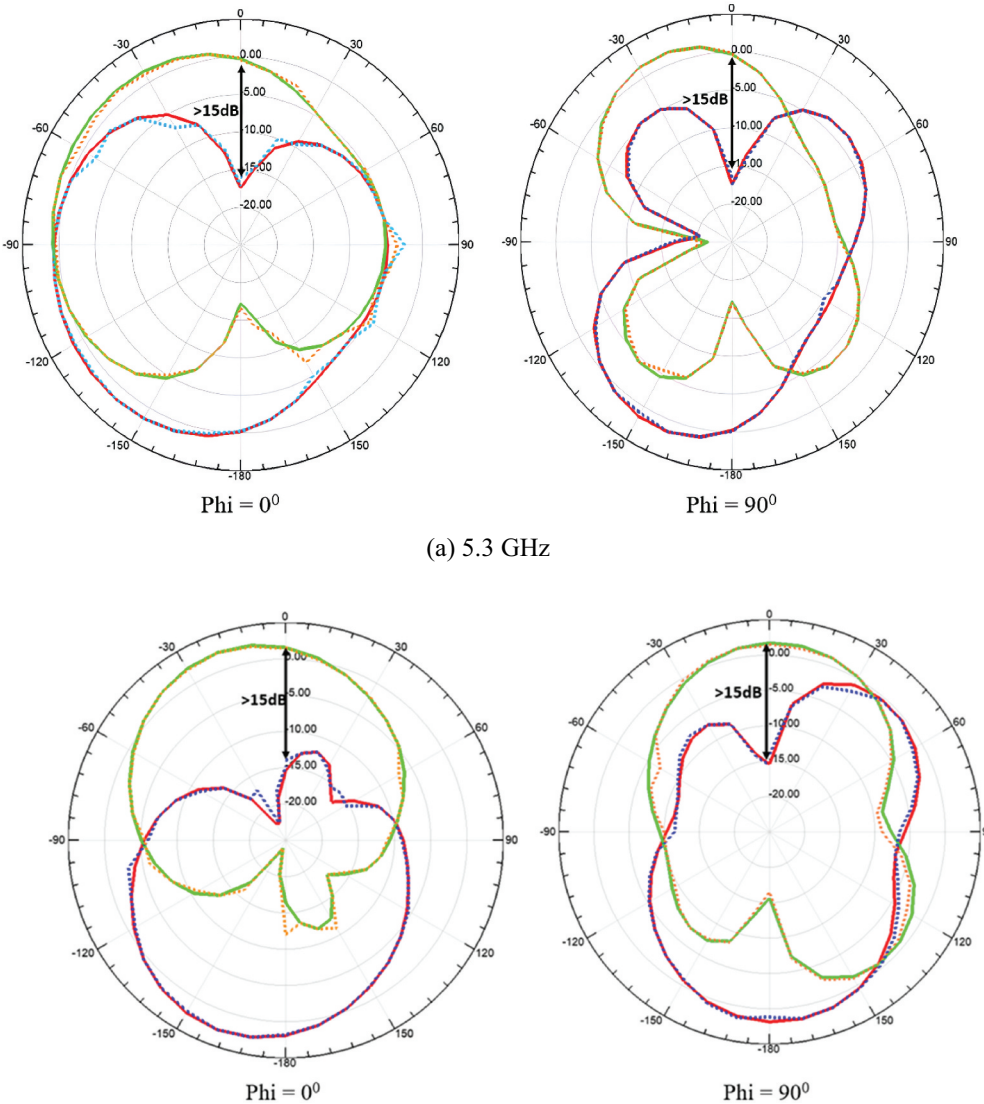


Figure 7. Radiation patterns of the proposed antenna at circularly polarized frequencies.

circular polarization (RHCP) pattern in the broadside direction at both resonating frequencies, 5.3 GHz and 7.4 GHz, respectively. In both the E-plane and H-plane, the isolation between RHCP and LHCP exceeds 15 dB, indicating that the antenna is designed to have a high level of cross-polarization discrimination (XPD). A further confirmation of the circular polarization characteristic of the proposed 2-port MIMO antenna is provided by the radiation pattern (Dwivedi et al. 2021).

Performance of the recommended MIMO antenna

For the suggested MIMO antenna to show its wide range of abilities, we tested it against a number of performance metrics. The Envelope Correlation Coefficient (ECC), Diversity Gain (DG), Total Active Reflection Coefficient (TARC), and Channel Capacity Loss (CCL) were some of these.

ECC can be evaluated either by S parameter or far field. In order to improve the MIMO antenna performance, low correction should be achieved. The range of ECC values is 0 to 1, where 0 denotes antennas that are entirely uncorrelated and 1 denotes antennas that are fully correlated ranging from 0 to 0.5 are acceptable (Raj et al. 2023).

The formula for ECC calculation by S parameter is given below (Khan et al. 2020).

$$\mathrm{ECC} = \frac{\left|S_{11}^{*}S_{21} + S_{21}^{*}S_{22}\right|^{2}}{\left(1 - \left(|S_{11}|^{2} - |S_{21}|^{2}\right)\right)\left(1 - \left(|S_{22}|^{2} - |S_{12}|^{2}\right)\right)} \tag{5}$$

Where the denoted complex conjugate of S_{11} and S_{21} are S_{11}^{*} and S_{21}^{*} correspondingly.

The far fields might be utilized to calculate ECC in order to obtain a more precise outcome (Birwal et al. 2021).

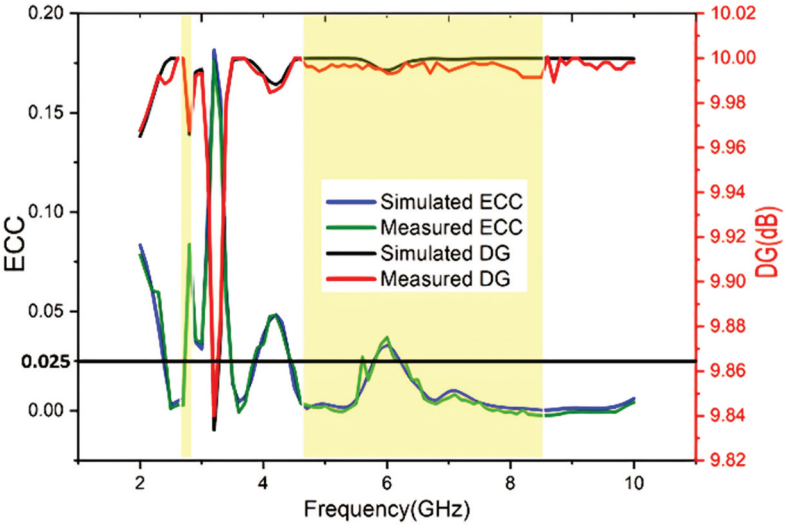
$$\mathrm{ECC} = \frac{\left|\int^{4} \pi[R_1(\theta,\Phi)\cdot R_2(\theta,\Phi)d\Omega]\right|^{2}}{\int^{4} \pi|R_1(\theta,\Phi)|^{2}d\Omega \int^{4} \pi|R_2(\theta,\Phi)|^{2}d\Omega} \tag{6}$$

Where R_i (θ, Φ) denotes the radiation pattern that occurs when port (i) gets excited whereas Ω represents the solid angle.

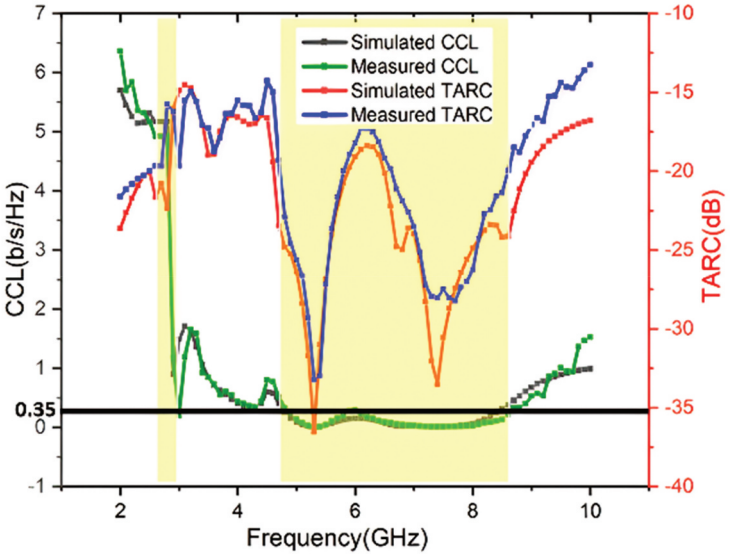
By employing DG, it is feasible to evaluate the performance of a MIMO antenna. Although a DG of 10 dB is considered to be the most favorable, values of 6 dB or higher are still regarded as satisfactory (Mondal et al. 2021). The diversity gain is calculated using the following approximate equation (Ojaroudi Parchin et al. 2020):

$$\mathrm{DG} = 10\sqrt{1 - (\mathrm{ECC})^{2}} \tag{7}$$

Figure 8(a) illustrates the simulated and measured values of ECC. They are consistently below 0.025 within our specified frequency ranges of 2.75 GHz to 2.86 GHz (band 1) and 4.8 GHz to 8.5 GHz (band 2). Additionally, the diversity gain (DG) exhibits exceptional performance in these two frequency bands, with DG values exceeding 9.98 dB throughout both bands. Thus, it shows that the antenna exhibits strong MIMO performance.



(a)



(b)

Figure 8. Proposed antenna (a) simulated and measured ECC and DG (b) simulated and measured TARC and CCL.

TARC is another crucial metric for evaluating the power that is returned when all the antenna elements in a MIMO system are energized simultaneously. It is determined by the ratio of the reflected power to the total incident power. For 2-port antennas in terms of S-parameters, it is given by the formula given below:

$$\mathrm{TARC} = \sqrt{\{(s_{11}+s_{12})^2+(s_{22}+s_{21})^2\}/2} \tag{8}$$

Another parameter, termed CCL, can be determined using the following formula to represent the maximum transmission rate (Sahu, Das, and Gangwar 2018).

$$CCL = -log_2 det(\alpha^R) \tag{9}$$

where $\alpha^{\mathrm{R}} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}$

$\alpha_{11} = 1 - \left(\mathrm{Is}_{11}\mathrm{I}^2 + \mathrm{Is}_{12}\mathrm{I}^2\right); \alpha_{22} = 1 - \left(\mathrm{Is}_{22}\mathrm{I}^2 + \mathrm{Is}_{21}\mathrm{I}^2\right); \alpha_{12} = -\left(s_{11}^{*}\mathrm{s}_{12} + s_{21}^{*}\mathrm{s}_{12}\right); \alpha_{21} = -\left(s_{22}^{*}\mathrm{s}_{21} + s_{12}^{*}\mathrm{s}_{21}\right)$

In practical MIMO systems, a TARC value of ≤ −10 dB signifies that a minimum of 90% of the input power is effectively transmitted from every port to the antennas, while an average CCL of less than 0.35 bps/Hz for a two-element MIMO system shows a lower level of correlation. Figure 8(b) shows that, for the entire frequency range of 2 GHz to 10 GHz, TARC is less than −10 dB, offering higher isolation in the operating ranges, while CCL is less than 0.35 bps/Hz from 4.7 GHz to 8.5 GHz, showing a good transmission rate.

Performance comparison

To show how unique the design suggested in this study is, the performance of the proposed 2-port MIMO antenna is compared with that of other two-port MIMO antennas that have recently been published, as shown in Table 2, where the proposed antenna has a smaller dimension compared to (Khan et al. 2023; Kumar et al. 2019, 2021; Neamah, Sabbagh, and Al-Rizzo 2023; Tiwari et al. 2021; Ullah et al. 2020). Also, most of the two-port antennas are showing either LP (Kumar et al. 2019; Tiwari et al. 2021) or CP (Jaiverdhan, M, and Yadav 2021; Khan et al. 2023; Kumar et al. 2021; Tiwari et al. 2021; Ullah et al. 2020; Zhang, Feng, and Khan 2022) behavior. However, the authors in (Neamah, Sabbagh, and Al-Rizzo 2023) designed a two-port antenna with dual-band characteristics showing LP and CP simultaneously but with a larger size and lower gain. This proposed work reduces the size by 15% and improves the maximum gain by 1.4 dB as compared to (Neamah, Sabbagh, and Al-Rizzo 2023). Additionally, it improves the maximum isolation in the first band to 27 dB, compared to 17 dB in (Neamah, Sabbagh, and Al-Rizzo 2023). As a consequence, it has improved device integration, signal range, data throughput, and communication reliability, making it an excellent candidate for contemporary wireless systems.

Conclusion

A CPW feed 2-port MIMO antenna for WLAN and C band Wireless applications has been proposed in this paper. The dimension of the antenna is 45.5×27 x 1.57 mm^3. The suggested 2×1MIMO antenna comprises modified square patch radiator with asymmetric

Table 2. Comparison to the recent published literature.

Ref. No.	Antenna's Dimensions(mm^3)/Substrate Specifications	Min. Isolation (dB)	Working Band (GHz)	ECC	Max. Gain (dBi)	Polarization (LP/CP)	3dB ARBW
(Kumar et al. 2021)	32.5 x 42 x 1/Textile (ε_r 1.34, tan δ 0.02)	>18	3.6-13	<0.02	5.7	CP	5.2-7.1
(Ullah et al. 2020)	99.7 x 33.5 x 0.813/RogersRO 4003 (ε_r 3.38, tan δ 0.0027)	>20	2.9-7.1	<0.003	3.5	CP	3.1-6.35
(Kumar et al. 2019)	72 x 56 x 0.8/FR-4 (ε_r 4.4, tan δ 0.02)	>24	2.24-2.90, 3.90–7.55	<0.04	2.5, 5.6	LP	-
(Tiwari et al. 2021)	46 x 30 x 1.6/FR-4 (ε_r 4.4, tan δ 0.02)	Band 1 > 17.21, Band 2 > 22.42	1.85-3.63, 5.07–7.96	<0.03	4.12, 4.78	LP	-
(Tiwari et al. 2021)	24 x 24 x 1.6/FR-4 (ε_r 4.4, tan δ 0.02)	>16	3.04-8.11	<0.004	6.11	CP	2.76
(Jaiverdhan, M, and Yadav 2021)	40 x 20 x 1.6/FR-4 (ε_r 4.4, tan δ 0.02)	>16	3.17-17.39	<0.005	4.41	CP	4.8-8.8
(Zhang, Feng, and Khan 2022)	25 x 51 x 0.8/FR-4 (ε_r 4.4, tan δ 0.02)	>18	3.1-13.5	<0.01	4	CP	4.7-6.1
(Khan et al. 2023)	40 x 70 x 1.6/FR-4 (ε_r 4.4, tan δ 0.02)	>30	4.89-6.85	<0.002	6.45	CP	5.41-6.57
(Neamah, Sabbagh, and Al-Rizzo 2023)	45 x 31.5 x 1.6/FR-4 (ε_r 4.4, tan δ 0.025)	Band 1 > 17, Band 2 > 26	3.45-4.7, 5.45-9	<0.009	3.4, 4	LP+CP	5.95-8.1
This Work	45.5 x 27 x 1.57/Rogers 5880 (ε_r 2.2, tan δ 0.0009)	Band 1 > 27, Band 2 > 22.5	2.75-2.86, 4.8 -8.5	<0.02	2.8, 5.4	LP+CP	4.7-5.6, 7.1–8.5

ground plane structure placed opposite to each other's direction with an NL in between them. Band 1 is from 2.75 GHz to 2.86 GHz showing linear polarization behavior with $S_{11} < -17$ dB and $S_{12} < -27$ dB showing excellent isolation in this band. Band 2 is from 4.8 GHz to 8.5 GHz in which 3 dB axial ratio bandwidth lies between (4.7–5.6) GHz and (7.1–8.5) GHz, respectively. The minimum isolation for the entire second band is 22.5 dB. The designed MIMO antenna exhibits a stable radiation pattern at 5.3 GHz and 7. 4 GHz. The peak realized gain of the entire antenna achieved is 5.4 dB, showing the better gain performance of the antenna with this compact size. The diversity performance such as ECC, DG, TARC and CCL of the proposed MIMO antenna were found and seen that it well within the acceptable range. The antenna was designed and evaluated. The measured and simulated results exhibit a high degree of correlation. The suggested MIMO antenna has found use in WLAN and C-band wireless applications, such as satellite communication, weather radar, 5 G communication, and remote sensing, due to its compact size and ability to transmit and receive signals in two different polarizations.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Behera, H. K., M. Midya, and L. P. Mishra. 2024. A compact dual-band dual-sense circular polarized wide slot antenna for WLAN applications. *Wireless Personal Communications* 136 (3):1735–47. doi: 10.1007/s11277-024-11357-z.

Benkhadda, O., M. Saih, K. Chaji, S. Ahmad, and A. Reha. 2023. A compact dual-band cpw-fed slot monopole antenna for WIFI, WLAN and WiMAX applications. *Arabian Journal for Science & Engineering* 48 (5):6089–98. doi: 10.1007/s13369-022-07268-5.

Birwal, A., S. Singh, B. K. Kanaujia, and S. Kumar. 2021. Mimo/diversity antenna with neutralization line for WLAN applications. *MAPAN* 36 (4):763–72, January. doi: 10.1007/s12647-020-00427-9.

Dwivedi, A. K., A. Sharma, A. K. Pandey, and V. Singh. 2021. Two Port circularly polarized MIMO antenna design and investigation for 5G communication systems. *Wireless Personal Communications* 120 (3):2085–99. doi: 10.1007/s11277-021-08461-9.

Gad, M. M., and M. O. Sallam. 2024. High gain antipodal meander line antenna for point-to-point WLAN/WiMAX applications. *Scientific Reports* 14 (1):5769. doi: 10.1038/s41598-024-54994-x.

Hu, Y., Y. M. Pan, and M. Di Yang. Circularly polarized MIMO dielectric resonator antenna with reduced mutual coupling. *IEEE Transactions on Antennas & Propagation* 69 (7):3811–20. Accessed January 8, 2021. doi: 10.1109/TAP.2020.3048502.

Huang, J., G. Dong, Q. Cai, Z. Chen, L. Li, and G. Liu. Dual-band MIMO antenna for 5G/WLAN mobile terminals. *Micromachines* 12 (5):489. Accessed April 26, 2021. doi: 10.3390/mi12050489.

Ishteyaq, I., and K. Muzaffar. 2022. Multiple input multiple output (MIMO) and fifth generation (5G): An indispensable technology for sub-6 GHz and millimeter wave future generation mobile terminal applications. *International Journal of Microwave and Wireless Technologies* 14 (7):932–48, September. doi: 10.1017/S1759078721001100.

Jaiverdhan, S., M. M, and R. P. Yadav. 2021. Broadband circularly polarized compact MIMO slot antenna based on strip and stubs for UWB applications. *Electromagnetics* 41 (3):185–95. doi: 10.1080/02726343.2021.1903205.

Jamshed, M. A., M. Ur-Rehman, J. Frnda, A. A. Althuwayb, A. Nauman, and K. Cengiz. Dual band and dual diversity four-element MIMO dipole for 5G handsets. *Sensors (Basel)* 21 (3):767. Accessed January 24, 2021. doi: 10.3390/s21030767.

Jenath Sathikbasha, M., and V. Nagarajan. 2020. Design of multiband frequency reconfigurable antenna with defected ground structure for wireless applications. *Wireless Personal Communications* 113 (2):867–92, Jul. doi: 10.1007/s11277-020-07256-8.

Khan, A., S. Geng, X. Zhao, Z. Shah, M. U. Jan, and M. A. Abdelbaky. Design of MIMO antenna with an enhanced isolation technique. *Electronics* 9 (8):1217. Accessed Jul 28, 2020. doi: 10.3390/electronics9081217.

Khan, I., K. Zhang, Q. Wu, I. Ullah, L. Ali, H. Ullah, and S. U. Rahman. 2023. A wideband High-isolation Microstrip MIMO circularly-polarized antenna based on parasitic elements. *Materials* 16 (1):103. doi: 10.3390/ma16010103.

Kumar, A., A. Quaiyum Ansari, B. Kumar Kanaujia, and J. Kishor. 2019. A novel iti-shaped isolation structure placed between two-port cpw-fed dual-band MIMO antenna for high isolation. *AEU - International Journal of Electronics & Communications* 104:35–43, ISSN 1434-841110.1016/j.aeue.2019.03.009.

Kumar, S., Nandan, D., Srivastava, K., Kumar, S., Singh, H., and Marey, M. 2021. Wideband circularly polarized textile MIMO antenna for wearable applications. IEEE Access, vol. 9, 108601–13, doi: 10.1109/ACCESS.2021.3101441.

Li, M., L. Jiang, and K. L. Yeung. A general and systematic method to design neutralization lines for isolation enhancement in MIMO antenna arrays. *IEEE Transactions on Vehicular Technology* 69 (6):6242–53. Accessed April 2, 2020. doi: 10.1109/TVT.2020.2984044.

Mondal, R., P. S. Reddy, D. C. Sarkar, and P. P. Sarkar. Investigation on MIMO antenna for very low ECC and isolation characteristics using FSS and metal-wall. *AEU-International Journal of Electronics and Communications* 135:153708. Accessed Jun 1, 2021. doi: 10.1016/j.aeue.2021.153708.

Neamah, W. A., H. M. A. Sabbagh, and H. Al-Rizzo. 2023. A compact two-element linearly and orthogonal circularly polarized MIMO antenna system for 5G cellular and WLAN/Wi-fi 6E application. IEEE Access, vol. 11, 96879–91, doi: 10.1109/ACCESS.2023.3312122.

Ojaroudi Parchin, N., H. Jahanbakhsh Basherlou, M. Al-Yasir Yi, A. Abdulkhaleq, M. A. Patwary, and R. Abd-Alhameed. A new CPW-Fed diversity antenna for MIMO 5G smartphones. *Electronics* 9 (2):261. Accessed February 4, 2020. doi: 10.3390/electronics9020261.

Parvathi, K. S., and S. R. Gupta. Novel dual-band EBG structure to reduce mutual coupling of air gap based MIMO antenna for 5G application. *AEU-International Journal of Electronics and Communications* 138:153902. Accessed August 1, 2021. doi: 10.1016/j.aeue.2021.153902.

Puri, V., and H. S. Singh. Design of an isolation improved MIMO antenna using metasurface based absorber for wireless applications. *Optik* 259:168963. Accessed Jun 1, 2022. doi: 10.1016/j.ijleo.2022.168963.

Raj, T., R. Mishra, P. Kumar, and A. Kapoor. Advances in MIMO antenna design for 5G: A comprehensive review. *Sensors (Basel)* 23 (14):6329. Accessed Jul 12, 2023. doi: 10.3390/s23146329.

Sahu, N. K., G. Das, and R. K. Gangwar. 2018. Dual polarized triple-band dielectric resonator-based hybrid MIMO antenna for WLAN/WiMAX applications. *Microwave and Optical Technology Letters* 60 (4):1033–41. doi: 10.1002/mop.31102.

Salamin, M. A., A. Zugari, M. Alibakhshikenari, C. H. See, R. Abd-Alhameed, and E. Limiti. Compact highly isolated dual-band 4-port MIMO antenna for sub-6 GHz applications. *Journal of Electromagnetic Waves and Applications* 37 (10–12):1023–43. Accessed August 13, 2023. doi: 10.1080/09205071.2023.2220145.

Simons, R. N. *Coplanar waveguide circuits, components, and systems*. New Jersey, USA: John Wiley & Sons. Accessed April 7, 2004.

Tiwari, R. N., P. Singh, B. K. Kanaujia, and P. Kumar. 2021. Compact circularly polarized MIMO printed antenna with novel ground structure for wideband applications. *International Journal of RF and Microwave Computer-Aided Engineering* 31 (8):e22737, August. doi: 10.1002/mmce.22737.

Tiwari, R. N., P. Singh, S. Pandey, R. Anand, D. K. Singh, and B. K. Kanaujia. 2021. Swastika shaped slot embedded two port dual frequency band MIMO antenna for wireless applications. *Analog Integrated Circuits and Signal Processing* 109 (1):103–13. doi: 10.1007/s10470-021-01923-x.

Turan, G., and H. Odabasi. 2023. A multiband quasi-Yagi antenna using stepped driver and slotted ground for 2.4/3.5/5.2 GHz bands applications. *Microwave and Optical Technology Letters* 65 (7):2031–38. doi: 10.1002/mop.33643.

Ullah, U., I. B. Mabrouk, and S. Koziel. 2020. Enhanced-performance circularly polarized MIMO antenna with Polarization/Pattern diversity. IEEE Access, vol. 8, 11887–95, doi: 10.1109/ACCESS.2020.2966052.

Ullah, U., I. B. Mabrouk, S. Koziel, and M. Al-Hasan. 2020. Implementation of Spatial/Polarization diversity for improved-performance circularly polarized multiple-input-multiple-output ultra-wideband antenna. IEEE Access, vol. 8, 64112–19, doi: 10.1109/ACCESS.2020.2984697.

Xing, H., X. Wang, Z. Gao, X. An, H. X. Zheng, M. Wang, and E. Li. Efficient isolation of an MIMO antenna using defected ground structure. *Electronics* 9 (8):1265. Accessed August 6, 2020. doi: 10.3390/electronics9081265.

Yuan, X. T., W. He, K. D. Hong, C. Z. Han, Z. Chen, and T. Yuan. Ultra-wideband MIMO antenna system with high element-isolation for 5G smartphone application. *Institute of Electrical and Electronics Engineers Access* 8:56281–89. Accessed March 19, 2020. doi: 10.1109/ACCESS.2020.2982036.

Zhang, L., Q. Feng, and M. K. Khan. 2022. Design of a novel circularly polarized MIMO antenna with enhanced isolation for ultra-wideband communication. *Applied Computational Electromagnetics Society Journal (ACES)* 37 (5):607–18. doi: 10.13052/2022.ACES.J.370510.

Neural Computing and Applications
https://doi.org/10.1007/s00521-024-10587-3

S.I.: MULTIMODAL DATA FUSION AND MACHINE LEARNING FOR SMART CITIES

Hyper-personalized employment in urban hubs: multimodal fusion architectures for personality-based job matching

Dipika Jain[1] · Saurabh Raj Sangwan[2] · Akshi Kumar[3]

Received: 1 November 2023 / Accepted: 1 October 2024
© The Author(s) 2024

Abstract
In the evolving landscape of smart cities, employment strategies have been steering towards a more personalized approach, aiming to enhance job satisfaction and boost economic efficiency. This paper explores an advanced solution by integrating multimodal deep learning to create a hyper-personalized job matching system based on individual personality traits. We employed the First Impressions V2 dataset, a comprehensive collection encompassing various data modalities suitable for extracting personality insights. Among various architectures tested, the fusion of XceptionResNet with BERT emerged as the most promising, delivering unparalleled results. The combined model achieved an accuracy of 92.12%, an R2 score of 54.49%, a mean squared error of 0.0098, and a root mean squared error of 0.0992. These empirical findings demonstrate the effectiveness of the XceptionResNet + BERT in mapping personality traits, paving the way for an innovative, and efficient approach to job matching in urban environments. This work has the potential to revolutionize recruitment strategies in smart cities, ensuring placements that are not only skill-aligned but also personality-congruent, optimizing both individual satisfaction and organizational productivity. A set of theoretical case studies in technology, banking, healthcare, and retail sectors within smart cities illustrate how the model could optimize both individual satisfaction and organizational productivity.

Keywords Multimodal deep learning · Smart cities · Personality detection · ChaLearn dataset · Urban employment strategies

1 Introduction

As we step into the embrace of smart cities, we're concurrently stepping into the age of the Fourth Industrial Revolution—a revolution that heralds the convergence of digital, biological, and physical innovations [1]. In this vibrant juncture, traditional employment paradigms are being rigorously challenged, as the nexus of technology and urban sophistication necessitates a fresh perspective on workforce placement. Traditionally, employment strategies, especially in bustling urban environments, prioritized tangible metrics such as qualifications, years of experience, and technical proficiencies [2]. Such metrics, while imperative, often failed to capture the comprehensive essence of a candidate. To illustrate, consider two candidates, both with stellar resumes showcasing impeccable technical skills and identical years of experience. Yet, beneath this surface uniformity lie stark differences in their personalities, work preferences, and interpersonal aptitudes. One might be an extrovert, thriving in collaborative scenarios, while the other might excel in solitary, deep analytical tasks. Traditional hiring metrics may overlook these disparities, but in the nuanced professional environments of today's urban hubs, such distinctions could be critical.

Modern urban hubs, with their diverse populations and equally diverse professional landscapes, offer vast opportunities. Yet, these opportunities come intertwined with challenges. In the quest to match the right individual with

✉ Akshi Kumar
Akshi.Kumar@gold.ac.uk

[1] Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India

[2] School of Computer Science and Engineering, Galgotias University, Greater Noida, India

[3] Department of Computing, Goldsmiths, University of London, London, UK

Published online: 04 December 2024

the right job role, cities and corporations have started realizing that the 'right match' goes beyond just aligning skills. Individual personalities, encompassing traits, preferences, motivations, and behaviours, play a pivotal role. For instance, an extroverted individual with a penchant for social interactions might excel in client-facing roles or team leadership positions. In contrast, an introverted individual, who finds solace in solitude and deep work, might be better suited for research or individual contributor roles.

As urban companies continue their relentless evolution, there is a palpable shift towards recognizing the importance of individual personalities in employment. Such hyper-personalized approaches discern that job satisfaction and optimal performance are not solely contingent on skill alignment but require a deeper resonance between job roles and individual personalities. To bring this concept closer to reality, consider the vibrant tech sector in cities like San Francisco or Seoul. Here, tech enterprises, ranging from fledgling startups to established giants, often grapple with finding the 'right fit' for their teams. A conventional hiring approach might yield technically proficient candidates, but a personality misalignment could culminate in reduced job satisfaction, and consequentially, diminished productivity. However, by integrating advanced profiling that goes beyond skills to encompass personality traits—like innovative thinking, communication styles, or adaptability—the hiring process can evolve. For instance, a role demanding constant stakeholder interaction might be best suited for someone with an extroverted, communicative personality, even if their technical prowess is slightly inferior to a more introverted counterpart. Further, imagine the potential of an AI-driven tool that can analyse a candidate's LinkedIn articles to assess their industry insights, their Twitter feed to gauge their communication style and topical interests, and their participation in online forums to understand their problem-solving approach. Such a tool, when fine-tuned ethically and with respect to privacy concerns, could provide recruiters with a goldmine of information, enabling hyper-personalized job matches that account for personality traits just as much as, if not more than, skills.

To realize the vision of hyper-personalized employment and achieve a comprehensive understanding of personality traits, an advanced methodology is paramount—one that can combine and interpret data from diverse sources. Guided by this objective, we turned to the First Impressions V2 dataset [3]. This dataset, with its expansive collection of 10,000 video clips and meticulously curated transcripts, stands as an exemplary source of multimodal data. Originating from a broad spectrum of YouTube HD videos, these clips provide more than just visual indicators. Averaging a duration of 15 s, the clips are classified according to the five-factor model, assessing personality traits such as Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness. They offer insight into vocal nuances, non-verbal behaviours, and various facets of human communication. Concurrently, the transcribed text presents a rich analysis of linguistic patterns, vocabulary preferences, and other linguistic markers indicative of personality attributes. The dataset has been crucial in exploring the relationship between audio-visual cues and personality perceptions.

Our primary objective is to develop a nuanced understanding of personality traits through multimodal data analysis and to apply this understanding in creating a sophisticated job matching system. This system is designed to align individual personality traits with the optimal job roles, enhancing both employee satisfaction and organizational productivity. Employing a comprehensive array of data processing techniques, including video and transcript analysis through state-of-the-art convolutional neural networks [4] and transformer architectures [5] like XceptionResNet and BERT, our methodological approach is robust and innovative. We focus on extracting and interpreting nuanced personality cues from video clips and textual data, aiming to provide a holistic personality assessment. In this research, a systematic approach was undertaken to provide an enriched understanding of video and transcript processing, as well as the models applied. Summary of the research steps undertaken is as follows:

(a) Video Processing

- The videos were pre-processed to extract 3 frames per video at equal intervals and audio in the WAV format.
- Audio data were transformed into images representing Mel spectrogram, chromagram, and spectral bandwidth, suitable for CNNs.
- Common audio and image processing tasks included noise reduction, resizing, cropping, etc., using libraries like OpenCV, PIL, and librosa.

(b) Transcript Processing

- Transcriptions were pre-processed using Tokenizers.
- They were cleaned to remove punctuations and irrelevant tokens.
- The processed transcripts were tokenized using BertTokenizer and DebertaTokenizer.

(c) Models Used

- For video (frames and audio as images) modality: InceptionV3, EfficientResNet, and XceptionResNet were used.
- For text modality (transcripts): BERT and DeBERTa were used.
- For multimodal data, fusion models were created combining features of InceptionV3, Efficient ResNet, and XceptionResNet with BERT.

Such a multifaceted approach, augmented by AI and machine learning, can revolutionize urban employment. Envision an intelligent system capable of analysing a candidate's public presentations, gauging their expertise, their communication styles on professional networks, and even their problem-solving discourse on platforms like GitHub or Stack Overflow. The significance of this study lies in its potential to revolutionize recruitment strategies, enabling a deep resonance between an individual's personality and their professional role. By advancing the capabilities of AI in personality assessment, we pave the way for more effective and satisfying employment in the fast-evolving urban job market.

The rest of the paper is organized as follows: firstly, the paper talks about the shift from skill-based to personality-driven approach, further section discusses the related work followed by the detailed description of the dataset. The methodology section, that is Sect. 4 of the paper, is emphasized on the deep neural architecture and the multimodal fusion models which further relates to the analysis of the results obtained with graphical representations in Sect. 5.

2 The urban shift: from skill-based to personality-driven employment

Modern urban centres are not just places of work; they are melting pots of diverse cultures, backgrounds, and personalities. Companies in these hubs are often involved in multifaceted projects, requiring teams with varied skills and personalities. Here is where the results of personality trait detection become pivotal. Take the tech industry, for example. While a developer's technical knowledge is undeniably crucial, their personality traits like adaptability, resilience, or inclination for teamwork can significantly influence their overall contribution to a project. A startup environment, characterized by rapid changes and multitasking, might prioritize adaptability and openness over other traits. In contrast, a well-established tech giant, with its structured workflows, might value consistency and detail-orientation.

Moreover, urban hubs are increasingly becoming the centres for global businesses, meaning employees often collaborate with teams across different parts of the world. Here, cultural sensitivity, a component of agreeableness, becomes indispensable. An individual's ability to understand, appreciate, and work harmoniously with colleagues from diverse backgrounds can be a game-changer. This is where the results from personality trait detection techniques shine. By analysing vast datasets, such as video interviews, we can derive accurate personality profiles of candidates. These profiles, when matched against a role's requirements, can result in better hiring decisions. For example, roles that require constant stakeholder interaction might prioritize extraversion, while positions involving deep research might value introversion and conscientiousness. Further, modern businesses aren't static; they evolve, pivot, and adapt. An employee's ability to grow with the company, learning new skills and adapting to new roles, is often influenced by their personality. Traits like openness to experience and adaptability can be strong indicators of an individual's potential for growth and versatility.

However, this shift towards personality-driven employment does not come without challenges. The foremost concern is ensuring that the process remains unbiased. Relying heavily on personality assessments might inadvertently lead to a monoculture, where only certain personality types are favoured, leading to a lack of diversity in thought and perspective. Moreover, while personality traits are crucial, they cannot completely replace skill assessments. A balance must be struck where skills and personality are both given their due importance. Over-relying on one at the expense of the other can lead to suboptimal hiring decisions.

The urban shift from skill-based to personality-driven employment is a testament to the complexities of modern work environments. As urban hubs continue to grow, in both size and diversity, understanding an individual's personality will become even more crucial. However, the future likely lies in a balanced approach, where skills and personality are both factored into hiring decisions. Personality trait detection offers a powerful tool in this paradigm, enabling companies to make more informed, holistic hiring choices. In conclusion, as urban hubs redefine employment, integrating the insights from personality trait detection can lead to more harmonious, productive workplaces. While the road ahead might have its challenges, the potential rewards—in terms of employee satisfaction, productivity, and business growth—are immense.

2.1 Impact of personality impressions on interview outcomes in personality-driven employment

The impact of personality impressions on interviews is crucial in the modern job market, especially in a personality-driven employment landscape. These impressions provide valuable insights that help employers make strategic hiring decisions that benefit both the individual and the organization in the long term. Personality traits not only influence how candidates are perceived by potential employers but also determine how well they fit into the company's culture and the specific job role. Here are the key points highlighting this importance:

- *First Impressions* Interviews are often the first point of contact between a candidate and an employer, where initial impressions are formed. Personality impressions during these interactions can greatly influence an interviewer's perception of a candidate's suitability for the role, often affecting the final hiring decision.
- *Cultural fit* Organizations look for individuals who not only have the necessary skills and qualifications but also fit well with their organizational culture. Personality impressions help employers assess whether candidates' values, work ethic, and behaviour align with the company's ethos.
- *Predicting job performance* Employers often use personality impressions as indicators of future job performance. Traits like conscientiousness, agreeableness, and emotional stability are linked to positive work outcomes and can be pivotal in predicting a candidate's success in the role.
- *Enhancing team dynamics* The personality impression can provide insights into how a candidate might interact with potential colleagues and contribute to team dynamics. Employers aim to select individuals whose personality traits will complement and enhance the existing team structure.
- *Reducing turnover* By assessing personality impressions, employers can make more informed hiring decisions that lead to longer-lasting employment relationships. Hiring individuals whose personalities are a good fit for the job and company can reduce turnover rates and associated costs.

3 Related work

Over the past decade, several studies have delved into the realm of personality trait recognition using various data modalities, such as text, audio, and visual inputs. In 2008, Pianesi et al. [6] conducted pioneering research on this topic, applying machine learning techniques to understand personality traits, including the Big Five model, during social interactions. This work has had significant implications across psychology, human–computer interaction, and the social sciences. Subsequent study in 2014 by Alam and Riccardi [7], explored the integration of multimodal data for personality analysis, utilizing advanced machine learning approaches.

In more recent years, research has focused on enhancing the interpretability and practical applications of personality analysis. In 2018, Escalante et al. [8] introduced an innovative method for modelling and explaining apparent personality traits using video data, increasing accuracy and transparency. In the same year, Gorbova et al. [9] synergized visual and textual data to improve personality trait analysis, demonstrating its superiority over baseline methods. In 2019, Junior et al. [10] conducted a comprehensive survey of vision-based techniques for evaluating apparent personality traits, consolidating existing knowledge and highlighting practical applications in human–computer interaction and social robotics. These studies collectively contribute to our understanding of how diverse data sources can be harnessed to gain insights into human behaviour and personality traits within social contexts, with wide-ranging implications in various fields.

Additionally, in 2019, Wingate et al. [11] explored how individuals form initial impressions, especially in situations involving deception or "faking." The study examined the attributions people make when encountering deceptive behaviour and discussed the practical implications of these early impressions in contexts like hiring decisions, interpersonal relationships, and legal proceedings. In 2020, Nørskov [12] and colleagues investigated the dynamics of human–robot interaction in job interviews, shedding light on the ethical concerns and psychological implications associated with robotic interviewers. Finally, in 2021, Aslan et al. [13] and Song et al. [14] continued to advance the field by combining models for improved accuracy and exploring self-supervised learning for automated personality trait recognition based on facial dynamics, highlighting practical applications in human–computer interaction and affective computing.

4 Methodology

In this research, a systematic approach was undertaken to provide an enriched understanding of video and transcript processing, as well as the models applied. The following subsections discuss the details of dataset and techniques used.

0	He's cutting it and then turn around and see t...	0.523364	0.488889	0.626374	0.552083	0.601942
1	Responsibility to house the organ I had been g...	0.345794	0.366667	0.472527	0.375000	0.582524
2	I actually got quite a few sets of black pens ...	0.252336	0.511111	0.406593	0.291667	0.485437
3	I ate a lot. I'd like a lot of foods. I rememb...	0.457944	0.377778	0.505495	0.489583	0.398058
4	Now I'll ask you guys to leave a question in t...	0.607477	0.622222	0.406593	0.489583	0.621359
...	...	...	...	...	...	...
9995	Du du du. No, that's not why I made this video...	0.289720	0.300000	0.208791	0.312500	0.135922
9996	They do it all the time.	0.719626	0.722222	0.670330	0.781250	0.572816
9997	Comfortable and I don't want anyone to feel un...	0.355140	0.677778	0.472527	0.395833	0.446602
9998	You're not giving yourself enough calories to ...	0.467290	0.622222	0.527473	0.645833	0.669903
9999	Eat enough carbs, eat more fats to get in more...	0.654206	0.588889	0.813187	0.635417	0.728155

Fig. 1 Transcripts of the dataset

4.1 The First Impression dataset

First Impression ChaLearn is a publicly available largest Big-5 personality dataset. It constitutes 10,000 videos labelled into various types of sentiments displayed by them. Using the videos, we have generated novel datasets of various modalities like frames, audios, and text.

The videos have an average duration of 15s, extracted from more than 3,000 HD YouTube videos of people speaking in English while facing the camera. Each video has five labels in the range [0,1] which depict the Big Five personality traits, namely—Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness. The videos are also accompanied by their transcripts, in the form of a pickle file. The transcript contains 435,984 words, averaging to 43 words per video, with a maximum of 85 words in a single transcript and 14,535 unique words (Figs. 1 and 2).

4.2 Pre-processing of video and textual modalities

For video processing, the initial steps involved the extraction of three frames from each video at consistent intervals, coupled with the retrieval of audio in the WAV format (Fig. 3). These audio data were subsequently converted into image formats, encapsulating representations like Mel spectrogram (Fig. 4), chromagram (Fig. 5), and spectral bandwidth (Fig. 6), making them apt for analysis via convolutional neural networks (CNNs). To ensure the accuracy and clarity of these image and audio outputs, a series of processing tasks were employed, such as noise

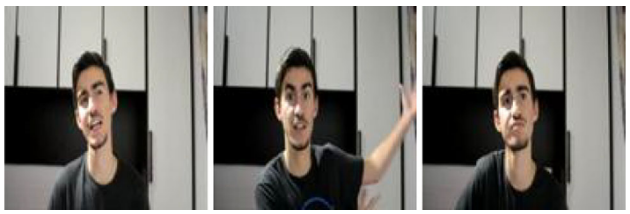


Fig. 2 Frames of the dataset

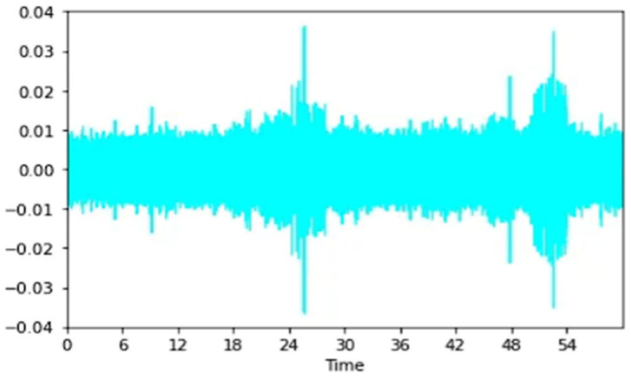


Fig. 3 Audio data in WAV format

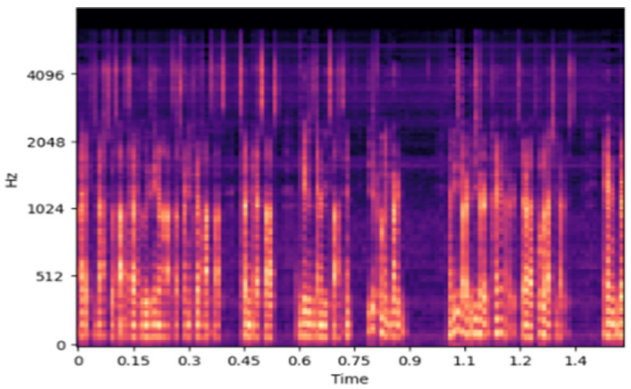


Fig. 4 Mel spectrogram

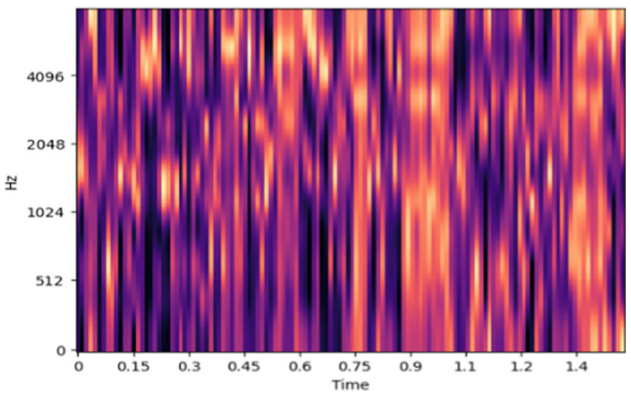


Fig. 5 Chromagram

reduction, resizing, and cropping. Renowned libraries like OpenCV, PIL, and librosa proved instrumental in this phase. A pre-processed audio data are shown in Fig. 7.

Transitioning to transcript processing, transcriptions of the videos were provided in pickle files with the annotations in the dataset itself and thus used directly for the models. The transcript data have been pre-processed using Tokenizers after filtering out punctuations and irrelevant tokens. The transcript contains 435,984 words, averaging to 43 words per video, with a maximum of 85 words in a

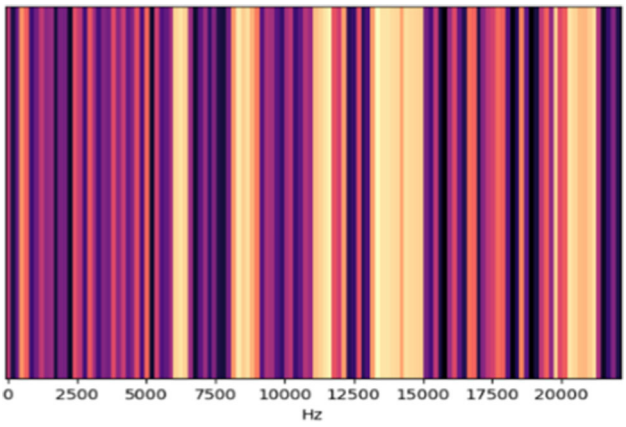


Fig. 6 Spectral bandwidth

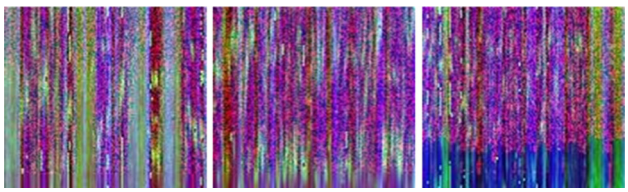


Fig. 7 (128,128,3) shaped audio images generated after pre-processing

single transcript and 14,535 unique words. Using the transcription and annotation pickle files, we have generated a CSV file which included a text transcript and the 5 labels of Big Five traits corresponding to the videos the transcript has been extracted from. The data-frame generated from the csv file is shown in Fig. 1. These data are tokenized using BertTokenizer and DebertaTokenizer, respectively, for both the models, after adding [CLS], [SEP] and [PAD] tokens to it.

4.3 Deep neural architectures

When it came to model application, a varied yet targeted approach was adopted. The video modality, which encompasses both frames and the transformed audio images, was analysed using models such as InceptionV3 [15], EfficientResNet [16], and XceptionResNet [17]. In contrast, the textual content derived from transcripts was put through transformer models like BERT [18] and DeBERTa [19]. Our recent paper [20] demonstrates the applicability of transformer-based models in understanding complex personality traits from textual data. Recognizing the potency of a combined approach, fusion models were developed, seamlessly integrating features from video-focused models with the textual capabilities of BERT, ensuring a comprehensive multimodal data analysis. With our processed video frames and audio feature images in hand, the next crucial step is data fusion. We fuse the processed video frames and audio feature images into a single input tensor. This integration of modalities allows our model to learn from both visual and auditory cues simultaneously.

4.3.1 InceptionV3

InceptionV3, an advanced convolutional neural network introduced by Google, is designed to improve computational efficiency and accuracy. It utilizes inception modules with various filter sizes (1×1, 3×3, 5×5) and pooling operations to efficiently capture features at different scales. The factorized 7×7 convolutions further reduce computational complexity, allowing for better performance with fewer parameters compared to traditional architectures. InceptionV3 has demonstrated remarkable results in computer vision tasks like image classification and object detection. Its versatility makes it suitable for various applications, and pre-trained models can be fine-tuned for specific tasks through transfer learning. However, InceptionV3's drawback lies in its high computational cost and memory requirements, making it challenging for resource-constrained environments. Additionally, its architecture might not be as effective for tasks involving extensive sequential data modelling, such as certain natural language processing tasks.

4.3.2 EfficientResNet

EfficientNet and ResNet are two prominent convolutional neural network architectures known for their performance and efficiency. EfficientNet scales width, depth, and resolution uniformly, balancing accuracy and computational cost. ResNet addresses deep network training issues with its residual or skip connections, allowing for deeper and more effective training. We've merged both architectures into EfficientResNet, a model that incorporates ResNet's skip connections into the EfficientNet design. The fused data, which now include information from both video frames and audio features, serve as the input to our EfficientResNet hybrid model. The final output is regression-based.

4.3.3 XceptionResNet

Xception (Extreme Inception) is a deep convolutional neural network architecture that builds upon the idea of the Inception architecture. It replaces the standard convolutional layers in Inception modules with depth-wise separable convolutions, which significantly reduces the number of parameters and computations while maintaining or even improving performance. This architecture aims to capture complex patterns in a more efficient manner. ResNet

(Residual Network), on the other hand, introduces residual connections that allow gradients to flow more easily during training, enabling the training of very deep networks. Residual connections involve adding the original input to the output of a layer, which helps in alleviating the vanishing gradient problem. This innovation has enabled the creation of extremely deep neural networks that were previously difficult to train. We have developed an innovative architecture called XceptionResNet, harnessing the collective strengths of Xception and ResNet designs to engineer a potent and resource-efficient neural network. By leveraging the individual advantages of Xception and ResNet paradigms, this hybridized approach engenders a unified model marked by heightened performance and computational efficiency.

4.3.4 BERT

BERT (Bidirectional Encoder Representations from Transformers) is a transformer-based neural network architecture that revolutionized natural language processing (NLP). Introduced in 2018, BERT addresses the contextual understanding challenge in NLP by training on a large corpus of text data using a masked language modelling objective. This allows BERT to capture deep contextual relationships between words, making it highly effective for a wide range of NLP tasks. The architecture of BERT consists of multiple layers of self-attention mechanisms, enabling it to capture intricate dependencies between words in both directions. By leveraging BERT's bidirectional contextual understanding, the processed transcripts can be analysed in-depth to discern subtle nuances in language use, emotional expressions, and linguistic patterns. This bidirectional approach empowers the model to capture not only what is explicitly stated but also the underlying context and relationships between words, allowing for a more comprehensive and accurate assessment of individuals' personality traits, including the Big Five (Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism). BERT's advanced natural language processing capabilities make it a valuable tool for uncovering the intricacies of human communication, enabling insights into the rich tapestry of personalities expressed through written or spoken language.

4.3.5 DeBERTa

DeBERTa (Decoding-enhanced BERT with Disentangled Attention) is an influential advancement in the realm of natural language processing (NLP), building upon the foundation of BERT. Introduced as A model architecture to address limitations in understanding global context, DeBERTa leverages disentangled self-attention mechanisms to enhance contextual comprehension. By untangling different aspects of language, DeBERTa excels in capturing long-range dependencies, making it well-suited for tasks requiring extensive sequential data modelling.

4.4 Multimodal fusion architectures

This section presents a comprehensive overview of the fusion models employed in personality trait prediction. In the preceding section, the deep neural architectures described are now employed in an integrated manner, incorporating the technique of weighted fusion. Weighted fusion is a valuable approach applied to amalgamate data from diverse modalities within the realm of multimodal data analysis. This technique grants the capability to allocate differing levels of significance to the contributions of each modality, contingent upon their pertinence and effectiveness in a given task. By doing so, it enhances the overall performance and resilience of multimodal systems, ensuring that the most pertinent and impactful information from various sources is effectively harnessed for comprehensive analysis and prediction. The first architecture, Inception V3 + BERT, combines InceptionV3 and BERT models to simultaneously process textual, image, and audio data. The video modality is processed using InceptionV3, while the text modality undergoes transformation by the BERT transformer. The concatenated outputs from these models pass through a series of densely connected layers to predict the Big Five traits. Similarly, the EfficientResNet + BERT architecture combines EfficientResNet and BERT for multimodal data processing. EfficientResNet handles image inputs, while BERT processes text inputs. These outputs are fused using a weighted fusion technique and subsequently processed through dense layers to predict personality traits. The final fusion architecture, XceptionResNet + BERT, utilizes Xception blocks with residual connections for image inputs and pretrained BERT for text data, ultimately employing weighted fusion and dense layers to predict personality traits. Figure 8 illustrates the XceptionResNet + BERT multimodal fusion architecture.

Incorporating findings from the First Impression dataset used in this study, we can leverage advanced AI models to decode subtle cues in candidates' expressions, speech patterns, and behaviours that signal underlying personality traits. This method enables a nuanced understanding of how a candidate's personality aligns with the cultural and functional demands of a job role. By applying a model trained on this dataset, recruiters can systematically evaluate and match candidates to positions where they are most likely to thrive, thereby optimizing job satisfaction and organizational productivity. This approach not only streamlines the recruitment process but also ensures a

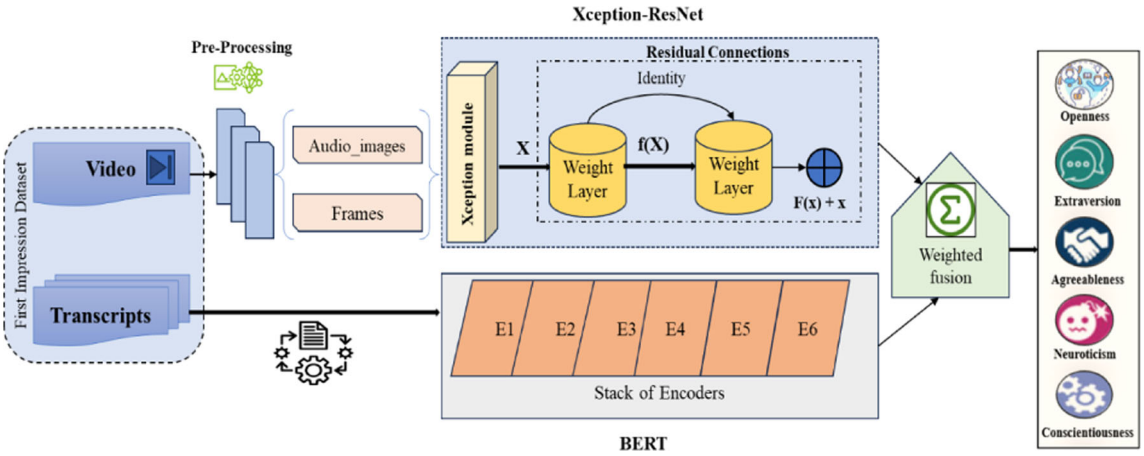


Fig. 8 XceptionResNet + BERT structure

deeper harmony between an individual's innate characteristics and their professional responsibilities, embodying the essence of personality-driven employment.

5 Results and discussions

This section presents a comprehensive analysis of the experimental results obtained from the multimodal models used for predicting personality traits. Evaluation metrics such as accuracy, F1-score, precision, and recall are employed to assess the performance of these models. The discussion highlights the model performance across different data modalities and provides insights into the effectiveness of each approach. The following hyperparameters (Table 1) have been tuned while training the models:

The following metrics have been used while analysing the models [21]:

- *Accuracy* The degree to which our predicted outputs conform with the true labels of the data. It has been calculated by first calculating the mean difference between the predictions and target labels, and subtracting it from 1, giving us a measure of how close our predictions were to the true labels.
- *R2 score* The R2 score, or coefficient of determination, measures the proportion of the variance in the dependent variable that is explained by the independent variables in a regression model. It is calculated by dividing the explained variance by the total variance and ranges from 0 to 1, where a higher value indicates a better fit of the model to the data.
- *Mean squared error (MSE)* It quantifies the average squared difference between predicted and actual values in a regression model, providing a measure of the model's prediction accuracy.

Table 1 Model hyperparameters

Model	Hyperparameter
BERT	Learning Rate = 5e-5
DeBERTa	Learning Rate = 1e-4
InceptionV3	Learning rate = 0.03
	Alternate ReLU and Tanh activation functions for Dense layers
	SGD optimizer with 0.9 momentum instead of Adam allows the gradient to not get stuck at a local minimum easily
EfficientResNet	Learning rate = 0.01 initially, with an exponential decay with weight decay = 0.3 over 500 steps (approximately 2.5 epochs)
	Sigmoid dense layer followed by a linear regression layer for output
	Weighted fusion of convolutional outputs from frames, audio data and transformers output of transcript with weights 0.1 for each frame, 0.4 for audio and 0.3 for transcript
XceptionResNet	Learning rate = 0.001
	Weighted fusion of convolutional outputs from frames, audio data and transformers output of transcript with weights 0.2 for each frame, 0.3 for audio and 0.1 for transcript

- *Root mean squared error (RMSE)* It is the square root of the mean squared error and is a better measure for the accuracy of a model.

These metrics collectively offer a comprehensive view of the performance of our multimodal models and enable us to draw meaningful conclusions regarding their effectiveness in personality trait prediction across various modalities.

5.1 Performance across modalities

We begin by analysing the performance of our models across various data modalities. Table 2 summarizes the results for different models and modalities:

From Table 2, we observe that the performance of the models varies across modalities. For text, DeBERTa performs poorly in terms of R2 score, while for audio and video, EfficientResNet and InceptionV3 perform well, respectively. When considering all modalities, XceptionResNet + BERT stands out with the highest accuracy, a high R2 score, and the lowest mean squared error and root mean squared error. It appears to be the best-performing model for the task involving all modalities. The bar chart in Fig. 9 illustrates the accuracy of various deep learning models in processing text, image, and frame data modalities for personality analysis, with EfficientResNet showing a leading performance.

5.2 Results on text modality

The results for two different models namely BERT and DeBERTa in terms of average accuracy and accuracy on various personality traits are in Table 3 given.

Both models demonstrate high average accuracy, with BERT scoring 88.09% and DeBERTa scoring 88.34%.

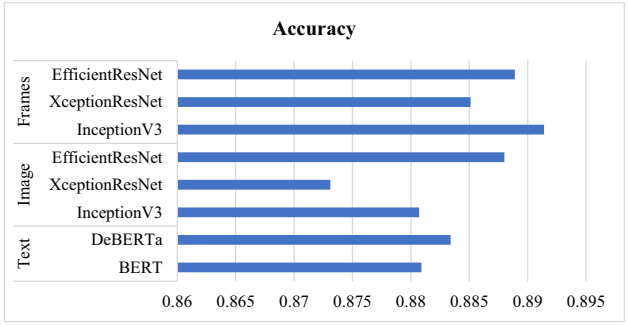


Fig. 9 Model comparison for personality analysis: accuracy of multimodal data processing

This suggests that both models perform well on the overall task of personality trait prediction from text data. Analysing the trait-wise accuracies, we find that Extraversion is the easiest trait to identify using the transcripts of the videos. This implies that individuals' spoken words can effectively predict whether they are extroverts or introverts. However, the other traits are relatively more challenging to identify from text data. Notably, Openness proves to be the most difficult to predict, likely requiring the assessment of body language and non-verbal cues.

5.3 Results on image modality generated from audio files

Moving on to the image modality derived from audio files, we summarize the results in Table 4:

In this modality, Extraversion remains the most predictable trait, closely followed by Openness. Conscientiousness appears to be the most challenging trait to identify. EfficientResNet appears to perform the best overall, with the highest average accuracy and the highest accuracy on most traits.

Table 2 Performance across modalities

Modalities	Model	Accuracy	R2 Score	Mean squared error	Root mean
Text	BERT	0.8809	− 2.60	0.0216	0.1469
	DeBERTa	0.8834	-8.31×10^9	0.0215	0.1467
Image	InceptionV3	0.8807	− 1.62%	0.0219	0.1478
	EfficientResNet	0.8880	9.97%	0.0193	0.1390
	XceptionResNet	0.8731	− 14.46%	0.0247	0.1569
Frames	InceptionV3	0.8914	15.37%	0.0182	0.1347
	EfficientResNet	0.8889	12.12%	0.0189	0.1374
	XceptionResNet	0.8851	4.64%	0.0205	0.1431
Fusion/ Decision models	InceptionV3 + BERT	0.8905	13.05%	0.0187	0.1366
	EfficientResNet + BERT	0.8876	10.55%	0.0194	0.1394
	XceptionResNet + BERT	**0.9212**	**54.49%**	**0.0098**	**0.0992**

The bold values indicate the best performing model

Table 3 Performance of text modality models

Model	Avg. Accuracy (%)	Accuracy on various traits				
		Extraversion (%)	Openness (%)	Agreeableness (%)	Neuroticism (%)	Conscientiousness (%)
BERT	88.09	89.24	87.25	87.85	87.81	88.28
DeBERTa	88.34	89.49	88.23	87.77	87.70	88.52

Table 4 Performance of audio-generated images models

Model	Avg. accuracy (%)	Accuracy on various traits				
		Extraversion	(%)Openness	(%)Agreeableness (%)	Neuroticism	(%)Conscientiousness (%)
InceptionV3	88.07	89.42	88.47	87.40	87.67	87.38
EfficientResNet	88.80	89.98	89.23	88.29	88.37	88.13
XceptionResNet	87.31	88.70	87.76	86.36	87.01	86.73

5.4 Results on frames generated from video files

For frames generated from video files, the results are summarized in Table 5:

In this modality, InceptionV3 delivers the best performance, achieving the highest average accuracy and excelling in most individual traits. Extraversion and Openness continue to be the easiest traits to identify from First Impressions, while traits like conscientiousness and neuroticism, which depend on one's approach to various life domains, remain challenging to judge from initial encounters.

5.5 Comparative analysis of the results

Inconsistent results are obtained when traits are analysed using multimodal decision models, as some of the models can analyse all the modalities together, while some of them fail to do so. This analysis (Table 6) shows that InceptionV3 + BERT Architecture is biased towards the video modality, hence showing results like those of Video models, while EfficientResNet + BERT architecture shows results closer to that of text models. The only model which could fit to all the modalities the best was XceptionResNet + BERT, which showed almost same accuracy on all labels except Neuroticism, thus making it the toughest trait to identify when we can fully assess the First Impression from a video, undertaking all its modalities.

Our study provides valuable insights into the effectiveness of multimodal models for predicting personality traits. The choice of data modality significantly impacts model performance, with multimodal fusion models showing superior accuracy. This suggests that combining textual, visual, and auditory information offers a holistic understanding of an individual's personality. The ease of predicting certain traits, such as Extraversion, from textual and visual data implies that linguistic and non-verbal cues play a pivotal role in personality assessment. On the other hand, traits like Openness, Conscientiousness, and Neuroticism may require a more comprehensive assessment, considering multiple modalities. The multimodal fusion architecture, XceptionResNet + BERT, stands out as a robust choice for personality trait prediction, with consistently high accuracy across all traits. However, careful consideration of the modality and task at hand is essential when choosing the most appropriate model architecture.

Table 5 Performance of video frames

Model	Avg. Accuracy (%)	Accuracy on various traits				
		Extraversion	(%)Openness	(%)Agreeableness (%)	Neuroticism	(%)Conscientiousness (%)
InceptionV3	89.14%	89.93	89.66	88.63	88.89	88.59
EfficientResNet	88.89%	89.99	89.28	88.35	88.61	88.21
XceptionResNet	88.51%	89.54	88.71	87.78	88.64	87.90

Table 6 Performance of multimodal decision models

Model	Avg. accuracy (%)	Accuracy on various traits				
		Extraversion	(%)Openness (%)	Agreeableness (%)	Neuroticism (%)	Conscientiousness (%)
InceptionV3 + BERT	89.05	89.94	89.56	88.46	88.78	88.49
EfficientResNet + BERT	88.76	89.56	89.09	88.65	88.08	88.39
XceptionResNet + BERT	92.12	92.35	92.24	92.16	91.31	92.51

Our system, leveraging the First Impression dataset, introduces several advancements in the field of personality-driven employment:

- *Enhanced accuracy in personality assessment* Utilizing multimodal data analysis, the system provides a more accurate and holistic view of an individual's personality traits, improving the precision of job-role matching.
- *Dynamic adaptability* By integrating real-time labour market analytics, the system can adapt to evolving job market trends, ensuring continuous alignment with current employment needs.
- *Scalable and comprehensive analysis* The architecture is designed to handle large datasets, allowing for a comprehensive analysis of personality traits across diverse job sectors.

5.6 Limitations and Ethical Concerns

While AI-driven personality analysis offers transformative potential for employment strategies, it is imperative to address the associated limitations and ethical concerns comprehensively:

- *Privacy and data protection* The use of AI for personality analysis raises significant privacy concerns, particularly regarding the collection, storage, and processing of sensitive personal data. Ensuring robust data protection measures and adhering to privacy regulations like GDPR are paramount to maintain individuals' trust and safeguard their personal information.
- *Fairness and bias* AI models can inadvertently perpetuate biases present in the training data, leading to unfair treatment of certain groups based on gender, race, ethnicity, or other characteristics. It is critical to implement measures to detect and mitigate such biases, ensuring that AI-driven personality assessments are fair and equitable across diverse populations.
- *Consent and transparency* Obtaining explicit consent from individuals whose data are used for personality analysis is essential. Furthermore, the processes and criteria used in AI-driven assessments should be transparent, allowing individuals to understand how their personality profiles are evaluated and used in employment decisions.
- *Regulatory compliance* Navigating the complex landscape of legal and ethical standards is a significant challenge. AI systems must be designed and operated in compliance with all applicable laws and ethical guidelines to avoid legal repercussions and maintain public confidence.
- *Moral and ethical responsibility* Beyond legal compliance, there is a moral obligation to use AI in a way that respects individual dignity and promotes positive outcomes for society. This includes careful consideration of how personality analysis can impact individuals' employment opportunities and career development.

In conclusion, while AI-driven personality analysis in employment contexts offers promising benefits, it is crucial to address these limitations and ethical concerns proactively. Doing so will not only enhance the fairness and effectiveness of these systems but also ensure their sustainability and acceptance in the long term.

6 Hyper-personalized employment in urban hubs: the role of personality traits across industries

The balance between traits and skills is crucial as shown in Fig. 10. Traits may provide the foundation for potential, while skills are often the practical tools through which one can execute tasks effectively. In a professional setting, both are important: traits can drive one's approach to work and interpersonal interactions, while skills are necessary to perform specific job functions. Together, they contribute to an individual's overall competency and potential for success in their career.

This section presents a series of case studies which underscore the universal importance of recognizing and nurturing the Big Five personality traits across various industries. In each sector, from technology to banking,

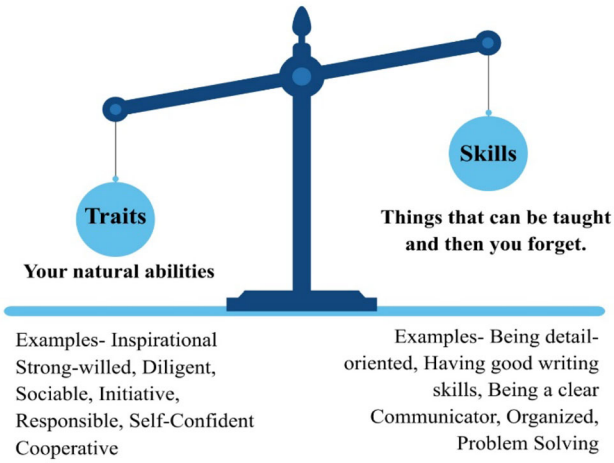


Fig.10 Future of urban employment: a balanced approach

healthcare to retail, the personalities of individuals significantly influence their effectiveness in their respective roles.

6.1 Case study 1: the role of personality traits in retail customer service and visual merchandising

In the dynamic landscape of the retail industry, success hinges not only on products and services but also on the human factor. Customer service and visual merchandising are two critical components that significantly impact a retail business's bottom line [22]. However, what often goes overlooked are the specific personality traits that make individuals excel in these roles. In this case study, we examine the relevance of the Big Five personality traits—Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness—in the contexts of Customer Service Representatives (CSRs) and Visual Merchandisers. The mapping of personality traits to retail roles is as follows:

- Extraversion

a) *CSR* Extraversion is valuable for CSRs. Enthusiastic and sociable CSRs engage customers warmly, creating a positive shopping experience.
b) *Visual merchandiser* While less critical, a touch of extraversion can benefit Visual Merchandisers in team collaborations and discussions about store aesthetics.

- Agreeableness

a) *CSR* Agreeableness is essential for CSRs. Friendliness, empathy, and a cooperative attitude are crucial for resolving customer issues and building rapport.
b) *Visual merchandiser* Agreeableness can help Visual Merchandisers collaborate effectively within the team and understand the preferences of diverse customers.

- Conscientiousness

a) *CSR* Attention to detail and reliability are vital for CSRs. They ensure accurate transaction processing and effective inventory management.
b) *Visual merchandiser* Conscientiousness plays a pivotal role in visual merchandising. Attention to detail ensures that displays are impeccable and on-brand.

- Neuroticism

a) *CSR* Low neuroticism is beneficial for CSRs. It helps them remain calm and composed, even in challenging situations.
b) *Visual Merchandiser* A moderate level of neuroticism can drive perfectionism in Visual Merchandisers, ensuring meticulous attention to visual details.

- Openness

a) *CSR* Openness is not as crucial for CSRs but can facilitate adaptability to various customer needs.
b) *Visual merchandiser* Openness to new design ideas and trends is vital for Visual Merchandisers to create innovative and visually captivating displays.

6.2 Case study 2: the role of personality traits in healthcare—nurses and hospital Administrators

The healthcare industry is a demanding and complex field that relies not only on technical skills but also on interpersonal abilities and emotional intelligence. In this case study, we examine how the Big Five personality traits—Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness—play a crucial role in the professions of Nurses and Hospital Administrators [23]. The mapping of personality traits to healthcare roles is as follows:

- Extraversion

a) *Nurse* Extraversion can benefit nurses in their interactions with patients and colleagues. A warm and sociable demeanour can help patients feel at ease and foster teamwork among healthcare staff.

b) *Hospital Administrator* While not as essential as for nurses, some level of extraversion can aid hospital administrators in effective communication with staff and patients, especially during crisis management.

- Agreeableness

a) *Nurse* Agreeableness is vital for nurses as it facilitates empathy and compassion when caring for patients. Patients often appreciate nurses who are kind, patient, and understanding.
b) *Hospital Administrator* Hospital administrators benefit from agreeableness when working with diverse healthcare teams and ensuring a harmonious working environment.

- Conscientiousness

a) *Nurse* Conscientiousness is a fundamental trait for nurses. Attention to detail, adherence to protocols, and organization are essential for providing high-quality patient care.
b) *Hospital Administrator* Hospital administrators rely heavily on conscientiousness to manage resources efficiently, maintain compliance with regulations, and uphold high standards of patient safety and care.

- Neuroticism

a) *Nurse* Lower levels of neuroticism can be advantageous for nurses as they often work in high-stress environments. Remaining calm and composed during critical situations is crucial.
b) *Hospital Administrator* Hospital administrators with lower neuroticism can handle the pressures of healthcare management effectively, ensuring the smooth operation of the hospital.

- Openness

a) *Nurse* Openness to new medical research and treatment methods can enhance a nurse's ability to provide the latest evidence-based care to patients.
b) *Hospital Administrator* Openness can be beneficial for hospital administrators when considering innovative healthcare practices, technologies, and management strategies.

6.3 Case study: the role of personality traits in banking—customer service representatives and risk analysts

The banking sector is a dynamic industry that relies not only on financial acumen but also on interpersonal skills and risk management [24]. This case study explores how the Big Five personality traits—Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness—play a pivotal role in the professions of Customer Service Representatives and Risk Analysts within the banking industry. The mapping of personality traits to banking roles is as follows:

- Extraversion

a) *Customer Service Representative* Extraversion can be a valuable trait for customer service representatives. An outgoing and approachable demeanour can enhance interactions with customers, leading to better customer satisfaction and loyalty.
b) *Risk Analyst* While not as crucial as for customer service representatives, some level of extraversion can aid risk analysts in effectively communicating their findings and insights to the bank's management or clients.

- Agreeableness

a) *Customer Service Representative* Agreeableness is vital for customer service representatives as it enables them to handle customer inquiries and complaints with empathy and patience, fostering positive customer relationships.
b) *Risk Analyst* Risk analysts benefit from agreeableness when working within a team to assess and mitigate financial risks, ensuring smooth collaboration.

- Conscientiousness

a) *Customer Service Representative* Conscientiousness is a fundamental trait for customer service representatives. Attention to detail, timeliness, and accuracy are essential when dealing with customers' financial transactions and inquiries.
b) *Risk Analyst* Conscientiousness is equally critical for risk analysts as it influences their ability to conduct thorough risk assessments, maintain meticulous records, and adhere to regulatory compliance.

- Neuroticism

a) *Customer Service Representative* Lower levels of neuroticism can be advantageous for customer service representatives as they often handle high-stress customer interactions. Remaining calm under pressure is vital.
b) *Risk Analyst* Risk analysts with lower neuroticism can objectively evaluate financial risks without being influenced by emotions or anxiety, ensuring more accurate risk assessments.

- Openness

a) *Customer Service Representative* While not as crucial as other traits, openness can benefit customer service representatives by enabling them to adapt to changing banking procedures and technologies.
b) *Risk Analyst* Openness can be beneficial for risk analysts when considering innovative risk assessment methodologies and staying updated with evolving financial markets.

6.4 Case study: the role of personality traits in the technology sector—software developers and UX designers

The technology sector is known for its innovation and dynamic nature. In this case study, we will examine how the Big Five personality traits—Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness—influence the work of Software Developers and User Experience (UX) Designers [25]. The mapping of personality traits to tech roles is as follows:

- Extraversion

a) *Software Developer* While software development often involves independent work and coding, some level of extraversion can be beneficial for effective collaboration with team members, especially in Agile environments.
b) *UX Designer* Extraversion can help UX designers when conducting user interviews and gathering feedback. Being able to engage with users and stakeholders can lead to better-designed products.

- Agreeableness

a) *Software Developer* Agreeableness can contribute to harmonious teamwork within development teams, fostering a positive and collaborative work environment.
b) *UX Designer* UX designers with high agreeableness can communicate more effectively with team members and stakeholders, facilitating the incorporation of feedback into the design process.

- Conscientiousness

a) *Software Developer* Conscientiousness is highly valuable for software developers. Attention to detail, commitment to deadlines, and thorough code reviews are essential for creating high-quality software.
b) *UX Designer* Conscientiousness ensures that UX designers meticulously plan and execute user research, usability testing, and design iterations, leading to user-friendly interfaces.

- Neuroticism

a) *Software Developer* Lower neuroticism can be advantageous for software developers as it helps them remain calm under pressure and adapt to rapidly changing technologies.
b) *UX Designer* UX designers with lower neuroticism can handle design critique and user feedback more constructively, ensuring the iterative design process remains productive.

- Openness

a) *Software Developer* Openness is valuable for software developers when exploring new programming languages, frameworks, and approaches to problem-solving.
b) *UX Designer* UX designers with openness are more likely to embrace innovative design concepts and adapt to emerging design trends, resulting in more creative and user-centric solutions.

These case study highlights the importance of recognizing and nurturing the Big Five personality traits that serve as invaluable indicators for role alignment and workforce optimization, empowering organizations across industries to achieve success and foster innovation.

6.5 Economic impact analysis

The deployment of our AI-driven hyper-personalized job matching system is anticipated to have substantial economic implications for the job market and organizational productivity.

- *Job Market Efficiency* By aligning candidates to roles that resonate with their innate traits and acquired skills, the system can significantly streamline the hiring process, reducing the time and resources spent on finding the right candidate for the job.
- *Employee Turnover Rates* With a more precise match between job roles and personal traits, employee satisfaction is likely to increase, leading to a reduction in turnover rates. This can result in considerable cost savings for organizations, as the expenses associated with recruiting and training new employees are lowered.
- *Overall Productivity* Employees who are well-matched to their roles are more likely to be engaged and motivated, enhancing productivity. This improved efficiency can contribute to the organization's bottom line and economic growth in the broader market.

6.6 Technical implementation: addressing scalability and logistical challenges

The technical execution of our hyper-personalized job matching system, while feasible, necessitates a detailed examination of scalability and logistical considerations to ensure successful real-world application:

- *Scalability Concerns* Technically, the system must be designed to efficiently handle varying scales of operation, from small enterprises to large multinational corporations. This involves optimizing the system architecture for load handling, data processing speed, and storage capacity, ensuring that it can expand or contract based on the organization's size and requirements.
- *Logistical Integration* From a logistical standpoint, integrating the system into existing human resources and IT infrastructures is paramount. This includes establishing seamless data pipelines, ensuring compatibility with current HR software, and providing mechanisms for continuous data update and retrieval.
- *Resource Allocation* Adequate allocation of computational and human resources is essential for maintaining system performance. This involves planning for the necessary hardware, software, and personnel to manage the system effectively and respond to any issues promptly.
- *System Maintenance and Support* Ongoing technical support and maintenance are crucial for long-term system stability and reliability. Regular updates, bug fixes, and system audits should be planned to ensure the system remains current and effective.
- *Training and Adaptation* To facilitate logistical adoption, comprehensive training programs for HR professionals and end-users are necessary. This will aid in smooth system integration and ensure that users are proficient in leveraging the system's capabilities.

Addressing these technical and logistical elements will not only enhance the system's practical deployment but also ensure its sustainable operation and scalability in diverse organizational environments.

7 Conclusion

In the context of smart cities, this research addresses the shift in employment strategies towards greater personalization, with the goal of enhancing job satisfaction and economic efficiency. The study utilizes multimodal deep learning techniques to create a hyper-personalized job matching system based on individual personality traits, drawing on the First Impressions V2 dataset. Among various model architectures, the fusion of XceptionResNet with BERT demonstrates exceptional performance, achieving a remarkable accuracy of 92.12% and other robust metrics. These empirical findings underscore the efficacy of the XceptionResNet + BERT approach in personality trait mapping, presenting an innovative and efficient method for job matching in urban environments. This research has the potential to revolutionize recruitment practices in smart cities by ensuring job placements that align not only with skills but also with individual personality traits, thereby optimizing both individual job satisfaction and organizational productivity. Theoretical case studies across diverse sectors further illustrate the model's potential to enhance both individual and organizational outcomes in smart city settings.

The limitations of this research lie in its dataset specificity, ethical considerations, interpretability, and potential biases, which should be mitigated in future studies. Future work should encompass real-world validation, multimodal enhancement, dynamic personality assessment, bias mitigation strategies, user-centred design, cross-cultural adaptation, and regulatory compliance to refine and extend the model's applicability. By addressing these limitations and pursuing these avenues, the research can contribute to the development of a more robust, fair, and globally adaptable personality-based job matching system, thus advancing the evolution of modern employment strategies within the dynamic landscape of smart cities and the Fourth Industrial Revolution.

Funding No funding was received for this research.

Data availability A publicly available dataset is used in this research; dataset will be made available on request.

Declarations

Conflict of Interest The authors have no competing interests related to the work submitted for publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Rose, G. (Ed.). (2016). *The fourth industrial revolution: A Davos reader*. Foreign Affairs.
2. Purcell KPJ (1998) In-sourcing, outsourcing, and the growth of contingent labour as evidence of flexible employment strategies. Eur J Work Organ Psy 7(1):39–59
3. Ponce-López V, Chen B, Oliu M, Corneanu C, Clapés A, Guyon I, Baró X, Escalante HJ, Escalera S. (2016). Chalearn lap 2016: First round challenge on first impressions-dataset and results. In: *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8–10 and 15–16, 2016, Proceedings, Part III 14* (pp. 400–418). Springer International Publishing.
4. Nirthika R, Manivannan S, Ramanan A, Wang R (2022) Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. Neural Comput Appl 34(7):5321–5347
5. Wang J, Li H, Wang L, Wu C (2023) A multimodal dialogue system for improving user satisfaction via knowledge-enriched response and image recommendation. Neural Comput Appl 35(18):13187–13206
6. Pianesi F, Mana N, Cappelletti A, Lepri B, Zancanaro M (2008). Multimodal recognition of personality traits in social interactions. In: Proceedings of the 10th international conference on Multimodal interfaces (pp. 53–60).
7. Alam F and Riccardi G. (2014). Predicting personality traits using multimodal information. In Proceedings of the 2014 ACM multi media on workshop on computational personality recognition (pp. 15–18).
8. Escalante HJ, Kaya H, Salah AA, Escalera S, Güçlütürk Y, Güçlü U, Baró X, Guyon I, Junior JC, Madadi M, Ayache S (2018). Explaining first impressions: Modeling, recognizing, and explaining apparent personality from videos. *arXiv preprint* arXiv:1802.00745.
9. Gorbova J, Avots E, Lüsi I, Fishel M, Escalera S, Anbarjafari G (2018) Integrating vision and language for first-impression personality analysis. IEEE Multimedia 25(2):24–33
10 Junior JCJ, Güçlütürk Y, Pérez M, Güçlü U, Andujar C, Baró X, Escalera S (2019) First impressions: A survey on vision-based apparent personality trait analysis. IEEE Trans Affect Comput 13(1):75–95
11. Wingate TG, Bourdage JS (2019) Liar at first sight? Early impressions and interviewer judgments, attributions, and false perceptions of faking. J Pers Psychol 18(4):177
12. Nørskov S, Damholdt MF, Ulhøi JP, Jensen MB, Mathiasen MK, Ess CM, Seibt J (2022) Employers' and applicants' fairness perceptions in job interviews: using a teleoperated robot as a fair proxy. Technol Forecast Soc Chang 179:121641
13. Aslan S, Güdükbay U, Dibeklioğlu H (2021) Multimodal assessment of apparent personality using feature attention and error consistency constraint. Image Vis Comput 110:104163
14 Song S, Jaiswal S, Sanchez E, Tzimiropoulos G, Shen L, Valstar M (2021) Self-supervised learning of person-specific facial dynamics for automatic personality recognition. IEEE Trans Affect Comput. 14(1):178–195
15. Demir A, Yilmaz F, Kose O (2019). Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3. In: *2019 medical technologies congress (TIPTEKNO)* (pp. 1–4). IEEE.
16. Canayaz M (2021) C+ EffxNet: a novel hybrid approach for COVID-19 diagnosis on CT images based on CBAM and EfficientNet. Chaos, Solitons Fractals 151:111310
17. Panthakkan A, Anzar SM, Jamal S, Mansoor W (2022) Concatenated Xception-ResNet50—A novel hybrid approach for accurate skin cancer prediction. Comput Biol Med 150:106170
18. Jain D, Kumar A, Beniwal R (2022, July). Personality BERT: A Transformer-Based Model for Personality Detection from Textual Data. In: *Proceedings of International Conference on Computing and Communication Networks: ICCCN 2021* (pp. 515–522). Singapore: Springer Nature Singapore.
19. He P, Liu X, Gao J, Chen W. (2020). Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint* arXiv:2006.03654.
20 Jain D, Kumar A (2024) AI unveiled personalities: profiling optimistic and pessimistic attitudes in Hindi dataset using transformer-based models. Expert Syst. https://doi.org/10.1111/exsy.13572
21 Handelman GS, Kok HK, Chandra RV, Razavi AH, Huang S, Brooks M, Asadi H (2019) Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. Am J Roentgenol 212(1):38–43
22. Huang JL, Ryan AM (2011) Beyond personality traits: a study of personality states and situational contingencies in customer service jobs. Pers Psychol 64(2):451–488
23. Ganu D, Kogutu CN (2014) Effect of the big five personality traits on job satisfaction and organizational commitment in the healthcare industry: the case of Kenya. Am J Health Sci (AJHS) 5(2):145–154
24. Rashid NM, Sah NFM, Ariffin NM, Ghani WSWA, Yunus NSNM (2016) The Influence of Bank's frontlines' personality traits on job performance. Procedia Econ Financ 37:65–72
25. Yilmaz M, O'Connor RV, Colomo-Palacios R, Clarke P (2017) An examination of personality traits and how they impact on software development teams. Inf Softw Technol 86:101–122

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

2024 IEEE International Conference on Communication, Computing and Signal Processing (IICCCS) | 979-8-3503-9075-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/IICCCS61609.2024.10763664

Interconnection of SDN node with conventional network using an ONOS controller

Samiullah Mehraban

Dept. Of CSE, Delhi Technological University, Delhi, India

Mehraban780@gmail.com

Rajesh K Yadav

Dept. Of CSE, Delhi Technological University, Delhi.

rkyadav@dtu.ac.in

Abstract

For future network design, different architectures worldwide have been proposed; among them, one of the finest network designs is Software-Defined Networking (SDN), which separates the network data and control layers and enables the network with more agility, programmability, and flexibility. However, because of economic and technical challenges, directly upgrading the whole network to SDN is challenging for many organizations, and the optimal solution is to upgrade conventional network devices incrementally to SDN nodes, which is the Hybrid SDN. In a hybrid SDN environment, SDN nodes interconnect with the traditional legacy network, which takes advantage of both network designs. In this paper, we studied the integration of SDN with conventional legacy networks. We explore how to peer and integrate SDN nodes with conventional networks using the SDN-IP application located on the top of the application layer of the ONOS controller. Through different experiments on both networks, IP-IP and IP-SDN, we found that compared to IP-IP networks, IP-SDN improves the overall network performance and provides more outstanding performance in maximization of throughput and minimization of delay.

Keywords: SDN, TE, IP-IP, IP-SDN, BGP

1. INTRODUCTION

Software-defined networking (SDN) is an emerging network design that manages complex network infrastructure by separating the data and control layers. SDNs are employed in almost every environment to facilitate efficient communication among computer networks, including wireless networks, cloud computing, mobile telephones, data centers, and the Internet of Things. SDN has many features that provide agility and flexibility to the network. SDN's primary concept is decoupling network layers to improve network flexibility and agility. The primary advantage of SDN is the ability to manage the data flow dynamically and efficiently. After a flow passes through an SDN switch, its forwarding direction may be easily adjusted; we refer to this type of traffic as programmable traffic. For large-scale networks, it is necessary to develop a network that can be programmed to meet the need for innovation in network access; SDN technology simplifies network administration and control operations through programmable devices and automation.

The SDN design has three main layers: The application layer hosts network applications designed to communicate with data layer devices. The control layer, the leading layer of the SDN is mainly responsible for controlling the data layer, and the last layer is data layer, which hosts network elements accountable for deploying packet-forwarding flows. In the control layer of the SDN network, there should be at least one controller; we can have multiple controllers to avoid a single source of failure. For the Wide-area network in different locations, various controllers can be deployed physically to accomplish the purpose of a centralized control plane. **Figure 1** presents the SDN fundamental architecture, consisting of three main layers: Data, Control, and Application.

Recently, SDN has been widely researched and progressively implemented in several types of networks, including Data Centers, campus networks, enterprise networks and IXP (Internet Exchange Point) networks.

With all the advantages and futures an SDN network provides, directly migrating to a Fully SDN network is challenging because of technical and economic challenges. Hence, migrating to a hybrid SDN is the best solution, but there are challenges in integrating SDN capabilities with the traditional IP network. Due to economic and operational factors, SDN technology is implemented incrementally alongside conventional network devices. Namely, with each network upgrade, only some number of traditional network devices will be upgraded to programmable switches that support SDN. In many complex network environments, ISPs (Internet Service Providers) incrementally upgrade some of conventional network devices to SDN to increase traffic management capabilities.

This coexistence of SDN alongside traditional networks gives rise to the hybrid SDN. Current hybrid SDNs fail to address the possible performance drawbacks caused by a centralized SDN controller. Deploying a proper controller is another issue that may cause significant delays in processing flow requests. Because of these factors, many researchers have proposed the incremental deployment of SDN nodes alongside traditional network devices, such as traffic engineering [1], flexible routing [2], link failure recovery [3], power saving [4], and safe updates [5].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 06,2024 at 07:03:38 UTC from IEEE Xplore. Restrictions apply.

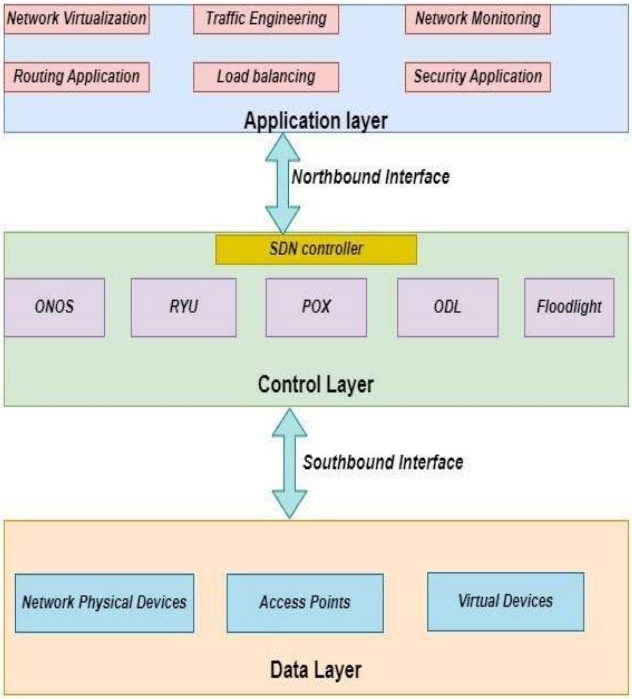


Figure 1. SDN architecture

One of the main elements of an SDN network is the controller; therefore, the controller must present features that ensure service continuity during failures and improve the overall network performance. Open Networking Operating Systems (ONOS) and Open Daylight controllers typically uses three controllers to provide robust services at a single location to guarantee network consistency. All the controllers communicate with each other. ONOS is an open-source controller with an application for integrating SDN networks with conventional legacy networks via BGP, that is, SDN-IP, located at the top of the application layer of the ONOS controller.

This paper investigates the interconnection of SDN nodes with conventional networks devices. We explore how to peer and integrate SDN with legacy networks using the SDN-IP application located on the top of the application layer of the ONOS controller. This application allows the SDN network to communicate with conventional network devices through BGP and explains how the controller converts the BGP information into open-flow entries. The Internet is a collection of autonomous systems (ASs) or interconnected networks.

On the other hand, on the internet, BGP is the standard EGP protocol designed to exchange routing among ASs. BGPv4 is the most generally deployed mechanism for peering ASs on the internet. So, a peering mechanism is necessary to integrate conventional networks with SDN-driven ASs; this interconnection is performed through an SDN-IP application, which operates on the top of the SDN controller.

2. Related Work

This section describes the existing research on the migration sequence and gradual deployment of SDN nodes in a hybrid SDN system. The gradual implementation of SDN nodes has received significant attention in recent times. Several aspects of a hybrid SDN may be enhanced during the migration process. In the hybrid SDN environment, multiple algorithms have been recently used to optimize different network characteristics.

D. Levin et al. [6] presented an optimization approach for determining the partial deployment of SDN, assuming that every traffic in the network passes through at least one SDN nodes. H. Lu et al. [7] developed a configuration technique to autonomously convert network settings from traditional legacy networks to Software-defined networks. C. Jin et al. [8] manipulated the forwarding entries in the conventional switches by implementing customized flow control methods. This method allowed the user to direct the forwarding flow down a specific path chosen by the user rather than following the path determined by the spanning tree protocol.

Jia et al. [9] suggested a method to gradually enhance SDN to optimize the amount of traffic that may be programmed while adhering to a specified budget restriction for the upgrades. Xu et al. [10] compared hybrid SDN by substituting existing conventional devices with SDN nodes or including new SDN nodes. Caria et al. [11] utilized the characteristics of network topologies and considered the network's centrality to modify switches in a hybrid SDN environment. Agarwal et al. [12] initially developed an optimization problem to achieve traffic engineering in SDN by partially deploying SDN nodes. S.Vissicchio et al.[2] implemented a centralized control mechanism for distributed OP routing by injecting carefully designed routing massages into OSPF; this approach improved the adaptability, variety, and dependability of Layer 3 routing.

Guo et al. [13] suggested modifying the connections and flow split ratio weights at the SDN nodes to perform load balancing in a specific hybrid SDN environment. Hong et al. [14] suggested a method to achieve many traffic objectives in the hybrid SDN environment. Nevertheless, none of the abovementioned studies interconnect SDN nodes with conventional legacy networks using ONOS controllers in a real network scenario. We have considered this issue as the interconnection of SDN nodes into conventional legacy networks with ONOS controllers using Mininet tools.

3. Hybrid SDN

In a hybrid SDN, conventional network devices coexist with SDN nodes and communicate differently to adjust, regulate, set up, and oversee network behaviour for the best possible user experience and performance. With the integration of both network devices, a hybrid SDN network takes the benefits of both the SDN and conventional networks, the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 06,2024 at 07:03:38 UTC from IEEE Xplore. Restrictions apply.

programmability and flexibility of SDN, and the robustness and familiarity of the traditional network; there is no need to convert all the legacy devices to SDN as both network devices work together, conventional network devices work on network robustness while SDN work on network optimization.

Conventional network devices can only support the OSPF (Open Shortest Path First) protocols. At the same time, SDN nodes can operate in hybrid mode by supporting OF (OpenFlow) protocol to communicate with SDN nodes and OSPF protocol to communicate with conventional network devices. Conventional legacy routers may forward traffic along the shortest path from origin to the destination, specified by the routing tables. At the same time, the SDN nodes can dynamically manage and divide the flows to multipath forwarding, which the SDN controller indicates. As a result, hybrid SDN combines the programmability and flexibility of a centralized network with the robustness and familiarity of conventional networks.

To interconnect SDN nodes alongside conventional legacy devices, we have simulated all the simulations using Mininet and Mini-Edit tools. We have created the topology and the hybrid SDN environment using these tools, and as a controller of SDN nodes, we have tested multiple controllers. Finally, we have selected the ONOS controller as the SDN controller for our hybrid SDN environment.

Mininet Simulator simulates the network environments [16]; for configuring legacy routers, we have used FRR (Free Range Routing) using VTY Shel to configure network protocols. To connect different networks in the conventional network and to integrate the legacy network with SDN AS (Autonomous Systems), we have configured BGP (Border Gateway Protocols) inside all the legacy routers, as it is necessary to have a peering system that makes the BGP protocol work. The integration between the legacy network and the SDN network is carried out through the SDN-IP. This program operates on top of the ONOS controller. The SDN-IP application allows SDN nodes to communicate with legacy routers using BGP protocols.

In this simulation experiment, the ONOS [17] controller is chosen to be the SDN controller, and the integration between the legacy network and the SDN network is carried out through the SDN-IP. This program operates on top of the ONOS controller.

First, we will open the Mininet tool to set up the network environment; with the MiniEdit tool, we can create and emulate network topologies. It also can configure network device interfaces such as IP addresses and gateways, and we can save the topologies.

To open the MiniEdit tool, we first open the Mininet tool and then execute this command to open the MiniEdit tool's graphical interface.

$ sudo ~/mininet/examples/miniedit.py

Some tools are inside the MiniEdit tool, like a host tool, select tool, switch tool, legacy router, legacy Switch, Netlink, and controller tools.

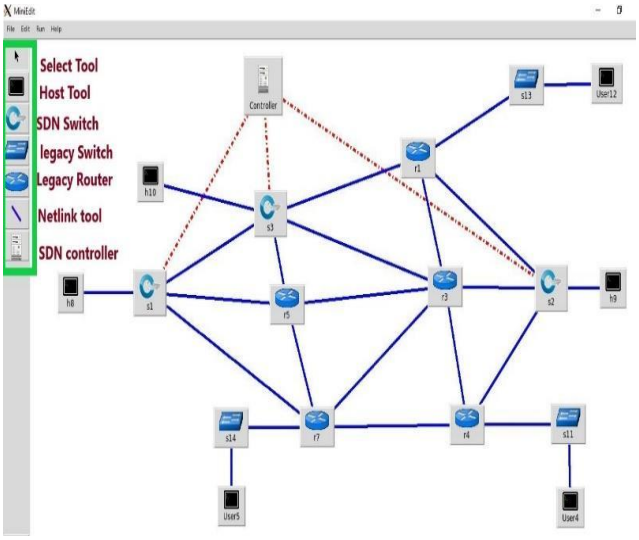


Figure .2 MiniEdit graphical interface

Select tool: used to select and move nodes from one area to another.

Host tool: Creates the nodes that conduct the host computer's functions.

Switch tool: This tool produces SDN-based OpenFlow switches that a controller manages.

Legacy switch: This tool creates a legacy Ethernet switch.

Legacy router tool: This tool generates a legacy router that does not require a controller.

NetLink tool: The NetLink tool creates linkages between nodes.

Controller tool: this tool is used to create a controller for the topology.

Once the topology is created, we will run it. To check the links on the command line, we will write links, and then we will assign the IP address according to the given topology; once we assign the IP address to the interfaces, we will check it. To verify whether the IP is assigned correctly, we can check it on each host by executing the “ifconfig” command.

We will configure and execute the commands to each single router to configure each legacy router. First, we will start the Zebra domain. It is a routing software with multiple servers which provides TCP/IP based routing protocols.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 06,2024 at 07:03:38 UTC from IEEE Xplore. Restrictions apply.

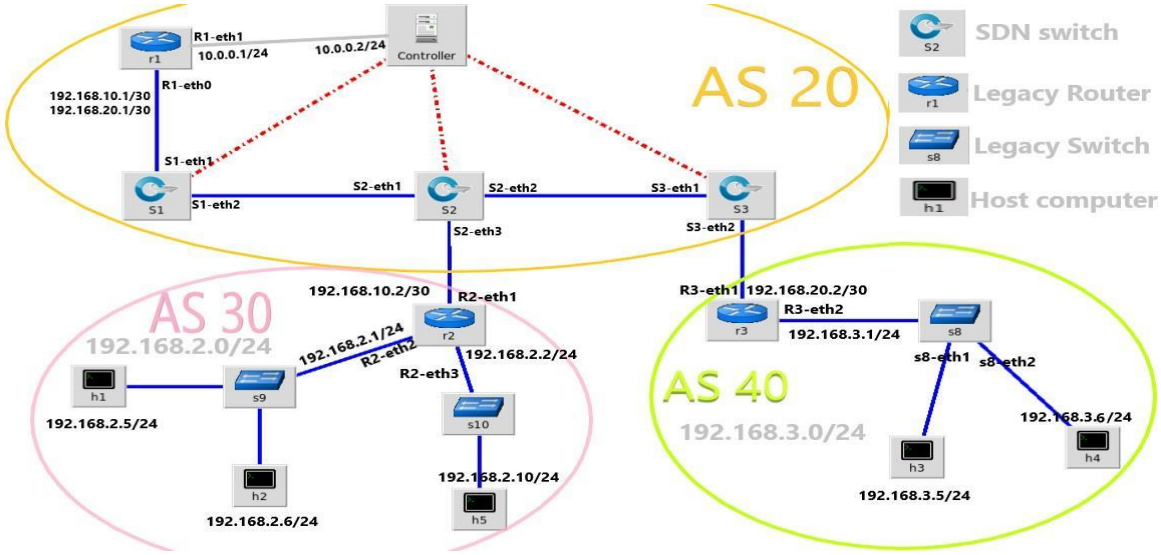


Figure 3 Hybrid SDN environment.

Figure 3. shows the topology, which we created by using the Mininet and the MiniEdit tools; as shown in **Figure 3**, we have assigned IP addresses to the interfaces of network devices, here we have three SDN nodes and three legacy routers, and we have a controller to manage SDN nodes; inside the legacy routers, we have configured BGP protocols with different AS, as shown in **Figure.3**. Because of the lack of space, we did not mention all the commands executed inside each network element. We have used Zebra domain, a routing program for several servers that offers TCP/IP-based routing protocols and FRR (Free Range routing) domain.

To connect different networks in the conventional network and to integrate the legacy network with SDN AS (Autonomous Systems), we have configured BGP (Border Gateway Protocols) inside all the legacy routers, as it is necessary to have a peering system that makes the BGP protocol work. and the integration between the conventional and SDN networks is carried out through the SDN-IP. This program operates on top of the ONOS controller.

Once all the configuration is done, we will save the topology and run the ONOS controller to connect it with our network topology and interconnect SDN nodes with the conventional legacy network devices.

We have already installed the ONOS controller; now, we will open it on a different terminal and run it. To integrate SDN nodes with the legacy network devices, activating IPv4 forwarding and the SDN-IP application is mandatory. To activate these applications on the ONOS controller terminal, we will run these commands:

First, we will activate the application of config file by executing this command

>app activate org.onosproject.config

then, we will activate the ipv4 forwarding by executing this command

>app activate org.onosproject.fwd

>app activate org.onosproject.proxyarp

for the integration, we will activate SDN-IP by executing this command

>app activate org.onosproject.sdnip

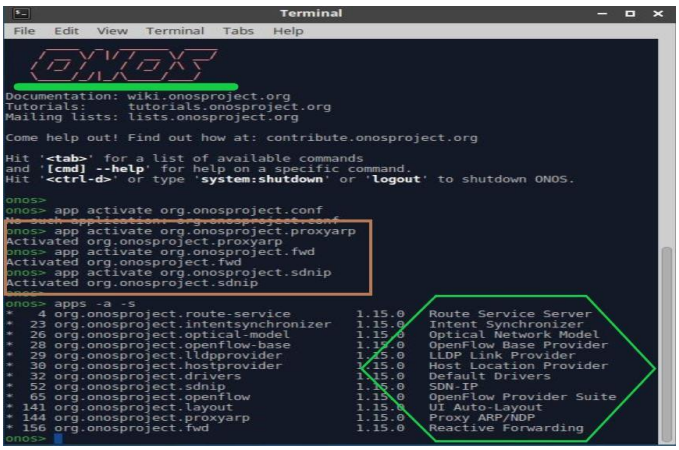


Figure .4 ONOS terminal and applications

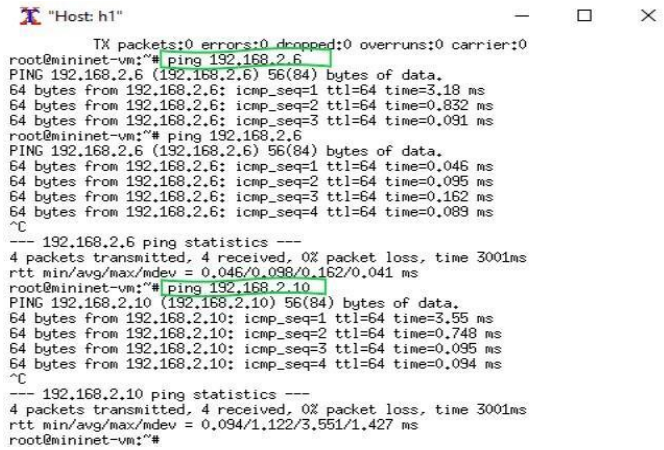


Figure .5 ping from h1 to h2 and h5 inside single AS.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 06,2024 at 07:03:38 UTC from IEEE Xplore. Restrictions apply.

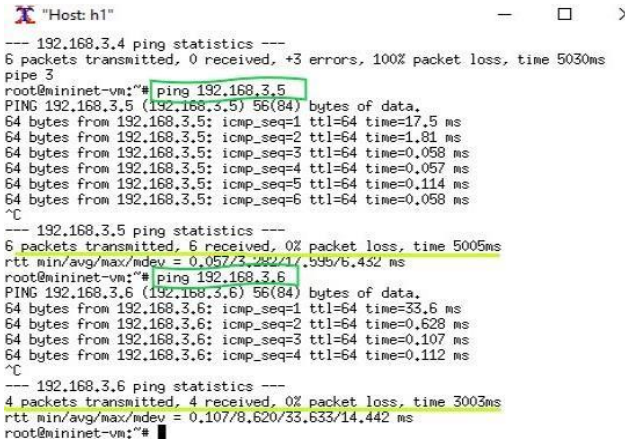


Figure .6 ping from h1 to h3 & h4

Figure 4 shows the ONOS controller terminal in which we have executed the commands for activating different applications. Once the application of ONOS is successfully activated, we will wait for some time as it will take some time, and then we will try to ping from one host to another. **Figures 5** and **6** show that the hosts are successfully pinging as the integration between SDN nodes and the conventional legacy devices has been done through SDN-IP, an ONOS controller application. Now, we will retrieve the data and analyze it to take measurements of the hybrid SDN network performance.

The data is retrieved in a real-time network using TCP (Transmission Control Protocol), a reliable protocol based on our network environment. The data includes background traffic, which is generated using the Iperf tool to measure network performance, and we have used the Wireshark tool to analyze the data as it passes through the network.

4. Performance Analysis of Hybrid SDN

Throughput

Throughput refers to the rate of successfully processed or transferred data through a network, device, or system divided by the observed duration of the time in seconds. It is usually measured and expressed in BPS (bits per second), in terms of data transfer rates.

The throughput obtained by each device in a network may not always correspond to the channel bandwidth; the network throughput varies based on the number of devices and the data flows within the network.

$$\text{Throughput} = \frac{amount\ of\ data\ received(bit)}{time\ delivery\ of\ data\ (second)} \qquad (1)$$

In a computer network, throughput refers to the quality of data delivered via a network in a given time frame. The factors influencing throughput in a network include latency, bandwidth, packet loss, and network congestion; with higher bandwidth, network throughput increases since it gives more capacity for the data to be transferred, while lower latency and minimum packet loss boost throughput by decreasing delays and assuring reliable data transmission.

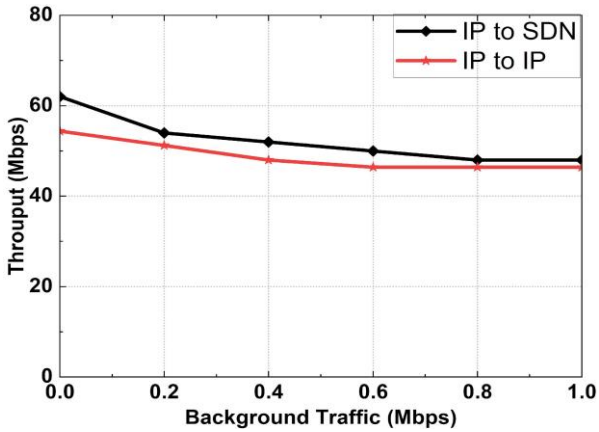


Figure 7 Throughput comparison

The throughput parameter measures the number of packets that may be transferred in a given time frame in a network. Larger packets lead to faster data transmission as they improve network quality and increase network throughput. In TCP traffic communication, one host acts as a customer and another as a server, as shown in Figure 7; in the case of IP to IP using TCP data protocol, we first measured throughput without adding background traffic, then we added background traffic, and we increased the background traffic; as a result, with the growth in the size of the network traffic, we found 5-20% decreases in the network throughput. This indicates that the presence and growth of background traffic negatively impact the network throughput in the case of IP-IP.

In the second case, IP-SDN, we again take network throughput measurement without background traffic. We found no decrease in the network throughput, and then we increased the size of background traffic, and the result showed that with the increased in the network traffic, the throughput value increased too by 5-10%. This means that the presence and increase of background traffic positively impact the network throughput in the case of IP-SDN, which is an excellent sign for organizations that want to migrate from a conventional network to a hybrid SDN and fully SDN network.

Delay

Delay is another critical parameter: the time to transfer data from the origin to the destination in a network. delays in a network can occur due to different variables, such as device quality, source-to-destination distance, and congestion. In a computer network, we can measure the delay of the network by this formula:

$$\text{Delay} = \frac{t2-t1}{\sum n} \qquad (2)$$

Variable “t2” indicates when a packet arrives on the destination node. The value of “t2” is calculated by subtracting the packet time sent from the sender, represented by the variable “t1”, and the total amount of data is defined by $\sum n$.

The delay value, which is the difference between “t2” and “t1,” is defined by the number of packets

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 06,2024 at 07:03:38 UTC from IEEE Xplore. Restrictions apply.

transmitted into the devices or network, so the volume of packets being sent influences the total delay.

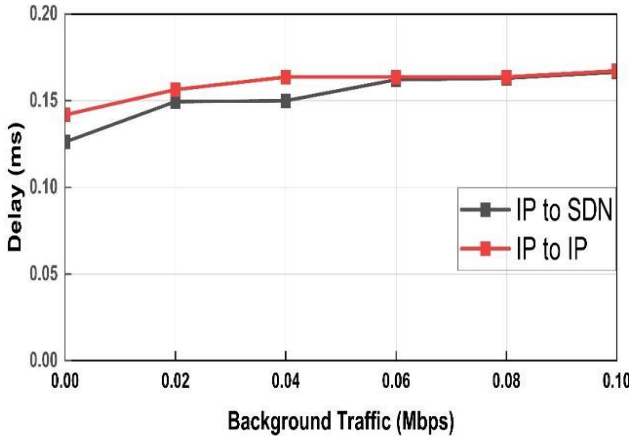


Figure 8 Delay comparison.

To measure the delay of the network in our environment, we have used Wireshark to collect information from each packet delivered from the source to the destination, as delay refers to the time a packet takes to be sent from the source to the recipient.

As you can see in Figure 8, for the delay measurement, we took and considered both the case IP-IP and IP-SDN, the same as before. We increased the background traffic gradually or increased the size of traffic data for both cases to measure the delay. As a result, we found that with the traffic increases, the delay value increases too in both cases, IP to IP and IP-SDN. More experiments and taking the average delay showed that with increasing gradual traffic, IP-SDN performs much better than IP-IP networks, as it gets much lower delay than IP-IP networks.

For the delay measurement in both cases, we experienced an increased delay with more background traffic. Still, the IP-SDN network processes extra traffic more effectively, resulting in a lower average delay than the IP-IP network.

5. Conclusion

As we know, SDN is a novel network design that provides tremendous network control, flexibility, and efficiency by decoupling data and control layers. However, because of some significant challenges, upgrading the entire conventional network to a complete SDN appears difficult. Hybrid SDN is a modern network design that integrates the flexibility and programmability of SDN with the familiarity and robustness of conventional networking. In this article, we investigated the interconnection of SDN nodes with conventional legacy devices in a hybrid SDN environment; we looked at how to interconnect SDN nodes with conventional network devices using ONOS as a controller of the SDN nodes. Through extensive simulation, we found that hybrid SDN is the optimal solution for organizations seeking a modern network infrastructure without network disruption and service downtime, as hybrid SDN can deliver a more scalable, flexible, and programmable network environment.

References

[1] *Agarwal, Sugam, Murali Kodialam, and T. V. Lakshman. "Traffic engineering in software defined networks." 2013 Proceedings IEEE INFOCOM. IEEE, 2013.*

[2] *Vissicchio, Stefano, et al. "Central control over distributed routing." Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication. 2015.*

[3] *Chu, Cing-Yu, et al. "Congestion-aware single link failure recovery in hybrid SDN networks." 2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015.*

[4] *Wang, Huandong, et al. "Saving energy in partially deployed software defined networks." IEEE Transactions on Computers 65.5 (2015): 1578-1592.*

[5] *Vissicchio, Stefano, et al. "Safe update of hybrid SDN networks." IEEE/ACM Transactions on Networking 25.3 (2017): 1649-1662.*

[6] *Levin, Dan, et al. "Panopticon: Reaping the {Benefits} of Incremental {SDN} Deployment in Enterprise Networks." 2014 USENIX Annual Technical Conference (USENIX ATC 14). 2014.*

[7] *Lu, Hui, et al. "Hybnet: Network manager for a hybrid network infrastructure." Proceedings of the Industrial Track of the 13th ACM/IFIP/USENIX International Middleware Conference. 2013.*

[8] *Jin, Cheng, et al. "Telekinesis: Controlling legacy switch routing with openflow in hybrid networks." Proceedings of the 1st ACM SIGCOMM Symposium on Software Defined Networking Research. 2015.*

[9] *Jia, Xuya, Yong Jiang, and Zehua Guo. "Incremental switch deployment for hybrid software-defined networks." 2016 IEEE 41st Conference on Local Computer Networks (LCN). IEEE, 2016.*

[10] *Xu, Hongli, et al. "Incremental deployment and throughput maximization routing for a hybrid SDN." IEEE/ACM Transactions on Networking 25.3 (2017): 1861-1875.*

[11] *Caria, Marcel, et al. "Divide and conquer: Partitioning OSPF networks with SDN." 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM). IEEE, 2015.*

[12] *Agarwal, Sugam, Murali Kodialam, and T. V. Lakshman. "Traffic engineering in software defined networks." 2013 Proceedings IEEE INFOCOM. IEEE, 2013.*

[13] *Guo, Yingya, et al. "Traffic engineering in SDN/OSPF hybrid network." 2014 IEEE 22nd international conference on network protocols. IEEE, 2014.*

[14] *Hong, David Ke, et al. "Incremental deployment of SDN in hybrid enterprise and ISP networks." Proceedings of the Symposium on SDN Research. 2016.*

[15] *Mehraban, Samiullah, and Rajesh Kumar Yadav. "Traffic engineering and quality of service in hybrid software-defined networks." China Communications 21.2 (2024): 96-121.*

[16] *Lantz, Bob, Brandon Heller, and Nick McKeown. "A network in a laptop: rapid prototyping for software-defined networks." Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics in Networks. 2010.*

[17] https://opennetworking.org/onos/ accessed 05/02/2024.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 06,2024 at 07:03:38 UTC from IEEE Xplore. Restrictions apply.

Applied Physics A (2025) 131:14
https://doi.org/10.1007/s00339-024-08143-w

Applied Physics A
Materials Science & Processing

Investigation of MoS_2-hydrogen interaction using in-situ X-ray diffraction studies

Ramesh Kumar[1] · Nitesh K. Chourasia[2,4] · Pawan K. Kulriya[2] · Mahesh Kumar[3] · Vinod Singh[1]

Received: 3 August 2024 / Accepted: 25 November 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
The present manuscript examines the significant effects of hydrogen (H_2) exposure on the structural properties of molybdenum disulfide (MoS_2) thin films through in-situ X-ray diffraction (XRD) analysis. Molybdenum (Mo) thin films were initially deposited using the electron beam (e-beam) deposition method and subsequently sulfurized via chemical vapor deposition (CVD) to obtain MoS_2 thin films. The quality of the MoS_2 films was optimized by varying the thickness of the Mo layer, sulfurization temperature, and the temperature of the Mo film. It was determined that crystalline MoS_2 thin films with an optimal thickness of 20 nm can be achieved through sulfurization at 220 °C, while maintaining the Mo thin film at 600 °C. Pressure-dependent hydrogenation of the MoS_2 thin films, as investigated by in-situ XRD, reveals an increase in crystallite size accompanied by a decrease in the relative intensity of the diffraction peaks with rising hydrogen pressure. Furthermore, a microstrain of approximately 6.3% is induced in the MoS_2 films upon exposure to 1% and 10% hydrogen. Notably, the MoS_2 thin films remain predominantly stable up to a hydrogen pressure of 400 mbar; however, they undergo abrupt transformations and become entirely amorphous when the hydrogen gas pressure is subsequently elevated to 800 mbar. These observations of hydrogen-induced crystalline-amorphous phase transformation in MoS_2 not only enhance the understanding of the interactions between MoS_2 and hydrogen but also have critical implications for the application of MoS_2 thin films in various devices.

Keywords Hydrogenation · MoS_2 · In-situ XRD · Phase transformation

1 Introduction

In recent decades, the realm of nanomaterials has witnessed an extraordinary surge in interest, with researchers exploring the diverse and intriguing properties of two-dimensional (2D) materials. Among these materials, molybdenum disulfide (MoS_2), a transition metal dichalcogenide (TMD), has emerged as a standout candidate for applications spanning electronics, optoelectronics, energy storage, photodetectors and catalysis [1–5]. This versatile 2D material has attracted significant attention due to its unique structural and electronic characteristics, making it an exciting platform for cutting-edge research and technological advancement [6, 7]. The interaction between 2D materials with various gases is an area of profound significance because it underpins their potential utilization in gas-sensing devices, catalytic reactions [8], and gas storage systems [9, 10]. Understanding how gases interact with MoS_2 at the atomic and structural level is also crucial for optimizing its performance in these applications. Globally, the matter of energy has become a cause for widespread apprehension, primarily driven by the depletion of fossil fuels. The escalation in energy demand is principally fueled by two paramount factors: the rapid expansion of the world's population and the concurrent elevation in the average standard of living [11]. In the contemporary context, the most urgent challenges faced by the society is navigating the endeavor to fulfill the growing energy requirements amid substantial limitations imposed by the diminishing reservoirs of traditional fuels such as coal, oil, and natural gas [12]. In addition, hydrogen gas (H_2), in particular, has garnered immense interest as a clean

✉ Vinod Singh
vinodsingh@dtu.ac.in

[1] Present Address: Department of Applied Physics, Delhi Technological University, New Delhi 110042, India

[2] School of Physical Sciences, Jawaharlal Nehru University, New Delhi 110067, India

[3] Department of Electrical Engineering, Indian Institute of Technology Jodhpur, Jodhpur, India

[4] Department of Physics, Faculty of Science, Shree Guru Gobind Singh Tricentenary University, Gurugram, Haryana 122505, India

Published online: 10 December 2024

and sustainable energy carrier. The global interest in H_2 energy has grown substantially primarily attributed to its remarkably high heat combustion rate. Beyond its impressive combustion capabilities, H_2 energy offers a crucial advantage by fostering independence from finite fossil fuel resources. Unlike non-renewable energy sources, H_2 combustion results in the generation of clean combustion products, contributing to a more environmentally sustainable energy landscape [13–15]. Today, H_2 gas is widely employed in various sectors, including oil refining, chemical industry, space transportation, medicine and aerospace. However, due to its high explosiveness and flammability when its concentration exceeds 4% of the environment volume (attributed to its low ignition energy of 0.017 mJ), even a minor H_2 leakage poses serious risks [14]. To mitigate potential hazards and ensure safe H_2 operations, there is a need for a highly sensitive, selective, fast-recovery, and responsive H_2 gas sensor to detect leakages promptly. MoS_2, with its unique surface properties and catalytic potential, can be used in various technological applications such as LDH-based heterostructures as electrode materials for water splitting, energy storage devices and H_2-related technologies [16, 17]. MoS2 can be prepared by various methods such as hierarchical NiCo-LDH@MoS_2/CuS heterostructure (HS) was prepared by facile two-step hydrothermal synthesis, A facile chemically exfoliated combined microwave-assisted method was used to synthesize MoS2·ZnO nanocomposites [17, 18]. Among them, preparation of MoS2 thin film by evaporation of Mo using e-beam technique followed by sulphuration in CVD techniques offers unique advantages such as this method provides high control over MoS2 film thickness and crystallinity, especially compared to simpler, more scalable methods like thermal decomposition. However, it is more complex and expensive than other methods. Much like graphene, MoS_2 is composed of layers arranged in a vertical stack [19]. Each individual layer is created through covalent bonding between molybdenum (Mo) and sulfur (S) atoms. The cohesion between adjacent layers is facilitated by the presence of relatively weak van der Waals forces. This structural similarity highlights the layered nature of MoS_2, where the combination of covalent bonds within layers and van der Waals forces between layers contributes to its unique properties and versatile applications [19]. The presence of weak van der Waals interactions in MoS_2 enables gas molecules to easily infiltrate and diffuse between its layers. Consequently, the resistance of MoS_2 can undergo significant changes as a result of the adsorption and diffusion of gas molecules within these layers. This property makes MoS_2 highly responsive to the presence of gases, offering potential applications in gas sensing and related fields where the material's electrical conductivity can be modulated by the interaction with different gas species. The interaction between MoS_2 and H_2 gas is governed by a multitude of factors, including temperature, pressure, and the structural integrity of MoS_2 [20, 21]. These factors intricately influence the adsorption, diffusion, and subsequent chemical reactions of H_2 on the MoS_2 surface [22, 37]. To delve into the intricacies of this interaction and gain a comprehensive understanding, in-situ characterization techniques have become invaluable tools. Among these techniques, in-situ X-ray diffraction (XRD) stands out as a powerful technique for monitoring structural changes in materials during gas–solid interactions [23]. By allowing real-time observation of atomic rearrangements and phase transformations, *in-situ* XRD offers unique insights into the dynamics and thermodynamics of these interactions [24]. To the best of our knowledge, there is no report on the structural properties of MoS2 on the exposure to the hydrogen gas. This article presents a systematic and detailed investigation into the MoS_2-H_2 gas interaction at different pressures using in-situ X-ray diffraction studies. In the case of nanostructured materials, there could be sample to sample variation during synthesis which leads to modification in the structural properties when they are exposed to hydrogen gas and later investigated using off-line XRD. Here, state-of-art *in-situ* XRD facility has been used in which all the studies has conducted in the same sample without changing experimental conditions. Our objective is to unravel the structural modifications that occur within MoS_2 under varying H_2 pressures and to elucidate the underlying mechanisms that govern this intricate interaction. This research is a crucial step towards realizing how MoS_2 behaves in the presence of H_2 gas, offering fundamental insights into the design, engineering, and optimization of MoS_2-based catalytic systems for H_2-related applications. Throughout this article, we will explore the current state of knowledge regarding the interaction between MoS_2 and H_2 gas, laying the foundation for our in-depth investigation. The unique properties of MoS_2, the significance of H_2 gas as an energy carrier, and the key challenges in harnessing MoS_2's potential for H_2-related applications is also discussed. In the subsequent sections, we will delve into the experimental methodology, data analysis, and results obtained from our in-situ X-ray diffraction studies on MoS_2- H_2 gas interactions at different pressures. The findings of this research not only enhance our fundamental knowledge of MoS_2's behavior but also provide valuable guidance for the design and development of advanced catalytic systems and H_2 storage materials, contributing to the ongoing efforts towards a sustainable energy future.

2 Experimental details

2.1 Preparation of Mo films

The electron beam evaporation (e-beam) deposition method was employed for preparation of high quality Mo thin films by using Mo target on Si/SiO_2 substrate. The key advantages of this technique lie in its ability to produce high-purity Mo films with excellent control over thickness and

microstructure, making it indispensable in various industries ranging from electronics to thin-film coatings and beyond [25, 26]. At first, the substrates were cleaned with trichloroethylene followed by acetone, and alcohol to remove contaminations and native oxide. The base vacuum of the order of 2×10^{-6} mbar was maintained with the help of rotary vane and turbo-molecular pump. The deposition rate is kept at a slow rate of 0.7 – 0.8 nm/sec for the uniformed and to achieve desired thickness. Three different thicknesses (5 nm, 10 nm, 20 nm) of Mo film were obtained by keeping the same deposition rate. The film deposited at the RT having black color which may be due to oxygen deficiency and a large number of defects. Therefore, these films were further annealed in oxygen ambient at temperature of 400 °C for duration of 1 h. After annealing, the color of films changed and they become totally transparent.

2.2 Sulfurization of Mo films

Sulfurization of Mo films is a fundamental process to transform deposited molybdenum (Mo) thin films into two-dimensional molybdenum disulfide (MoS_2) layers. For this purpose, a double zone chemical vapor deposition (CVD) method was employed for the sulfurization. The sulfurization of thin film of 20 nm thickness were grown on Si/SiO_2 substrates via double zone sulfurization using an ambient-pressure CVD technique as shown in Fig. 1.

Flushing of the quartz tube using Argon gas stream, followed by continuous Ar flow for 15 min, was performed. This reduces the oxygen content in the reactor prior to the sulfurization process. Before the experiment, a quartz boat containing 1.0 gm Sulphur powder (Sigma Aldrich, 99.999%) was placed upstream in the heating ring, and the Si/SiO_2 substrate covered with pre-exfoliated Mo film was placed face up on another quartz boat and located in the middle of the tubular furnace, as shown in Fig. 1. The distance between the two quartz boats is about 150 mm. After that, the tubular furnace was heated to 600 °C at a rate of 20 °C/min. When the temperature of the tubular furnace rose to 470 °C, the Sulphur powder was heated to 200 °C through the heating ring at a rate of 20 °C/min. During this process, the Ar flow rate remained at 80 SCCM. After the growth time of 30 min, the furnace was cooled down naturally to room temperature with Ar flow of 100 SCCM.

2.3 Characterizations

XRD along with Raman spectroscopy was used to probe the structural properties of the as prepared and hydrogenated MoS_2 films. XRD patterns have been recorded using Bruker AXS D8 Advanced in-situ XRD facility equipped with Cu K_α source with wavelength 1.54 Å (40 kV and 40 mA power). The *in-situ* XRD setup has provision for creation of

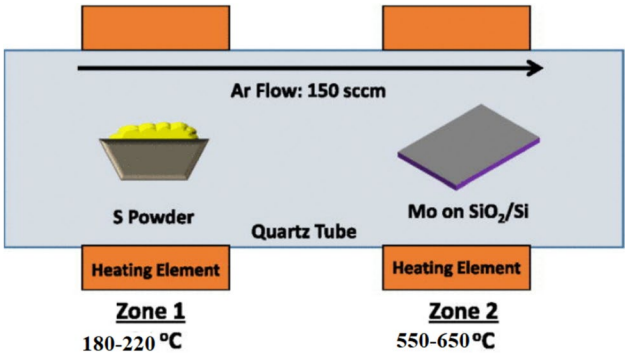


Fig. 1 Sulfurization of Mo films using double zone CVD technique

high vacuum using rotary vane pump and turbo-molecular pump [24]. A gas injection facility is also integrated to injection the H_2 gas in the controlled manner which allows us to record diffraction pattern in the controlled gas atmosphere. A schematic diagram of the *in-situ* XRD setup is shown in the Fig. 2. First, the MoS_2 films were mounted on sample holder and a high vacuum order of 1×10^{-5} mbar was created. After that, diffraction pattern of the MoS_2 films was recorded in the vacuum. Now, MoS_2 films was exposed to different concentrations of H_2 gas (3%) mixed with noble gas Argon (Ar), one by one after different intervals of pressure ranging from 50 to 900 mbar and diffraction pattern of the same sample is recorded after each pressure without changing position of sample. Since, all diffraction patterns are recorded on the same sample so effect of the sample to sample variation is ruled out. This process also allows us to compare of intensity of the diffraction peaks with change in the pressure of the H_2 gas. The data were recorded in locked coupled mode in the scan range of 2θ from 10º to 60º with step size of 0.02º at a scan speed of 0.25º/min.

Raman spectroscopy was performed on pure and MoS_2 samples exposed with different H_2 concentrations using Renishaw Invia Raman microscope with Argon laser as source (514.5 nm wavelength) for the scan range of 100–3200 cm^{-1}. The topography of the MoS_2 films including pristine were studied using Digital Instruments Nanoscope IIIa atomic force microscope (AFM) in the tapping mode.

3 Results and discussions

3.1 Structural properties of the as-prepared MoS_2 thin film

Glancing angle X-ray diffraction (GIXRD) studies was carried out to understand the effect of film thickness, molybdenum temperature and sulfurization temperature on

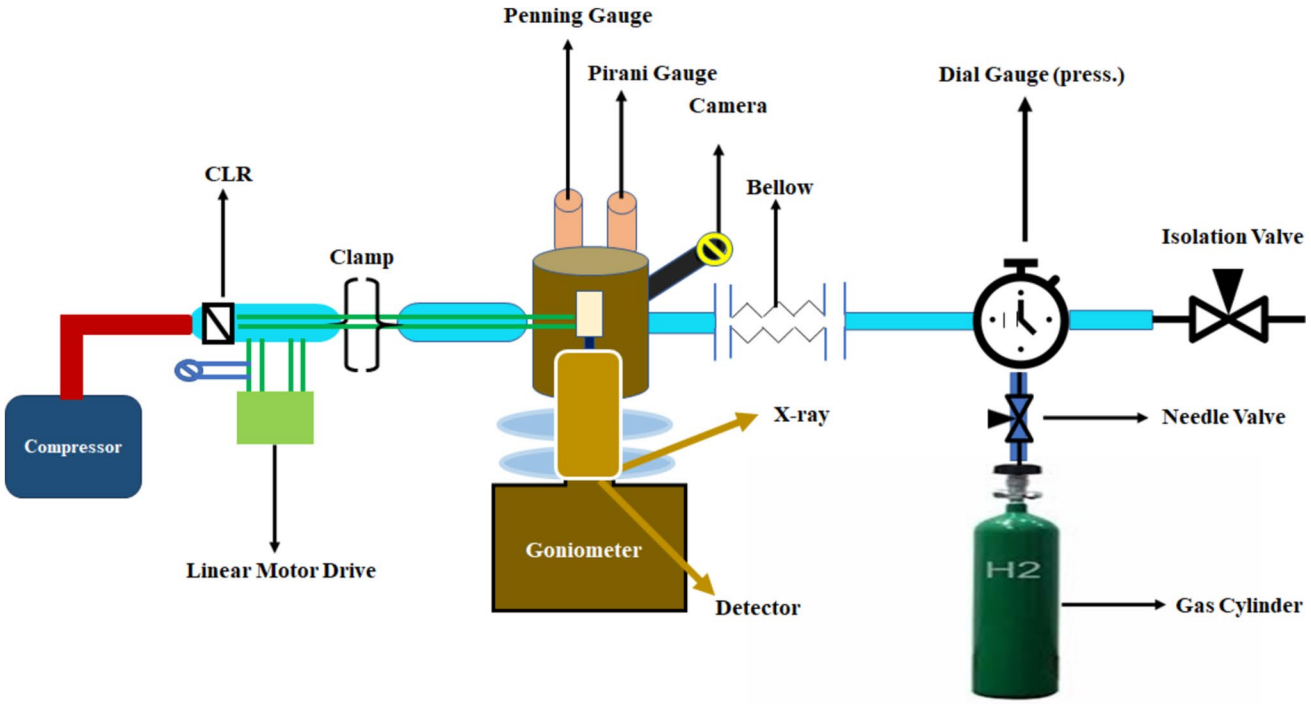


Fig. 2 Schematic diagram of the in-situ XRD setup

crystallinity and phase formation in the molybdenum thin film deposited on Si/SiO_2 substrate.

3.1.1 Effect of Mo film thickness

Thin film of molybdenum having thickness of 5 nm and 10 nm films has been prepared using electron beam evaporation techniques. In order to understand the effect of film thickness on MoS_2 phase formation, these films were exposed to various synthesis conditions of varied molybdenum and sulfur temperatures (molybdenum at 550 °C, 600 °C, 650 °C and sulfur at 180 °C, 200 °C, 220 °C, 300 °C). Figure 3 shows the GIXRD pattern of MoS_2 thin films of the different thickness of (a) 5 nm, (b) 10 nm and (c) 20 nm, prepared by keeping molybdenum and sulfur at different temperature (molybdenum at 550 °C, 600 °C, 650 °C and sulfur at 180 °C, 200 °C, 220 °C, 300 °C), respectively. It is clear from the GIXRD plot of Fig. 3a that MoS_2 phase does not form in the case of 5 nm for all process temperatures except Mo 600 °C and Sulfur 200 °C. When film thickness is increased to 10 nm then sharp peaks are visible at 2θ value of 14.3°, 25.4°, 39.0° and 58.4° corresponding to (002), (004), (103) and (110) planes, respectively. The peaks well matched with JCPDS No. 37–1492, which confirms the successful synthesis of polycrystalline hexagonal-phase MoS_2 in the case of 10 nm films. This also shows that crystallinity is improved with increase in the film thickness. It may be noted that crystalline phase was not formed when the sulfurization was done at 300 °C by keeping Mo at 600 °C. This could be due to suppress of Mo/S interaction at the elected temperature. When thickness of the Mo film was subsequently increased to 20 nm then high quality MoS_2 phase was formed in all the experimental condition used in the present study. In addition, the GIXRD analysis clearly exhibits that the MoS_2 thin film (20 nm) corresponding to synthesis parameter (Mo temperature 650 °C and sulfurization temperature 220 °C) is the optimum condition to get high quality film due to presence of all peaks related to MoS_2 only. It may be noted that other synthesis condition of MoS_2 thin film (20 nm) film also contains polycrystalline MoS_2 phase but they have lesser intensity peaks as well as some impurity peaks.

3.1.2 Effect of process temperature

In order the understand the detailed investigation of process temperature on the structural properties and crystalline quality of the films, the XRD pattern of Mo (20 nm) synthesized at varying (a) Mo temperature and (b) sulfurization temperature are plotted in Fig. 4. The process temperature effect can be studied by either varying the temperature of Mo film or the temperature of sulphur powder while keeping other at fixed temperature during synthesis of MoS_2 thin film. In the first case, the sulfurization temperature was kept at 220 °C and Mo thin film temperature was varied from 550 to 650 °C. When the annealing temperature of Mo film

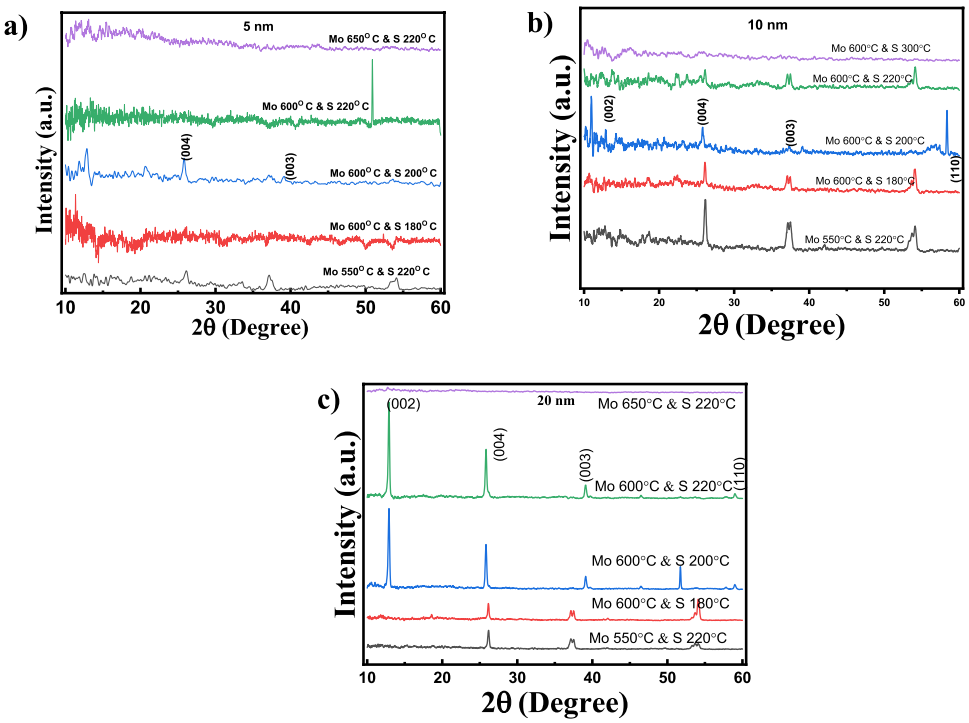


Fig. 3 GIXRD analysis of MoS_2 thin films having **a** 5 nm, **b** 10 nm and **c** 20 nm thickness

was 550 °C then few peaks of the MoS_2 can be clearly seen [Fig. 4a]. As the annealing temperature of Mo was subsequently increased to 600 °C then single phase MoS_2 film was synthesized. It contains all the peaks corresponding to hexagonal phase of the MoS_2. After that, the annealing temperature of the Mo was further increased to 650 °C then surprisingly all the peaks were disappeared and amorphous phase was formed. This indicates that the crystalline MoS_2 phase does not formed due to less interaction of sulphur at elevated temperature. Thus, highly crystalline single phase MoS_2 film can be synthesized by sulfurization at 220 °C by keeping Mo thin films at 600 °C. In order to understand the effect of sulfurization temperature on quality of the film, it was varied from 180 to 220 °C by keeping Mo thin films at the temperature of 600 °C. XRD pattern displayed in Fig. 4b shows that intensity of the diffraction peaks are increasing with increase in the sulfurization temperature. Highly crystalline thin films of MoS2 can be synthesized at the sulfurization temperature of 220 °C by keeping Mo thin films at 600 °C.

Thus, based on the results presented above, one can conclude that optimum process parameters such as film thickness (20 nm), Mo thin film temperaure (600 °C) and sulfurization temperature (220 °C) are required for getting good quality thin films of the MoS_2.

3.2 Surface morphology

Scanning electron microscopy (SEM) was used for determining the surface morphology of as-prepared MoS_2 thin films (20 nm) grown under optimized conditions. A typical SEM image of the MoS_2 thin film is shown in Fig. 5a.

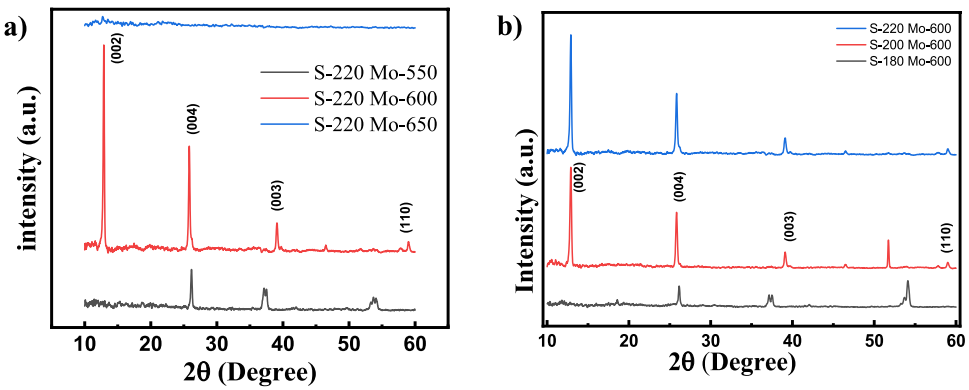


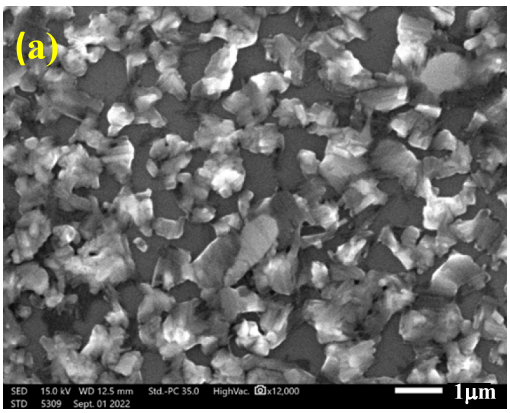
Fig. 4 MoS_2 thin film synthesis with varying **a** sulfurization temperature **b** Mo film temperature

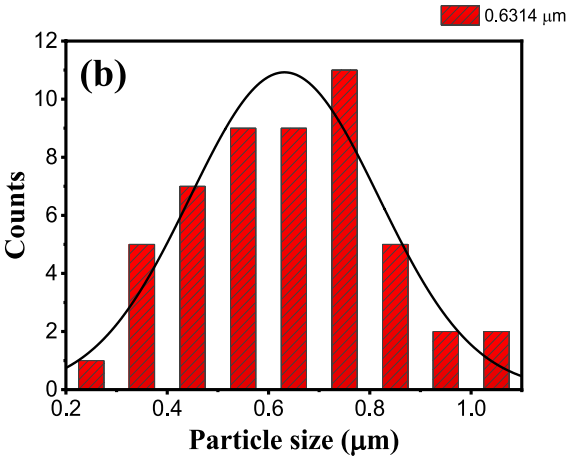
As can be seen from the image, the MoS_2 nanosheets were uniformly grown on the substrate and it appears smooth, homogeneous and without topography contrast. The particle size from the SEM images is estimated using imageJ software. The Fig. 5b represents particle size distribution histogram which gives an estimation of average particle size. The calculated average particle size is 0.63 ± 0.20 µm. The compositions of MoS_2 thin film were investigated using energy dispersive spectroscopy (EDS) and tabulated in the Fig. 5c. The elemental distribution shows that sample primarily consists of Mo, O, S and Si. The extra Si element is also observed due to the Si/SiO_2 substrate.

Atomic force microscope (AFM) technique was also used to compare and analyze the surface characteristics of Mo and MoS_2 thin film with different thickness. Figure 6 shows the AFM images of Mo thin films of different thickness of (a) 5 nm, (b) 10 nm and (c) 20 nm. The surface roughness of these films was calculated using Nasoscope analysis 1.7 software and found to be 0.85 nm, 21 nm and 4.1 nm for 5 nm, 10 nm and 20 nm thick films, respectively. From the AFM image, one can conclude that the 20 nm thin film is of superior quality as compared to other films. In addition, all the MoS_2 thin film of different thickness have very low root-mean-squared (RMS) surface roughness as shown in Fig. 6d–f. The low roughness minimizes surface defects, carrier scattering and ensuring a consistent and uniform interaction with analytes. The detailed of the film parameter has been shown in Table 1.

As one can be seen from the AFM image and Table 1, the roughness of the film having 5 nm thicknesses is lower than 20 nm thick film. Although, roughness of the 20 nm thick film is higher but it is used for further *in-situ* structural investigation because its crystallinity is very good and display all the characteristics peaks of MoS_2, which can be seen from GIXRD.

Formation of single phase MoS_2 and presence of the short rang order is further probed by the Raman spectroscopy technique. It is a powerful technique used to analyze the vibrational modes of materials, providing valuable information about their structure and composition. In the case of MoS_2 films, Raman spectroscopy not only offers insights into the quality and characteristics of the deposited material but also provides information about the





(c)

Display name	Standard data	Quantification method	Result type
Spc_001	Standardless	ZAF	Metal
Element	Line	Mass%	Atom%
O	K	23.35±0.39	35.89±0.60
Si	K	71.32±0.56	62.45±0.49
S	K	0.56±0.16	0.43±0.12
Mo	L	4.78±0.50	1.22±0.13
Total		100.00	100.00
Spc_001			Fitting ratio 0.0593

Fig. 5 **a** SEM image of MoS_2 thin film prepared on Si/SiO_2 substrate **b** size distribution histogram **c** elemental distribution of MoS_2 thin film

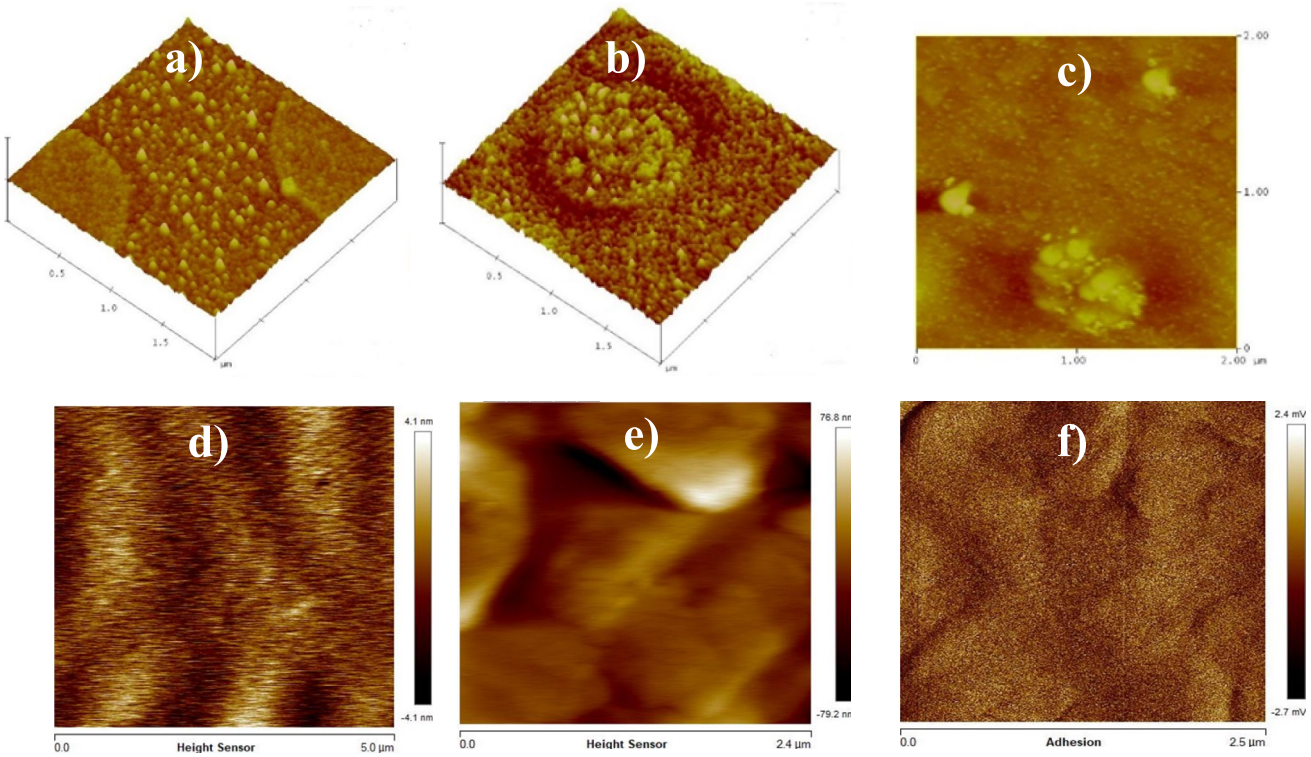


Fig. 6 AFM image of Mo thin film with the thickness of, **a** 5 nm, **b** 10 nm, and **c** 20 nm and MoS_2 thin film with the thickness of **d** 5 nm, **e** 10 nm, **f** 20 nm

Table 1 The variation in the roughness of MoS_2 thin films with film thickness

S.N	Thickness of film (nm)	Roughness of film (nm)
1	5	0.85
2	10	2.1
3	20	4.1

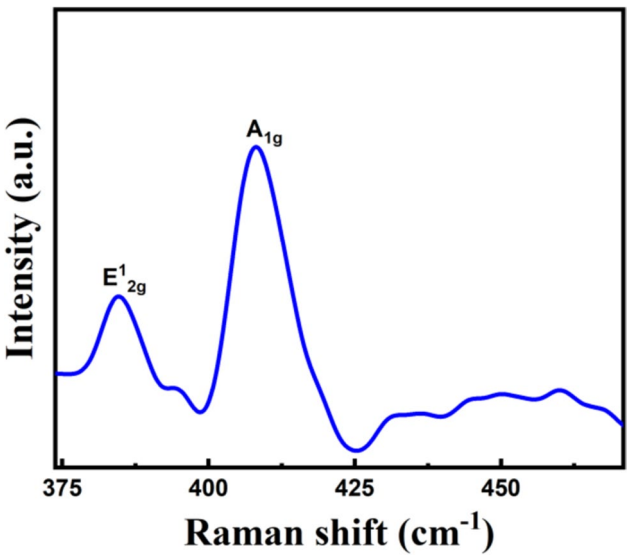


Fig. 7 Raman spectra of the MoS_2 thin films of 20 nm thicknesses grown by vapor phase sulfurization of Mo thin films

phonon modes in MoS_2, which are related to the vibrational motion of atoms within the crystal lattice.

Raman spectrum of the MoS_2 thin films (t = 20 nm) is shown in the Fig. 7. It exhibits the distinctive signal of 2H-MoS_2, with no apparent traces of polymer residue, carbon contamination or oxides.

Raman spectrum comprises signature peaks E^1_{2g} at the wavenumber of 385 cm^{-1} corresponding to the in-plane vibrational modes of the MoS_2 lattice (in-plane vibration between Mo and S atom) and A_{1g} at the wavenumber of 405 cm^{-1} represents the out-of-plane vibrations of the MoS_2 layers (out-of-plane expansion between S atom) [26, 27, 33]. Generally, the A_{1g} mode gives the significant information about the van der Waals force among the adjacent MoS_2 layers, because of out-of-plane interaction with the adjacent atoms [28, 34]. In this context, the presence of two distinct Raman peaks suggests that the MoS_2 film exhibits hexagonal coordination (2H) rather than the trigonal prismatic configuration (1 T) as observed in mechanically exfoliated MoS_2 [29, 32]. Additionally, the frequency difference of these two peaks is found to be 20 cm^{-1}, which can be ascribed a few layers of MoS_2 thin film [30, 32]. The well-matched

characteristics peaks is a clear evidence of the formation of the impurity free single phase MoS_2.

3.3 Hydrogenation of MoS_2: pressure effect

To understand MoS_2-H_2 kinetic, an *in-situ* XRD studies was performed on MoS_2 thin film (20 nm) by exposing it to pure H_2 gas a pressure varying from vacuum to 1000 mbar. The diffraction was recorded in the vacuum and different hydrogen pressure at the measurement parameters of the diffractometer and without changing position of sample. Figure 8a shows *in-situ* XRD patterns of as-prepared MoS_2 thin films in the vacuum and on exposure to H_2 environment by varying its pressure between vacuum to 1000 mbar. It shows that few new peaks were appeared even on film exposed to a very low pressure of H_2 (50 mbar). This clearly indicates very high sensitivity of MoS_2 towards H_2 gas at the room temperature. When the H_2 pressure is further increased from 50 to 100 mbar, then intensity of the diffraction peaks are decreased. But, observation of new peaks and phase change was not observed event pressure of the H gas increased to 200 mbar. As H_2 gas pressure is subsequently increased up to 400 mbar then drastic reduction in the diffraction peaks is observed. This indicates that a large lattice strain and defects are produced by hydrogen if the pressure is $P > 400$ mbar. The production of the excessive defects generates structural distortion which leads to amorphization of the MoS_2. As the H_2 pressure is further increased to 1000 mbar, diffraction peak intensity corresponding to different planes tends to deteriorate and it vanishes completely at 1000 mbar pressure. On the basis of the diffraction data, one can infer that MoS_2 remains predominantly stable within the H_2 pressure range from vacuum to 400 mbar. However, beyond this pressure range, MoS_2 thin films undergo significant and abrupt transformations and it completely loses it phase at 1000 mbar pressure. To understand the H_2 pressure induced lattice micro-strain in the MoS_2 films, the inter-planer spacing of pure and hydrogenated MoS_2 was calculated. The variation in the inter-planer spacing (d) with H_2 pressure is shown in Fig. 8b. It exhibits that the inter-planer spacing increases with increase in the H_2 pressure which confirms the production of the lattice strain due to hydrogenation. Due to increase in hydrogen pressure from vacuum to 1000 mbar, the MoS_2 (002) inter planar spacing was observed. It was noticed that inter-planar spacing is a function of *in-situ* hydrogen pressure. Before exposure to hydrogen gas the inter-planar spacing of (002) plane was observed $c = 12.25$ Å such as khan et al. XRD study shows hexagonal structure of few-layers of MoS_2 with observed (0 0 2) plane, with lattice parameter is ≈12.28 A° [29, 37]. Subsequently, *in-situ* hydrogen pressure was increased to 50 mbar and corresponding inter-planar spacing was observed $d = 11.87$ Å, which was measured using XRD

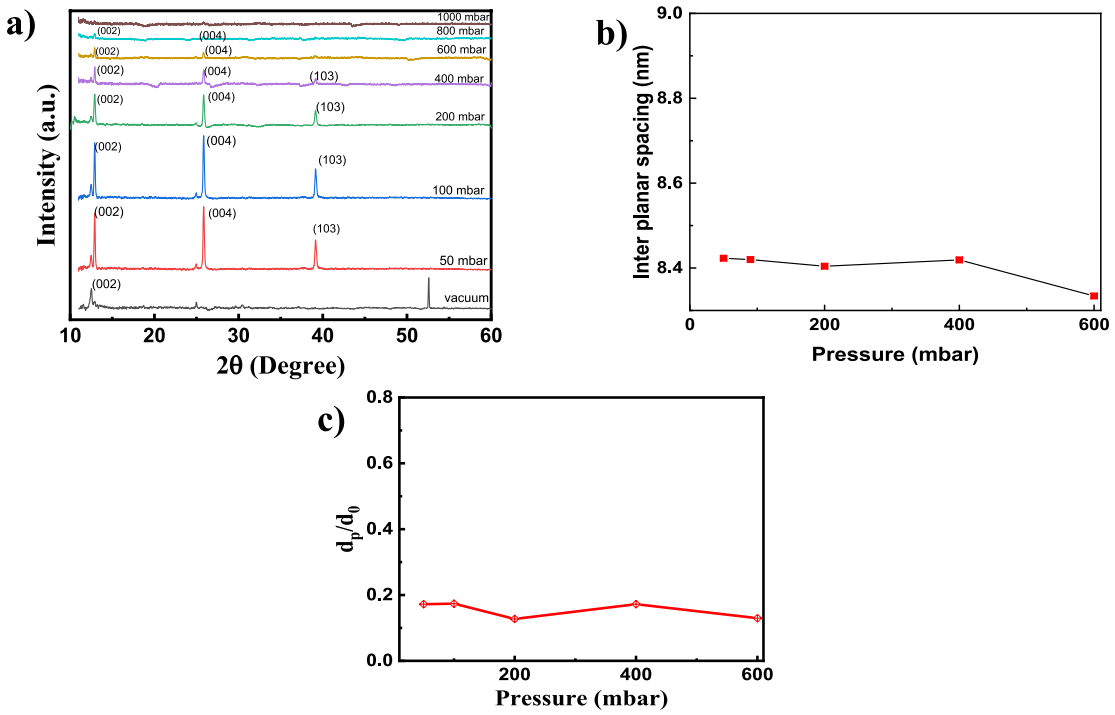


Fig. 8 **a** *In-situ* XRD pattern of MoS_2 thin film (20 nm) recorded under vacuum and exposure to different H_2 gas pressure. The variation of **b** interplaner spacing **d** and **c** relative change in the inter planar spacing $(d_p - d_0)/d_0$ at different H_2 gas pressure

technique. The in-situ hydrogen pressure was increased up to 1000 mbar. The inter-planar spacing at 800 mbar and 100 mbar hydrogen pressure are d = 11.89 Å and 13.58 Å, respectively. This shows that the inter-planar spacing was almost constant up to 800 mbar hydrogen pressure. Further, there was a drastic increase in inter-planar spacing when pressure was increased from 800 to 1000 mbar which could leads to phase change in the MoS2. This shows that there is a significant increase in inter-planar spacing due to hydrogen pressure. So, this threshold level of hydrogen pressure significantly affects the crystalline properties of MoS_2 structure. Thus, hydrogen pressure of 800 mbar can be considered as a transition pressure.

Apart from that relative change in the inter planer spacing $(d_p - d_0)/d_0$ was also calculated. Here, d_0 and d_P are inter planar spacing in vacuum and at a specific hydrogen pressure respective. The variation in the $(d_p - d_0)/d_0$ with pressure is shown in Fig. 8c. The results indicates that as pressure in increasing the inter planar spacing is also increasing.

The crystalline size of MoS_2 films exposed at various H_2 partial pressures was estimated from the XRD data presented in the Fig. 8. It was calculated using Scherrer equation which is expressed as:

$$D = \frac{K\lambda}{\beta\cos(\theta)}$$

where, D is the average crystallite size., K is the shape factor, typically around 0.9, λ is the wavelength of the X-ray radiation used for the diffraction, β is the full width at half maximum (FWHM) of the diffraction peak and θ is the Bragg angle. Figure 9 shows variation in the crystallite size with H_2 partial pressure for diffraction peaks appeared at the (a) 14.3°, (b) 25.4° and (c) 39.4°, corresponding to (002), (004) and (103) planes, respectively. It is evident that the crystallite size of MoS_2 thin films is increasing with H_2 gas pressure for all three crystalline planes which may be due to hydrogen-induced recrystallization. Hydrogen being a reducing agent can promote the removal of oxygen-containing functional groups and defects from the MoS_2 surface which leads to a more pristine and ordered atomic arrangement. Hydrogen induced reduction process may also facilitate the coalescence and growth of crystallites, promoting larger crystal sizes. Additionally, H_2 exposure can influence the energetics of grain boundaries, encouraging the migration and coarsening of grains [29–31].

In order to understand the influence of hydrogenation on the crystallinity of the MoS_2 films, change of relative intensity (I/I_0) of all the characteristics peaks of MoS_2 with H_2 gas pressure is determined. Here, I_0 and I are the intensity of diffraction peak recorded in vacuum and at a particular pressure of H_2. Figure 10a shows the variation in

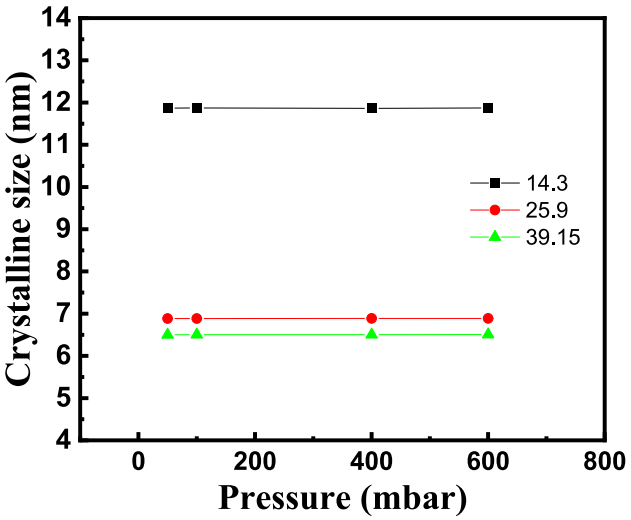


Fig. 9 Change in the crystallite size vs partial pressure calculated for the diffraction peaks appeared at **a** 14.3°, **b** 25.4° and **c** 39.4°

the relative intensity (I/I_0) with H_2 gas pressure. It shows that relative intensity of three characteristics peaks is reduced as the pressure of the H_2 gas is increased. This clearly shows that crystallinity of the MoS_2 is deteriorating on the exposure to H_2 gas. It may be noted that effect of hydrogenation on the peak appearing at the diffraction angle of 14.3° is negligible. The variation in relative intensity (I/I_0) for peaks appearing at the 25.4° and 39.15° are fitted using exponential function.

Apart from this, to understand the H_2 induced preferred orientation, change in relative intensity corresponding to (004) plane and (103) plane with the variation of the pressure is also calculated. Figure 10b displays a change in the relative plane intensity $I_{(004)}/I_{(103)}$ with partial pressure of H_2 It displays that the $I_{(004)}/I_{(103)}$ is decreased with increase in the pressure of H_2 gas. This suggest that there is preferred orientation along the (103) plan of the MoS_2. In order to validate H2 storage materials, a detail understanding comes from these studies [8, 26, 35]. This confirms that the H_2 makes a great impact through interaction on the crystal structure of MoS_2 thin film.

3.4 Conclusion

In summary, synthesis process parameters optimization and hydrogenation properties of the molybdenum thin films of different thickness prepared using e-beam deposition technique followed by sulfurization using CVD system are reported. Optimization of viz, film thickness, sulfurization temperature by keeping Mo thin film at fixed temperature and Mo film temperature by keeping fixed sulfurization temperature has been carried out for getting

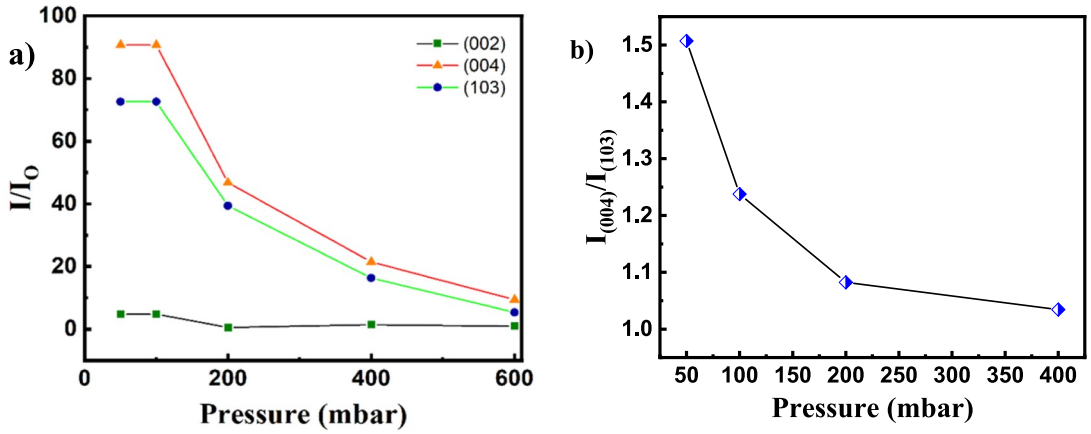


Fig. 10 **a** XRD relative intensity vs pressure 14.3°, 25.4° and 39.0°, **b** Relative plane intensity $I_{(004)}/I_{(103)}$ vs pressure (mbar)

highly crystalline MoS_2 film and confirmed its structural, micro-structural and compositional analysis using GIXRD, SEM, EDX, Raman spectroscopy techniques. The structural properties investigations confirms that crystalline MoS_2 thin films having optimum thickness of 20 nm can be prepared by sulfurization at 220 °C while keeping Mo thin film at 600 °C. The pressure dependent hydrogenation of the MoS_2 thin films investigated using *in-situ* XRD technique suggests that MoS_2 thin film remains predominantly stable within the pressure range from vacuum to 400 mbar. However, it undergoes significant and abrupt transformations when hydrogen gas pressure is subsequently increased. The MoS_2 thin films become completely amorphous at 800 mbar pressure. Thus, observation of order–disorder phase transformation has not only deepens the understanding of MoS_2-H_2 interactions but also important for utilization of the MoS_2 thin films in the devices.

Acknowledgements This work was supported by "Science and Engineering Research Board," India. Nitesh K. Chourasia acknowledges the "Science and Engineering Research Board" for National Post-Doctoral Fellowship (PDF/2021/001490). One of the authors, Ramesh Kumar is thankful to Delhi Technological University for providing Institute Research Fellowship.

Author contributions Ramesh Kumar- Original manuscript writing, investigation, characterization, and synthesis. Nitesh K. Chourasia- Editing and drafting. Pawan K. Kulriya- Formal analysis and visualization. Mahesh Kumar- Conceptualization, supervision, and validation. Vinod Singh- Resources, supervision, and validation.

Funding Science and Engineering Research Board, PDF/2021/001490, Nitesh K. Chourasia.

Data availability Data will be available on request.

Declarations

Conflict of interest All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report.

References

1. A. Sharma, G. Gupta, Recent development and prospects for metal Selenide-based gas sensors. Mater. Sci. Eng. B **290**, 116333 (2023)
2. A. Dodda, D. Jayachandran, A. Pannone, N. Trainor, S.P. Stepanoff, M.A. Steves et al., Active pixel sensor matrix based on monolayer MoS2 phototransistor array. Nat. Mater. **21**, 1379–1387 (2022)
3. T. Wang, S. Chen, H. Pang, H. Xue, Y. Yu, MoS2-based nanocomposites for electrochemical energy storage. Adv. Sci. **4**, 1600289 (2017)
4. Y. Huang, Y. Sun, X. Zheng, T. Aoki, B. Pattengale, J. Huang et al., Atomically engineering activation sites onto metallic 1T-MoS2 catalysts for enhanced electrochemical hydrogen evolution. Nat. Commun. **10**, 982 (2019)
5. P.M. Pratik, V. Patel, C.K. Sumesh, MoS2/WSe2 nanohybrids for flexible paper-based photodetectors. Nanotechnology **32**(31), 315709 (2021)
6. X. Bao, Q. Ou, Z.Q. Xu, Y. Zhang, Q. Bao, H. Zhang, Band structure engineering in 2D materials for optoelectronic applications. Adv. Mater. Technol. **3**, 1800072 (2018)
7. H. Xu, J. Zhu, Q. Ma, J. Ma, H. Bai, L. Chen et al., Two-dimensional MoS2: Structural properties, synthesis methods, and regulation strategies toward oxygen reduction. Micromachines **12**, 240 (2021)
8. G. Li et al., All the catalytic active sites of MoS2 for hydrogen evolution. J. Am. Chem. Soc. **138**(51), 16632–16638 (2016)
9. R. Kumar, X. Liu, J. Zhang, M. Kumar, Room-temperature gas sensors under photoactivation: From metal oxides to 2D materials. Nano Micro Lett. **12**, 1–37 (2020)

10. J. Chen, N. Kuriyama, H. Yuan, H.T. Takeshita, T. Sakai, Electrochemical hydrogen storage in MoS2 nanotubes. J. Am. Chem. Soc. **123**, 11813–11814 (2001)
11. N. Rusman, M. Dahari, A review on the current progress of metal hydrides material for solid-state hydrogen storage applications. Int. J. Hydrog. Energy **41**, 12108–12126 (2016)
12. I. Jain, Hydrogen the fuel for 21st century. Int. J. Hydrog. Energy **34**, 7368–7378 (2009)
13. S. Moon, K. Kim, J. Kim, A study on public acceptance of hydrogen portfolio standard for renewable energy expansion: Comparative analysis of hydrogen production mix plans. Int. J. Hydrog. Energy **49**, 538–552 (2024)
14. W. Chen, L. Sui, X. Long, J. Liao, S. Wang, X. Wei, Anatase TiO2 aerogel with high specific surface areas and porous network structures for ultra-fast response hydrogen sensor. Int. J. Hydrog. Energy **50**, 973–991 (2024)
15. P.M. Pratik, C.K. Sumesh, MoS2 nanosheets on Cu-foil for rapid electrocatalytic hydrogen evolution reaction. J. Electroanal. Chem. **912**, 116270 (2022)
16. S. Kumar, A. Mirzaei, A. Kumar, M.H. Lee, Z. Ghahremani, T.-U. Kim et al., Nanoparticles anchored strategy to develop 2D MoS2 and MoSe2 based room temperature chemiresistive gas sensors. Coord. Chem. Rev. **503**, 215657 (2024)
17. S. Yu, S. Song, R. Li, B. Fang, The lightest solid meets the lightest gas: An overview of carbon aerogels and their composites for hydrogen related applications. Nanoscale **12**, 19536–19556 (2020)
18. P. Chauhan, J.L. Dattatray, V. Patel, P. Sahatiya, C.K. Sumesh, Hierarchical NiCo-LDH@MoS2/CuS composite as efficient trifunctional electrocatalyst for overall water splitting and asymmetric supercapacitor. Electrochimica Acta **469**, 143197 (2023)
19. M. Patel, P. Pataniya, H. Vala, C.K. Sumesh, One-dimensional/two-dimensional/three-dimensional dual heterostructure based on MoS2-modified ZnO-heterojunction diode with silicon. J. Phys. Chem. C **123**(36), 21941–21949 (2019)
20. Kapatel, Sanni, C. K. Sumesh (2016) One pot sono-chemical synthesis of 2D layered MoS2 nanosheets. AIP Conference Proceedings. Vol. 1728. No. 1. AIP Publishing
21. G. Eda, H. Yamaguchi, D. Voiry, T. Fujita, M. Chen, M. Chhowalla, Photoluminescence from chemically exfoliated MoS2. Nano Lett. **11**, 5111–5116 (2011)
22. A.V. Agrawal, R. Kumar, G. Yang, J. Bao, M. Kumar, M. Kumar, Enhanced adsorption sites in monolayer MoS2 pyramid structures for highly sensitive and fast hydrogen sensor. Int. J. Hydrog. Energy **45**, 9268–9277 (2020)
23. U.J. Cho, D. Jang, Y. Jeon, T. Kim, B. Jo, R. Kim et al., A palladium-deposited molybdenum disulfide-based hydrogen sensor at room temperature. Appl. Sci. **13**, 10594 (2023)
24. R. Kronberg, M. Hakala, N. Holmberg, K. Laasonen, Hydrogen adsorption on MoS 2-surfaces: A DFT study on preferential sites and the effect of sulfur and hydrogen coverage. Phys. Chem. Chem. Phys. **19**, 16231–16241 (2017)
25. X. Liu, W. Yang, Z. Liu, Recent progress on synchrotron-based in-situ soft X-ray spectroscopy for energy materials. Adv. Mater. **26**, 7710–7729 (2014)
26. E.J. Kautz, D.K. Schreiber, A. Devaraj, B. Gwalani, Mechanistic insights into selective oxidation and corrosion of multi-principal element alloys from high resolution and in situ microscopy. Materialia **18**, 101148 (2021)
27. L.G. De Arco, Y. Zhang, A. Kumar, C. Zhou, Synthesis, transfer, and devices of single-and few-layer graphene by chemical vapor deposition. IEEE Trans. Nanotechnol. **8**, 135–138 (2009)
28. M. Patel, P. Pataniya, H. Vala, C.K. Sumesh, One-dimensional/two-dimensional/three-dimensional dual heterostructure based on MoS 2-modified ZnO-heterojunction diode with silicon. J. Phys. Chem. C (2019). https://doi.org/10.1021/acs.jpcc.9b05134
29. M. Khan, S. Kumar, A. Mishra, I. Sulania, M.N. Tripathi, A. Tripathi, Study of structural and electronic properties of few-layer MoS2 film. Mater. Today Proc. **57**, 100–105 (2022)
30. P.K. Kulriya, M. Kumar, J. Singh, D.K. Avasthi, Hydrogen pressure dependent in-situ electrical studies on Pd/C composite. Int. J. Hydrog. Energy **42**, 3399–3406 (2017)
31. P. Kulriya, F. Singh, A. Tripathi, R. Ahuja, A. Kothari, R. Dutt et al., Setup for in situ X-ray diffraction study of swift heavy ion irradiated materials. Rev. Sci. Instrum. (2007). https://doi.org/10.1063/1.2804171
32. C. Yim, M. O'Brien, N. McEvoy, S. Riazimehr, H. Schäfer-Eberwein, A. Bablich et al., Heterojunction hybrid devices from vapor phase grown MoS2. Sci. Rep. **4**, 5458 (2014)
33. H. Li, Q. Zhang, C.C.R. Yap, B.K. Tay, T.H.T. Edwin, A. Olivier et al., From bulk to monolayer MoS2: Evolution of Raman scattering. Adv. Func. Mater. **22**, 1385–1390 (2012)
34. K.-G. Zhou, F. Withers, Y. Cao, S. Hu, G. Yu, C. Casiraghi, Raman modes of MoS2 used as fingerprint of van der Waals interactions in 2-D crystal-based heterostructures. ACS Nano **8**, 9914–9924 (2014)
35. L. Lei, D. Huang, G. Zeng, M. Cheng, D. Jiang, C. Zhou et al., A fantastic two-dimensional MoS2 material based on the inert basal planes activation: Electronic structure, synthesis strategies, catalytic active sites, catalytic and electronics properties. Coord. Chem. Rev. **399**, 213020 (2019)
36. L. Ma, D.N. Nath, E.W. Lee, C.H. Lee, M. Yu, A. Arehart et al., Epitaxial growth of large area single-crystalline few-layer MoS2 with high space charge mobility of 192 cm2 V− 1 s− 1. Appl. Phys. Lett. (2014). https://doi.org/10.1063/1.4893143
37. D. Marinov, J.F. de Marneffe, Q. Smets, G. Arutchelvan, K.M. Bal, E. Voronina et al., Reactive plasma cleaning and restoration of transition metal dichalcogenide monolayers. npj 2D Mater. Appl. **5**, 17 (2021)

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

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

J Mater Sci: Mater Electron (2024) 35:2250

Investigations on structural, optical and mechanical parameters of Pr^{3+}-doped oxyfluoro antimony borate glasses

Ch. Pravallika[1], K. Swapna[1,*], Sk. Mahamuda[1], M. Venkateswarlu[1], and A. S. Rao[2]

[1] *Department of Physics, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India*
[2] *Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi 110042, India*

Received: 10 September 2024
Accepted: 19 November 2024

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

ABSTRACT

This study investigates the physical, structural, optical and mechanical properties of Oxyfluoroantimony borate (OFSbB) glasses doped with Pr^{3+} ions, emphasizing the impact of these properties on potential optical applications. The X-Ray Diffraction, Transmission Electron Microscopy and Energy Dispersive X-ray analysis confirms the amorphous nature and glass composition, while optical absorption spectra indicate multiple electronic transitions from $^3H_4$ ground state associated with Pr^{3+} ions. This study also explores bonding parameters, optical bandgap characteristics, and Judd-Oflet analysis, highlighting the ionic nature of bonding in the glasses. Fluorescence analysis demonstrates the emission characteristics of Pr^{3+} ions, with 484 nm ($^3P_0$ to $^3H_4$) and 610 nm ($^1D_2$ to $^3H_4$) transitions showing potential for laser applications. The mechanical properties, assessed through Vickers microhardness tests reveal increased hardness and elastic moduli with higher Pr^{3+} concentrations, indicating increased material strength and rigidity. Overall, OFSbBPr1.0 glass emerges as promising candidate for laser host materials due to its favorable optical and mechanical properties.

1 Introduction

Materials doped with rare earth (RE) ions have shown potential in imaging, solid-state lasers, sensors and up-converters applications. Among various hosts, glass hosts containing RE ions are easier to incorporate than crystalline materials, making them ideal for luminescent applications. Because of their unique electronic structure characterized by partially filled 4f and 5d electron orbitals. The resulting higher emission efficiencies are due to reduced phonon energy, isolation of rare earth ions, broad emission bands, and unique 4f-4f and 4f-5d electronic transitions [1–6]. Among RE ions, Trivalent Praseodymium (Pr^{3+}) ions exhibit closed shells and a + $4f^2$ configuration in their electronic configuration. For Pr^{3+} ions, electron–electron interactions result in L–S terms, which generate mixed states with different J momenta. Pr^{3+} ions exhibits metastable states for simultaneous laser emission towards green, blue, and red light, and this mixture violates Lande intervals [7–9]. Additionally, it has been shown that increasing the concentration of doped RE

Address correspondence to E-mail: kswapna80@gmail.com

https://doi.org/10.1007/s10854-024-13945-y
Published online: 11 December 2024

ions causes concentration quenching at certain energy levels of Pr^{3+} ions. Mostly, Pr^{3+} doped glasses have a lasing effect of the $^1D_2 \rightarrow {}^3H_4$ transition around 600 nm [10, 11]. Thus, glasses doped with Pr^{3+} ions and having low phonon energy may be suitable for luminescent devices. Recently, M.K. Komal Poojha et al. [12], studied the spectroscopic characteristics of lead borotellurite glasses doped with alkali/alkaline modifiers. The glasses are good for use as solid-state laser materials, especially for tunnel roads and blue laser applications. Jyoti Dahiya et al., [13] investigated the Praseodymium-doped LiF–ZnO–Bi_2O_3–B_2O_3 glass prepared using a melt-quenching method designed for photonics uses. The XRD study confirmed the glassy nature of prepared glass samples. Because LiF has a bigger ionic size and a lower molecular weight, its IR spectra revealed an increase in molar volume and a drop in density. UV–Vis-NIR studies indicated the materials potential for photonic applications. LiF doping significantly modified local structure and ligand environment, generating more NBOs. S. Mao et al. [14], Using a melt-quenching process zinc phosphate glasses with varying Pr^{3+} ions doping concentrations emerged with the objective to examine their structural and optical characteristics. The glasses showed deep blue and yellow emission with increasing doping content, strong emission at 600 nm and 1034 nm, and increased luminescence intensity with increasing mole ratio. The prepared glasses could be used in optoelectronic devices like orange-red or near-infrared lasers. J. Anjaiah et al. [15], investigated the dielectric and conductivity properties of 30Li_2O-10CdO-(60-x) B_2O_3:xLn_2O_3 (CdB) glass with different types of trivalent rare earth ions like Nd, Pr, Sm and Eu. They observed that among other rare earths, the CdBPr glass exhibited the most excellent conductivity. Praseodymium-doped borate glasses exhibit increased optical non-linearity and a broad emission spectrum ranging from ultraviolet to infrared [16]. The incorporation of RE^{3+} ions enhance the luminescence intensity and durability of borate-based glasses. These distinctive features render them highly desirable in optoelectronics particularly in improvement of amplifiers for optics and SSL devices [17]. The incorporation of fluorides such as sodium fluoride (NaF) and lithium fluoride (LiF) into borate (B_2O_3) enhances glass stability and inhibits moisture. Furthermore, the fluorides diminish phonon energy hence reducing non-radiative decay loss caused by multiphonon relaxations [18]. The most significant infrared cut-off edge diminishes non-radiative loss, decreases the non-linear refractive index and enhances quantum efficiency [19–21]. Antimony oxide (Sb_2O_3) alters borate doped glasses by forming stable units and generates a random network thereby improving physical, optical and structural properties. This enables fluorescence yield, enhanced lasing, increasing transparency and RE ion solubility [22, 23]. Aluminum oxide (Al_2O_3) improves the mechanical strength and optical properties of glass by acting as intermediate. It enhances the fluorescence of RE ions by splitting them into Al-O-RE bonds, resulting in a distance between RE ions [24, 25]. Hence, the glass is prepared by using the B_2O_3, Sb_2O_3, LiF, NaF, Al_2O_3 and Pr_6O_{11} chemical constituents.

In the present study, 60B_2O_3-10Sb_2O_3-10Al_2O_3-10NaF-(10-x) LiF (OFSbB) doped Pr^{3+} ions with different concentrations of 0.1,0.5,1.0,1.5, and 2.0 mol% glasses were synthesized. Various spectroscopic characterization studies like XRD, EDS, TEM were used to understand amorphous nature of the prepared glass. By using absorption spectra, J–O theory studied. However, the J-O hypothesis of Pr^{3+} is inconclusive because of the little energy disparity between the (initial excited configuration) $4f^15d^1$ and the (ground configuration) $4f^2$ [26]. The difficulty in modeling the hypersensitive transition $^3H_4 \rightarrow {}^3P_2$ arises from the significant fluctuation in oscillator strengths observed in computed measurement [27, 28]. By studying the luminescence and laser properties like total branching ratios(β_{cal}), Effective bandwidth ($\Delta\lambda_{eff}$), stimulated emission cross-section (σ_{SE}) and so on for obtained transitions the OFSbBPr glasses are useful for luminescent devices.

2 Experimental section

2.1 Synthesis

The melt-quenching method is used for the synthesis of glass samples. Analytical reagent grade chemicals, B_2O_3, Sb_2O_3, Al_2O_3, NaF, LiF and Pr_6O_{11} from Sigma Aldrich with a purity of 99.99% are mixed according to the composition 60B_2O_3-10Sb_2O_3-10Al_2O_3-10NaF-(10-x) LiF- xPr_6O_{11} where x represents the concentration of Pr^{3+} ions (0.1, 0.5, 1.0, 1.5, 2.0 mol%). The glasses that were made with varying concentrations of Pr are referred to as OFSbBPr0.1, OFSbBPr0.5, OFSbBPr1.0, OFSbBPr1.5, and OFSbBPr2.0. The compounds were measured in

terms of their molecular weight using an electronic balance. Using agate mortar and pestle, the chemical compounds are ground into fine powder after being weighed. Using an electric furnace, the 8 g of powdered material is heated at 1200 °C for 10 min in silica crucibles. Stirring the mixture every three minutes kept it homogenous. To reduce the thermal stress brought on by sudden cooling, the liquid metal is solidified in heated brass molds and plates, which are then annealed at 500 °C for two hours. The synthesized glasses exhibited a thickness of 2 mm and had transparency which are shown in the Fig. 1

2.2 Sample characterization

The glass sample thickness has been determined using a screw gage. The refractive index and density were determined using the Brewster angle method and the Archimedes principle. The XRD analysis of titled glasses are performed using an PANalytical EMPERYAN diffractometer. The EDX analysis of the samples are determined using an EDS- Oxford max 50 detector. The TEM analysis of the glass are characterized using Talos F200S G2 instrument. The mechanical strength of the glasses is measured using an Economet VH-1MDX METCO microhardness tester. The prepared glasses at room temperature, spectra of optical absorption in the 300–2600 nm wavelength range were examined using 1 nm-resolution JASCO V-670 spectrometer. For glasses doped with Pr^{3+}, photoluminescence excitation (also known as PLE) and photoluminescence emission (also known as PL) were investigated across the 200–800 nm wavelength range. The spectrofluorometer used for the observations was the JASCO FP-8300, which was excited by a 150W xenon lamp. The wavelength resolution was attained at 1 nm.

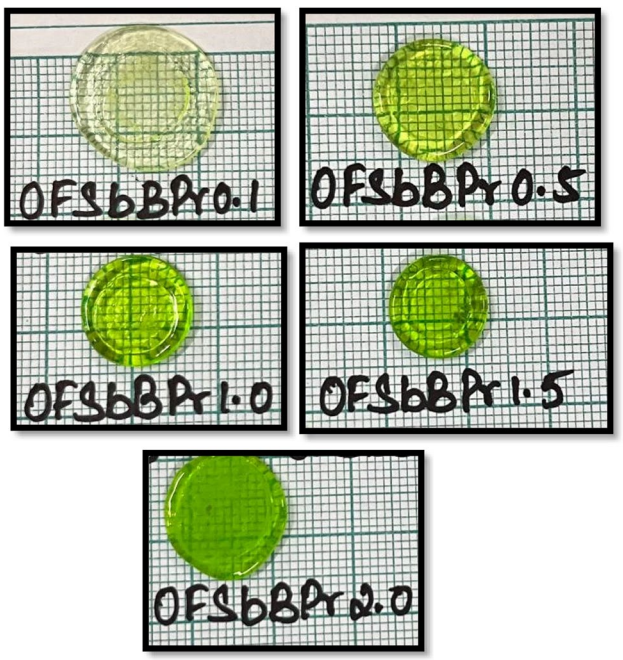


Fig. 1 Photographs of various concentrations of Pr^{3+} ions doped OFSbB glasses

3 Results and discussions

3.1 The physical parameters

Physical parameters significantly impact the optical characteristics of OFSbB glasses doped with Pr^{3+} ions. The Physical parameters of Pr^{3+} ions doped OFSbB glasses are given in Table 1. Density (ϱ) of resulting glass increased due to the substitution of Pr_6O_{11}, which has a higher molecular weight for LiF. Among these properties, the refractive index(n) is very important because it determines the appropriateness of the material for optical applications. The determination of J-O parameters and laser parameters is also significantly influenced by n. Figure 2 illustrates the variation of refractive index and density observed in OFSbB glasses doped with Pr^{3+} ions. Rare earth ions play two functions: they establish networks and modify them. Adding this ion to specific glass compositions increases the ϱ in the structure. The inclusion of RE ions reduces glass compaction ϱ values [29]. Figure 3 illustrates the variation of (a) polaron radius(r_p), interionic distance(r_i), (b) field strength (F_s) and (c) molecular polarizability(α_m) with Pr^{3+} ions concentration in OFSbB glasses. When Pr^{3+} ions were added, the interionic distance (r_i) and polaron radius (r_p) decreased steadily. With addition of Pr^{3+} ions, which possess an ionic radius of 0.99 A° into the lattice results in a decrease in the interionic distance as compared to LiF ions. As the ion shrinks, the bond between the electrons of neighboring atoms becomes stronger and the polaron radius decreases. Furthermore, r_i and r_p decrease when the number of ϱ of Pr^{3+} ions increase. Also, the strength of the electric field around the Pr^{3+} ion increases as the average distance between praseodymium (Pr–O) and oxygen decreases. The change in interionic radius, polaron radius, field strength, and molecular polarizability. In addition, the molar refraction (R_m), and electrical polarization (α_m) increasing

Table 1 Physical parameters of Pr^{3+} ions doped OFSbB glasses

Physical properties	OFSbBPr0.1	OFSbBPr0.5	OFSbBPr1.0	OFSbBPr1.5	OFSbBPr2.0
Refractive index (n)	2.381	2.365	2.353	2.329	2.325
Density (g/cc)	2.302	2.32	2.341	2.369	2.384
Average molecular weight (g)	84.234	88.216	93.194	98.171	103.148
Pr^{3+} ion concentration ($\times 10^{22}$ ions/cm^3)	0.127	0.138	0.251	0.36	0.462
Mean atomic volume (g/cm^3/atom)	8.810	9.129	9.523	9.949	10.277
Dielectric constant (Ɛ)	5.669	5.593	5.536	5.424	5.405
Optical dielectric constant (P $\partial t/\partial p$)	4.669	4.593	4.536	4.424	4.405
Reflection losses (R)	0.166	0.164	0.162	0.159	0.158
Molar refraction Rm (cm^{-3})	22.277	23.001	23.963	24.694	25.739
Polaron radius (A°)	3.782	3.679	3.014	2.673	2.460
Interionic distance (Å)	9.384	9.128	7.480	6.633	6.104
Molecular electronic polarizability ($\times 10^{-23}$ cm^3)	0.114	0.104	2.728	3.153	3.975
Field strength ($\times 10^{15}$ cm^2)	2.097	2.216	3.301	4.197	4.955

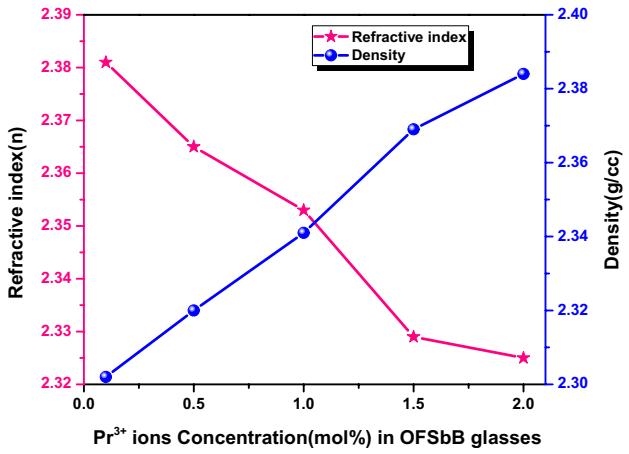


Fig. 2 Variation of refractive index and density with Pr^{3+} ions concentration in OFSbB glasses

with Pr_6O_{11} concentration indicates an increase in NBOs in the glass. Observing the low field strength values shows that the synthesized glasses have high solubility of RE ions [11, 18, 29, 30].

3.2 Structural analysis

3.2.1 X-Ray Diffraction (XRD) and Transmission electron microscopy (TEM) analysis

To confirm the amorphous nature of the prepared glasses the XRD spectra analysis was carried out and it was shown in Fig. 4. As shown in Fig. 4a board hump observed in the XRD spectra without any sharp peak confirms the amorphous nature of the as-prepared glasses. TEM was measured for the OFSbBPr1.0 glass sample which is illustrated in Fig. 5. Since the sample micrographs showed a hallow and no diffraction rings and verifying the amorphous nature.

3.2.2 Energy dispersive X-ray (EDX) analysis

To examine the components of the produced glass samples, the EDX analysis was performed. The EDX spectra of Pr^{3+}-doped OFSbB1.0 glass are shown in Fig. 6 and Table 2. The EDX spectrum shows the glass composition chemicals like Boron (B), Fluoride (F), Praseodymium (Pr), sodium (Na), Oxygen (O), Aluminum (Al), Antimony (Sb). The EDX spectrum also shows the presence of additional elements like carbon (C) because the carbon substrate is used in the analytical method which causes the carbon contamination peak [31]. However, lithium was not detected due to its low energy ($K\alpha < 1$ keV) which has difficulty in detecting mild elements such as Li [32] and concern about the absorption capacity of the specimen [33]. The inset image in the spectrum depicts the SEM image with 200μm magnification. The non-existence of grains in SEM image specifies that OFSbBPr glass is non-crystalline.

3.3 The optical absorption spectra analysis and bonding parameters

Figure 7 illustrates the absorption and transmission spectrum of un-doped OFSbB glass. The absorption spectra of OFSbB glass samples doped with Pr^{3+} ions are illustrated in Fig. 8 which is recorded in

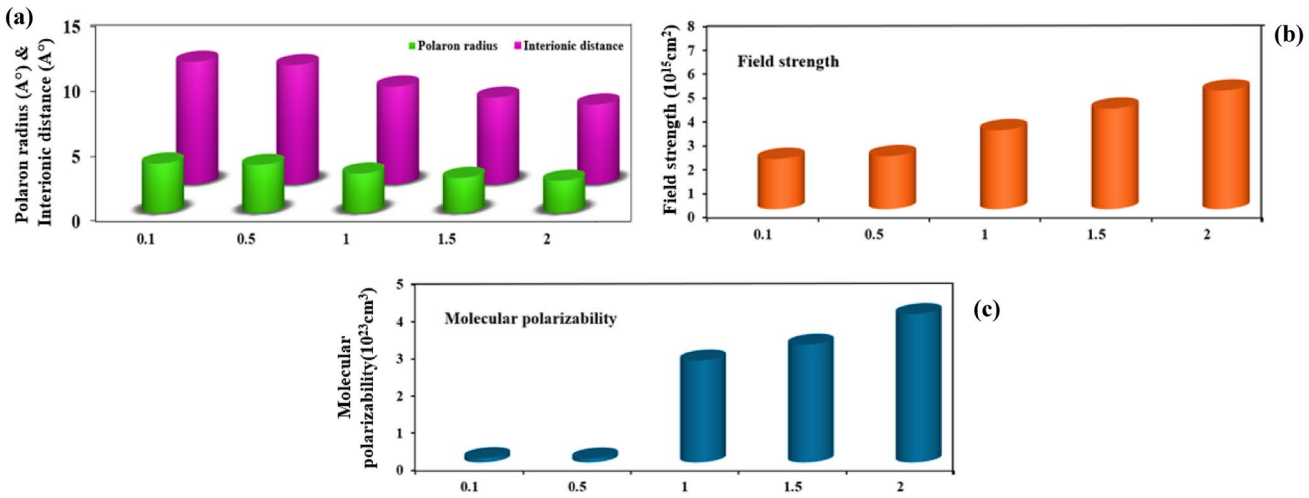


Fig. 3 **a** Variation of Polaron radius(r_p), Interionic distance(r_i), **b** Field strength (F_s) and (c) Molecular polarizability (α_m) with Pr^{3+} ions concentration in OFSbB glasses

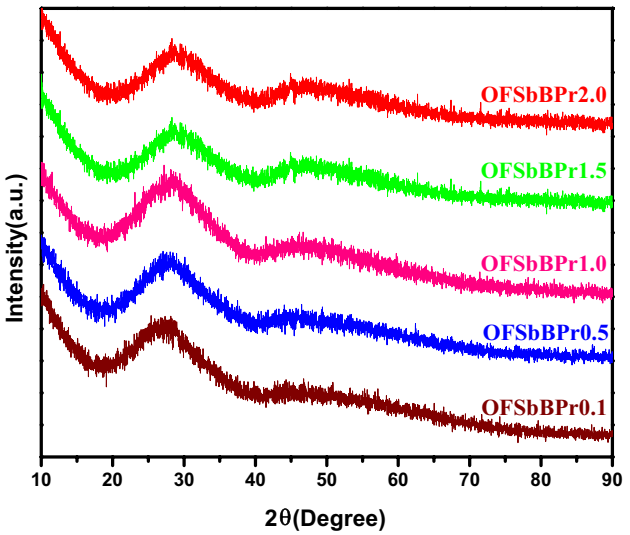


Fig. 4 The XRD spectra of Pr^{3+} ions doped OFSbB glasses

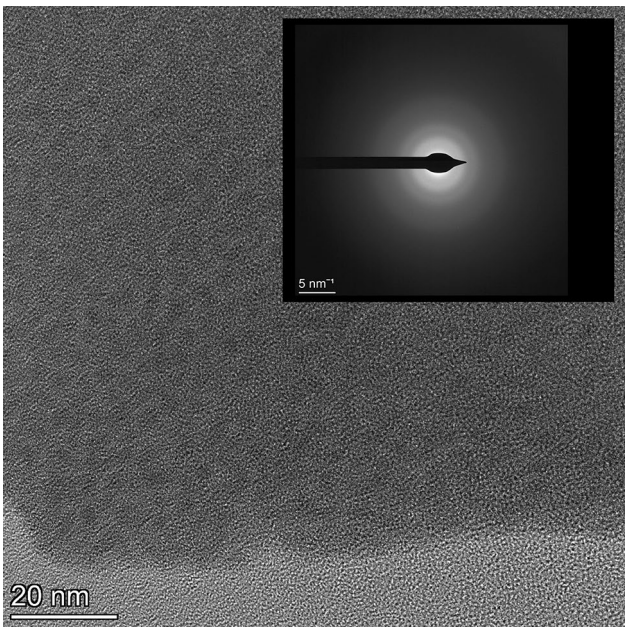


Fig. 5 The TEM micrograph of OFSbBPr1.0 glass (Inset image shows the SAED pattern of OFSbBPr1.0 glass)

the 400–2000 nm range. From the results, OFSbBPr glasses show seven absorption bands resulting from the transitions in the $4f^2$ arrangement of the Pr^{3+} ion, these changes mainly involve the transition from the $^3H_4$ initial state to different higher states, which may be attributed to the $4f^2$–$4f^2$ band. Three sets of bands are identified in the violet-blue region of the visible the range of wavelengths (444 nm, 470 nm, and 483 nm) in the middle, with an isolated band at 590 nm. In the infrared region, there are three absorption bands at 1234 nm, 1486 nm, and 1913 nm. These absorption peaks are equivalent to electronic transitions, $^3H_4 \rightarrow {}^3P_2$, $^3P_1$, $^3P_0$, $^1D_2$, $^1G_4$, ($^3F_3 + {}^3F_4$) and $^3F_2$.

The Carnal et al., [34] study reports the identification of absorption bands to discrete transitions. Of these transitions, only two are transitions: $^3H_4 \rightarrow {}^3P_2$ and the hypersensitive $^3H_4 \rightarrow (^3F_3 + 3F_4)$. They are highly dependent on the ligands in their environment and in accordance with the selection principle $\Delta S = 0$, $\Delta L < 2$, and $\Delta J \leq 2$, $\Delta S = 0$, $\Delta L < 2$, and $\Delta J \leq 2$ [28, 35].

Glass hosts exhibit a stronger interaction with 4f orbitals compared to crystalline hosts. Additionally, the

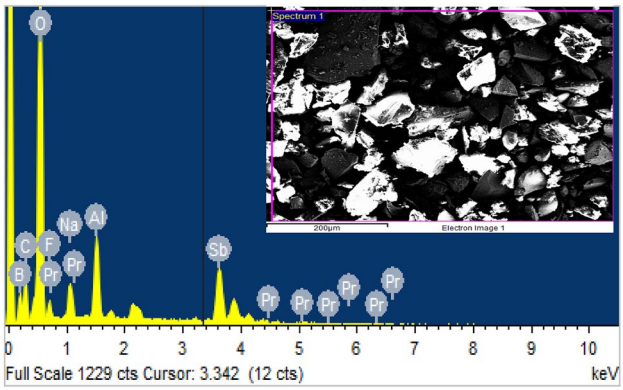


Fig. 6 EDX spectrum of $Pr.^{3+}$ ions doped OFSbB1.0 glass (Inset shows the SEM image of OFSbBPr1.0 glass)

Table 2 EDX compositional analysis of Pr^{3+} ions doped OFSbB1.0 glass

1.0 Pr_6O_{11} mol%		
Element	Weight	Atomic (%)
B K	13.27	25.67
C K	6.86	11.94
O K	36.07	47.14
F K	3.51	3.87
Na K	2.27	2.07
Al K	4.65	3.6
Sb L	32.23	5.54
Pr L	1.14	0.17
Totals	100.00	

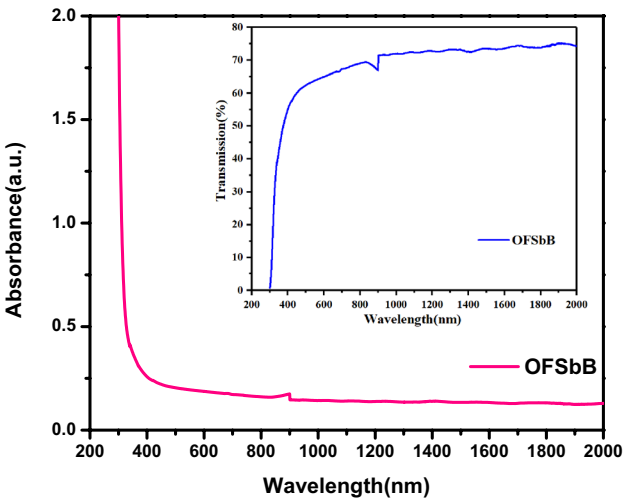


Fig. 7 The absorption spectrum of OFSbB glass (Inset shows the transmission spectrum of OFSbB glass)

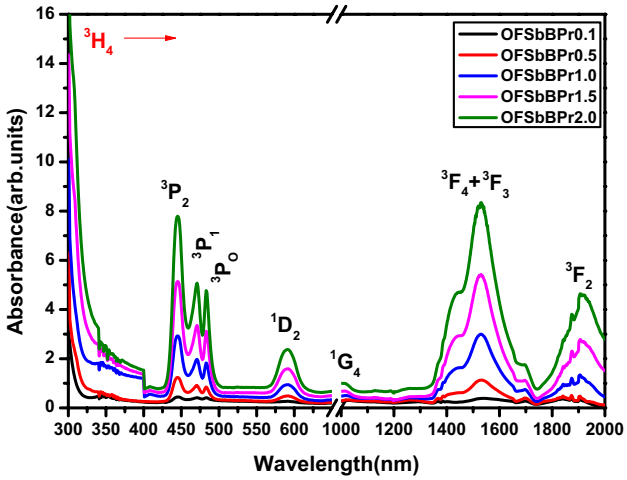


Fig. 8 Absorption spectra of Pr^{3+} ions doped OFSbB glasses in Vis–NIR region

energy domains of ions have a greater magnitude in glass hosts compared to either aqua ions or crystalline hosts. The obtained absorption spectra data were analyzed to determine the type of bond between Pr and O in the prepared sample glasses. Modifications in environment of the RE ions, namely the Nephelauxetic ratio (β), may explain the slight shifts in the absorption bands peak areas. Using a formula, the bonding parameters (δ) and β were determined.

$$\beta = \frac{\vartheta_a}{\vartheta_c} \tag{1}$$

where, ϑ_c represents the measured wavenumber of a specific transition associated with complex ion; ϑ_a indicates the wavelength representing the corresponding aquo-ion transition. Depending on whether their values are positive or negative, the bonding factors identify if the bond is ionic or covalent. The δ were used to calculate using the equation below [12, 35]

$$\delta = \frac{1 - \overline{\beta}}{\overline{\beta}} \tag{2}$$

The bonding parameters of the OFSbBPr glasses are evaluated based on the shifts detected in the resulting optical absorption spectra and the average value of β which is $\overline{\beta}$. The calculated results are shown in Table 4. All the prepared OFSbBPr glass samples had bonding parameter values that were determined to be negative, confirming the existence of ionic bonding.

3.4 Optical bandgap analysis

A crucial analysis for determining the absorption characteristics of optically active materials is the optical bandgap (E_g). There are two types of energy bandgap, Indirect and Direct. Where there is no electric field, a direct bandgap was observed when a minimal energy of the band associated with conduction (CB) and the highest energy of the band associated with valence (VB) coincide at the same momentum. As a result, the bandgap photon absorbs the smallest quantity of energy needed to transfer an electron across the VB to the CB. On the other hand, an electron cannot be moved from VB to the CB with the energy provided by a single photon that has the same energy as the bandgap. A difference in the momentum of two types of materials varies the distinct behavior of electron movement. The Pr^{3+} ions doped OFSbB optical bandgap are evaluated using Mott and Davis expression [36]

$$\alpha(\vartheta) = \frac{A\left(h\vartheta - E_g\right)^n}{h\vartheta} \tag{3}$$

In this context, the symbol "n" denotes the exponent that characterizes the mode of electrical transition, namely whether it is indirect or direct, inside k-space. Studying the principal absorption cut-off at the shorter wavelength, which is usually in the ultraviolet provides valuable insights into the band structure & optical shifts in both non-crystalline material and crystalline material. The change in the absorption coefficient (α) with the energy of photons($h\nu$) is seen in several non-crystalline systems, as shown in Tauc's plot [37]. Figure 9 represents the Tauc's plot of a energy versus $(\alpha h\nu)^{1/2}$ and b energy versus $(\alpha h\nu)^2$ for the Pr^{3+} ions doped OFSbB glasses (Inset 9(a) shows energy versus $(\alpha h\nu)^{1/2}$ and 9(b) shows energy versus $(\alpha h\nu)^2$). The calculated values of the direct and indirect bandgaps of un-doped and Pr^{3+} ions doped OFSbB glasses are shown in Table 3. The result clearly indicates that the direct bandgap is greater than the indirect bandgap for all the synthesised glasses. The presence of Pr_6O_{11} causes an increase in the amount of NBOs, which subsequently towards the conduction band. Consequently, NBOs minimize the change amongst the valence and conduction bands, resulting in a decrease in the energy gap [38]. Considering that the indirect transition occurs at lower energy than the direct transition, all the prepared samples are classified as indirect or non-direct materials.

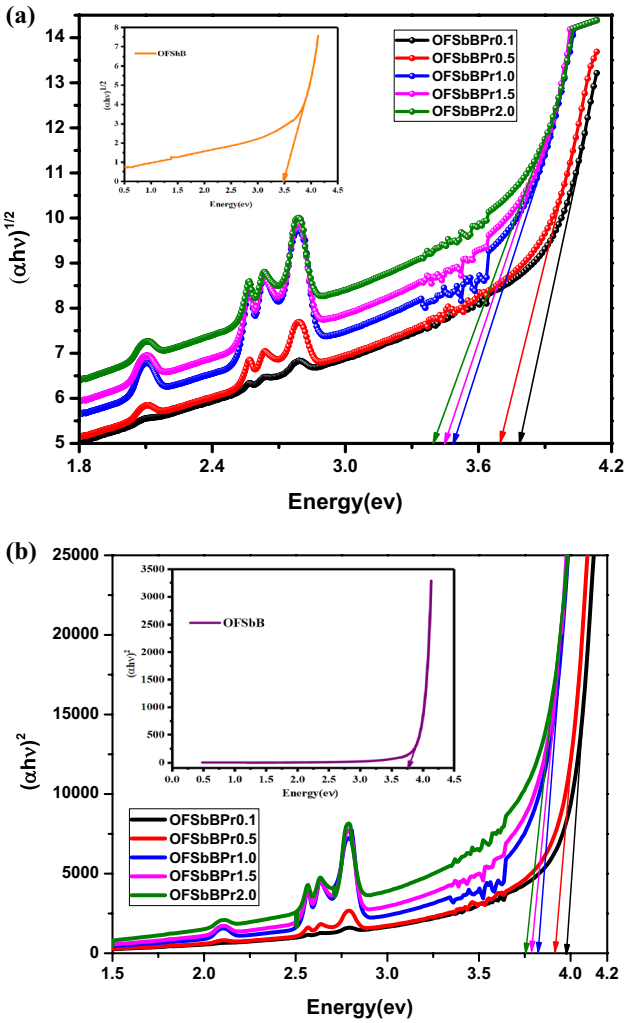


Fig. 9 The Tauc's plot of **a** energy versus $(\alpha h\nu)^{1/2}$ and **b** energy versus $(\alpha h\nu)^2$ for the Pr^{3+} ions doped OFSbB glasses (Inset 8(a) shows energy versus $(\alpha h\nu)^{1/2}$ and 8(b) shows energy versus $(\alpha h\nu)^2$)

Table 3 Direct and Indirect bandgap energy (E_g) of the un-doped and Pr^{3+} ions doped OFSbB glasses

Glass code	Direct bandgap (E_g)	Indirect bandgap (E_g)
OFSbB	3.76	3.50
OFSbBPr0.1	3.98	3.79
OFSbBPr0.5	3.91	3.7
OFSbBPr1.0	3.82	3.49
OFSbBPr1.5	3.79	3.45
OFSbBPr2.0	3.75	3.4

Table 4 Experimental oscillator strength(f_{exp}) (10^{-6}), calculated oscillator strength(f_{cal}) (10^{-6}) with δ_{rms} (rms deviation) and bonding parameters (δ) of Pr^{3+} ions doped OFSbB glasses

Transition	OFSbBPr0.1		OFSbBPr0.5		OFSbBPr1.0		OFSbBPr1.5		OFSbBPr2.0	
$^3H_4 \rightarrow$	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}
$^3P_2$	3.8400 ± 0.0001	6.9003 ± 1.2915	6.9624 ± 0.0001	7.7794 ± 0.0094	8.7795 ± 0.0095	12.2606 ± 0.0006	5.1688 ± 0.0008	10.8354 ± 0.0003	4.4478 ± 0.0008	7.4147 ± 0.0007
$^3P_1$	1.1430 ± 0.0001	0.8206 ± 0.2885	1.9287 ± 0.0001	1.5530 ± 0.0029	2.5292 ± 0.0092	2.2843 ± 0.0003	1.7202 ± 0.0002	1.4259 ± 0.0009	0.7519 ± 0.0004	0.600 ± 0.0003
$^3P_0$	1.0380 ± 0.1710	0.8084 ± 0.2839	1.1592 ± 0.0092	1.5298 ± 0.0098	1.9762 ± 0.0062	2.2502 ± 0.0001	1.0551 ± 0.0004	1.4046 ± 0.0006	0.4067 ± 0.0003	0.5910 ± 0.0003
$^1D_2$	0.7471 ± 0.0005	2.0070 ± 0.3670	0.5337 ± 0.0003	2.2962 ± 0.0003	0.8465 ± 0.0067	3.6118 ± 0.0017	0.6186 ± 0.0006	3.1830 ± 0.0006	0.1908 ± 0.0009	2.1723 ± 0.0002
$^1G_4$	0.2976 ± 0.0003	0.4880 ± 0.0907	0.3322 ± 0.0004	0.5557 ± 0.0057	0.6623 ± 0.0026	0.8748 ± 0.0047	0.4861 ± 0.0005	0.7729 ± 0.0005	0.0975 ± 0.0002	0.5287 ± 0.0008
$^3F_4 + ^3F_3$	7.5890 ± 0.0003	7.2688 ± 0.1628	10.5720 ± 0.0001	10.4449 ± 0.004	16.7660 ± 0.006	16.4186 ± 0.0086	14.9590 ± 0.0001	14.4637 ± 0.0002	10.1590 ± 0.0001	9.8667 ± 0.0007
$^3F_2$	1.9632 ± 0.0001	1.9701 ± 0.0001	4.0230 ± 0.0001	4.0280 ± 0.008	5.9554 ± 0.0054	5.9730 ± 0.0030	4.9152 ± 0.0002	4.9425 ± 0.0004	3.1399 ± 0.0009	3.1548 ± 0.0009
δ_{rms}	1.268		0.767		1.693		2.367		1.366	
δ	− 0.8534		− 0.8532		− 0.8542		− 0.857		− 0.8532	

3.5 Judd-Oflet analysis

The measurement of oscillator strength is commonly used as a means of quantifying the magnitude of an absorption band. The following formula can be applied to calculate the experimental oscillator strength of an f-f transition in an absorption spectrum (f_{exp}).

$$f_{exp} = 4.318 \times 10^{-9} \int \varepsilon(\nu) d\nu \quad (4)$$

The wavelength at which a material can absorb light is represented by its molar absorptivity, which is abbreviated as Ɛ. Table 4 depicts the values of the RMS deviation, δ, and calculated oscillator strength (f_{cal}) (10^{-6}) of OFSbB glasses doped with Pr^{3+} ions.

The theoretical oscillator strength may be determined by using the J-O theory to an f-f transition [39, 40]. The probability of changing from an initial state to a higher state is represented by the f_{cal} value for an obtained electric dipole (ED) shift, according to the theory mentioned above.

$$f_{cal} = \frac{8\pi^2 mc\nu}{3h(2J+1)} \frac{(n^2+2)^2}{9n} \Sigma_{\lambda=2,4,6} \Omega_\lambda |< \psi J || \psi' J' >|^2 \quad (5)$$

Here, the dipole-to-dipole corrections, the Lorentz localized field change may be represented as $(n^2+2)^2/9n$. In the initial state, the total of the angular momentum is represented by the variable 'J,' and the deformation ($^{2S+1}L_J$) of the original state is represented by the variable 2J + 1. The symbol 'ν' denotes the wavenumber associated with the transition. The square of the charge displacement caused by the induced ED transition is measured by the J-O intensity parameters, Ω_J (where J = 2,4,6). When moving from ΨJ to Ψ'J', the twice reduced matrix components are taken into consideration in the intermediate coupling approximation, which is represented by the term $||U^\lambda||^2$ [34].

The J-O theory and the experimental oscillator strength act as the base for the least square fit method, which is employed to calculate J-O and f_{exp}. The J-O data show that the prepared glasses follow the pattern of ($\Omega_6 > \Omega_2 > \Omega_4$). The Ω_2 parameter, which relates to the asymmetry of the immediate environment of the RE^{3+} ions, represents the covalent properties of Pr^{3+} ions and ligand anions. In comparison to Ω_4 and Ω_6, the Ω_2 values are often lower. The parameters Ω_4, Ω_6 which determine the intensity are interconnected with the fundamental characteristics of the substance

such as its viscosity and dielectric properties [35, 39, 41]. Variations in the RE ion vibrational frequencies associated with the binding atoms also impact these characteristics. The OFSbBPr1.0 glass exhibits a higher level of rigidity in its Pr–O bonds and the also Pr^{3+} ions exhibit a higher degree of symmetry, which is indicated by the increased magnitude of Ω_6. The J-O values of the OFSbBPr glasses are shown and compared with the other reported glasses, as indicated in Table 5 [42–44]. The J-O intensity parameters for the titled glasses follow same trend which is like the trends observed in PbO-Al_2O_3-SiO_2:5Pr_2O_3 [42], PZLBPr [43] glasses, whereas the dissimilar trends have been observed in BPGBPr [44]. The results for the present OFSbB1.0 glass is greater than have more ionic nature.

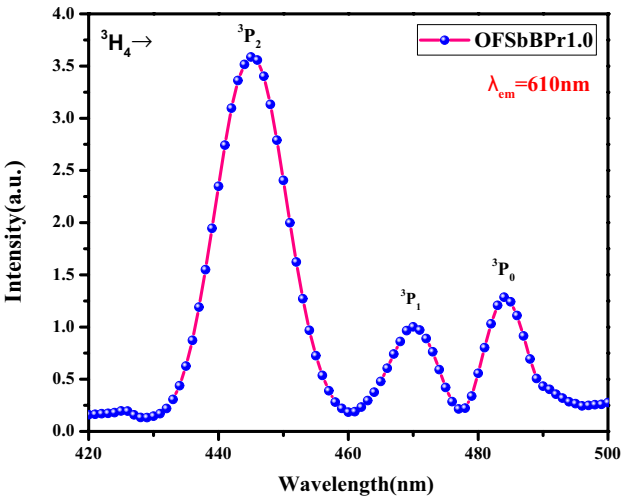


Fig. 10 Excitation spectrum of Pr^{3+} ions doped OFSbB glasses under the emission of 610 nm

3.6 Fluorescence analysis

The fluorescence characteristics of Pr^{3+} ions doped OFSbB glasses can be determined by obtaining information about the excitation level (λ_{exc}). When studying PL spectra of optically active materials doped with RE ions, the excitation wavelength is a key parameter. The excitation spectra of OFSbB glasses doped with Pr^{3+} ions under 610 nm emission are shown in Fig. 10. At 420–520 nm, the spectra show three distinct bright states associated with transitions from $^3H_4$ (ground level) to $^3P_2$, $^3P_1$, and $^3P_0$ (upper level). Non-radiative energy transfer may rapidly convey excitation energy from the $^3P_2 \rightarrow {}^3P_1$ & $^3P_0$ states due to minimum energy difference among the stimulated levels of the Pr^{3+} ion. Due to succession of radiative transitions, the excited state $^3P_0$ is depopulated, and the Pr^{3+} ion is transferred to lower energy levels. Hence, the photoluminescence properties of OFSbBPr glasses are studied using the 445 nm excitation band, which is a strong and relates to the change from $^3H_4$ to $^3P_2$ [25].

The emission spectra of Pr^{3+}-doped OFSbB glasses are shown in Fig. 11 at an excitation wavelength of 445 nm. From the emission spectra, nine bands are observed $^3P_0 \rightarrow {}^3H_4$, $^3P_1 \rightarrow {}^3H_5$, $^3P_0 \rightarrow {}^3H_5$, $^3P_0 \rightarrow {}^3H_6$, $^1D_2 \rightarrow {}^3H_4$, $^3P_0 \rightarrow {}^3F_2$, $^3P_1 \rightarrow {}^3F_3$, $^3P_1 \rightarrow {}^3F_4$, and $^3P_0 \rightarrow {}^3F_4$ The Pr^{3+} ions undergo rapid excitation after the $^3H_4$ to the $^3P_2$ higher state. Within the non-radiative relaxation process, there is a transfer of energy from the higher $^3P_2$ excited state to lower $^3P_0$ excited state. Both luminous states, $^3P_0$ and $^3P_1$, attain thermal equilibrium even at ambient conditions, but having a tiny energy difference between them. As a result, radiative relaxation takes place from both the excited states $^3P_0$ and $^3P_1$. Pr^{3+} doped glasses often display emission transitions within the spectral range of 580–640 nm, specifically the $^1D_2 \rightarrow {}^3H_4$ and $^3P_0 \rightarrow {}^3H_6$ in orange region. From the spectra, emission quenching has

Table 5 Judd-Oflet parameters ($\times 10^{-20}$ cm^2) of the Pr^{3+} ions doped OFSbB glasses along with other reported glasses

Glass system	Ω_2	Ω_4	Ω_6	Trend	References
OFSbBPr0.1	0.0883 ± 0.006	0.3793 ± 0.1331	3.6950 ± 0.7464	$\Omega_6 > \Omega_4 > \Omega_2$	Present work
OFSbBPr0.5	1.0625 ± 0.0001	0.7419 ± 0.0001	4.2251 ± 0.0005	$\Omega_6 > \Omega_2 > \Omega_4$	Present work
OFSbBPr1.0	1.5456 ± 0.0001	1.1346 ± 0.0001	6.9442 ± 0.0002	$\Omega_6 > \Omega_2 > \Omega_4$	Present work
OFSbBPr1.5	1.3575 ± 0.0001	0.7105 ± 0.00005	6.2363 ± 0.0005	$\Omega_6 > \Omega_2 > \Omega_4$	Present work
OFSbBPr2.0	0.9171 ± 0.0001	0.3009 ± 0.0001	4.3462 ± 0.00005	$\Omega_6 > \Omega_2 > \Omega_4$	Present work
PbO-Al_2O_3-SiO_2:5Pr_2O_3	0.871	0.910	2.103	$\Omega_6 > \Omega_2 > \Omega_4$	[42]
PZLBPr	0.018	1.641	1.816	$\Omega_6 > \Omega_2 > \Omega_4$	[43]
BPGBPr	0.67	1.99	3.75	$\Omega_6 > \Omega_4 > \Omega_2$	[44]

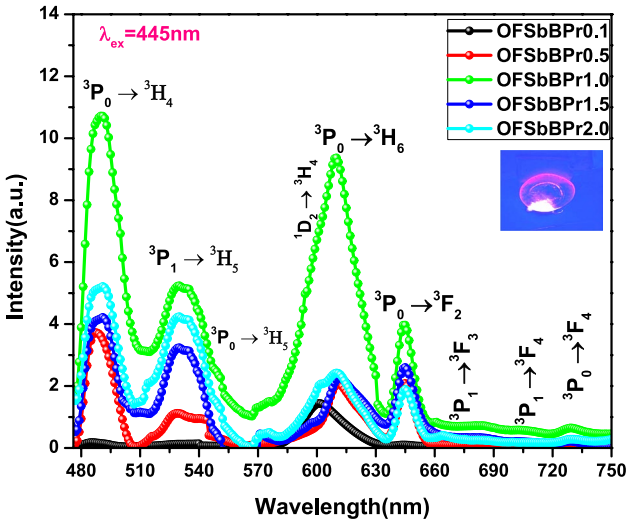


Fig. 11 Emission spectra of Pr^{3+} ions doped OFSbB glasses under the excitation of 445 nm (Inset shows the photograph of synthesized glass sample under 445 nm light source)

been observed for OFSbBPr1.0 glass for all peaks due to concentration quenching which plays a significant role in the multiphonon emission [11, 12]. Among these emission transitions, the transitions at 490 nm ($^3P_0$ to $^3H_4$) and 610 nm ($^1D_2$ to $^3H_4$,$^3P_0 \rightarrow {}^3H_6$) exhibited higher intensity. In some glasses, the green–blue emission is more dominant than the red fluorescence due to greater radiative relaxation from $^3P_0$ to the $^1D_2$ state [9, 25, 26].

The radiative properties of Pr^{3+} ions doped OFSbB glasses, like A_T, A_R, β_r, σ_{SE} radiative lifetime(τ_R), and Effective bandwidth ($\Delta\lambda_p$) of all emission transitions, were evaluated to predict their lasing performance. The calculation was carried out applying the J-O intensity values as well as the refractive index. A_R is the probability of spontaneous emission per second. The below equation is used to calculate A_R for emission transitions [45, 46]

$$A_R(\psi J, \psi' J') = \frac{64\pi^4 \vartheta}{3h(2J+1)}\left[\frac{n(n^2+2)}{9} S_{ed} + n^3 S_{md}\right] \quad (6)$$

Here, $\psi J, \psi' J'$ represent the primary and ending states, respectively. S_{ed} and S_{md} are the primary strengths for electric and magnetic dipole transitions, respectively, v (cm^{-1}) is the energy difference between the states.

The A_T is a sum of A_R for an excited states which is calculated in the overall terms [45, 46]

$$A_T(\psi J) = \sum_{\psi J'} A_R(\psi J, \psi' J') \quad (7)$$

The β_r is the one of significant parameter, used to expect the lasing emission characteristics [45, 46]

$$\beta_r = \frac{A_R(\psi J, \psi' J')}{A_T(\psi J)} \quad (8)$$

The τ_R of the emission state is [45, 46]

$$\tau_R = \frac{1}{A_T(\psi J)} \quad (9)$$

Finally, the σ_{SE} another important parameter because it quantifies the speed at which energy is extracted from the optical medium [45, 46]

$$\sigma_{SE} = \frac{\lambda_p^4 A_R(\psi J, \psi' J')}{8\pi c n^2 \Delta\lambda_p} \quad (10)$$

Here, λ_p denotes peak maximum, $\Delta\lambda_p$ stands for the Full-Width at Half-Maximum (FWHM) and n represents the refractive index.

Table 6 shows the computed radiative parameters of OFSbB glasses doped with Pr^{3+} ions. In addition to these, optical gain parameters (σ_{SE} x τ_R) (10^{-25}) (in cm^2 s), Effective bandwidth ($\Delta\lambda_p$) (in nm), radiative lifetime(μs), stimulated emission cross-sections (σ_{SE}) (10^{-20}) (in cm^2), gain bandwidth (σ_{SE} x $\Delta\lambda_p$) (10^{-25}) (in cm^3), and for the $^3P_0 \rightarrow {}^3H_4$ and $^1D_2 \rightarrow {}^3H_4$ emission transitions for the OFSbBPr glasses and other reported glasses [18, 26, 30, 45, 47] are tabulated in the Table 7 that are equally crucial for identifying laser active material. It is quite probable that OFSbBPr1.0 glass, out of all the prepared glasses, will undergo the $^3P_0 \rightarrow {}^3H_4$ transition. The relative intensities of the emission bands produced at a certain stimulated energy level are evaluated using branching ratios. Due to its ability to characterize the probability of obtaining stimulated emission. The β_r of around 50% is used to define the efficiency of laser radiation. The $^3P_0 \rightarrow {}^3H_4$ has the largest magnitude of β_{exp} among all transitions and the OFSbBPr1.0 glass has higher values for the $^1D_2 \rightarrow {}^3H_4$ and $^3P_0 \rightarrow {}^3H_4$ transitions (about 47% and 41%). Similarly, the σ_{SE} of OFSbBPr1.0 is higher compared to other prepared glasses which may be useful for laser applications. To achieve the laser host material with the highest stability, the (σ_{SE} x τ_R) and (σ_{SE} x $\Delta\lambda_p$) remain larger. Therefore, RE-doped glass can also work as

Table 6 The radiative parameters of radiative transition probability (A_R) (s^{-1}), branching ratio(β_{cal}), and stimulated emission cross-section (σ_{se}) (10^{-20} cm^2) for the emission transitions of Pr^{3+} ions doped OFSbB glasses

Transition	Parameters	OFSbBPr0.1	OFSbBPr0.5	OFSbBPr1.0	OFSbBPr1.5	OFSbBPr2.0
$^3P_1 \rightarrow {}^3F_4$	A_R	1103.06 ± 38.69	2017.68 ± 0.27	2796.91 ± 41.7	1772 ± 0.23	738.07 ± 49.24
	β_{cal}	0.0430 ± 0.015	0.0470 ± 0.0001	0.0455 ± 0.0005	0.035 ± 0.0001	0.024 ± 0.064
$^3P_1 \rightarrow {}^3F_3$	A_R	1488.97 ± 7.75	8389.31 ± 0.22	11,342.08 ± 0.62	9356 ± 43.99	5860.89 ± 19.802
	β_{cal}	0.0580 ± 0.030	0.1953 ± 0.0001	0.1844 ± 0.0006	0.1881 ± 0.00005	0.1914 ± 0.0368
$^3P_1 \rightarrow {}^3H_5$	A_R	11,636 ± 7.61	14,632.04 ± 10.38	21,519.11 ± 10.01	17,311.22 ± 13.41	10,617.20 ± 12.865
	β_{cal}	0.4537 ± 0.027	0.3407 ± 0.0001	0.3498 ± 0.0005	0.3480 ± 0.00004	0.3467 ± 0.022
$^3P_0 \rightarrow {}^3F_4$	A_R	1212.83 ± 42.54	2218.47 ± 0.3	3075.24 ± 45.85	1949.37 ± 0.25	811.52 ± 54.148
	β_{cal}	0.0534 ± 0.015	0.0530 ± 0.0001	0.0520 ± 0.0006	0.0413 ± 0.00001	0.0285 ± 0.0068
$^3P_0 \rightarrow {}^3F_2$	A_R	1102.14 ± 85.17	12,402.76 ± 0.0001	16,665.08 ± 60.38	14,537.72 ± 80.32	9655.52 ± 31.93
	β_{cal}	0.0485 ± 0.034	0.297 ± 0.0001	0.2818 ± 0.0021	0.3083 ± 0.0001	0.3385 ± 0.1885
$^3P_0 \rightarrow {}^3H_6$	A_R	13,567.59 ± 27.42	14,511.00 ± 17.18	22,029.95 ± 13.33	19,650.09 ± 19.85	13,463.2 ± 26.72
	β_{cal}	0.5970 ± 0.139	0.3483 ± 0.0003	0.3725 ± 0.0016	0.4168 ± 0.00005	0.4721 ± 0.2631
$^3P_0 \rightarrow {}^3H_4$	A_R	6551.55 ± 22.98	11,983.88 ± 1.61	16,612.07 ± 24.76	10,530.23 ± 1.33	4383 ± 29.250
	β_{cal}	0.2883 ± 0.085	0.2877 ± 0.0001	0.2809 ± 0.0003	0.2233 ± 0.00005	0.153 ± 0.3674
$^1D_2 \rightarrow {}^3H_4$	A_R	2252.89 ± 41.20	2488.10 ± 2.74	3755.08 ± 0.57	3299.92 ± 3.36	2228 ± 33.781
	β_{cal}	0.7021 ± 0.0701	0.4722 ± 0.0002	0.494 ± 0.00001	0.5080 ± 0.00007	0.523 ± 0.086

Table 7 Effective bandwidth ($\Delta\lambda_{eff}$) (nm), stimulated emission cross-sections (σ_{SE}) (10^{-20}) (cm^2), gain bandwidth (σ_{SE} x $\Delta\lambda_{eff}$) (10^{-26}) (cm^3), radiative lifetime (τ_R) (μs) and optical gain parameters (σ_{SE} x τ_R) (10^{-25}) (cm^2 s) for the $^3P_0 \rightarrow {}^3H_4$ and $^1D_2 \rightarrow {}^3H_4$ emission transitions for the Pr^{3+} ions doped OFSbB glasses and other reported glasses

Glass code	$^3P_0 \rightarrow {}^3H_4$					$^1D_2 \rightarrow {}^3H_4$					References
	$\Delta\lambda_{eff}$	σ_{SE}	σ_{SE} x $\Delta\lambda_{eff}$	τ_R	σ_{SE} x τ_R	$\Delta\lambda_{eff}$	σ_{SE}	σ_{SE} x $\Delta\lambda_{eff}$	τ_R	σ_{SE} x τ_R	
OFSbBPr0.1	7.142	1.32	0.94	352	0.46	1.428	1.25	0.69	253	14.6	Present Work
OFSbBPr0.5	4.545	3.58	1.63	218	0.78	3.846	0.31	0.77	211	6.57	Present Work
OFSbBPr1.0	3.846	5.79	2.23	154	0.89	2.631	1.56	1.12	159	24.95	Present Work
OFSbBPr1.5	4.166	3.34	1.39	128	0.42	3.571	0.61	1.02	147	9.00	Present Work
OFSbBPr2.0	4.166	1.36	5.65	105	0.14	2.777	1.30	0.65	139	18.13	Present Work
0.75PrZTFB	8.431	6.706	58.007	20	13.41	12.906	0.858	11.073	183	9.5	[19]
MgTP	17	9	2	14	12.6	24	~0	–	204	–	[27]
PTBPr10	22.56	4.03	0.91	39	4.84	26.82	1.38	0.23	388	12.77	[31]
LiFBMg:Pr	702	5.78	4057	122	75.14	772	0.93	717.96	13	113.46	[46]
Pr-LP	14.58	5.53	80.633	23	12.71	599.5	0.484	7.021	864	41.818	[48]

an optical fiber. Out of all the prepared samples, the OFSbBPr1.0 glass has the largest (σ_{SE} x τ_R) and (σ_{SE} x $\Delta\lambda_p$) which is shown in Table 7. The OFSbBPr1.0 glass is a good host material that can emit a reddish-orange laser.

3.7 Emission decay curves

With an excitation of 445nm and emission wavelengths at 490 and 610nm, respectively. The decay curves derived from the excited levels of $^3P_0$ and $^1D_2$ have been recorded which is demonstrated in Fig. 12a and b. A single exponential trend is appropriately described by the following equation, which is exact for all the glass decay curves:

$$I = A\exp\left(\frac{-t}{\tau}\right) \tag{11}$$

where, "I" denotes the luminescence intensity at a specific time "t", 'A' for the weighing parameter, while the decay component is denoted as 'τ'. Based on the analysis of decay profiles, it has been shown that an exponential decrease in decay curves is observed as the concentration of Pr^{3+} rises, and decay curves begin to exhibit an exponentially decreasing behavior. The movement of excitation energy from donors to acceptors is the cause of this behavior. Moreover, it is shown that the $^3P_0$ and $^1D_2$ state durations exhibit prominent variation because of the population mechanisms such as multiphonon and cross-relaxation (CR) processes. The observed exponential decrease in the lifetime of both fluorescence decay curves when the concentration rises suggests the presence of a non-radiative energy (NR) transfer or multiphonon relaxation process occurring between Pr^{3+} ions [48]. Meanwhile, the phonon energy is smaller than the energy gap (~ 3754cm^{-1}) between the $^3P_0$ and $^1D_2$ levels, so multiphonon relaxation is not possible. This includes the selected host matrix, borate (~ 1400 cm^{-1}). It follows that CR among Pr^{3+} ions in the metal–ligand ion are the mechanism responsible for energy transmission. Figure 13 demonstrates the partial energy mechanism of Pr^{3+} ions doped OFSbB glasses along with non-radiative (NR) and cross-relaxation channels. The possible cross-relaxation in the $^3P_0$ ground level is A: ($^3P_0 \rightarrow {}^3H_6$: $^3H_4 \rightarrow {}^1D_2$), B: ($^3P_0 \rightarrow {}^1G_4$: $^3H_4 \rightarrow {}^1G_4$), C: ($^3P_1 \rightarrow {}^1D_2$: $^3H_4 \rightarrow {}^3H_6$) and similarly the cross-relaxation in the $^1D_2$ ground level is D: ($^1D_2 \rightarrow {}^1G_4$: $^3H_4 \rightarrow {}^3F_4$) and E: ($^1D_2 \rightarrow {}^3F_4$: $^3H_4 \rightarrow {}^1G_4$) [47]. The absence of resonance in the energy levels has been determined through an analysis of the CR channels associated with $^3P_0$ and $^3P_1$. The transfer of energy between these levels requires the bridging of the gap by one or more phonons. A radiative energy (ET) mechanism is visible in the relaxation channel linked to $^1D_2$ between the transitions of $^1D_2$ to $^3H_4$ and $^3H_4$ to $^1D_2$ in the Pr^{3+} ions [49, 50]. The non-radiative rates of $^3P_0$ and $^1D_2$ are calculated using the following expression:

$$W_{NR} = \frac{1}{\tau_{exp}} - \frac{1}{\tau_R} \tag{12}$$

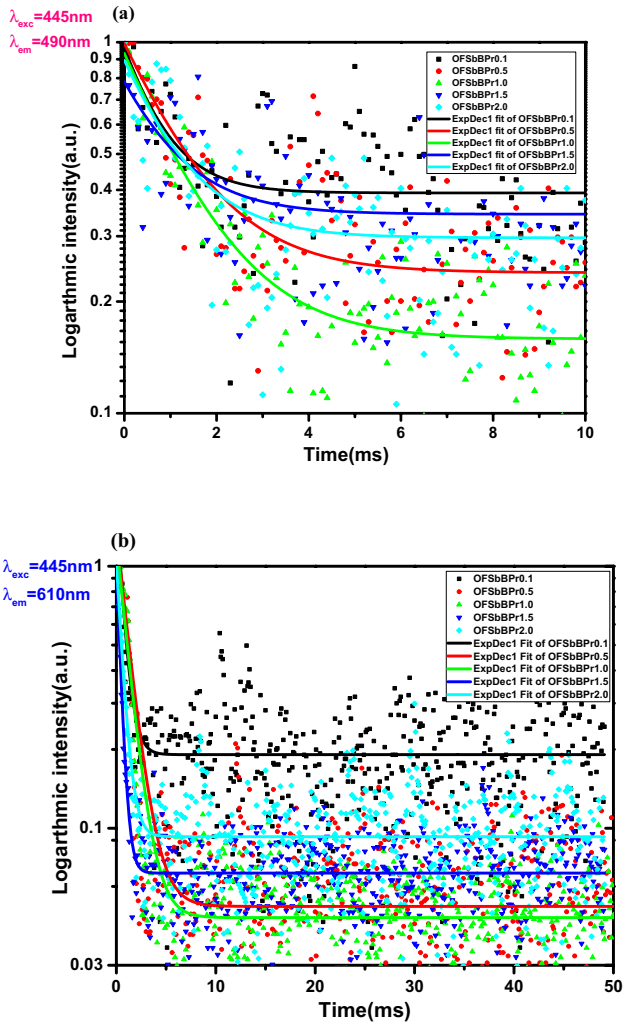


Fig. 12 **a** The decay profiles of $^3P_0$ excited level of the Pr^{3+} ions doped OFSbB glasses. **b** The decay profile of $^1D_2$ excited level of the Pr^{3+} ions doped OFSbB glasses

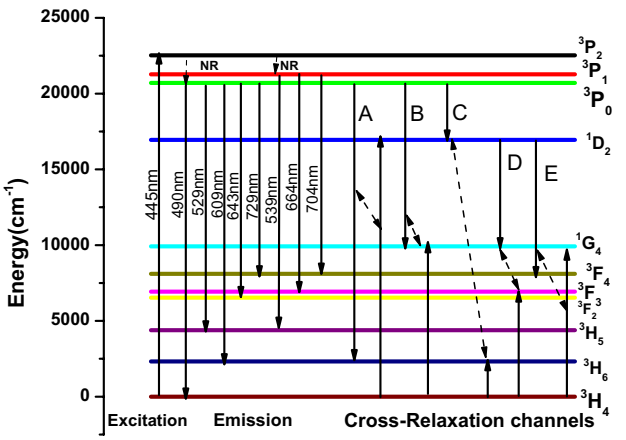


Fig. 13 Partial energy mechanism of Pr^{3+} ions doped OFSbB glasses along with the Non-radiative (NR) and Cross-Relaxation channels [A: ($^3P_0 \rightarrow {}^3H_6$: $^3H_4 \rightarrow {}^1D_2$), B: ($^3P_0 \rightarrow {}^1G_4$: $^3H_4 \rightarrow {}^1G_4$), C: ($^3P_1 \rightarrow {}^1D_2$: $^3H_4 \rightarrow {}^3H_6$) D: ($^1D_2 \rightarrow {}^1G_4$: $^3H_4 \rightarrow {}^3F_4$) and E: ($^1D_2 \rightarrow {}^3F_4$: $^3H_4 \rightarrow {}^1G_4$)]

where the τ_{exp} and τ_R represent the lifetimes obtained from experiments and calculations, respectively. The W_{NR} values for $^1D_2$ and $^3P_0$ levels are shown in Table 9. Quantum efficiency (η) in an emission state is determined by various factors, including σ_{SE} concentration of RE ions, lifetimes of metastable states, and the ligand field effect. The following equation can be used to determine the η:

$$\eta = \frac{\tau_{exp}}{\tau_R} \text{x } 100 \tag{13}$$

In Table 8, the calculated values of the Pr^{3+} ions doped OFSbB glasses are presented. The $^3P_0$ laser level of OFSbBPr1.0 (83%) glass is higher compared with $BTPbMFPr_{1.0}$ (40%) [12], TWPLaPr3 (13%) [18], LBTAF (19%) [51], TeO_2-BaF_2-NaF (21.1%) [52] and PbO-H_3BO_3-TiO_2-AlF_3 (19%) [53] glasses.

3.8 Mechanical properties

The Vicker's Microhardness (VMH) test has significant importance in enhancing comprehension of a material's mechanical properties, including yield strength, fracture behavior, and cracking temperature [54]. This technique gages a material's ease of indentation which offers important information about how tough it is. It is often used to evaluate the durability of fragile materials like glass. The VMH test involves scratching the surface of the material being tested using a diamond indenter shaped like a pyramid. The Vickers hardness number is calculated by measuring the applied force and the size of the resultant indentation. Glass Vickers microhardness can be changed based on its properties and composition. Vicker's test is used to know the material long-term strength and toughness [55] of the as-prepared glass samples. A 500 g force is applied to each sample for 10 s of OFSbBPr glasses. It has been observed that the hardness values increase with the rise of the Pr^{3+} ions concentration. This is because samples with larger concentrations have higher densities, which strengthens their resistance to external forces. The density (mass per unit volume) of the samples has a direct impact on the hardness value. Figure 14 illustrates the Vickers-Hardness vs Density of the Pr^{3+} ions doped OFSbB glasses.

Using the Mackenzie and Makishima theory, the elastic moduli of the OFSbBPr glasses are computed [56, 57]. According to this theory, elastic moduli are expressed as a function of packing density (V_t), which is the volume density of binding energy. By comparing the actual ionic volume of the glass with the smallest theoretical volume taken up by ions, we can determine the dissociation energy (G_t) per unit volume. To determine Young's modulus of crystalline oxide, use the following equation.

$$E = 2\alpha U / r_0^3 \tag{14}$$

Here, α denotes the Madelung constant, U stands for the electrostatic attraction energy, and r_o stands for the interatomic distance. Achieving a significant Madelung constant because of the disordered glass structure. Makishima and Mackenzie multiplied the packing density factor multiplied the V_t by the G_t per unit volume of oxide [58]. As a result, Young's modulus (E) is expressed as follows:

$$E = 2G_tV_t = 2\frac{\rho}{M}\sum G_iX_i \sum V_iX_i \tag{15}$$

Here, packing density and dissociation energy is expressed as,

$$V_t = (\rho/M)\,\Sigma_i V_i X_i \tag{16}$$

Table 8 Radiative lifetime (τ_R)(μs) and Radiative experimental lifetime (τ_{exp}) (μs), quantum efficiency (η) and non-radiative decay rates (W_{NR}) for the $^3P_0 \rightarrow {}^3H_4$ and $^1D_2 \rightarrow {}^3H_4$ emission transitions for the Pr^{3+} ions doped OFSbB glasses

Glass code	$^3P_0 \rightarrow {}^3H_4$				$^1D_2 \rightarrow {}^3H_4$			
	τ_{exp}	τ_R	η	W_{NR}	τ_{exp}	τ_R	η	W_{NR}
OFSbBPr0.1	144	352	40.909	4104	159	253	62.845	2337
OFSbBPr0.5	135	218	61.926	2820	146	211	69.194	2110
OFSbBPr1.0	129	154	83.766	1258	124	159	77.987	7530
OFSbBPr1.5	90	128	70.312	3299	111	147	75.51	2206
OFSbBPr2.0	45	105	42.857	12,698	95	139	68.345	3332

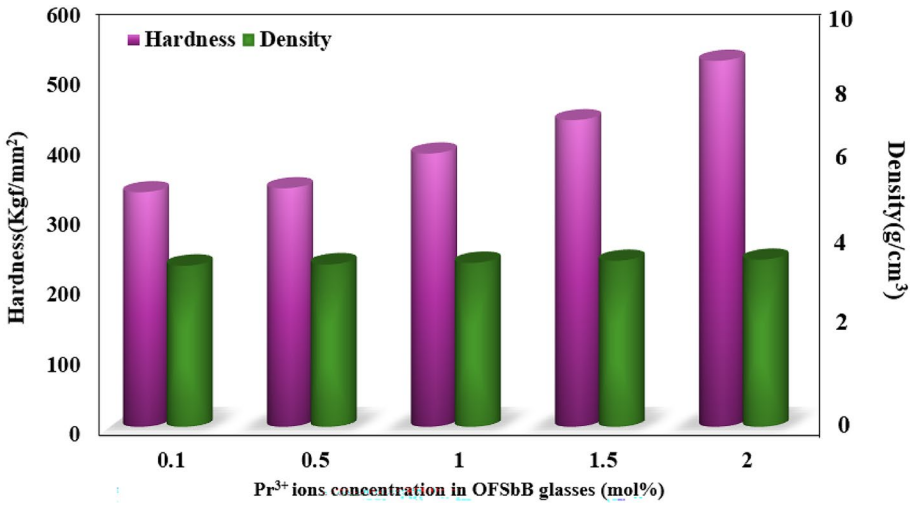


Fig. 14 The Vickers-Hardness vs Density of Pr^{3+} ions doped OFSbB glasses

$$G_i = \left(\rho_i/M_i\right)\left[X\Delta H_f(M,\ gas) + Y\Delta H_f(O,\ gas) - \Delta H_f\left(M_xO_y crystal - (X+Y)\ RT\right]\right. \quad (17)$$

Here, ϱ_i stands for the compound's density, M_i for the molecular mass of that molecule, and X_i for the molar fraction of the i_{th} component. Furthermore, T denotes the ambient temperature, which is determined to be 273 + 25 °C = 298 K, R stands for the gas constant, which is 8.314 kJ/mol, and ΔH_f is the molar heat of formation obtained from the standard states of the oxide and gaseous atoms.

By determining enthalpies of each component (using Hess's cycle), the dissociation energy of the B_2O_3, Sb_2O_3, Al_2O_3, NaF, LiF, Pr_6O_{11} are obtained.

The first step in determining the V_t factor of an oxide M_x,O_y glass is to calculate the V_i factor, which reflects the individual density factor for each oxide, which is related to the packing density.

$$V_i = \left(4\pi N_A/3\right)\left(xr_m^3 + yr_o^3\right) \quad (18)$$

Therefore, the total packing density factor is written as a function of r_m, r_o and N_A. Where r_m stands for the cation's ionic radius, N_A for Avogadro's number and N for the anion's ionic radius.

$$V_t = \left(\frac{\rho}{M}\right)\sum V_i x_i \quad (19)$$

To determine bulk modulus(K), use the following formula,

$$K = 1.2V_tE = 2.4V_t^2G \quad (20)$$

The shear modulus(S) is calculated using the following equation,

$$S = 3EK/(9K - E) \quad (21)$$

The Poisson's ratio(μ) is determined by [29, 59, 60],

$$\mu = \frac{E}{2S} - 1 \quad (22)$$

Figure 15 represents the variation of the V_t and G_t of the Pr^{3+} ions doped OFSbB glasses. The V_t values are increasing from 0.417 to 0.487 cm^{-3}. This indicates increased material density. It also leads to increased material compactness and rigidity and the G_t of the samples decreases from 31.97 to 31.54 kJ/cm^3 due to the replacing the LiF with Pr_6O_{11}. Figure 16 and Table 9 represents the elastic moduli parameters (S, K, E and μ) and observed that values are increasing as dopant increases. The inclusion of dopants into the glass network causes the glass to become more rigid. Furthermore, the presence of dopants increases the elastic modulus, indicating a rise in the materials strength. With increased interaction, the samples exhibit a higher degree of atom bonding inside the glass network, resulting in better strain resistance [61, 62].

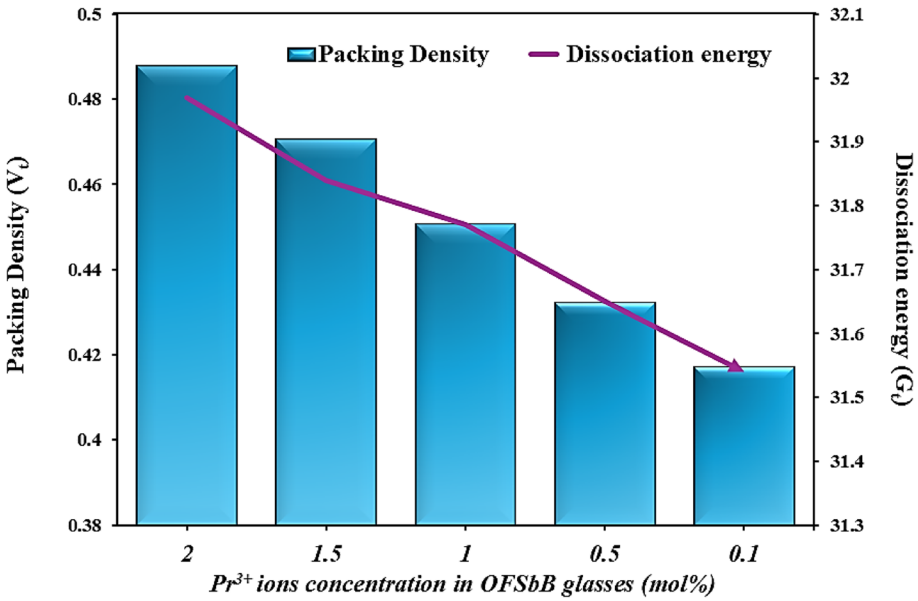


Fig. 15 The variation of the Packing density (V_t) and Dissociation energy (G_t) of the Pr^{3+} ions doped OFSbB glasses

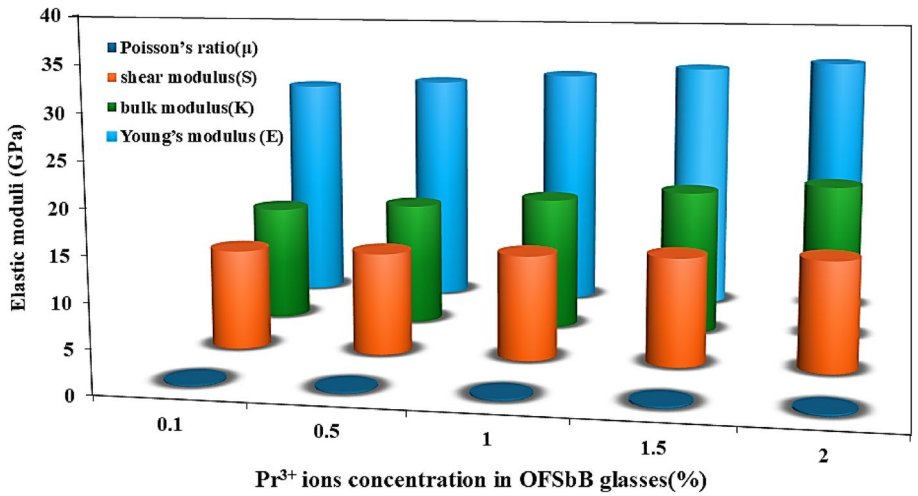


Fig. 16 The elastic moduli theoretically (according to Mackenzie and Makishmia model) of the Pr^{3+} ions doped OFSbB glasses

Table 9 Calculated Young's modulus (E)(GPa), bulk modulus (K)(GPa), Shear modulus (S)(GPa) and Poisson's rate (μ) of Pr^{3+} ions doped OFSbB glasses

Glass code	E	K	S	μ
OFSbBPr0.1	26.669	13.348	11.426	0.1670
OFSbBPr0.5	27.509	14.261	11.671	0.1785
OFSbBPr1.0	28.631	15.481	12.012	0.1917
OFSbBPr1.5	29.789	16.822	12.361	0.2048
OFSbBPr2.0	30.764	18.004	12.657	0.2152

4 Conclusion

The optical and luminous properties of Pr^{3+}: Oxy-fluoro Antimony Borate glasses doped at different concentrations were synthesised and analysed. The EDX spectra confirms the existence of all elements. The Vickers hardness test demonstrates that glasses possess sufficient strength to withstand a force of 500 g and we observed high elastic moduli indicate high mechanical stability of glasses. By analyzing the optical absorption spectra, we were able to find the optical bandgap values. The addition of Pr^{3+} ions alter the configuration of metal ions and ligands in the glass matrix, resulting in a decrease in measured

values. The prepared glass samples have a predominant trend of $\Omega_6 > \Omega_2 > \Omega_4$. A higher value of Ω_6 indicates that the Pr^{3+} ions in the OFSbBPr1.0 glass have a higher degree of symmetry and a higher degree of rigidity in their Pr–O bonds. The prepared glasses, emission spectra are excited at a wavelength of 445 nm and exhibit nine bands ranging from the initial level $^3H_4$ to the higher state $^3P_2$. In bluish-green and strong reddish-orange areas, the emission transitions with increased intensity were noticed at 490 nm ($^3P_0$ to $^3H_4$) and 610 nm ($^1D_2$ to $^3H_4$,$^3P_0 \rightarrow {}^3H_6$). Out of all the synthesized glass samples, OFSbBPr1.0 glass has high values of A_R, (σ_{SE} x τ_R), (σ_{SE} x $\Delta\lambda_p$) and great luminescence efficiency which allows them to release the light when exposed to UV or visible light, which suggests for the laser and display applications.

Author contributions

Ch. Pravallika: Conceptualization, Formal analysis, writing & editing, Investigation, K. Swapna: Supervision, writing & editing, Investigation and writing –review, Sk. Mahamuda: Formal analysis, writing editing, review, M. Venkateswarlu: Formal analysis, writing editing, review, A.S.Rao: Formal analysis, writing editing, review.

Funding

The authors K.Swapna and Sk. Mahamuda,thankful to DST & DST-SERB for sanctioning FIST (SR/FST/PS-1/2018/35), PURSE project (SR/PURSE/2023/196) and CRG Project (CRG/2022/003549), also thankful to the management of Koneru Lakshmaiah Education Foundation for the Full-Time Fellowship awarded to the Research Scholar, Ch. Pravallika (Regd. No: 2102520001) for her doctoral work.

Data availability

The research data that support the findings of this study are available on request from the corresponding author Dr.K.Swapna. The data are not publicly available due to restrictions, e.g., their containing information that could compromise the privacy of research participants.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. M. Shoaib, I. Khan, G. Rooh, S.M. Wabaidur, M.A. Islam, N. Chanthima, S. Kothan, I. Ullah, A. Ahad, J. Kaewkhao, J. Lumin. **247**, 118884 (2022)
2. H.J. Lozykowski, W.M. Jadwisienczak, I. Brown, J. Appl. Phys. **88**, 210 (2000)
3. Y.Y. Zhang, B.J. Chen, E.Y.B. Pun, H. Lin, Physica B Cond. **404**, 1132 (2009)
4. I. Khan, G. Rooh, R. Rajaramakrishna, N. Srisittipokakun, H.J. Kim, S. Kothan, J. Kaewkhao, K. Kirdsiri, J. Lumin. **222**, 117136 (2020)
5. K. Swapna, S. Mahamuda, A. Srinivasa Rao, S. Shakya, T. Sasikala, D. Haranath, G. Vijaya Prakash, Spectrochim. Acta A Mol. Biomol. Spectrosc. **125**, 53 (2014)
6. S.K. Mahamuda, K. Swapna, M. Venkateswarlu, A. Srinivasa Rao, S. Shakya, G. Vijaya Prakash, J. Lumin. **154**, 410 (2014)
7. R. Balda, A. Mendioroz, J. Fernández, M.A. Arriandiaga, L.S. Griscom, J.L. Adam, Opt Mater (Amst) **16**, 249 (2001)
8. A.A. Kaminskii, *Laser crystals* (Springer, Berlin, 1981)
9. M.V. Sasikumar, B. Rajeswara Reddy, S. Babu, A. Balakrishna, Y.C. Ratnakaram, J. Taibah Uni. Sci. **11**, 593 (2017)
10. J. Hegarty, D.L. Huber, W.M. Yen, Phys. Rev. B **25**, 5638 (1982)
11. N. Deopa, A.S. Rao, S.K. Mahamuda, M. Gupta, M. Jayasimhadri, D. Haranath, G.V. Prakash, J. Alloy. Compd. **708**, 911 (2017)
12. M.K. KomalPoojha, K.A. Naseer, P. Matheswaran, K. Marimuthu, E. El Shiekh, Opt Mater (Amst) **143**, 114289 (2023)
13. J. Dahiya, P. Phogat, A. Hooda, and S. Khasa, in 66TH DAE SOLID STATE PHYSICS SYMPOSIUM (AIP Publishing, 2024), p. 020065.
14. S. Mao, C. Ta, H. Wen, R.A. Talewar, Resul. Opt. **12**, 100429 (2023)
15. J. Anjaiah, Glass Phys. Chem. **49**, 584 (2023)
16. E.K. Abdel-Khalek, E.A. Mohamed, A. Ratep, S.M. Salem, I. Kashif, J. Non-Cryst, Solids **441**, 58 (2016)
17. N.T. El-Shamy, E.M. Mahrous, S.K. Alghamdi, M.J. Tommalieh, E.A. Rabiea, H.M. Abomostafa, D.E. Abulyazied,

A.S. Abouhaswa, Inorg. Chem. Commun. **164**, 112437 (2024)
18. P. Suthanthirakumar, Ch. Basavapoornima, K. Marimuthu, J. Lumin. **187**, 392 (2017)
19. D.E. Day, J. Non-Cryst. Solids **21**, 343 (1976)
20. J.O. Isard, J. Non-Cryst. Solids **1**, 235 (1969)
21. C.B. Annapurna Devi, S. Mahamuda, K. Swapna, M. Venkateswarlu, A. Srinivasa Rao, G. Vijaya Prakash, J. Non-Cryst. Solid. **498**, 345 (2018)
22. T. Satyanarayana, I.V. Kityk, M. Piasecki, P. Bragiel, M.G. Brik, Y. Gandhi, N. Veeraiah, J. Phys. Condens. Matter **21**, 245104 (2009)
23. M. Iezid, F. Goumeidane, A. Abidi, M. Poulain, M. Legouera, P. Syam Prasad, M. Środa, P. Venkateswara Rao, Opt. Mater. (Amst) **120**, 111422 (2021)
24. G. Lakshminarayana, R. Yang, M. Mao, J. Qiu, Opt. Mater. (Amst) **31**, 1506 (2009)
25. Ch. Basavapoornima, C.R. Kesavulu, T. Maheswari, W. Pecharapa, S.R. Depuru, C.K. Jayasankar, J. Lumin. **228**, 117585 (2020)
26. A. Balakrishna, S. Babu, V. Kumar, O.M. Ntwaeaborwa, Y.C. Ratnakaram, Spectrochim. Acta A Mol. Biomol. Spectrosc. **170**, 167 (2017)
27. J.A. Medeiros Neto, D.W. Hewak, H. Tate, J. Non-Cryst. Solid **183**, 201 (1995)
28. V.H. Rao, P.S. Prasad, K.S. Babu, Opt Mater (Amst) **101**, 109740 (2020)
29. K. R. Vighnesh, B. Ramya, S. Nimitha, A. Wagh, M. I. Sayyed, E. Sakar, H. A. Yakout, A. Dahshan, and S. D. Kamath, Elsevier BV (2020).
30. S. Mitra, S. Jana, J. Phys. Chem. Solids **85**, 245 (2015)
31. K.S.R.K. Reddy, K. Swapna, S.K. Mahamuda, M. Venkateswarulu, A.S. Rao, Opt. Mater. (Amst) (2021). https://doi.org/10.1016/j.optmat.2020.110615
32. J.A. Bearden, Rev. Mod. Phys. **39**, 78 (1967)
33. L. Vijayalakshmi, K.N. Kumar, P. Hwang, G. Kaur, Ceram. Int. **43**, 10881 (2017)
34. W.T. Carnall, P.R. Fields, K. Rajnak, J. Chem. Phys. **49**, 4424 (1968)
35. Sk. Mahamuda, K. Swapna, A. Srinivasa Rao, T. Sasikala, and L. Rama Moorthy, Physica B Condens 428, 36 (2013).
36. A. Lösche, Krist. Techn. 7, (1972).
37. J. Tauc, in Amorphous and Liquid Semiconductors, edited by J. Tauc (Springer US, Boston, MA, 1974), pp. 159–220.
38. A. Kumar, Ravina, Anu, V. Kumar, N. Deopa, R. Punia, and A. S. Rao, Mater. Lett.: X 19, 100210 (2023).
39. G. Lakshminarayana, K.M. Kaky, S.O. Baki, S. Ye, A. Lira, I.V. Kityk, M.A. Mahdi, J. Alloys Compd. **686**, 769 (2016)
40. M. Venkateswarlu, M. V. V. K. S. Prasad, K. Swapna, Sk. Mahamuda, A. S. Rao, A. M. Babu, and D. Haranath, Ceram. Int 40, 6261 (2014).
41. B. Burtan, J. Cisowski, Z. Mazurak, B. Jarzabek, M. Czaja, M. Reben, I. Grelowska, J. Non-Cryst, Solids **400**, 21 (2014)
42. K. Bhargavi, V. Sudarsan, M. G. Brik, M. Sundara Rao, Y. Gandhi, P. Nageswara Rao, and N. Veeraiah, J. Non-Cryst. Solids 362, 201 (2013).
43. B. Charfi, K. Damak, R. Maâlej, M. S. Alqahtani, K. I. Hussein, A. M. Alshehri, A. M. Hussain, B. Burtan-Gwizdala, M. Reben, and E. S. Yousef, Materials (Basel) **15**, (2022).
44. A. Herrera, C. Jacinto, A.R. Becerra, P.L. Franzen, N.M. Balzaretti, J. Lumin.Lumin. **180**, 341 (2016)
45. D. V. R. Murthy, B. C. Jamalaiah, T. Sasikala, L. Rama Moorthy, M. Jayasimhadri, K. Jang, H. S. Lee, S. S. Yi, and J. H. Jeong, Physica B Condens 405, 1095 (2010).
46. I. Pal, A. Agarwal, S. Sanghi, M.P. Aggarwal, J. Alloys Compd. **509**, 7625 (2011)
47. M.V.V. Kumar, K.R. Gopal, R.R. Reddy, G.V.L. Reddy, N.S. Hussain, B.C. Jamalaiah, J. Non-Cryst, Solids **364**, 20 (2013)
48. V. Naresh, B.S. Ham, J. Alloys Compd. **664**, 321 (2016)
49. B. Di Bartolo, in Radiationless Processes, edited by B. DiBartolo and V. Goldberg (Springer US, Boston, MA, 1980), pp. 39–87.
50. X. Wu, W.M. Dennis, W.M. Yen, Phys Rev. B Condens Matter **50**, 6589 (1994)
51. S. Babu, P. Rajput, Y.C. Ratnakaram, J. Mater. Sci. **51**, 8037 (2016)
52. F. Zhang, Z. Bi, A. Huang, Z. Xiao, J. Lumin.Lumin. **160**, 85 (2015)
53. B. C. Jamalaiah, J. Suresh Kumar, A. Mohan Babu, L. Rama Moorthy, K. Jang, H. S. Lee, M. Jayasimhadri, J. H. Jeong, and H. Choi, J. Lumin.129, 1023 (2009).
54. J.A. Duffy, M.D. Ingram, J. Non-Cryst, Solids **144**, 76 (1992)
55. V. Ravi Teja, M. Sreenivasulu, and V. K. Chavan, J Mater Sci: Mater Electron 35, 166 (2024).
56. A. Makishima, J.D. Mackenzie, J. Non-Cryst, Solids **17**, 147 (1975)
57. A. Makishima, J.D. Mackenzie, J. Non-Cryst, Solids **12**, 35 (1973)
58. I. Z. Hager, Springer Science and Business Media LLC (2002).
59. R. Laopaiboon, J. Laopaiboon, S. Pencharee, S. Nontachat, C. Bootjomchai, J. Alloys Compd. **666**, 292 (2016)

60. Zagazig University and A. Abd El-Moneim, (2019). Reference, Phys. Chem. Glasses: Eur. J. Glass Sci. Technol. B, 60: 146
61. N. Elkhoshkhany, E. Syala, E. Sayed Yousef, Resul. Phys. **16**, 102876 (2020)
62. R.A. Tafida, S. Thakur, M.Y. Onimisi, S. Garba, S.B. Adamu, I.G. Shitu, I.I. Lakin, Opt. Quant. Electron. **56**, 964 (2024)

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

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

International Journal of Machine Learning and Cybernetics
https://doi.org/10.1007/s13042-024-02455-2

ORIGINAL ARTICLE

KRAFS-ANet: A novel framework for EEG-based stress classification using channel selection and optimized ensemble stacking

Shikha Shikha[1,2] · Divyashikha Sethia[2,3] · S. Indu[1,2]

Received: 5 June 2024 / Accepted: 5 November 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
Mental stress poses a widespread societal challenge, impacting daily routines and contributing to severe health problems. The earlier studies have utilized Electroencephalograms (EEG) for stress classification; however, the computational demands of processing data from numerous channels often hinder the translation of these models to wearable devices. This paper proposes KRAFS-ANet, a novel framework designed for enhanced stress classification using EEG data on wearable devices. KRAFS-ANet framework incorporates two major novel components to achieve high accuracy with a lightweight design: (1) it strategically employs channel selection using Normal Mutual Information and Recursive Feature Elimination (NMI+RFE) to identify the most informative channels and (2) it uses ensemble stacking techniques integrate bagging K-Nearest Neighbour (KNN), bagging Random Forest (RF), and bagging Support Vector Machine (SVM) with an Artificial Neural Network (ANN) meta-classifier. The study conducts comprehensive experiments on the stress-based MAT dataset and further validates the framework on the stress-based SAM40 and anxiety-based DASPS datasets to demonstrate its effectiveness. KRAFS-ANet achieved the highest accuracies and F1-scores of 98.63% and 98.82% on the MAT dataset, 97.25% and 97.24% on the SAM40 dataset, and 94.92% and 95.15% on the DASPS dataset, respectively. This framework advances the practical application of EEG-based stress detection for portable devices such as wearables and smartphones. It enables real-time monitoring and interventions to enhance mental health in daily life, thus proving its efficacy in real-world scenarios.

Keywords Mental stress · Electroencephalography · Artificial Neural Network · Ensemble stacking · Channel selection · Machine learning · Deep learning

1 Introduction

Mental stress is a widespread problem affecting a large part of the population, and it is a major cause of several serious health issues. Researchers and psychologists offer an early diagnosis of mental stress to prevent its detrimental effects or before it progresses to chronic levels. The primary method for assessing mental stress is subjective, including self-report questionnaires such as the Perceived Stress Scale [28]. However, subjective procedures are not feasible for real-life applications because they are time-consuming, inconvenient, and frequently disregarded until related health issues become apparent. In contrast, objective methods, such as Electroencephalography (EEG), are considered promising approaches for creating practical applications that allow individuals to assess their mental state without expert involvement [17, 26]. Developing such an application requires an effective EEG analysis method, such as using only the selected channels and features most relevant to the mental state task. This approach results in enhanced classification performance, reduced system complexity, and increased diagnoses' convertibility [3, 25].

Each channel may contain multiple features when dealing with numerous EEG channels, significantly increasing

✉ Divyashikha Sethia
divyashikha@dtu.ac.in

Shikha Shikha
shikha_2k19phdco502@dtu.ac.in

S. Indu
s.indu@dtu.ac.in

[1] Computer Science and Engineering, Delhi Technological University, Bawana Road, New Delhi 110042, Delhi, India

[2] Software Engineering, Delhi Technological University, Bawana Road, New Delhi 110042, Delhi, India

[3] Electronics and Communication Engineering, Delhi Technological University, Bawana Road, New Delhi 110042, Delhi, India

Published online: 19 November 2024

the feature vector space. Various feature selection techniques has been adapted to optimize the number of features without compromising EEG channels [37, 38]. However, implementing feature selection solely in multi-EEG channels proves advantageous for high accuracy in a laboratory setting. Nonetheless, its practical efficiency diminishes in home-based applications or daily use due to increased setup time and user discomfort. Consequently, identifying relevant channels for source localization becomes paramount for developing a more user-friendly and applicable stress assessment device.

In various studies, channel selection relies on neuroscience expertise in stress classification, correlating data from specific brain regions with distinct tasks. For instance, the prefrontal region is associated with cognitive processes, while central lobes are linked to motor imagery tasks in Brain-Computer Interface (BCI) systems [41]. Various EEG channel selection methods, including ReliefF [62], Maximum Relevance Minimum Redundancy (mRMR) [60], T-test, Analysis of Variance (ANOVA) [35], and Normalized Mutual Information (NMI) [53] have been introduced in emotion classification to reduce channels while maintaining classification performance. However, the channel selection approach in stress classification is still under exploration.

In the mental stress assessment domain, the existing literature, including studies on widely recognized datasets such as Mental Arithmetic Task (MAT) [63], shows a notable emphasis on achieving high accuracy, often reaching up to 99% [40, 51]. Despite these advancements, translating high-accuracy models to real-world wearable device applications presents significant challenges. Advanced models like Convolutional Neural Networks (CNNs) [4] and Long Short-Term Memory networks (LSTM), including Bidirectional LSTMs [14], are capable of effectively handling complex patterns in EEG data; however their practical deployment is often constrained by high computational demands and extensive resource requirements. These limitations pose a challenge for wearable technologies, necessitating user comfort and device autonomy and demanding secure and prompt data handling without reliance on external processing resources.

To address these challenges, ensemble techniques such as bagging and stacking have demonstrated significant improvements in accuracy and robustness by leveraging multiple predictive models across various domains [20, 24]. Inspired by the success of these advanced modelling techniques, this study proposes a new optimized framework called KRAFS-ANet, which utilizes ensemble stacking. KRAFS-ANet integrates bagging K-Nearest Neighbour (KNN), bagging Random Forest (RF), and bagging Support Vector Machine (SVM) with Artificial Neural Network (ANN) meta-classifiers. This framework aims to streamline and optimize the model by employing a strategic channel selection approach using Normal Mutual Information (NMI) and Recursive Feature Elimination (RFE).

The study comprehensively explores bagging and stacking methodologies in the context of stress classification through experimentation when designing the proposed KRAFS-ANet framework. It utilizes machine learning and its prediction probability-based features to further improve deep learning model performance. This study observes that existing studies have achieved impressive results individually, but combining the strengths of multiple models and utilizing prediction probability-based features can further elevate classification performance. To the best of our knowledge, no previous research has explored ensemble stacking with channel selection specifically for stress classification using EEG.

The key contributions of this work are as follows:

1. To conduct a comprehensive experiment to design a model, exploring machine and deep learning, bagging, and various ensemble stacking techniques on the EEG-based MAT [63] dataset.
2. To propose the KRAFS-ANet framework, an optimized ensemble stacking approach for automated stress classification using EEG signals, incorporating an optimal channel selection process with the NMI+RFE approach.
3. To compare previous state-of-the-art work and validate the performance of the proposed framework on EEG-based SAM40 [19] stress dataset and DASPS [8] anxiety dataset.

1.1 Roadmap

The structure of this paper is as follows. Section 2 presents the related work. Section 3 describes the experimental methodology employed in the study, which includes a detailed description of the data acquisition, preprocessing, and the proposed KRAFS-ANet framework design for developing a stress detection system. Section 4 presents the experimental results, and section 5 comprehensively discusses the results and comparison with existing studies. Section 6 presents the limitations, and lastly, Sect. 7 explains the concluding remarks on this paper and potential future research directions.

2 Related work

Numerous studies have leveraged the publicly available EEG-based MAT dataset [63] from Physionet, consisting of 36 participants and recorded using the Neurocom EEG 23-channel system in.edf format. Sharma et al. [41] utilized the MAT dataset [63] and extracted entropy-based EEG features. The authors used channel recordings from the brain's

frontal and anterior frontal lobes and employed the F1-score feature selection approach. Further, they employed an evolutionary-inspired approach for SVM parameter optimization and achieved the highest accuracy of 94.01% and 97.25% with 4-second and 8-second window sizes, respectively.

Furthermore, Attallah et al. [6] utilized the MAT dataset [63] and achieved 99.7% accuracy with 23 channels using cubic SVM. In another study, Xiong et al. [59] employed an evolutionary feature selection algorithm for cognitive load classification of EEG and Electrocardiography (ECG) signal features. The authors obtained the highest accuracy of 97.2% using SVM. Similarly, Salankar et al. [39] incorporated EEG and ECG signal time domain and frequency domain features. The authors employed Multilayer perceptron (MLPNN) for multichannel classification and achieved the highest accuracy of 82.57%.

More recently, Yedukondalu et al. [61] utilized MAT [63] dataset for cognitive load classification, in which the authors extracted entropy-based features from the Intrinsic Mode Functions (IMFs). They employed the $C - SSA + BHHO + KNN$ technique and achieved a classification accuracy of 95.28%.

Comparatively, Malviya et al. [32] applied Discrete wavelet transform (DWT) to derive five physiological EEG frequency bands and 20 channels. The authors further input the preprocessed data into CNN to extract relevant features. Further, they stacked BiLSTM with CNN and achieved the highest accuracy of 99.20%. However, the computational demands associated with these methods may limit their scalability for real-time applications and large datasets.

Fatimah et al. [16] presented a methodology for automatic stress detection using a single-channel EEG signal. The authors utilized SVM, Decision Tree (DT), and Quadratic Discriminant Analysis (QDA) and achieved the highest accuracy of 98.6% using QDA.

Furthermore, several studies have utilized private datasets for stress classification using EEG signals. Mustafa et al. [36] employed a wrapper method for feature selection and used several machine learning models, including Naive Bayes (NB), SVM, Logistic Regression (LR), RF, KNN, Bagging, MLP, and Adaboost for stress classification. The authors utilized the MAT task and achieved the highest classification accuracy of 91.52% using Adaboost. Similarly, Hafeez et al. [21] manually selected 10 channels for stress classification. The authors formed brainwave images from theta, alpha, and beta frequency bands and achieved the highest accuracy of 90.64% using CNN. Further, Kaminska et al. [27], and Akella et al. [1] utilized all 32 channels and achieved the highest accuracy of 96.42% and 87.5% using SVM and CNN, respectively.

Table 1 compares the reviewed literature, where most of the studies collectively contribute to the field of mental stress classification, showcasing diverse methodologies and achieving remarkable accuracies. For instance, Malviya et al. [32] achieved nearly 100% accuracy using MLPNN. However, to the best of our knowledge, no research has been conducted on classifying stress using channel selection and ensemble stacking techniques, which can provide lightweight stress classification frameworks for wearable devices. On the other hand, various studies have been conducted based on neuroscience expertise for channel selection in stress classification. Therefore, existing channel selection may limit scalability and introduce subjectivity, impacting adaptability for broader stress detection applications. Accordingly, this study aims to explore the MAT [63] dataset to design an efficient framework with a channel selection algorithm by comprehensive experiments involving Machine Learning (ML), Deep Learning (DL), bagging, and various ensemble stacking techniques. Additionally, the paper further validates the efficacy of the proposed framework using the EEG-based SAM40 [19] and DASPS [8] dataset.

3 Materials and methods

The study aims to design a novel EEG-based KRAFS-ANet framework for automated stress classification, as shown in Fig 1. Table 1 compares the reviewed literature, where most of the studies collectively contribute to the field of mental stress classification, showcasing diverse methodologies and achieving remarkable accuracies. For instance, Malviya et al. [32] achieved nearly 100% accuracy using MLPNN. However, to the best of our knowledge, no research has been conducted on classifying stress using channel selection and ensemble stacking techniques, which can provide lightweight stress classification frameworks for wearable devices. On the other hand, various studies have been conducted based on neuroscience expertise for channel selection in stress classification. Therefore, existing channel selection may limit scalability and introduce subjectivity, impacting adaptability for broader stress detection applications. Accordingly, this study aims to explore the MAT [63] dataset to design an efficient framework with a channel selection algorithm by comprehensive experiments involving Machine Learning (ML), Deep Learning (DL), bagging, and various ensemble stacking techniques. Additionally, the paper further validates the efficacy of the proposed framework using the EEG-based SAM40 [19] and DASPS [8] dataset.

3.1 Dataset description

This study utilizes the MAT [63] dataset as the primary dataset and further validates the framework on the SAM40 [19] dataset as a secondary dataset, as described below:

Table 1 Existing studies on Stress Classification using EEG [63]

Author	Dataset	Channel selection	Features	Feature selection	Algorithm	Accuracy (%)	Limitations
Sharma et al. [41]	MAT [63]	Frontal and anterior frontal lobe channels		F1-score	WOA based SVM tuning	97.25	Focuses on specific channels
Attallah et al. [6]	MAT [63]	–	Frequency Domain and Statistical	PCA	RF, KNN, Cubic SVM, SVM, LDA	LDA: 98.6, SVM: 94.5, Cubic SVM: 99.7, KNN: 99, RF: 98.91	Reported accuracies have not been validated on another dataset.
Xiong et al. [59]	MAT [63]	–	Frequency Domain	SBS and PSO	Quadratic SVM, KNN, DT	SVM_quadratic: 80.6, SVM_rbf: 80.6, KNN: 91.7, DT: 86.1	Study is limited to machine learning algorithm.
Salankar et al. [39]	MAT [63]	–	Time Domain and Frequency Domain	–	MLPNN	MLPNN: 82.57	Limited insights into the relevance of individual features.
Yedukondalu et al. [61]	MAT [63]	–	Entropy Features	Evolutionary Algorithms	KNN, SVM	BGWO: 94.96, BHHO: 96.88, BDE: 96.42	Study is limited to machine learning algorithm.
Fatimah et al. [16]	MAT [63]	Analyses for each channel	Entropy Features	Kruskal Wallis Test	LR, SVM, KNN	SVM: 98.6, LR: 95.8, KNN: 93.1	Broader range of features could be explored.
Mathur et al. [33]	MAT [63]	–	Graph Features	–	GSP, RBF, Pearson	GSP: 98.6, RBF: 88.7, PC: 93	Limited sample size of secondary dataset. - Authors did not explore ensemble methods.
Fatimah et al. [16]	MAT [63]	Analyses for each channel	Frequency Domain	–	SVM, DT, QDA	QDA: 98.6	Single-channel bias limits its generalizability
Malviya et al. [32]	MAT [63]	19 channels	DWT based Frequency Domain	CNN	BILSTM	*CNN + BILSTM*: 99.20	No validation on secondary dataset. Computationally expensive for wearables.
Mustafa et al. [36]	Private	–	Time and Frequency Domain	Wrapper Method	NB, SVM, LR, RF, KNN, Bagging, MLP, Adaboost	Adaboost: 91.52	No validation on secondary dataset.
Hafeez et al. [21]	Private	Manually Selected	Frequency Domain	–	LSTM, RESNET, CNN	LSTM: 70.67, ResNet-50: 73.44, CNN-BWI 90.64	Less sample size
Kaminska et al. [27]	Private	32	Frequency Domain	–	KNN, RF, SVM, MLPNN, CNN	KNN: 91.7, RF: 92.86, SVM: 96.42,MLP: 96.42, CNN:87.5	Considered only band-power FD features. Could have explored other FD features.
Akella et al. [27]	Private	32	Frequency Domain	–	SVM, Adaboost, LDA, ridge classifiers, RF, KNN and DBF	SVM: 83, ADB: 53, LDA: 75, RC: 51, RF: 81, KNN: 85, DBF: 13, CNN:87.5	Ensemble stacking could be utilized to improve the accuracy.

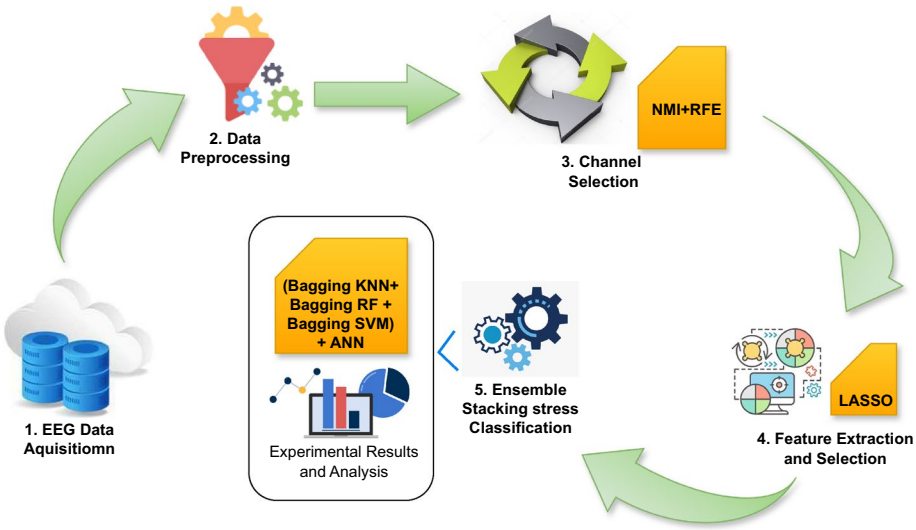


Fig. 1 KRAFS-ANet Framework: EEG-based Stress Classification Methodology with Channel Selection and Ensemble Stacking

1. **MAT** [63] **dataset:** The National Technical University of Ukraine provides the MAT dataset, encompassing EEG recordings from 36 students engaged in mental arithmetic tasks. The arithmetic activities involve serial subtraction of two numbers, and the EEG acquisition involves 23 channels. Each recording includes artefact-free EEG segments, comprising resting state data with closed eyes and EEG data during mental counting. The study employs data from 20 electrodes provided in a.mat file, sampled at 500Hz and arranged in a 10–20 pattern across the scalp.
2. **SAM40** [19] **dataset:** The dataset comprises EEG data from 40 subjects where 14 females and 26 males, and the mean age is 21.5 years. EEG recordings were obtained during three tasks involving the Stroop colour-word test, solving arithmetic questions, identifying symmetric mirror images, and a state of relaxation. The primary objective of this experiment is to monitor short-term stress responses elicited in individuals during the performance of the mentioned cognitive tasks. Individual tasks lasted for 25 s and were repeated for three trials. EEG data is recorded using a 32-channel Emotiv Epoc Flex gel kit and segmented into non-overlapping epochs of 25 s based on the various tasks performed by the subjects.
3. **DASPS** [8] **dataset:** The dataset consists of raw EEG data in.edf format collected from 23 subjects (10 men and 13 women). EEG signals were recorded using a 14-channel Emotiv EPOC wireless EEG headset at a sampling rate of 128 Hz. The dataset also includes preprocessed data in.mat format. During the recording sessions, participants were exposed to six distinct scenarios while keeping their eyes closed and minimizing gestures. Each scenario comprised two segments: a 15-second period in which a psychotherapist described the anxiety-inducing scenario, followed by a 15-second period where participants recalled the same scenario. The entire experiment included a 5-minute Hamilton Pre-Stimuli phase for each participant, followed by the recitation and recall of all six scenarios. Upon completion, participants evaluated their emotional responses, including valence and arousal, using the Self-Assessment Manikin (SAM).

Table 2 shows the summary of the dataset details. Furthermore, to preprocess the obtained EEG signals, the study incorporates Independent Component Analysis (ICA) with the FastICA technique, eliminating physiological artefacts. In addition to this, the *MNE 4*[1] Python package further preprocesses the EEG signals. It involves the application of a 5th-order Butterworth Bandpass filter to the raw EEG signal, focusing on a frequency range of 4–44 Hz. This step enhances data quality by eliminating unwanted frequencies. Additionally, this study uses the Synthetic Minority Oversampling Technique (SMOTE) to mitigate potential class imbalance effects, ensuring a more balanced dataset and avoiding biases in the model's performance assessment.

3.2 Channel selection

Brain signal acquisition through multiple EEG channels allows for a high spatial resolution and temporal precision by capturing signals from various brain locations. However, reducing the number of channels is crucial to

[1] https://mne.tools/stable/generated/mne.preprocessing.ICA.html.

Table 2 Stress and Anxiety Dataset Details

Dataset	Total EEG Channels	Sub.	Frequency Rate	Task	Classes
MAT [63]	20	36	500 Hz	Mental Arithmetic Task	Stress, Non Stress
SAM40 [19]	32	40	128 Hz	Mental Arithmetic Task Stroop Task Mirror Image Task	Stress, Non Stress
DASPS [8]	14	23	128 Hz	Face to face stimuli: Recitation of anxiety-inducing situation by psychotherapists and recall of situation by participants.	Anxiety, Non Anxiety

simplifying the setup, minimizing discomfort for subjects, and enhancing the system's efficiency for real-world applications. While reducing channels can decrease system complexity, more importantly, it facilitates quicker and more comfortable deployments, particularly in commercial wearable devices designed for continuous mental stress monitoring.

Thus, the approach to channel selection is important in managing the high dimensionality of feature vector spaces, thereby improving the system's performance. These enhancements are pivotal for developing wearable devices that offer robust diagnostic capabilities and accurate treatment options for mental stress [18, 42]. This study utilizes Normalized Mutual Information (NMI) to uncover intricate, non-linear relationships between EEG channels and stress labels, providing a comprehensive measure of information sharing. Furthermore, NMI channel selection has also proven effective in emotional classification studies [53], highlighting its capability to capture complex interactions within EEG data. Furthermore, Recursive Feature Elimination (RFE) is used to refine the channel selection process iteratively, reducing redundancy and optimizing the model for better generalization by retaining the most informative channels.

3.2.1 Normalized mutual information (NMI)

Mutual Information (MI) quantifies the degree of mutual dependence between two variables or the amount of information shared. Let us consider a discrete random variable X, which has distinct states $x1$, $x2$, ..., xn with corresponding probability distribution $px(x1)$, $px(x2)$, ..., $px(xn)$. By Shannon's entropy concept, Eq. 1 defines the information entropy of the random variable X:

$$H(X) = -\sum_{i=1}^{n} p_x(x_i) log_2 p_x(x_i) \tag{1}$$

Similarly, Eq. 2 defines the information entropy of random Y:

$$H(Y) = -\sum_{i=1}^{n} p_y(y_i) log_2 p_y(y_i) \tag{2}$$

Here, $p_x(x_i)$ denotes the probability distribution function for the i^{th} random state of X, and $p_y(y_i)$ signifies the probability distribution for the i^{th} random state of Y. Equation 3 illustrates the joint Shannon's entropy of X and Y:

$$H(X,Y) = -\sum_{i=1}^{n} p_{x,y}(x_i, y_i) log_2 p_{x,y}(x_i, y_i) \tag{3}$$

Where $p_{x,y}(x_i, y_i)$ stands for the joint probability function of X and Y. Equation 4 expresses the mutual information between two channel signals [10]:

$$MI(X,Y) = H(X) + H(Y) - H(X,Y), \tag{4}$$

MI normalizes the Normalized Mutual Information (NMI) to yield a value between 0 (indicating independence) and 1 (indicating strong dependence). This normalization is achieved through Eq. 5, which is formulated as:

$$NMI(X,Y) = \frac{MI(X,Y)}{H(X) + H(Y)} \tag{5}$$

In the equation, H (X) and H (Y) denote the marginal entropies, while H (X, Y) represents the joint entropy of variables X and Y.

MI possesses two key features that set it apart from other measures of dependency: its ability to quantify any relationship between variables and its consistency under spatial transformations [30].

In the case of channel selection, when the mutual information between two random variables is substantial, and the value is 1, which indicates a close relationship between them. Conversely, MI is zero when two random variables are entirely independent.

3.2.2 Recursive feature elimination (RFE)

RFE iteratively assigns ranks to each channel by eliminating those with the least impact on the SVM classifier [38, 50].

It computes the importance score of each channel and successively removes the channel with the lowest score from the subset [38]. This iterative elimination process continues until the desired number of channels is retained [12]. The Eq. 6 computes the channel importance score.

$$S(i) = \sum_{r=1}^{n} (k_{min} - k_{ir}) \tag{6}$$

$S(i)$ represents the channel score i, n denotes the total number of channels, k_{min} is the minimum rank value, and k_{ir} is the score of channel i in the set k. The channel with the lowest score, identified as the weakest, is systematically eliminated during each algorithm step.

3.2.3 NMI-RFE channel selection

The study considers preprocessed EEG signal data and calculates NMI for each EEG channel for the stress labels in the channel selection process. The channels with higher NMI values are considered more informative for stress classification. As a result, it arranges the normalized mutual information of all the channels in descending order. Further, it selects the top channels with high mutual information value as informative channels.

Additionally, this study calculates average classification accuracy using the Random Forest (RF) algorithm. RF is specifically selected due to its notable performance during preliminary experiments without channel selection, where it has achieved the highest accuracy, as shown in Table 5. Figure 2 illustrates a marginal alteration in the average classification accuracy for the MAT [63], SAM40 [19], and DASPS [8] datasets based on the number of channels. The findings indicate that selecting twelve channels for the MAT [63] dataset, eighteen channels for the SAM40 [19] dataset, and ten channels for the DASPS [8] dataset would be optimal. The study further refines the channel selection using the RFE algorithm across all three datasets. As a result, the final selection consists of nine channels for the MAT [63] dataset, twelve for the SAM40 [19] dataset, and six for the DASPS [8] dataset.

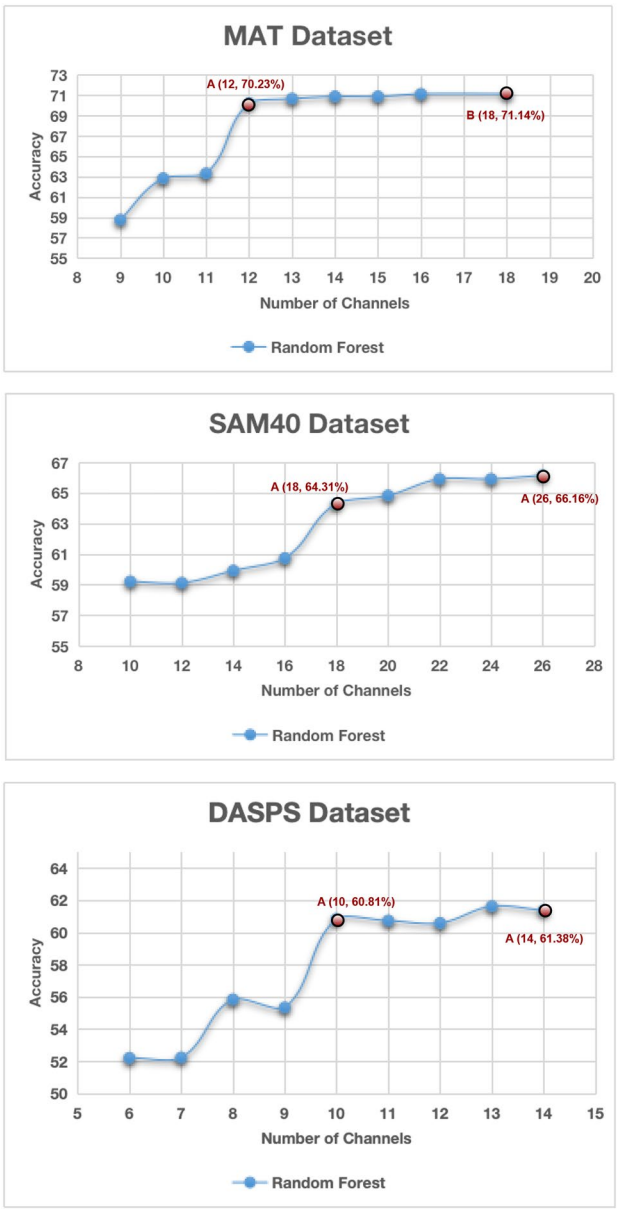


Fig. 2 Number of channels and accuracy rate using NMI in RF-based stress classification

3.3 Feature extraction

The study employs a diverse set of features from multiple domains to effectively discriminate between different levels of mental stress and improve classification accuracy. Time-domain features include Hjorth parameters related to activity, mobility, and complexity, Higuchi features, peak-to-peak amplitude, and statistical features. Frequency-domain features encompass the relative power of theta (4–8 Hz), alpha (8–12 Hz), beta (12–35 Hz), band power of theta/alpha, alpha/beta, and spectral entropy. These features are widely utilized in EEG studies and have demonstrated their efficacy in stress and anxiety classification [28, 43]. Table 3 provides a comprehensive overview of all features utilized in this study. Feature-based MinMax normalization is applied to the extracted features to maintain consistent scaling.

3.4 Bagging

Bagging, or Bootstrap Aggregating, is an ensemble method frequently employed in supervised machine learning for classification and regression tasks. Breiman et al. [56] in

Table 3 Summary of EEG extracted features employed on selected channels

Domain	Feature name	No of Features	Formula
	Hjorth Parameters	3	Activity = (y(i)) Mobility = $\sqrt{\frac{var'(y(i))}{Activity}}$ Complexity = $\sqrt{\frac{Mobility(\frac{(d(y))}{dt})}{Mobility(y(t))}}$
	Higuchi Fracxtal Dimension	1	HFD(k) = $\frac{log(N)}{log(\frac{N}{k.L(k)})}$
	Peak to peak amplitude	1	P.T.P = $peak_{high} - peak_{low}$
Time Domain	Mean	1	Mean = $\frac{1}{N}\sum_{i=1}^{N} x_i$
	Variance	1	Variance = $\frac{1}{N}\sum_{i=1}^{N}(x_i - Mean)^2$
	Standard Variation	1	Standard Deviation = $\sqrt{variance}$
	Skewness	1	Skewness = $\frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - Mean)^2)}{(\frac{1}{N}\sum_{i=1}^{N}(x_i - Mean)^2)^{\frac{3}{2}}}$
	Kurtosis	1	Kurtosis = $\frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - Mean)^4)}{(\frac{1}{N}\sum_{i=1}^{N}(x_i - Mean)^2)^2} - 3$
	Relative powers of theta (4–8 Hz), alpha (8–12 Hz), beta (12–35 Hz)	3	RP = $\frac{power(band)}{power(allbands)} * 100$
Frequency Domain	Band power ratio of $\frac{BP(theta)}{BP(alpha)}, \frac{BP(alpha)}{BP(beta)}$	2	ta = $\frac{thetapower}{alphapower}$ ab = $\frac{alphapower}{betapower}$
	Spectral entropy	1	SE = $-\sum_{f=4}^{f=N} P\bar{S}D(F)log(P\bar{S}D(F))$

1996 introduced this technique, which aims to enhance accuracy and mitigate overfitting. The method utilizes bootstrapping sampling, where diverse subsets of the training data are randomly generated with a replacement approach, thereby reducing variance. These subsets are then trained in parallel using multiple weak classifiers, which can be of the same or different types. Subsequently, it aggregates the predictions of each classifier and calculates their average, depending on the task type [47]. In regression cases, averaging the models' outcomes is termed *soft voting*. In contrast, in classification cases, it is referred to as *hard voting*, where the focus is on the class with the majority votes [48].

This study employs *hard voting*, as depicted in Eq. 7, where c denotes the class that achieves the majority vote, K is the number of classes, and f_{bagged} represents the bagged estimate function [22].

$$\hat{G}_{bag}(x) = argmax_k \hat{f}_{bag}(x) \tag{7}$$

Moreover, bagging is particularly selected for this study because it complements the use of strong learners like KNN and SVM. In contrast, boosting is effective for weak learners by focusing on misclassified instances. However, KNN and SVM do not benefit significantly from this process due to their inherently strong predictive power [55]. Instead, bagging reduces variance by training these classifiers on different subsets of data, allowing them to maintain their stability and leverage their strengths. Furthermore, bagging's parallel training approach is well-suited for real-time applications such as wearable devices, ensuring efficient and scalable model training.

3.5 Ensemble stacking

Stacking is an ensemble framework that employs a meta-model, where a new classifier combines predictions from multiple individual base learners to make predictions for the target variable. In 1992, Wolpert et al. [57] introduced stacking to reduce bias and variance, thereby enhancing predictive accuracy effectively. The stacking framework consists of two primary layers. The first layer involves multiple basic learning algorithms, while the second layer consists of the meta-learner, which combines the outputs. Initially, the dataset is partitioned into training and validation sets using k-fold cross-validation. Specifically, k-1 folds train the base models, and the remaining one-fold validates. Equation 8 presents how the basic learning algorithms produce a probability distribution for predicting each class, where c represents the class value, and $P^M(c_i \mid x)$ signifies the likelihood that x belongs to c_i [49].

$$P^M(x) = (P^M(c_1 \mid x), P^M(c_2 \mid x), ..., P^M(c_m \mid x)) \tag{8}$$

3.6 Experimental framework

Figure 3 outlines the visual representation of the proposed KRAFS-ANet framework. Subsequently, the following enumerates the various experimental phases:

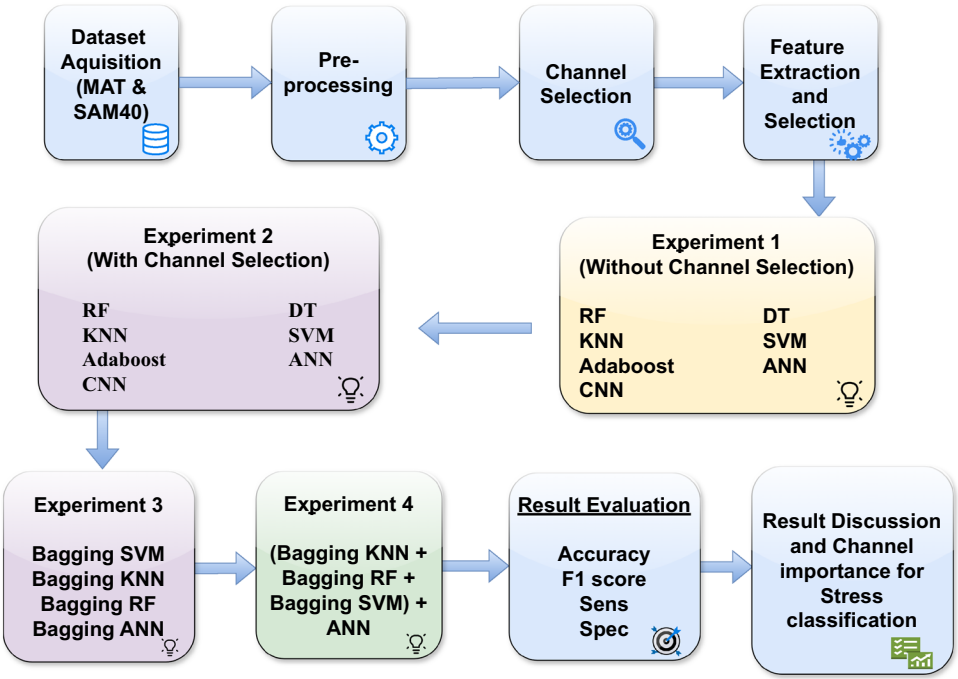


Fig. 3 KRAFS-ANet: Proposed Stress Classification Framework

1. The experimental phase starts with utilizing data from the publicly available MAT [63], SAM40 [19] and DASPS [8] datasets.
2. The raw dataset undergoes preprocessing to eliminate noise and artefacts.
3. Subsequently, channel selection is performed using the NMI+RFE algorithm on the preprocessed data. This approach selects 9 channels from the original 20 in the MAT [63] dataset, 12 channels from the initial 32 in the SAM40 [19] dataset, and 6 channels from the original 14 in the anxiety-based DASPS [8] dataset
4. After channel selection, the study applies a feature selection algorithm to both the selected and full set of channels across all datasets. It uses the Least Absolute Shrinkage and Selection Operator (LASSO) feature selection algorithm to reduce dimensions further. Further, the study employs data normalization using the MinMax scaler from the *sklearn* library.

   (a) *Experiment 1* includes employing machine and deep learning models on all channels' extracted and selected features. ML algorithms include SVM, K-NN, RF, DT, and Adaboost. In contrast, DL algorithms encompass Convolutional Neural Networks (CNN) and Artificial Neural Networks (ANN). The experiment uses GridSearchCV with stratified 10-fold cross-validation for model optimization.

   (b) *Experiment 2* involves applying ML and DL models specifically on the extracted and selected features of selected channels. Further, it compares the performance of machine and deep learning models with and without channel selection. The study observes superior performance on selected channels. RF, SVM, and KNN notably perform best in machine learning, while ANN stands out in deep learning. Therefore, this work continues experiments with RF, SVM, and KNN in ML and ANN in DL. It selects ANN due to its lightweight feature; being a universal approximator, it is less complex and achieves comparable accuracy rates even with smaller datasets. Further, it is suitable for deployment on resource-constrained devices while maintaining desired performance.

   (c) *Experiment 3* introduces bagging ensemble models, utilizing the best-performing models from both ML and DL on selected channels. Each classifier undergoes individual optimization through hyperparameter tuning on the selected channels via GridSearchCV, resulting in improved accuracy.

   (d) *Experiment 4* presents a novel ensemble Stacking model, combining bagging KNN, bagging RF, and bagging SVM. The ANN serves as a meta-classifier and undergoes optimization using GridSearchCV. Notably, the study maintains a constant random state variable of 42 throughout all experiments. The model achieves peak accuracies of 98.63% on the MAT [63] dataset, 97.25%

on the SAM40 [19] dataset, and 94.92% on the DASPS [8] dataset.

4 Experimental results

This section presents the results of each experiment designed to derive the proposed KRAFS-ANet framework for stress classification using EEG.

4.1 Performance measure

The study evaluates the experimental performance based on four key performance measures: accuracy (ACC), F1-score (F1), sensitivity (Sens), and specificity (Spec). The accuracy represents the overall correctness of the model predictions, considering both true positive (TP) and true negative (TN) predictions. The F1-score balances precision and recall, providing a unified metric that considers false positives and negatives. Sensitivity refers to the model's capacity to correctly identify all actual positive instances. In contrast, specificity gauges the model's proficiency in correctly identifying negative instances among all actual negative instances. Collectively, these measures offer a thorough evaluation and provide valuable insights into the models' performance in capturing stress-related patterns. Equations 9, 10, 11, and Eq. 12 represents the formula for calculation accuracy, F1-score, sensitivity and specificity, respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision \dot{+} Recall} \tag{10}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

Moreover, this study assessed the model's performance through 10-fold cross-validation and the confusion matrix. The 10-fold cross-validation involves dividing all samples into ten equal parts, considering each as a test dataset. This approach ensures the generation of reliable and stable models. Additionally, the confusion matrix commonly employed for evaluating the performance of classification algorithms provides an intuitive representation of classification errors.

4.2 Parameter optimization

The hyperparameter tuning process employed the GridSearchCV method to identify the optimal hyperparameters to maximize accuracy. This method systematically searches through a predefined grid of hyperparameter values, evaluating each combination through stratified 10-fold cross-validation. The selected combination that yields the best results is then finalized, contributing to improved model performance. Table 4 presents the selected hyperparameters for each algorithm.

4.3 Experiment 1: before channel selection

In experiment 1, Table 5 presents the performance of ML and DL classification algorithms on the MAT [63], SAM40 [19] and DASPS [8] datasets before applying channel selection. For the MAT [63] dataset, RF, KNN, and SVM demonstrate superior performance among machine learning algorithms, achieving accuracies of 86.81%, 77.23%, and 79.75%, respectively. Adaboost and DT obtain the lowest accuracy rates of 74.03% and 74.1%, respectively. Among deep learning models, ANN has the best performance, attaining an accuracy of 87.23%, and CNN provided an accuracy of 78.05%.

The study observes a similar trend in the SAM40 [19] dataset, where RF, KNN, and SVM outperform other machine learning algorithms, with accuracies of 84.44%,

Table 4 Selected Hyperparameters of Machine Learning and Deep Learning Models

Model	Selected Hyperparameters
RF	criterion: 'gini', max_depth: '20', max_features: 'sqrt', min_samples_leaf: '1', min_samples_split: '2', n_estimators: '100'
DT	criterion: 'gini', max_depth: '8', min_samples_leaf: '1', min_samples_split: '2'
KNN	metric: 'minkowski', n_neighbors: '5', p: '2', weights: 'uniform'
SVM	C: '1.0', degree: '3', gamma: 'scale', kernel: 'rbf'
Adaboost	learning_rate: '1.0', n_estimators: '150'
ANN	activation: 'tanh', learning_rate: '0.001', hidden_layer_sizes: '(100, 50)', optimizer: 'adam'
CNN	Filters: '32'; Activation Function:'Relu' and 'tanh'; pooling layer:'MaxPooling'; Dropout rate: '0.5'; Dense Layer Size: '32'

82.9%, and 76.5%, respectively. The deep learning algorithm, ANN, exhibited the best results with an accuracy of 85.58%. Adaboost and DT consistently exhibit the lowest performance among the models, achieving accuracies of 66.25% and 68.72%, respectively. Following closely, CNN obtains a moderate accuracy of 71.5% and an F1-score of 71.12%. Additionally, Fig. 4 illustrates the 10-fold cross-validation results of the test sets for three datasets.

For the DASPS [8] dataset, the study observes similar performance where SVM and RF achieve the highest accuracies of 84.98% and 83.17%, respectively, while ANN demonstrates comparable performance with an accuracy of 83.32%. KNN shows a moderate accuracy of 75.84%, while Adaboost and DT perform less effectively, obtaining an accuracy of 69.01% and 67.64%, respectively. CNN consistently provides a lower accuracy of 70.37%, similar to its performance in the other two datasets.

4.4 Experiment 2: after channel selection

In Experiment 2, the study applies the NMI-RFE algorithm to the preprocessed datasets for optimal channel selection.

Table 5 Experiment 1: Classification performance before channel selection

MAT [63] Dataset					SAM40 [19] Dataset				DASPS [8] Dataset			
Model	Acc. (%)	F1 (%)	Sens (%)	Spec (%)	Acc (%)	F1 (%)	Sens (%)	Spec (%)	Acc (%)	F1 (%)	Sens (%)	Spec (%)
RF	**86.81**	**85.58**	**87.69**	**83.56**	**84.44**	**84.78**	**85.65**	**83.26**	**83.17**	**82.84**	**82.42**	**84.90**
DT	74.1	73.46	73.5	75.36	68.72	69.01	69.12	68.3	67.64	66.83	66.34	68.02
KNN	**77.23**	**76.83**	**76.82**	**79.61**	**82.9**	**83.86**	**84.54**	**82.17**	**75.84**	**76.15**	**76.93**	**75.38**
SVM	**79.75**	**80.62**	**81.62**	**84.58**	**76.5**	**77.08**	**75.61**	**77.47**	**84.98**	**84.99**	**83.74**	**85.5**
Adaboost	74.03	76.48	79.45	71.59	66.25	65.49	66.49	66.02	69.01	70.93	70.34	69.73
ANN	**87.23**	**87.75**	**89.83**	**84.53**	**85.58**	**85.71**	**87.81**	**83.41**	**83.32**	**83.83**	**83.17**	**84.23**
CNN	78.05	79.98	80.02	78.03	71.5	71.12	69.59	73.54	70.37	70.09	69.4	72.62

Best performing model results are highlighted in bold

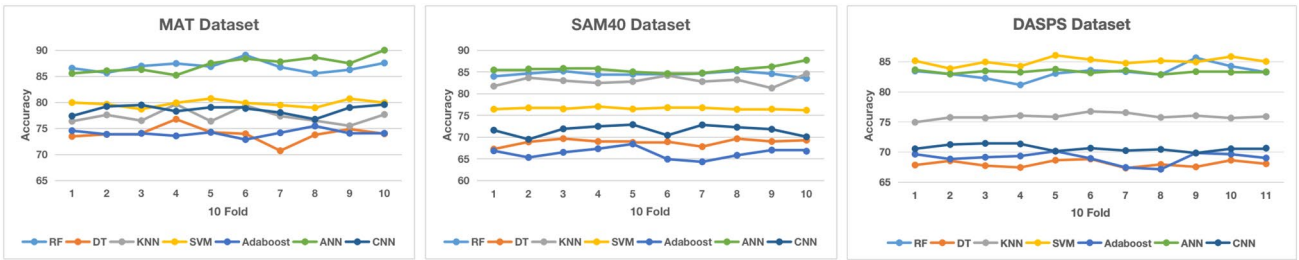


Fig. 4 Experiment 1: 10-fold cross-validation on classification algorithms before channel selection

Table 6 Experiment 2: Classification performance after channel selection

MAT [63] Dataset (9 channels)					SAM40 [19] Dataset (12 channels)				DASPS [8] Dataset (6 channels)			
Model	Acc (%)	F1 (%)	Sens (%)	Spec (%)	Acc (%)	F1 (%)	Sens (%)	Spec (%)	Acc (%)	F1 (%)	Sens (%)	Spec (%)
RF	**90.76**	**90.82**	**89.32**	**91.77**	**91**	**90.85**	**92.25**	**89.81**	**87.26**	**87.25**	**86.35**	**88.17**
DT	78.02	78.04	77.43	76.36	79.3	79.56	80.12	77	74.34	74.30	73.14	73.24
KNN	**82.55**	**82.83**	**82.96**	**80.44**	**84.48**	**85.69**	**88.9**	**80.57**	**81.29**	**81.18**	**82.79**	**81.80**
SVM	**82.8**	**82.78**	**81.83**	**81.9**	**89.4**	**89.4**	**90.34**	**97.82**	**85.67**	**85.53**	**83.65**	**85.68**
Adaboost	78.67	77.06	79.1	77.22	70.5	71.32	74.72	66.83	71.27	72.54	71.30	71.84
ANN	**86.85**	**86.92**	**87.92**	**85.77**	**90.72**	**90.62**	**91.76**	**89.72**	**85.37**	**85.36**	**84.38**	**85.87**
CNN	78.58	78.82	78.23	79.86	78.17	78.43	78.69	77.66	74.67	74.61	74.06	75.19

Best performing model results are highlighted in bold

As a result, the experiment selects nine and twelve optimal channels for the MAT [63] and SAM40 [19] datasets, respectively. For the DASPS [8] dataset, six optimal channels are selected. The EEG features are subsequently extracted, and the LASSO feature selection algorithm is applied to identify the most informative features for the selected channels.

Table 6 presents the classification performance of the ML and DL models after channel selection, highlighting the substantial improvement in accuracy. In the MAT [63] dataset, RF emerges as the top-performing model, achieving the highest accuracy and F1-score of 90.76% and 90.82%, respectively. ANN exhibits the second-highest performance, with an accuracy and F1-score of 86.85

Further, for the SAM40 [19] and DASPS [8] datasets, channel selection results in improved performance across most models. Random Forest achieves the highest accuracy in both datasets, increasing from 84.44% to 91.00% in SAM40 [19] and reaching 87.26% from 83.27% in the DASPS [8] dataset. KNN and SVM consistently show substantial performance improvements, while ANN stands out among the deep learning models, with accuracies of 90.72% in SAM40 [19] and 85.37% in the DASPS [8] dataset. Although DT and Adaboost show some improvement post-channel selection, they remain the lowest-performing models. Furthermore, Fig. 5 illustrates the 10-fold cross-validation results after channel selection approach.

Notably, ANN consistently outperforms CNN across all datasets in this study. This outcome can be attributed to the fact that both models utilize pre-extracted features rather than raw data. CNN is most effective when applied to raw data or images, where they can learn hierarchical spatial features through convolutional layers [7]. The utilisation of pre-extracted features limits CNN's ability to leverage its spatial feature extraction capabilities. Therefore, ANN is better suited for processing these features, allowing it to effectively capture and model the relevant information for stress classification. Consequently, ANN achieves higher accuracy compared to CNN in this research.

4.5 Experiment 3: proposed bagging model

In recent times, ensemble techniques have gained significant attention from researchers due to their ability to enhance the performance of conventional algorithms. Tables 5 and 6 consistently demonstrate that RF, SVM, KNN, and ANN yield optimal results for EEG-based stress classification across all datasets. Consequently, in experiment 3, the study builds bagging models for these high-performing algorithms identified in the initial experiments.

Table 7 presents the classification performance of the bagging models on the MAT [63], SAM40 [19], and DASPS [8] datasets. In the MAT [63] dataset, the bagging RF model exhibits superior performance with an accuracy of 92.25%, followed by bagging SVM and ANN with accuracies of 90.18% and 87.23%, respectively. Similarly, in the SAM40 [19] dataset, bagging SVM outperforms other models, yielding the highest accuracy of 92.85%, while bagging RF and ANN also provide comparable performances of 91.63% and 91.36%, respectively. For the DASPS [8] dataset, bagging RF achieves the highest accuracy of 90.45%, followed by bagging ANN with 88.82%. Bagging SVM and KNN also demonstrate noteworthy performance, with accuracies of

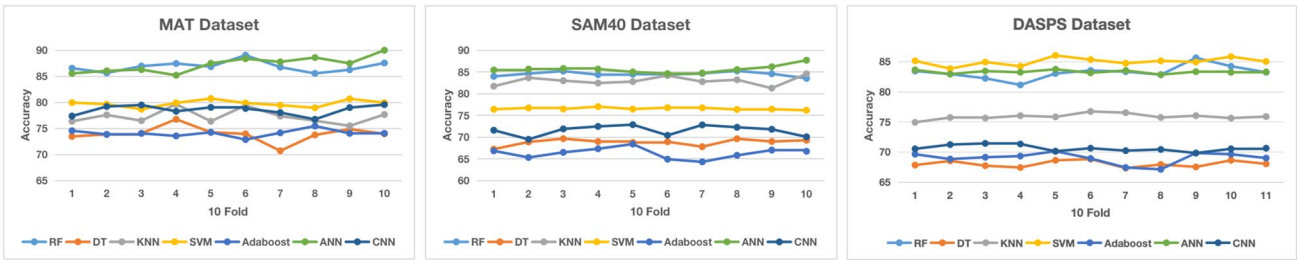


Fig. 5 Experiment 2: 10-fold cross-validation on classification algorithms after channel selection

Table 7 Experiment 3: Classification performance of the bagging models on the MAT [63] and SAM40 [19] datasets

MAT [63] MAT Dataset					SAM40 [19] MAT Dataset				DASPS [8] MAT Dataset			
Model	Acc (%)	F1 (%)	Sens (%)	Spec (%)	Acc (%)	F1 (%)	Sens (%)	Spec (%)	Acc (%)	F1 (%)	Sens (%)	Spec (%)
Bagging RF	92.25	92.32	92.54	94.32	91.63	91.48	92.6	90.72	90.45	90.77	89.53	91.45
Bagging KNN	84.62	84.41	86.34	83.3	87.61	87.91	89.36	85.82	85.77	86.17	84.50	87.18
Bagging SVM	90.18	90.44	92.31	91.54	92.85	92.8	91.61	94.1	86.79	86.43	84.49	89.07
Bagging ANN	87.23	86.75	89.83	84.53	91.36	91.2	89.35	93.37	88.82	89.44	91.02	87.44

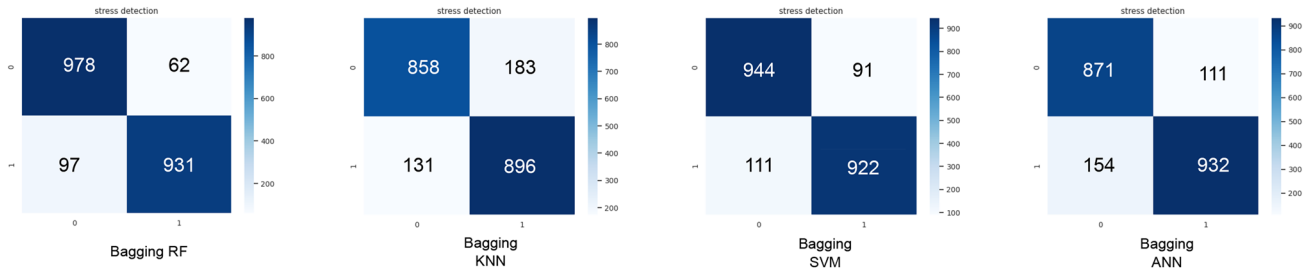


Fig. 6 Experiment 3: Confusion matrix of proposed bagging approach for MAT [63] dataset

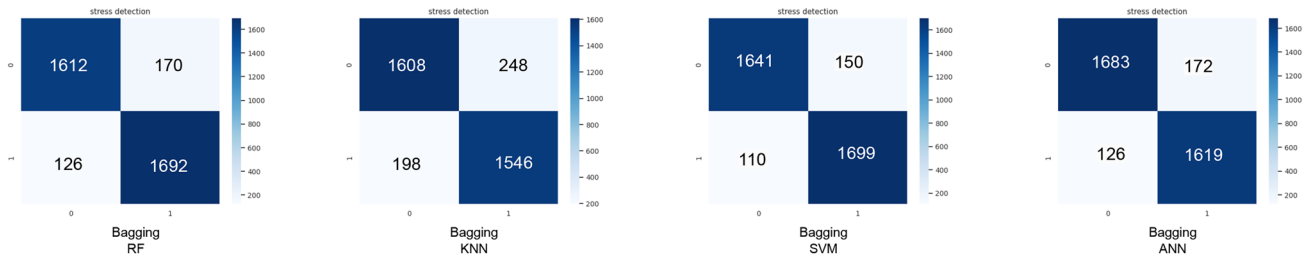


Fig. 7 Experiment 3: Confusion matrix of proposed bagging approach for SAM40 [19] dataset

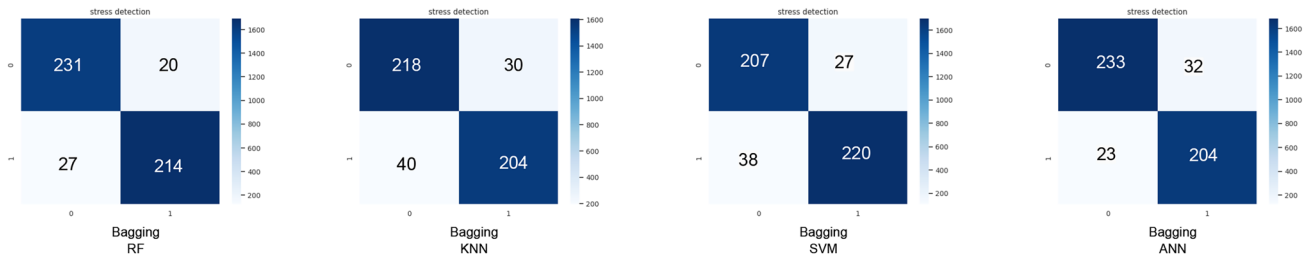


Fig. 8 Experiment 3: Confusion matrix of proposed bagging approach for DASPS [8] dataset

86.79% and 85.77%, respectively. Additionally, Figs. 6, 7, and 8 illustrate the confusion matrices, providing insights into the consistency and robustness of the proposed bagging models across the three datasets.

4.6 Experiment 4: proposed KRAFS-ANet framework

Stacking differs from bagging because it unifies heterogeneous algorithms into a singular model using a meta-classifier and fuses multiple predictive strengths into a single model. The study constructs a stacking model in the fourth experiment by merging the four most effective models from previous experiments, each with its optimal hyperparameters. Initially, it employs parallel stacking with bagging KNN, bagging RF, and bagging SVM to generate prediction probability-based features (PBF). Subsequently, these features are concatenated and fed into an ANN deep learning algorithm, which is the meta-classifier due to its superior performance compared to bagging ANN. GridSearchCV optimizes the meta-classifier for maximum accuracy. Figure 9 illustrates the proposed ensemble stacking model structure for stress classification.

Table 8 Experiment 4: Proposed KRAFS-ANet Framework performance on MAT [63] and SAM40 [19] datasets

Proposed Framework	Acc (%)	F1 (%)	Sens (%)	Spec (%)
MAT [63] Dataset (9 Channels)				
KRAFS-ANet	98.63	98.82	98.23	99.11
SAM40 [19] Dataset (12 Channels)				
KRAFS-ANet	97.25	97.24	97.39	97.12
DASPS [8] Dataset (6 Channels)				
KRAFS-ANet	94.92	95.15	95.70	94.07

Table 8 presents the classification performance of the proposed KRAFS-ANet framework on the MAT [63] dataset, along with its validation on the SAM40 [19] and DASPS [8]

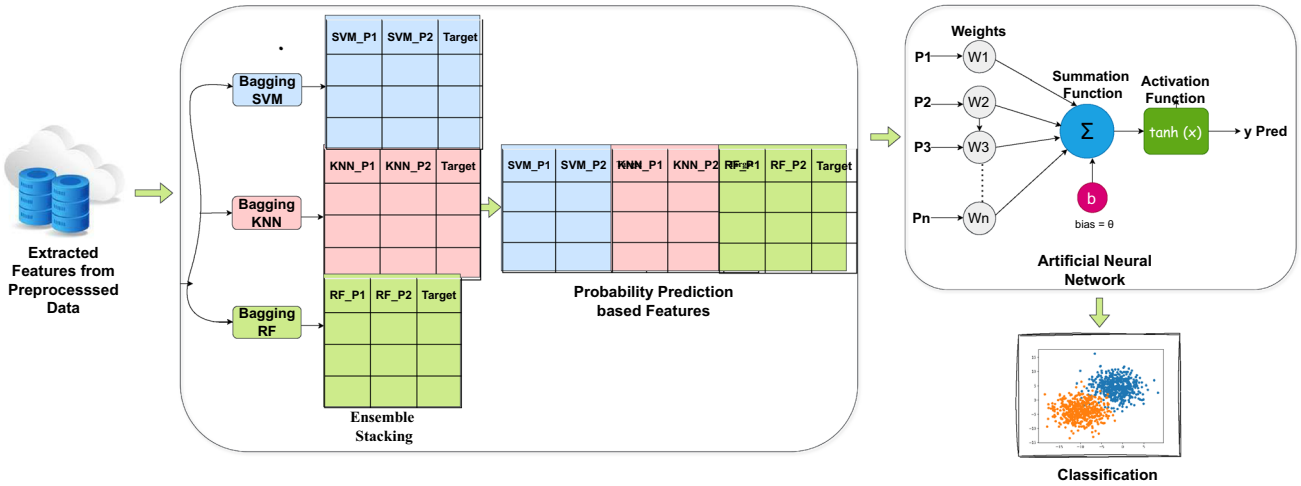


Fig. 9 Experiment 4: Proposed KRAFS-ANet Framework for Stress Classification

secondary datasets. On the MAT dataset, the model achieves an accuracy of 98.63%, an F1-score of 98.82%, a sensitivity of 98.23%, and a specificity of 99.11%, demonstrating its robustness in stress classification. The model is further validated on the SAM40 [19] and DASPS [8] datasets, achieving an accuracy of 97.25% and 94.92% and F1-scores of 97.24% and 95.15%, respectively. These results confirm the model's ability to generalize effectively across both secondary datasets.

Additionally, Fig. 10 complements the analysis, illustrating the proposed ensemble stacking approach's Confusion Matrix and ROC curve, providing insights into the model's classification performance and ability to discriminate stress levels.

5 Discussion

Recent research has focused on machine learning approaches for physiology-based stress classification. However, the challenge lies in effectively applying these accurate models in real-world scenarios, emphasizing the need for lightweight and adaptable models. This paper addresses the issue by proposing a novel and optimized ensemble stacking-based KRAFS-ANet framework for efficient stress classification. It performs comprehensive experiments incorporating machine learning, bagging, and various ensemble stacking techniques coupled with a channel selection approach. This paper evaluates the proposed KRAFS-ANet framework on the EEG-based MAT [63] dataset and validates it on the SAM40 [19] and DASPS [8] dataset.

Table 9 compares the performance of the selected classifiers across various evaluation metrics for all four experiments, highlighting the contributions of each base model. The results from experiment 2 indicate improvements compared to experiment 1, particularly notable due to the inclusion of channel selection using NMI+RFE. The NMI+RFE algorithm has two main benefits. Firstly, it ensures the preservation of relevant channels. It eliminates redundancy and optimizes the model's capacity to capture distinctive stress-related patterns within EEG signals. This process enhances the signal-to-noise ratio, making a more precise and effective stress classification model.

The study identifies the most informative channels by recognizing the non-stationary nature of EEG data and variations in individual brain activation patterns to stress stimuli. For the MAT [63] dataset, which involves arithmetic tasks, selected channels include Fp1, F7, F8, Fp2, O2, P3, T3, Fz, and T6. Conversely, the selected channels for the SAM40 [19] dataset include F2, FT8, FT10, F8, PO10, CP5, P8, C3, FCz, CP1, which encompasses arithmetic tasks, Stroop tests, and mirror image tests. Similarly, for the DASPS [8] dataset, which involves reciting anxiety-inducing situations and recall tasks. The final selected channels are AF4, F7, T7, T8, FC6, and O2, representing the most informative regions for anxiety classification.

Figure 11 illustrates the predominant activation of the brain's right hemisphere during stress and anxiety, particularly the frontal, temporal, parietal, and occipital lobes. Figure 11a highlights the activation in regions such as the frontal (Fp1, F7, F8, Fp2, Fz), temporal (T3, T6), parietal (P3), and occipital (O2) areas for the MAT [63] dataset. Likewise, previous research has also observed predominant activation of the brain's right hemisphere during stress, especially for tasks like MAT, aligning with the activation patterns observed in the frontal [41], temporal [51], parietal, and occipital regions [15].

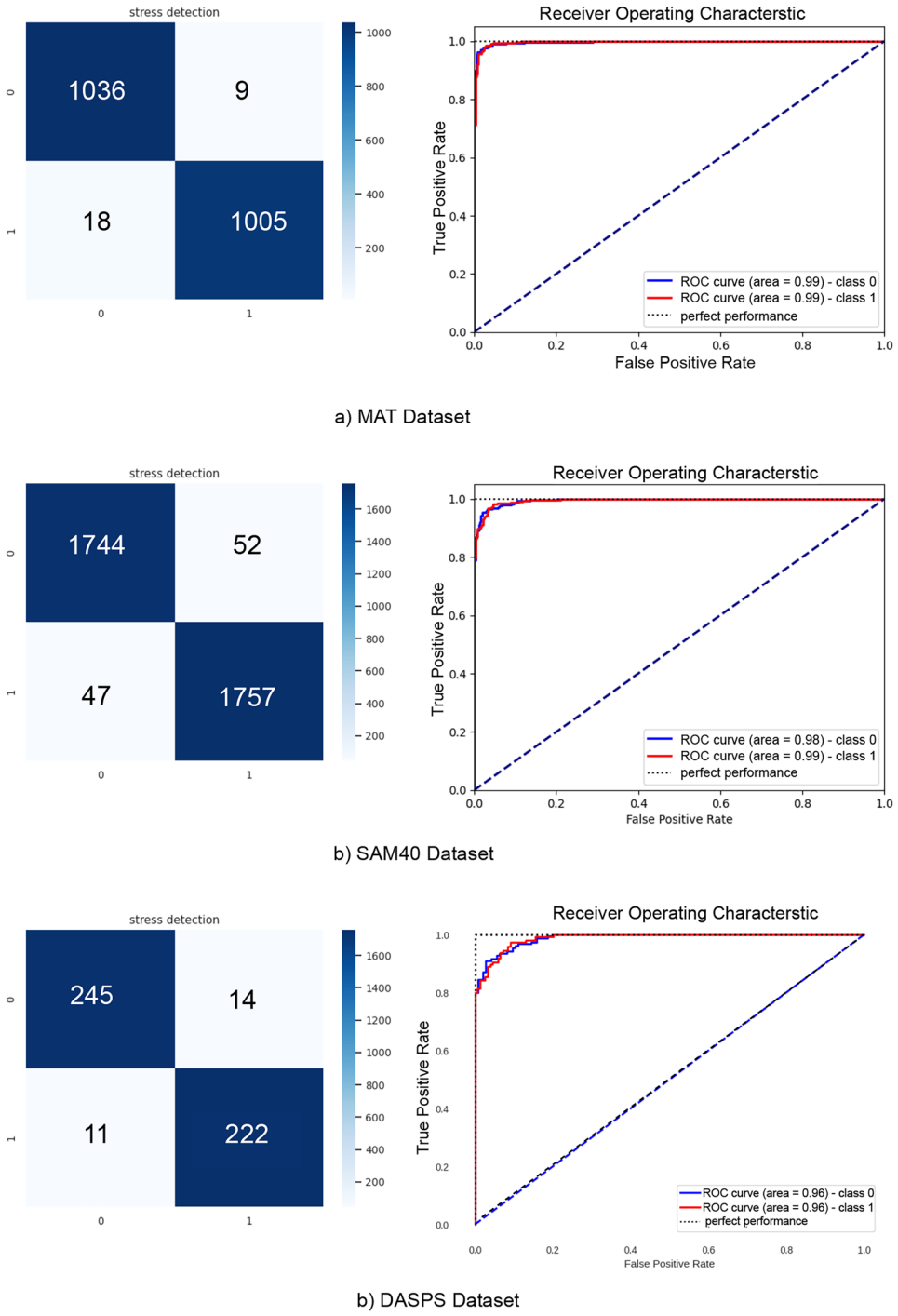


Fig. 10 Experiment 4: Confusion Matrix and ROC curve of the proposed KRAFS-Anet framework for stress classification

Figure 11b shows Stress-related activation in the SAM40 [19] dataset, which is primarily associated with the frontal and right hemisphere, encompassing regions such as the frontal (F2, F8), frontotemporal (FT10, FT8), parietal-occipital (PO10), and central-parietal (CP5, P8, CP1) areas. SAM40 [19] dataset includes distinct tasks and suggests the involvement of both the frontal and central lobes, particularly in the right hemisphere [13, 52].

Figure 11c illustrates similar findings for the DASPS [8] dataset for anxiety classification. The selected channels also exhibit significant activation in the right hemisphere, particularly in regions such as the frontal (AF4, F7), temporal (T7, T8), and occipital (O2) lobes. These findings align with the stress-related activation patterns observed in the other datasets, confirming the involvement of both frontal and temporal regions during stress. The selection of channels like FC6, T7, and O2 further supports the notion that stress

Table 9 Classification performance of the final models in each experiment on MAT [63] and SAM40 [19] and DASPS [8] datasets

Experiment	Classifier	MAT [63]			SAM40 [19]			DASPS [8]		
		Train Acc (%)	Test Acc (%)	F1 (%)	Train Acc (%)	Test Acc (%)	F1 (%)	Train (%)	Test (%)	F1 (%)
Experiment 1	RF	91.44	86.81	85.58	88.95	84.44	84.78	85.15	83.17	82.84
	KNN	83.02	77.23	76.83	87.22	82.9	83.86	78.99	75.84	76.15
	SVM	84.80	79.75	80.62	81.9	76.5	77.08	88.51	84.98	84.99
	ANN	91.17	87.23	87.75	90.24	85.58	85.71	87.46	83.32	83.83
Experiment 2	RF	94.84	90.76	90.82	94.82	91	90.85	91.71	87.26	87.25
	KNN	86.47	82.55	82.83	89.54	84.48	85.69	84.2	81.29	81.18
	SVM	85.93	82.8	82.78	93.57	89.4	89.4	87.34	85.67	85.53
	ANN	89.53	86.85	86.92	95.15	90.72	90.62	90.17	85.37	85.36
Experiment 3	Bagging RF	95.58	92.25	92.32	95.88	91.63	91.48	92.15	90.45	90.77
	Bagging KNN	87.44	84.62	84.41	92.51	87.61	87.91	88.86	85.77	86.17
	Bagging SVM	92.12	90.18	90.44	96.1	92.85	92.8	90.37	86.79	86.43
	Bagging ANN	89.52	87.23	86.75	94.39	91.36	91.2	91.52	88.82	89.44
Experiment 4	**Proposed KRAFS-ANet Framework**	99.17	98.63	98.82	99.75	97.25	97.24	97.57	94.92	95.15

responses vary across different tasks but consistently engage specific brain regions across datasets. Thus, these findings reveal that recognizing mental stress and anxiety for different tasks involves distinct combinations of EEG channels.

Moreover, the results of experiment 3 indicate that the bagging ensemble significantly enhances accuracy across all stress and anxiety classification datasets. In the MAT [63] dataset, bagging RF demonstrates robust performance, achieving a moderate improvement in accuracy. Bagging KNN also contributes positively, although its enhancements are less pronounced. Notably, bagging SVM exhibits substantial improvements across all performance measures, likely due to the ensemble's ability to refine decision boundaries. Conversely, bagging Artificial Neural Network (ANN) show only a slight increase of 0.38%, reflecting stability but limited gains in performance relative to the other models. The SAM40 [19] dataset similarly experiences slight improvements across all machine and deep learning algorithms, with each bagged model enhancing classification performance.

The DASPS [8] dataset further underscores the effectiveness of the bagging ensemble approach, with bagging RF achieving a notable accuracy improvement of approximately 3.5%. Bagging KNN, SVM, and ANN demonstrate performance increases by 4.5%, 1.3%, and 3.45%, respectively, compared to their standalone models. These results illustrate how the bagging method leverages the complementary strengths of each classifier, enhancing overall classification accuracy. Specifically, RF's capacity to handle complex, non-linear relationships and KNN's localized classification insights contributed significantly to the ensemble's performance.

In experiment 4, the proposed KRAFS-ANet framework utilized a stacking approach by combining bagging RF, bagging SVM and bagging KNN, with bagging ANN as a meta-classifier. This stacking approach allows for the aggregation of the individual strengths of each bagged classifier, further enhancing performance. The results of experiment 4 demonstrate superior performance across all datasets. The MAT dataset achieves the highest accuracy of 98.63% and an F1-score of 98.82%. Similarly, the SAM40 and DASPS [8] datasets obtain the highest accuracies of 97.25% and 94.92%, with F1-scores of 97.24% and 95.15%, respectively. These findings underscore the efficacy of the ensemble stacking method and highlight the critical role of data-driven channel selection in improving EEG-based stress classification across diverse datasets.

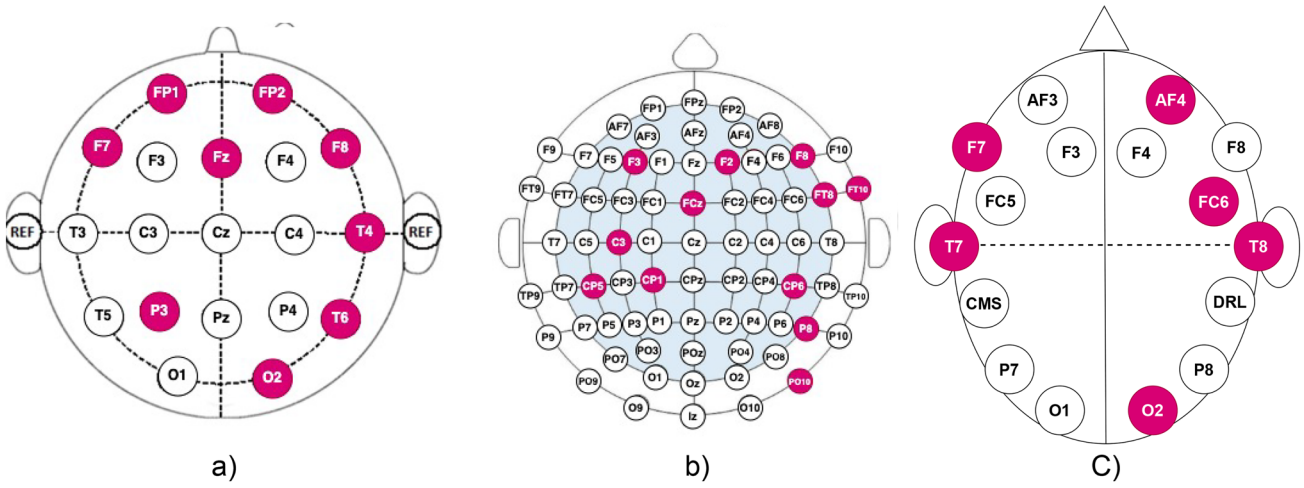


Fig. 11 Selected EEG channel location after NMI+RFE based channel selection a) 9 channels selected for MAT [63] dataset, b) 12 channels selected for SAM40 [19] dataset, c) 6 channels selected for DASPS [8] dataset

5.1 KRAFS-ANet framework: lightweight and efficient

The KRAFS-ANet framework incorporates several key approaches that contribute to its lightweight and efficient design:

1. *Channel Selection:* The framework integrates a channel selection technique that selects the most informative channels, reducing the input data size while maintaining high accuracy. This optimization helps lower the computational complexity and memory usage.
2. *Feature Selection:* The framework utilizes LASSO feature selection techniques to extract the most relevant features, which further minimizes the data size and results in lowering the computational complexity and improving model interpretability.
3. *Bagging and Parallelization:* Bagging, involving parallelization, reduces the effective training time by allowing multiple base classifiers to be trained simultaneously. It ensures that despite the complexity introduced by the ensemble approach, the model remains computationally efficient.

Time and memory efficiency: The study utilizes execution time and peak memory usage metrics to measure the lightweight design of the proposed KRAFS-ANet framework. These metrics help to demonstrate how efficiently the model operates, particularly in environments with limited computational resources.

1. *Execution Time:* The total time taken by the model or algorithm to complete its process from start to finish, measured in seconds.

Table 10 Performance Comparison of the Proposed KRAFS-ANet Framework and the Base Ensemble Stacking Model to highlight the impact of channel selection approach on accuracy, F1-score, data usage and execution time for the MAT [63], SAM40 [19], and DASPS [8] datasets

Model	Dataset	NMI+RFE Channel Selection	Number of channels	Total Features	Selected Featuresusing LASSO	Acc (%)	F1 (%)	Execution Time (sec)	Data Usage (MB)
Bagging(KNN+RF+SVM)+ANN	MAT [63]	×	20	320	158	92.48	93.32	2394.34	1266.53
Proposed KRAFS-ANet Framework	MAT [63]	✓	9	144	16	98.**63**	98.**82**	878.**67**	836.**31**
Bagging(KNN+RF+SVM)+ANN	SAM40 [19]	×	32	512	160	90.59	92.01	3129.94	1881.43
Proposed KRAFS-ANet Framework	SAM40 [19]	✓	12	192	37	97.**25**	97.**24**	1370.**32**	1130.**15**
Bagging(KNN+RF+SVM)+ANN	DASPS [8]	×	14	224	72	91.24	90.51	190.44	396.66
Proposed KRAFS-ANet Framework	DASPS [8]	✓	6	96	17	94.**92**	95.**15**	103.**28**	236.**07**

2. *Peak Memory:* The maximum amount of memory (RAM) used by the model during its execution, measured in mebibytes (MiB).

Table 10 presents a performance comparison between the proposed KRAFS-ANet Framework, which incorporates channel selection and feature selection, and the base ensemble stacking model without channel selection. This comparison highlights the impact of the channel selection approach on accuracy, F1-score, data usage, and execution time across the MAT [63], SAM40 [19], and DASPS [8] datasets. For the MAT [63] dataset, accuracy increases by 6.2%, while execution time decreases by 63.3% and memory usage drops by 34%. In contrast, the SAM40 [19] dataset sees an accuracy improvement of 7.3%, along with a 56.2% reduction in execution time and a 40% decrease in memory usage. Similarly, the DASPS [8] dataset observes a 3.68% increase in accuracy with NMI+RFE, while execution time reduces by 46.7% and memory usage decreases by 40.4%. These optimizations not only reduce computational overhead but also enhance accuracy and F1 scores across all datasets.

5.2 Comparison with existing studies

Table 11 compares the proposed approach with existing studies based on channel selection and stacking. In a previous study on the MAT dataset [63], Sharma et al. [41] utilized seven frontal and anterior frontal lobe channels and employed the Fisher score feature selection algorithm. The authors achieved an accuracy of 97.25% for an 8-second window using SVM with the Whale Optimization Algorithm (WOA). It highlights the importance of the frontal region in stress detection, given its association with emotional and cognitive processes. However, limiting channel selection to the frontal lobe may overlook critical information from other brain regions involved in stress responses.

Similarly, Fatimah et al. [15, 16] conducted two studies on the MAT [63] dataset, analyzed single-channel performance and achieved the highest accuracy of 98.6% using Logistic Regression (LR) and Quadratic Discriminant Analysis (QDA), respectively. The authors have shown that "C3" is the most informative channel. While their approach demonstrates that high accuracy can be achieved with a simple model, nevertheless relying on a single channel, they may miss the broader, complex brain activity involved in stress responses.

On the other hand, Malviya et al. [32] utilized DWT to preprocess the 19-channel EEG signal and fed the preprocessed signal into a stacked CNN + BiLSTM model. The CNN extracts the relevant features and inputs into the BiLSTM model. They achieved a high accuracy of 99.20% using CNN + BiLSTM, showing the potential of deep learning with multi-channel data. However, the high computational demands and resource requirements limit its practical use, especially in wearable technologies that demand efficient, real-time data processing and user comfort.

In another study, Mathur et al. [34] and Sim et al. [46] explored the SAM40 [19] dataset, achieving the highest accuracies of 79.08% and 71.57% with Rustboost and a stacking algorithm, respectively.

In our previous work [43] explored DASPS [8] dataset with 14 channels for anxiety classification. The authors utilised 14 channels and achieved the highest accuracy of 83.93% using Stacked Sparse Autoencoders (SSAE). In our another study [44], investigated the different windows size on DASPS dataset with 14 channels. The authors utilised genetic based feature selection algorithm and achieved the highest accuracy of 98.61% using GB. Nonetheless, it is important to note that no channel selection was performed in these studies, which may limit the effectiveness of the models. While Aldayel et al. [2] employed a manual channel selection approach with 10 channels based on affected brain regions, achieving an accuracy of 87.5% using Random Forest (RF). However, this method may introduce subjectivity and limit the model's adaptability to different datasets or individual differences.

To our knowledge, no previous research has explored channel selection with ensemble stacking specifically for stress classification using EEG signal. The study compares with state-of-the-art models employed in emotion classification to assess the efficacy of the proposed approach, as shown in Table 11.

Prior studies by Xie et al. [58] and Aquino et al. [5] utilized the DEAP dataset with 32 channels, employing diverse stacking approaches. Xie et al. [58] proposed an ensemble stacking approach involving LGB, XGB, and RF models for EEG-based emotion classification, achieving accuracy rates of 79.06% for valence and 77.19% for arousal, with feature selection using Linear Discriminant Analysis. Further, Aquino et al. [5] presented an EEG-based stacking approach combining CNN, SVM and DT models, achieving a remarkable accuracy of 94.6% with 32 channels.

In contrast, the proposed KRAFS-ANet framework reduces the number of channels from 20 to 9 on the MAT [63] dataset using NMI + RFE and employs an effective feature selection technique. The framework was also validated on the SAM40 [19] dataset, where 12 channels are selected, achieving an accuracy of 97.25% for stress classification. While this approach may result in a slight decrease in accuracy compared to Malviya et al. [32], it significantly improves computational efficiency and adaptability for real-world applications. Adapting effective channel and feature selection with an innovative ensemble stacking approach ensures that the KRAFS-ANet framework outperforms

Table 11 Performance comparison with the existing studies

Author	Domain	Dataset	Channel Selection Approach	Number of Channels	Feature Domain	FSA	Algorithm	Acc (%)
Sharma et al. [41]	Stress	MAT [63]	Frontal and anterior frontal lobe channels	7	Time-frequency domain	Fisher score	WOA + SVM	4 s - 94.01, 8 s - 97.25
Malviya et al. [32]	Stress	MAT [63]	–	19	Frequency domain	CNN	CNN+BiLSTM	CNN+BLSTM: 99.20
Fatimah et al. [15]	Stress	MAT [63]	Analysis on different channels performance	1	Frequency domain	Kruskal Wallis test	LR, SVM, KNN	SVM: 98.6, LR: 95.8, KNN: 93.1
Fatimah et al. [16]	Stress	MAT [63, 63]	Analysed results for each channel	1	Frequency domain	No	SVM, DT, QDA	QDA: 98.6
Mathur et al. [34]	Stress	SAM40 [19]	–	17	Frequency domain	Kruskal- Wallis Test	DT, KNN, SVM	DT: 87.30, KNN: 92.38, SVM: 93.38
Sim et al. [46]	Stress	SAM40 [19]	–	32	Time-frequency domain (DWT)	PCA, PCA+ICA, LDA	Rusboost and Adaboost	Rustboost: 79.08, Adaboost: 72.42
Hemakom et al. [23]	Stress	Kaggle [23]	–	8	Frequency domain	ANOVA F-value	NB+LR+AB+KNN +RF+SVM+ RBF-SVM	Male:71.57, Female:62.60, Mixed: 64.08
Shikha et al. [44]	Anxiety	DASPS [8]	–	14	Time, Frequency and domain	Genetic Algorithm	DT, KNN, XGBoost, GB	GB: 98.61
Aldayel et al. [2]	Anxiety	DASPS [8]	Manually selected based on most affected brain regions	10	Time, Frequency and Time-Frequency domain	-	KNN, Adaboost, GB,RF, LDA, SVM	RF: 87.5
Shikha et al. [43]	Anxiety	DASPS [8]	–	14	Time, Frequency and domain	RFE	DT, RF Stacked Sparse Autoencoder (SSAE)	SSAE: 83.93
Chatterjee et al. [11]	Emotion	Kaggle [9]	–	4	-	-	RF+LGBM+GB	Stacking: 99.55
Aquino et al. [5]	Emotion	DEAP [29]	–	32	-	-	CNN+SVM+DT	Stacking: 94.6
Xie et al. [58]	Emotion	DEAP [29]	–	32	Entropy	Linear Discriminant Analysis	(LGB+XGB+RF) +XGB	Stacking: Valence − 79.06, Arousal − 77.19
Proposed work	Stress and Anxiety	MAT [63], SAM40 [19] and DASPS [8]	NMI+RFE	MAT: 9 SAM40: 12 DASPS: 6	Time and frequency domain	Lasso	(Bagging RF + Bagging KNN + Bagging SVM) + ANN	MAT: 98.63, SAM40: 97.25 DASPS: 94.92

existing stress classification methods comprehensively and efficiently.

6 Limitations and future work

Despite the proposed ensemble stacking-based KRAFS-ANet framework, results emphasize the superiority of the primary and secondary datasets with accuracy rates of 98.63% and 97.25%, respectively. However, it still has some limitations.

1. The proposed framework uses the time and frequency domain features. Future studies may consider features from other domains, such as the time-frequency domain.
2. The channel selection algorithms can be further explored by incorporating other evolutionary algorithms such as Particle Swarm Optimization (PSO) [31], BAT algorithm [54], Genetic Algorithm(GA) [45], Whale Optimization Algorithm (WOA) [25]. Other heuristic optimization methods can reduce the occurrence of the "curse of dimensionality" and increase the classification performance.
3. Future studies may consider ensemble stacking comprehensive experiments on other physiological signals such as ECG, Galvanic Skin Response (GSR), Heart Rate Variability (HRV), and Heart Rate (HR) for stress classification.
4. The KRAFS-ANet framework can be validated with channel selection techniques in other domains, such as emotion classification and epilepsy detection.

7 Conclusion

In conclusion, the study attempts to balance accuracy and computational efficiency, introducing a novel ensemble stacking-based KRAFS-ANet framework. This framework integrates K-Nearest Neighbour (KNN), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN) meta-classifiers. The framework utilizes the NMI+RFE algorithm to choose a specific number of channels, focusing on the most informative ones. LASSO feature selection is also applied to remove negatively impacting features and enhance classification performance, ensuring the framework is lightweight and adaptable for practical use. The study encompasses four experiments involving machine learning algorithms, such as SVM, KNN, DT, RF, and Adaboost, and deep learning algorithms, including CNN and ANN. These experiments are conducted on the MAT dataset and concurrently validated on the SAM40 and DASPS datasets. Experiments 1 and 2 involve all classifiers with or without channel selection, while experiments 3 and 4 employ two ensemble approaches, utilizing the best-performing models. Notably, the model's accuracy significantly improves in experiment 2, demonstrating the efficacy of channel selection. The empirical results highlight the superiority of the proposed ensemble stacking model with accuracy rates of 98.63%, 97.25% and 94.92% on the MAT, SAM40 and DASPS datasets, respectively. In future, the proposed ensemble stacking framework may contribute to wearable technology for stress management in real-time applications.

Funding The author declared that no funding was received for this research and publication.

Data availibility Data sharing is not applicable to this article as no datasets were generated during the current study. The datasets referenced in the current study are cited with their base papers, and their links are also given in the paper.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

Consent for publication The authors claim that none of the material in the paper has been published or is under consideration for publication elsewhere.

References

1. Akella A, Singh AK, Leong D, Lal S, Newton P, Clifton-Bligh R, Mclachlan CS, Gustin SM, Maharaj S, Lees T et al (2021) Classifying multi-level stress responses from brain cortical EEG in nurses and non-health professionals using machine learning auto encoder. IEEE J Transl Eng Health Med 9:1–9
2. Aldayel M, Al-Nafjan A (2024) A comprehensive exploration of machine learning techniques for EEG-based anxiety detection. PeerJ Comput Sci 10:e1829
3. Alotaiby T, El-Samie FEA, Alshebeili SA, Ahmad I (2015) A review of channel selection algorithms for EEG signal processing. EURASIP J Adv Signal Process 2015:1–21
4. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L (2021) Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8:1–74
5. Aquino LS, dos Santos Coelho L (2021) Stacking ensemble learning approaches applied to emotional state classification
6. Attallah O (2020) An effective mental stress state detection and evaluation system using minimum number of frontal brain electrodes. Diagnostics 10(5):292
7. Badr Y, Tariq U, Al-Shargie F, Babiloni F, Al Mughairbi F, Al-Nashash H (2024) A review on evaluating mental stress by deep learning using EEG signals. Neural Comput Appl 36:12629–12654
8. Baghdadi A, Aribi Y, Fourati R, Halouani N, Siarry P, Alimi A (2021) Psychological stimulation for anxious states detection

based on EEG-related features. J Ambient Intell Humaniz Comput 12:8519–8533
9. Bird JJ, Ekart A, Buckingham CD, Faria DR (2019) Mental emotional sentiment classification with an EEG-based brain-machine interface. In: Proceedings of the International Conference on Digital Image and Signal Processing (DISP'19)
10. Cahill ND (2010) Normalized measures of mutual information with general definitions of entropy for multimodal image registration. In: Biomedical Image Registration: 4th International Workshop, WBIR 2010, Lübeck, Germany, July 11–13, 2010. Proceedings 4. Springer, pp 258–268
11. Chatterjee S, Byun YC (2022) EEG-based emotion classification using stacking ensemble approach. Sensors 22(21):8550
12. Chen XW, Jeong JC (2007) Enhanced recursive feature elimination. In: Sixth international conference on machine learning and applications (ICMLA 2007). IEEE, pp 429–435
13. Davidson R (1979) Frontal versus parietal EEG asymmetry during positive and negative affect. Psychophysiology 16(2):202–203
14. DiPietro R, Hager GD (2020) Deep learning: Rnns and lstm. Handbook of medical image computing and computer assisted intervention. Elsevier, pp 503–519
15. Fatimah B, Javali A, Ansar H, Harshitha B, Kumar H (2020) Mental arithmetic task classification using fourier decomposition method. In: 2020 International Conference on Communication and Signal Processing (ICCSP). IEEE, pp 0046–0050
16. Fatimah B, Pramanick D, Shivashankaran P (2020) Automatic detection of mental arithmetic task and its difficulty level using EEG signals. In: 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE. pp 1–6
17. Gao Q, Yang Y, Kang Q, Tian Z, Song Y (2022) EEG-based emotion recognition with feature fusion networks. Int J Mach Learn Cybern 13(2):421–429
18. Gaur P, McCreadie K, Pachori RB, Wang H, Prasad G (2021) An automatic subject specific channel selection method for enhancing motor imagery classification in EEG-BCI using correlation. Biomed Signal Process Control 68:102574
19. Ghosh R, Deb N, Sengupta K, Phukan A, Choudhury N, Kashyap S, Phadikar S, Saha R, Das P, Sinha N et al (2022) SAM 40: Dataset of 40 subject EEG recordings to monitor the induced-stress while performing stroop color-word test, arithmetic task, and mirror image recognition task. Data Brief 40:107772
20. Gollapalli M, Alansari A, Alkhorasani H, Alsubaii M, Sakloua R, Alzahrani R, Al-Hariri M, Alfares M, AlKhafaji D, Al Argan R et al (2022) A novel stacking ensemble for detecting three types of diabetes mellitus using a Saudi Arabian dataset: pre-diabetes, t1dm, and t2dm. Comput Biol Med 147:105757
21. Hafeez MA, Shakil S (2024) EEG-based stress identification and classification using deep learning. Multimed Tools Appl 83(14):42703–42719
22. Hastie T, Tibshirani R, Friedman JH, Friedman JH (2009) The elements of statistical learning: data mining, inference, and prediction, vol 2. Springer
23. Hemakom A, Atiwiwat D, Israsena P (2023) ECG and EEG based detection and multilevel classification of stress using machine learning for specified genders: a preliminary study. Plos one 18(9):e0291070
24. Hoc H.T, Silhavy R, Prokopova Z, Silhavy P (2023) Comparing stacking ensemble and deep learning for software project effort estimation. IEEE Access
25. Hussien AG, Oliva D, Houssein EH, Juan AA, Yu X (2020) Binary whale optimization algorithm for dimensionality reduction. Mathematics 8(10):1821
26. Jain P, Yedukondalu J, Chhabra H, Chauhan U, Sharma LD (2024) EEG-based detection of cognitive load using vmd and lightgbm classifier. Int J Mach Learn Cybern 15:4193–4210
27. Kamińska D, Smółka K, Zwoliński G (2021) Detection of mental stress through EEG signal in virtual reality environment. Electronics 10(22):2840
28. Katmah R, Al-Shargie F, Tariq U, Babiloni F, Al-Mughairbi F, Al-Nashash H (2021) A review on mental stress assessment methods using EEG signals. Sensors 21(15):5043
29. Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I (2011) Deap: a database for emotion analysis; using physiological signals. IEEE Trans Affect Comput 3(1):18–31
30. Kullback S (1997) Information theory and statistics. Courier Corporation
31. Li Z, Qiu L, Li R, He Z, Xiao J, Liang Y, Wang F, Pan J (2020) Enhancing BCI-based emotion recognition using an improved particle swarm optimization for feature selection. Sensors 20(11):3028
32. Malviya L, Mal S (2022) A novel technique for stress detection from EEG signal using hybrid deep learning model. Neural Comput Appl 34(22):19819–19830
33. Mathur P, Chakka VK (2022) Graph signal processing based cross-subject mental task classification using multi-channel EEG signals. IEEE Sens J 22(8):7971–7978
34. Mathur P, Kaistha S, Chakka VK (2023) Mental task induced stress detection using multi-variate weighted visibility graph (mv-wvg) from EEG signals. In: 2023 IEEE 20th India Council International Conference (INDICON). IEEE. pp 1265–1270
35. Muhammad F, Al-Ahmadi S (2022) Human state anxiety classification framework using EEG signals in response to exposure therapy. Plos One 17(3):e0265679
36. Mustafa MU, Buzdar SA et al (2023) Machine learning approach for multi-class stress assessment with electroencephalography signals. J Popul Ther Clin Pharmacol 30(19):1464–1478
37. Rakhmatulin I (2020) Review of EEG feature selection by neural networks. Int J Sci Bus 4(9):101–112
38. Remeseiro B, Bolon-Canedo V (2019) A review of feature selection methods in medical applications. Comput Biol Med 112:103375
39. Salankar N, Koundal D, Chakraborty C, Garg L (2023) Automated attention deficit classification system from multimodal physiological signals. Multimed Tools Appl 82(4):4897–4912
40. Salankar N, Qaisar SM (2022) EEG based stress classification by using difference plots of variational modes and machine learning. J Ambient Intell Humaniz Comput 14(12):16347–60
41. Sharma LD, Bohat VK, Habib M, Ala'M AZ, Faris H, Aljarah I (2022) Evolutionary inspired approach for mental stress detection using EEG signal. Expert Syst Appl 197:116634
42. Shen J, Zhang X, Huang X, Wu M, Gao J, Lu D, Ding Z, Hu B (2020) An optimal channel selection for EEG-based depression detection via kernel-target alignment. IEEE J Biomed Health Inform 25(7):2545–2556
43. Shikha Agrawal M, Anwar M.A, Sethia D (2021) Stacked sparse autoencoder and machine learning based anxiety classification using EEG signals. In: Proceedings of the First International Conference on AI-ML Systems, pp 1–7
44. Shikha S, Sethia D, Indu S (2024) Investigating the effect of sliding window length for EEG-based anxiety classification. In: 2024 Asia Pacific Conference on Innovation in Technology (APCIT). IEEE, pp 1–6
45. Shon D, Im K, Park JH, Lim DS, Jang B, Kim JM (2018) Emotional stress state detection using genetic algorithm-based feature selection on EEG signals. Int J Environ Res Public Health 15(11):2461

46. Sim DY, Chong C (2023) Effects of dimension reduction methods on boosting algorithms for better prediction accuracies on classifications of stress EEGs. In: 2023 6th International Conference on Electronics and Electrical Engineering Technology (EEET). IEEE, pp 49–54
47. Simske S (2019) Introduction, overview, and applications. Meta-analytics pp 1–98
48. Talia D, Trunfio P, Marozzo F (2016) Introduction to data mining. Data analysis in the cloud. Morgan Kaufmann, pp 1–25
49. Tawsifur R, Khandakar A, Abir FF, Faisal MAA, Hossain MS, Podder KK, Abbas TO, Alam MF, Kashem SB, Islam MT et al (2022) QCovSML: a reliable COVID-19 detection system using CBC biomarkers by a stacking machine learning model
50. Torres-Valencia C, Álvarez-López M, Orozco-Gutiérrez Á (2017) SVM-based feature selection methods for emotion recognition from multimodal data. J Multimodal User Interfaces 11:9–23
51. Varshney A, Ghosh SK, Padhy S, Tripathy RK, Acharya UR (2021) Automated classification of mental arithmetic tasks using recurrent neural network and entropy features obtained from multi-channel EEG signals. Electronics 10(9):1079
52. Wang Q, Sourina O (2013) Real-time mental arithmetic task recognition from EEG signals. IEEE Trans Neural Syst Rehabil Eng 21(2):225–232
53. Wang ZM, Hu SY, Song H (2019) Channel selection method for EEG emotion recognition using normalized mutual information. IEEE Access 7:143303–143311
54. Wang Z, Zhang Z, Wang W (2019) Emotion recognition based on framework of badeba-svm. Math Probl Eng 1:9875250
55. Wickramaratna J, Holden S, Buxton B (2001) Performance degradation in boosting. In: Multiple Classifier Systems: Second International Workshop, MCS 2001 Cambridge, UK, July 2–4, 2001 Proceedings 2. Springer, pp 11–21
56. Witten IH, Frank E, Hall MA, Pal CJ (2017) Ensemble learning. Data mining, vol 4. Elsevier Amsterdam, The Netherlands, pp 479–501
57. Wolpert DH (1992) Stacked generalization. Neural Netw 5(2):241–259
58. Xie O, Liu ZT, Ding XW (2018) Electroencephalogram emotion recognition based on a stacking classification model. In: 2018 37th Chinese Control Conference (CCC). IEEE, pp 5544–5548
59. Xiong R, Kong F, Yang X, Liu G, Wen W (2020) Pattern recognition of cognitive load using EEG and ECG signals. Sensors 20(18):5122
60. Xu H, Wang X, Li W, Wang H, Bi Q (2019) Research on EEG channel selection method for emotion recognition. In: 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, pp 2528–2535
61. Yedukondalu J, Sharma LD (2023) Cognitive load detection using circulant singular spectrum analysis and Binary Harris Hawks Optimization based feature selection. Biomed Signal Process Control 79:104006
62. Zhang J, Chen M, Zhao S, Hu S, Shi Z, Cao Y (2016) Relieff-based EEG sensor selection methods for emotion recognition. Sensors 16(10):1558
63. Zyma I, Tukaev S, Seleznov I, Kiyono K, Popov A, Chernykh M, Shpenkov O (2019) Electroencephalograms during mental arithmetic task performance. Data 4(1):14

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

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

EAI Endorsed Transactions
on Internet of Things
Research Article EAI.EU

Major vulnerabilities in Ethereum smart contracts: Investigation and statistical analysis

Mohammad Pishdar[1], Mahdi Bahaghighat[2,*], Rajeev Kumar[3], Qin Xin[4]

[1]Computer Engineering Department, Bu-Ali Sina University, Hamedan, Iran
[2]Computer Engineering Department, Imam Khomeini International University, Qazvin, Iran
[3]Department of Computer Science and Engineering, Delhi Technological University, Delhi, India
[4]Faculty of Science and Technology, University of the Faroe Islands, Faroe Islands

Abstract

The general public is becoming increasingly familiar with blockchain technology. Numerous new applications are made possible by this technology's unique features, which include transparency, strong security via cryptography, and distribution. These applications need certain programming tools and interfaces to be implemented. This is made feasible by smart contracts. If the prerequisites are satisfied, smart contracts are carried out automatically. Any mistake in smart contract coding, particularly security-related ones, might have an impact on the project as a whole, available funds, and important data. The current paper discusses the flaws of the Ethereum smart contract in this respect. By examining publically accessible scientific sources, this work aims to present thorough information about vulnerabilities, examples, and current security solutions. Additionally, a substantial collection of current Ethereum (ETH) smart contracts has undergone a static code examination to conduct the vulnerability-finding procedure. The output has undergone assessments and statistical analysis. The study's conclusions demonstrate that smart contracts have several distinct flaws, including arithmetic flaws, that developers should be more aware of. These vulnerabilities and the solutions that can be used to address them are also included.

Keywords: Blockchain Security, Smart Contract Security, Cryptocurrency Security, Smart contracts Attacks

Received on 14 02 2024, accepted on 17 12 2024, published on 18 12 2024

Copyright © 2024 M. Pishdar *et al.*, licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetiot.5120

*Corresponding author. Email: bahaghighat@eng.ikiu.ac.ir

1. Introduction

Blockchain is a kind of distributed database that combines data replication with chain-based cryptography to offer distribution, high security (because of encryption and increased availability), and transaction transparency. Blockchain technology has gained popularity in recent years among programmers, IT firms, and even users of virtual worlds as a leading solution for building distributed databases and employing collective agreements. Multiple networks have been made available to the general public as a blockchain platform following the technology competitions and the presentation of various blockchain solutions in line with the better development of applications. We can list some of the most well-known of these networks as Bitcoin, Ethereum, Solana, and Avalanche. To offer their users better services, each blockchain has published a variety of protocols in this area. We might mention different consensus systems between the nodes of a distributed network of information as examples of these protocols [1-3].

Several blockchain projects have created a middle and high-level layer as a user interface for developers after developing the necessary hardware platform and communication protocols. This work will make it easier for blockchain network developers to create a wide range of applications and handle the intricate technical and specialized requirements of the blockchain's foundational layer. This platform is known as smart contracts [1-2]. Programmers can use data structures, functions, and different object-oriented systems in smart contracts, as well as a number of unique functions and libraries that are accessible over the blockchain network. In other words, they

may quickly record information or query it on the infrastructure blockchain network using the appropriate technologies [2-4].

The automatic execution of smart contracts without human interaction is another aspect. It will be feasible to execute it if specific requirements are given for its operation in the blockchain network. Any programming error the developer makes can impact the application program, user data, financial resources linked to the program, and even the users' financial resources. A smart contract flaw can occasionally impact the blockchain network infrastructure [2-5]. Smart contracts feature a number of security flaws, some unique to this kind of network and some derived from broader ideas and applied to this particular kind of technology [6-9].

One of the first and most active blockchain networks is Ethereum, which powers smart contracts using a unique programming language called Solidity. It is why this study's focus is on Ethereum and the aforementioned language [10].

After reviewing related literature for this study, we concluded that the following questions still require a response. The current research has attempted to answer these questions in this regard.

- Which particular security flaws exist in smart contracts?
- What is a thorough description of these circumstances that includes examples?
- Is it possible to classify vulnerabilities?
- What methods are available to address these vulnerabilities?
- Which vulnerabilities are verified through the use of various techniques for static code analysis and a comparison of the outcomes in smart contracts?
- Do these vulnerabilities occur more frequently in some smart contracts than others? Or are some more prevalent?

For this investigation, a huge variety of sources were initially examined. The results were made to attain sufficient value and comprehensiveness by analyzing a pruning operation about the Ethereum (ETH) blockchain network, the unique smart contract vulnerabilities, and the Solidity programming language. The second stage involved looking for flaws in 100 smart contracts active on the publicly accessible Ethereum network until the study time, the number of major transactions, and the financial resources associated with four static code review tools. After evaluation and agreement, the outcomes were presented as a statistical analysis of the outputs. An extensive investigation of the specific Ethereum smart contract vulnerabilities has been attempted in this study. The list below shows the main innovations of this research.

- A thorough examination of particular Ethereum smart contract security flaws with pertinent examples and solutions
- Using several automatic tools to perform static analysis on a collection of smart contracts and provide statistical analysis of vulnerabilities found.
- Establishing a category for smart contract weaknesses.
- A description of fundamental blockchain ideas and comparison data

The smart contract security problem inspired us to write this paper. After creating smart contracts, many developers become aware of security issues, but because of their unchangeable structure, they must construct a new contract in order to address them. This is quite expensive. We have attempted to review the most significant issues and current solutions in this study by looking at the security issues with smart contracts and statically analyzing the codes of a sample of them. To determine the status of these vulnerabilities in actual contracts, we additionally analyze a set of contracts' weaknesses in terms of static code review. The rest of this paper organized as follows:

Section 2 describes related works. Section 3 introduces known vulnerabilities and their detection process, and also covers countermeasures. Based on the flaws described in Section 3, a statistical analysis of a dataset of Ethereum smart contracts has been presented in Section 4. Conclusions and recommendations are included in Section 5.

2. Related works

Related work falls into one of three categories:

- **Category 1: Research that only addresses one or a few specific vulnerabilities does not provide a comprehensive view of the security vulnerabilities of smart contracts.**

In [11] Wöhrer and his colleagues examined six security patterns in Ethereum smart contracts. Of course, it should be highlighted that these instances use the same smart contract security flaws that are described in this study in a different form. In [12], the researchers investigated re-entrancy vulnerability in a data set of smart contracts. This review consists of two review sections with several automated tools and manual reviews. According to the research results, the difference in these cases is very high, and this shows that automatic tools have many false positive reports. In [13], researchers have introduced various possible vulnerabilities in smart contracts. In this research, a wide range of more general vulnerabilities from the level of smart contracts have been proposed. Failure to provide examples and solutions for vulnerabilities is one of the problems of this research. Researchers in this study have also investigated some vulnerability detection methods.

- **Category 2: Research that covers a decent number of existing vulnerabilities but does not include providing solutions or examples in sufficient detail.**

Researchers have presented nine security vulnerabilities in reference [14] without offering a fix or outlining the detection mechanisms in these situations. Then, False/True positive and False/True negative criteria were developed to evaluate the abovementioned vulnerabilities. Without discussing the fix, researchers in research [15] published a list of Ethereum smart contract vulnerabilities (some briefly and others in detail). After assessing a vulnerability and offering a solution, they covered several automatic checking technologies. In [16], researchers created a dataset of Ethereum smart contracts and labeled its vulnerabilities with the outputs of five automatic detection tools. There are no descriptions of vulnerabilities or recommendations for solutions in this study. Furthermore, many of these contracts are no longer in use because they have expired. The research's methodology entails scanning a set of smart contracts that were amassed by another study [17] and classifying the data according to the results. In [18], the research method served as the study of publicly available scientific sources and their elimination to achieve the research objective. With a sample contract and very brief explanations. Researchers outline three areas of smart contract weaknesses in [19]: blockchain, code, and Ethereum virtual machine. In this approach, some vulnerabilities are described with examples, while in others, merely mentioning the subject is sufficient. It should be highlighted that no remedy was discovered in this study, and other more general and relevant smart contract vulnerabilities have also been suggested. In the research [20], the researchers have investigated a wide range of vulnerabilities related to smart contracts in the form of a general classification in terms of the impact and severity of the vulnerability. Of course, many of these cases are not specific to smart contracts, and a wide range of software also faces them. Limited explanations in describing vulnerabilities, examples and related solutions are another feature of this research. Researchers have also examined developers' views and provided some general solutions for diagnosing security problems.

- **Category 3: Research that has concentrated more on static code analysis has not taken any action to improve the accuracy of the static analysis output or statistically assess the findings, in addition to not giving enough details about each vulnerability.**

The study's authors [21] have also looked at thirteen other smart contract weaknesses and given corresponding solutions, limited to the Ethereum blockchain and following the current weaknesses through the review of English-language scholarly sources. An overview of vulnerability detection tools and their comparison was then completed. In the study [22], the researchers presented, reviewed, and contrasted the techniques for identifying these vulnerabilities in static and dynamic methods (the research's primary focus) after providing a very brief description of 11 major smart contract flaws. In the research [23], the researchers have investigated a set of Ethereum smart contracts (related to the SmartBugs tool) with nine automatic static vulnerability-checking tools. In this research, only the warnings related to automatic tools have been investigated, and no proper effort has been made to increase the accuracy of the results. For this reason, the number of vulnerable contracts is reported to be very high. In the research [24], after pointing out the vulnerabilities of smart contracts, the researchers introduced the available automatic detection tools and categorized and compared them. In this comparison process, things such as the speed and accuracy of diagnosis, the amount of error in diagnosis, and the types of interfaces that can be used have been examined. According to the research methods utilized in three studies [16,24,25], valid scientific papers were found through an academic search, and they were then culled based on the requirements of the study questions.
The second and third category studies may not be able to provide a proper view for the reader due to insufficient information. In the third category, the accuracy of the results can also be very poor because the output of error tools is usually accompanied by many erroneous reports. Therefore, it is necessary to provide appropriate analyses and sufficient information in addition to paying attention to the issue of increasing accuracy.

3. Research background

Some useful and crucial concepts related to smart contract vulnerabilities are provided in this section.body.

3.1. Blockchain network

Blockchain is a decentralized digital collective agreement that uses one-way encryption technology or hashing to stop information from being changed or manipulated. Due to the limited public access to this disseminated information, this network also exhibits transparency. These networks maintain duplicate copies of transaction information—information recorded in the network and made up of numerous blocks—between dispersed nodes. The likelihood of data loss will be significantly decreased in this approach. Some blockchain terminology is defined in the sentences that follow [8], [24-25].

3.1.1. Blocks

The blockchain network collects transactions in the form of blocks over a predetermined period. The following data [24-25] is saved in a block next to the transactions mentioned above **:**
Magic number: used to identify a block as a certain digital currency's network component**.**

- **Block header**: provides details about the block.

- **Block size**: specifies the maximum block size for writing a given data **.**
- **Transaction Counter**: A transaction counter is a number that indicates how many transactions have been stored in a block**.**
- **Transactions**: Lists each transaction that took place within a block.
- **Version**: Displays the current version of the digital currency.
- **Preceding block hash**: Contains the preceding block's header's hash (an encrypted number)**.**
- **Root Merkle Hash**: The current block comprises the hash codes of the transactions in the Merkel tree, a tree structure designed to minimize storage capacity and transaction calculations.
- **Time**: includes the time stamp of the blockchain block submission**.**
- **Bit**: Contains information describing the complexity of finding the nonce number and the difficulty of the target hash function **.**
- **Nonce**: A random number that the miner must find to validate and close the block.

An abstract block structure from the blockchain network is shown in Table 1**.**

Table 1. Block structure [21]

Block Version	0200000
Parent Block Hash	b6ff0b1b1680a2867a30ca44d34d9e8910d334be b48ca0c000000000000000
Merkle tree Root	9d10aa52ee949386ca9385695f04ede270dda208 10decd12bc9b048aaab31471
Time Stamp	24d95a54
nBits	30c31b18
Nounce	Fe9f0864
Transaction Counter	
TX1 TX2 TX3 ……TXn	

3.1.2. Validator

According to the decentralized collective consensus protocol, the systems in the blockchain network are in charge of validating and logging network transactions. The network validators are rewarded for completing these duties. It should be highlighted that any actions taken by validators that violate the blockchain's rules are detected and removed by the decentralized collective consensus protocol [25-27].

3.1.3. Consensus protocols

In the blockchain, a decentralized consensus protocol allows different nodes to agree on a network element (blocks). The blockchain network's nodes all concur to record the same data according to mutual understanding. Without the assistance of a centralized organization, this activity is carried out decentralized. We will look at the most well-known decentralized collective consensus protocols introduced [25-26].

- The POW (Proof of Work) protocol pits high-powered systems against one another in a race to crack tricky computational riddles. The first miner to find the solution, or, to put it another way, a hash compatible with the prior blocks, sends it to the network for confirmation by other nodes. It is better to describe the hashing method to comprehend this problem briefly. There is a form of one-way encryption used in the hashing method. This means that if X is present, a specific hashing technique can be used to quickly determine the value of h(x), although this problem requires a lot of computation and is expensive. The POW protocol reward miners to use trial and error to develop h(x) solutions compatible with the earlier blocks. This allows

  the miner who solves h(x) to broadcast it to the network first, where other miners can quickly verify it and validate its accuracy [26-28].
- The POS protocol (proof of stake) eliminates the prior form's flaw of requiring high processing costs and electrical resources from the verifiers. One of the network nodes is selected randomly (in a transparent procedure), with no longer any competition amongst the block verifiers, to validate the block. People with a higher amount in these two parameters have more chances in this selection process, which is based on the number of shares staked and the length of time the staked capital is locked. Age, a protocol parameter that describes the nodes' waiting time, is present in this protocol. This value will be zero if a validator is chosen to confirm the block,

increasing the likelihood that a fair procedure will involve waiting for validators. A validator receives a bonus in the form of the network's native currency after validating a block and submitting it to the network [26-27], [29].

- Proof-of-authority protocol: This protocol, which aims to utilize fewer computational resources than the POS protocol, is an alternative. This approach bases agreement on Proof-of-authority. This means that validators are chosen using a strict method, and then the accuracy of those validators is evaluated using a monitoring standard. A validator loses his validity and is removed from the system if his performance falls short of the required level. As a result, the right to validation is directly impacted by standard compliance. The identity of the validator, his history, and the amount of investment are all examined during the selection process [30-33].
- The enhanced POS Consensus Protocol is DPoS (Delegated Proof of Stake). By putting their tokens (the principal infrastructure blockchain token) into an investment pool, users in this system can cast votes for validators which will approve and create the block. Naturally, more deposits increase power in a way that prevents the system from being dominated by one or a few actors. Depending on the proportion of tokens invested, the validator will distribute a portion of the profit to the voters if the block is written correctly. By garnering more votes, more trustworthy and accepted validators are given more authority to change the block. It should be highlighted that only a few validators can vote and are chosen to write blocks. Naturally, this list is flexible [34-37].
- PBFT protocol (Practical Byzantine Fault Tolerance) Even with faults in specific dispersed nodes, this system may reach a consensus. Assume that the PBFT protocol allows a distributed system with 3f+1 nodes to have a maximum of error nodes (Byzantine nodes). In other words, when 2f+1 nodes agree on a message, the system as a whole will likewise agree. In this system, the process of republishing and confirming these messages in the form of a "preparation message" in the network is carried out by a set of pre-defined messages. These messages are called the first preparation message that is transmitted from a node called Primary Node to several additional nodes called Replicator. Validator nodes can advertise their ability to write the current block by broadcasting the "confirmation" message after receiving 2f+1 messages compatible with their sent item. It is done by rebroadcasting preparation messages. When the first node receives 2f+1 compatible messages of this type, It sends the confirmation message in the following step, known as the commitment step, writes the block, and modifies the system state [38-40].

The comparison of the blockchain consensus protocols is shown in Table 2 [41-43]. Table 2 indicates that every methodology has pros and cons of its own. Some of them have been improved from this perspective and outperformed POW in terms of speed, energy consumption, and scalability. Of course, in these cases, security weaknesses, power distribution, or increased communication overhead are seen.

3.1.4. Blockchain generations

In general, three different generations are defined in the blockchain. In the first generation (the first form of blockchain), network miners expended much energy and money to confirm decentralized transactions without a central processing unit. A prolonged process that was completed peer-to-peer and without the use of middlemen. The Bitcoin cryptocurrency project, of course, had fewer modifications from this generation than the first, but second-generation blockchains have undergone major alterations since the first generation. Deploying apps with smart contracts and asset management on the blockchain became possible with the second-generation blockchains. The second generation's key features also include the ability to issue blockchain shares (along with money and smart wallets). In comparison to the prior generation, transaction speeds also significantly improved in this generation [44-48].

Low scalability and security issues were caused by high prices and slow speed. For this reason, new blockchains emerged to solve this problem under the title of the third generation. The critical features of this generation of blockchains are simplicity in data access, better scalability, and higher speed. The third generation of blockchains made it possible for blockchains to connect with one another, which was also impossible in the prior two generations. The layer definition is another characteristic of the third generation of blockchain technology, which has considerably increased both the speed and security of the system [44-46].

3.2. Smart contracts

On the blockchain network, smart contracts are deployed and carried out as automated applications that are programmed. If the required conditions (described in the code) are established in the blockchain network, this contract can operate autonomously. Naturally, smart contracts must pay a GAS fee or network usage fee to register information in the network to prevent excessive consumption of blockchain network resources. The information in the blockchain network is disseminated among all users, as was indicated in the previous section. It is no longer possible to update information once it has been published. In other words, once a smart contract has been executed and has reached its expiration, it cannot be

stopped. Gas fees are particularly important in reducing resource usage. To put it another way, they are employed to stop malicious users from abusing the blockchain network's limited resources by creating fictitious transactions to take up available memory, processing power, and other resources. Additionally, by using this strategy, network management can get financial resources that they can utilize to grow, enhance performance, or better administer the network. [1-2], [5], [49-50].

After programming in the blockchain and using the VM (or executive virtual machine), the codes for a smart contract are performed. Then, to interact with it, users must carry out a specific transaction. With the aid of this transaction, users sign the smart contract and validate its usage within the blockchain network using their private key and asymmetric encryption. After the smart contract has been signed and verified, users can interact with it and use it (for instance, purchasing the tokens provided by the smart contract using tokens associated with the native blockchain network). Usually, specialist programming languages like Solidity or Serpent are used to create smart contracts. Of course, common languages such as Python are also active in this field. These kinds of languages offer particular functionality utilized by blockchain in addition to the standard characteristics provided by programming languages (concepts like variables, functions, classes, computing operations, and strings) [1-2], [5]. Some of the most effective uses for smart contracts in use today are depicted in Figure 1.

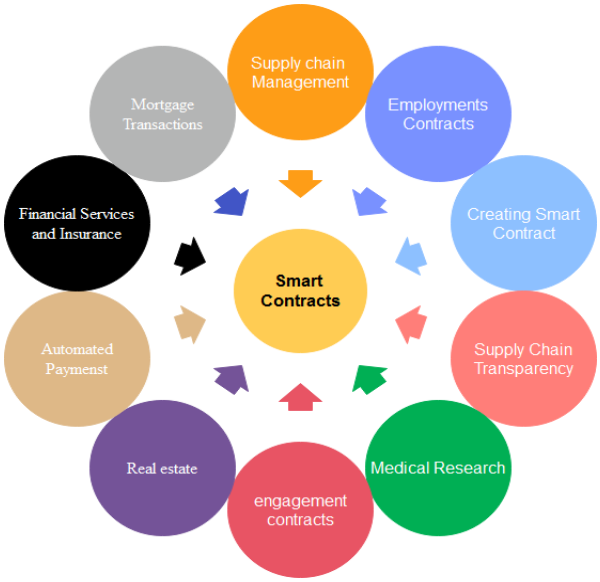


Figure 1. Some of the most important applications of smart contracts today [51]†

† All figures generated by Edraw Max Tool

Table 2. Comparison of consensus protocols

Consensus Protocol	Advantages	Disadvantages
POW	• High network security • Creating more employment	• High energy consumption • The need for high-processing equipment • High cost • The possibility of network control falling into the hands of a specific group, especially in smaller networks • Ability to control the network by a specific group • Moderate scalability
POS	• lower cost • Less energy consumption • Higher speed	• The possibility of greater vulnerability • Increase concentration • Ability to control the network by a specific group • Moderate scalability
DPOS	• Cost optimization and energy consumption • Increased processing speed	• Vulnerability to some network attacks due to increased communication
POA	• Lower transfer cost • Good security • High scalability	• Hard implementation • Good protection against denial of service attacks • Less distribution
PBFT	• Low energy consumption • High processing speed • Low cost	• High communication overhead • Vulnerability to some network attacks due to increased communication • vulnerable to denial of service attacks • High communication complexity

Table 3. shows the characteristics of different blockchain generations comparatively

Blockchain	Advantages	Sample projects	Disadvantages
Generation 1	Distributed transactions without intermediaries	Bitcoin	High cost, power consumption, low speed, poor scalability, and security issues
Generation 2	Better speed than 1st generation, distributed, transactionless transaction, smart contracts, smart money, and wallet.	Ethereum, Ethereum Classic, and Neo	Low scalability, security issues, lack of proper communication between blockchains
Generation 3	Distributedness, transaction without intermediaries, smart contracts, money and smart wallet, inter-blockchain communication, simplicity in data access, layered structure, better .security, and higher speed	Cardano, Algorand	Security issues, privacy issues, and .lack of global policies

3.3. Solidity programing language

One of the most well-known programming languages for smart contracts is called Solidity. This high-level, object-oriented language is compatible with Ethereum, Solana, and Binance Smart Chain, among other blockchains. In the Solidity programming language, a smart contract is made up of both data and code. Copies of the smart contract are placed on different network nodes, and it is given a public address (connected to public key cryptography). After that, users register their transaction for network execution by calling the contract using this address. Solidity smart contracts are carried out on the Ethereum network's virtual machine. With vast memory and processing resources, this virtual machine is a full Turing computer for mathematical calculations [16], [52-53].

3-4. Security Vulnerabilities

A vulnerability in smart contracts refers to a kind of security weakness that can satisfy the three rules in the list below. The existence of vulnerability will also be violated if any one of these three situations does not present [54]

- a flaw in the system that puts information security at risk.
- The potential for an attacker to access the pertinent defect
- The potential for an attacker to exploit that defect.

4. Vulnerabilities in smart contracts

In this section, we present the vulnerabilities in smart contracts along with the appropriate countermeasures. This section mostly addresses smart contract vulnerabilities. Things that are universal from a broader idea have not been included.

4.1. Re-entrancy vulnerability Check

4.1.1. Introduction of reentrance vulnerability

Reentrance vulnerability describes an external contract's ability under a cyber attacker's command to call contract functions. It may result in the depletion or intrusion of Ethereum (smart contract inventory). In other words, this vulnerability occurs if a smart contract uses the call, share, and send functions to transfer flow control to another smart contract, then updates the status of the primary contract using the callback function (in this case, the smart contract status may be "Not completed"). The smart contract is incomplete under the circumstances above (the callback function does not return the expected value), and the calling contract may run further smart contracts or functions based on the written codes. Pay attention to Sender's smart contract code [7] ,[55-57] for a better understanding.

Code snippet 1: Reentrancy Vulnerability

```
pragma solidity >=0.4.22 <0.6.0;
contract Sender {
uint public amount;
address payable public sender;
address payable public receiver;
 constructor() public payable {
 sender = msg.sender;
 amount = msg.value;
}
function send(receiver) payable {
receiver.call.value(value).gas(20317)();
}
}
contract Receiver {
uint public balance = 0;
function () payable {
balance += msg.value;
}
}
```

Ethereum is transmitted to the recipient address, and the callback function is activated when this contract's send function is used. To implement this function, the average consumption of GAS is 2300 [58], which is received even in the incorrect implementation. While the amount of gas in the transfer and call functions is altered or left unrestricted, they are similar to this function in other ways. It should be noted that GAS is not deducted in the transfer function when a problem occurs. code snippet 2 has a related vulnerability [55].

Code snippet 2: Reentrancy Vulnerability 2

```
function transferBalance(address receiver, uint
amount)
public {
require(balances[msg.sender] >= amount);
receiver.transfer(amount);
/* flow control transferred before the sender'
s balance is updated before an event
is emitted. Potentially the start of
trouble. */
balances[receiver] -= amount;
LogTransactions(msg.sender,receiver, amount);
                                                }
```

As can be seen, this contract's balance is transferred through the transfer function. After executing this function, the user's inventory status will be updated (in actuality, the contract status will change). By calling this method and modifying the callback function in this scenario, the malicious smart contract can empty the Ethereum balance of the smart contract [55]. When two functions or smart contracts share a state, a reentrancy attack can still succeed. For instance, the code below demonstrates this. Because of the value of balance[msg.

Sender] is not set to zero in this contract after the reentrancy attack. The transfer function in the code snippet 3 can use this variable to send additional Ethereum by calling the call function in the withdraw function [55]. The general layout of the reentrancy assault is depicted in Figure 2.

Code snippet 3: Reentrancy Vulnerability 3

```
mapping (address => uint) private balance;
function transfer(address to, uint amount) {
if (balance[msg.sender] >= amount){
balance[to] += amount;
balance[msg.sender] -= amount;
}
}
function withdraw() public {
uint amount = balance[msg.sender];
require(msg.sender.call.value(amount)());

/* At this point, the caller's code is
executed and can call transfer() */
balance[msg.sender] = 0;
```

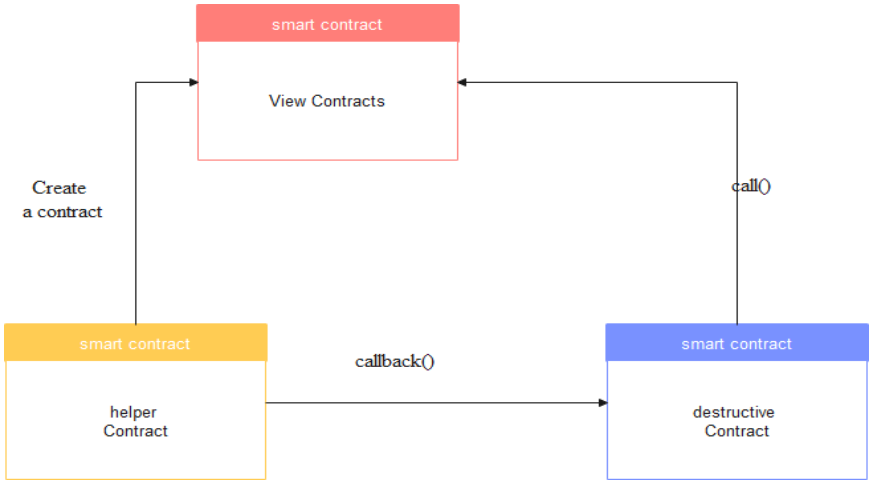


Figure 2. The overall structure of the Reentrancy attack [59]

4.1.2. The Solution to Reentrancy vulnerability

Changing the status or value of the influencing variables should be transferred before invoking the call, transmit, or send functions, as was stated, to address this issue. The reentrancy attack can no longer be successfully undertaken in this situation [55].

4.2. Investigating the vulnerability of access control in smart contracts

4.2.1. Access control vulnerability

User roles and privileges within an application are restricted using access controls. This notion in smart contracts can be connected to governance and crucial logic, including token issuance, proposal voting, money withdrawals, suspension and contract upgrades, ownership changes, etc. The access control mechanisms used by intelligent contracts have some flaws. The following can be listed as weaknesses among them [7], [9], [60].

Non-Validation of Modifiers: In critical functions, validating modifiers is crucial. The owner can be changed, tokens and existence can be transferred, and contracts can be stopped and canceled, among other things, thanks to modifier functions. Modifiers' access levels to key functions must be verified before use to avoid risks like money loss or contract termination.

Incorrect Modifier names: A Modifier's name or even the name of a function may be written differently than what is specified in the validation library due to programmer error. As a result, neither function nor modifier is subject to the modifier process anymore. Which, depending on performance, may result in a loss of cash or a change of ownership.

Having too many roles defined: Enabling users to have an excessive number of roles could result in an access control vulnerability.

Take a look at the example below for a better understanding.An access control problem that allowed anyone to burn Hosp tokens led to the hacking of the Hospowise project a while back. This smart contract's

susceptible token burning function is related to the burn function [61].

function _burn (account, amount); public; burn(address account, uint256 amount);

Put another way. A hacker might buy any token and then use the public write function. This function burns all of the Hospo tokens on the UniSwap exchange. The attacker can now trade his tokens for the Ethereum cryptocurrency because the token's value has increased [57].

4.2.2. Access control vulnerability solution

The process needing access control permissions must be reviewed to address this kind of vulnerability. This can be used by smart contract programmers with specialized access management libraries. The name of one of these libraries is OpenZeppelin. This library's Openzeppline's Ownable subsection offers modifiers like "only owner" to ensure that a function is called by its owner. It should be noted that other access control management sub-sections in OpenZeppelin employ modifiers or functions like "hasRole" to determine whether a user has the authority to invoke a function [62]. There are also a few appropriate models for controlling access in smart contracts [63].

4.3. Investigating arithmetic vulnerabilities

4.3.1. Introduction of arithmetic vulnerabilities

Smart contracts are extremely vulnerable to integer overflows. Unsigned integers are typically used in smart contracts, and most developers work with simple integer types (often only signed integers). Many code paths can become carriers for information theft or service denial when there is a buffer overflow [9-10], [64,65].

The below smart contract's ability to remove balances serves as a vivid illustration of this vulnerability. This function should permit withdrawals if the balance is greater than zero (and more significant than the minimum transaction cost). It is true even though it is possible to withdraw larger sums of money via the integer overflow vulnerability. In other words, the withdraw () function's check result is always positive, allowing the attacker to withdraw more than the permitted amount. Consider the following function [65] to gain a better understanding of the problem.

Code snippet 4: Arithmetic vulnerabilities

```
function withdraw(uint _amount) {
        require(balances[msg.sender] - _amount > 0);
        msg.sender.transfer(_amount);
        balances[msg.sender] -= _amount;
}
```

This method's sole criteria to determine whether the function caller has permission to withdraw funds balance is [msg. Sender] - _amount > 0. a procedure that enables free withdrawal in case of a buffer sequence vulnerability and can be bypassed.

The code snippet 5 has an arithmetic flaw and is connected to the Beauty Chain blockchain. A susceptible function named batch transfer is employed in the

Code snippet 5: Arithmetic vulnerabilities 2

BeautyChain smart contract, and it is in this function that the vulnerability mentioned above exists [66].

```
function batchTransfer(address[] _receivers, uint256
_value) public whenNotPaused returns (bool) {
uint cnt = _receivers.length;
uint256 amount = uint256 (cnt) * _value;
require(cnt > 0 && cnt <= 20);
require (_value > 0 && balances[msg.sender] >=
amount);
balances [msg.seneder] =
balances[msg.sender].sub(amount);
for (uint i=0; i<cnt; i++) {
balances[_receivers[i]] =
balances[_receivers[i]].add(_value);
Transfer(msg.sender, _receivers[i], _value)'
}
return true;
 }
}
```

The local variable value code determines the sum by multiplying the cnt and _value variables. A 256-bit integer may be used as the second parameter in this multiplication, designated as _value. Due to the batchTransfer() function's ability to accept two inputs, a hacker might overflow the amount variable and set it to zero by passing an extremely large number in the _value parameter. All checks on colored lines will succeed by setting this variable to zero, and the subsequent line's subtraction won't matter. Finally, the code will dump numerous recipients' balances with a very big value without incurring any fees to the attacker's account [66].

4-3-2. Solution of arithmetic vulnerabilities

Rewriting the susceptible sections of the code using the supplied libraries is a security fix for this vulnerability. SafeMath is one of these libraries, and it can be used to address arithmetic flaws. The safe functions add sub, mu, and div were used to construct the library's four core activities [67]. After including the SafeMath library in your smart contract, swap out all four major operations with the functions mentioned above to address the vulnerability. These alterations are listed below.

Code snippet 6: Arithmetic vulnerability Solution

a * b becomes a.mul(b)
a / b becomes a.div(b)
a - b becomes a.sub(b)
For example, the add function in this library is as follows.

```
function add(uint256 a, uint256 b) internal pure returns
(uint256) {
    return a + b;
  }
```

You can see that the result is returned as a pure function and uint256. It is guaranteed not to read or modify the status information in the pure function.

4.4. Not checking return values in low-level calls

4.4.1. Challenge return values in low-level calls

The availability of low-level Solidity functions like call(), call code (), delegate call (), and transmit() is one of the aspects of programming languages for smart contracts. In smart contract coding, it is often preferable to employ alternatives to low-level calls whenever available. Incorrect calling of these functions could result in the contract's security being jeopardized. These functions in Solidity behave entirely differently from the rest when it comes to accounting for errors. In other words, the error in a function of this kind is not even broadcast and does not result in a full return from the present execution. In other words, these functions only ever yield the logical value false, after which the Ethereum virtual engine is still executing the code. Consider the example in [9], [68] to better understand the problem.

Using the recipient's wallet address is the most straightforward approach to creating a contract to send Ethereum to another address in this blockchain network. The following conditional check (Code snippet 7) is a game component on a board where the winner receives money in the form of Ethereum digital currency [68].

Code snippet 7: Not checking return values in low-level calls

```
if (gameHasEnded && !( prizePaidOut ) ) {
  winner.send(1000); // send a prize to the winner
  prizePaidOut = True;
}
```

The issue is that the sending procedure—for instance, the send function in the Code snippet above—might not succeed. The prizePaidOut variable will be set to True, and the winner won't get any money. This value shouldn't be set to true because the correct transfer has not yet been made. The leading cause of this problem is that the programmer neglected to examine the send function's output. There are two situations in which the winner.send() function might not work as intended. First, an exception is made when the winning address relates to a contract rather than a user account. In this instance, the failure to mail the award is the winner's responsibility because they provided an erroneous address. In the second scenario, other contract programs (which have already been executed in the transaction) can use a finite resource called "Callstack" that is available on the Ethereum virtual machine. This operation will fail (and the winner's award will be forfeited) if this resource is used up before transmitting Ethereum to the winner's address. The smart contract must be appropriately secured to protect the winner. It should be mentioned that depending on the type of smart contract code, the results of this vulnerability may be extremely dangerous. For instance, by altering the values of the variables without providing any inventory, a cyberattacker could purposefully improve his access level to the smart contract [68-69].

4.4.2. Resolving return values in low-level calls

Two ideas are mentioned below [68] as a means of addressing this vulnerability:

Examining the results of low-level functions: This method uses vulnerable functions to check the output linked to vulnerable spots before using it if there are no issues. The following code snippet fragment provides a viable solution for the previously mentioned example using the same methodology.

Code snippet 8: Not checking return values in low-level call solution

```
if (gameHasEnded && !( prizePaidOut ) ) {
  if (callStackIsEmpty()) {
  if (winner.send(1000)){
    prizePaidOut = True;
}
  else throw;
} else throw;
}
```

In this example, the stack's emptiness is verified using the callStackIsEmpty method, and only if there are no issues is the sending procedure started and the prizePaidOut variable set. You should modify your code to ensure that the outcomes of unsuccessful sending are distinguished from other scenarios [68].

Utilizing exceptions Making use of exceptions is another method to address this problem. By specifying an exception in this method, the programmer can take control of unusual circumstances, including the results of not inspecting the return values in low-level calls. In other words, when the circumstances mentioned above occur, the program's execution flow departs from normal mode, and the code relating to the exception is run.

4.5. Investigation of denial of service attacks

4.5.1. Denial of service attacks

A denial of service attack is one in which the attacker sends the victim's system numerous input requests to deny service to the main users. Smart contracts may be permanently shut down following a denial-of-service

assault, in contrast to web services that can occasionally recover from these attacks. Denial of service attacks can be implemented in smart contracts in a variety of ways. The following list of instances can be listed among these [7], [70-71].

- Malicious behavior when receiving a transaction
- Artificial increase of Gas Fee necessary to calculate a function
- Abuse of access controls to access confidential parts of the smart contract

Take a look at the example below for a better understanding. The code snippet 9 relates to the blockchain game King of the Ether, which lets players buy access to the president through payments made in Ethereum to other players. The access transfer process will now fail if the other user is a smart contract, and the stated smart contract will always have the president's access authorization. Based on this, a denial of service attack has been conducted due to faulty access authorization management [70].

Code snippet 9: DOS attack

```
function becomePresident() payable {
   require(msg.value >= price); // must pay the price to
become president
   president.transfer(price);   // we pay the previous
president
   president = msg.sender;      // we crown the new
president
   price = price * 2;           // we double the price to
become president
}
```

Additionally, the smart contract may be terminated if the GAS Fee for a transaction exceeds the maximum amount allowed in a block because other network transactions will not be able to be accepted. A smart contract may unintentionally cause it to occur, as well as online hackers. Consider a cyber attacker who creates many destination addresses to steal modest amounts from the victim's smart contract. The following function is where this kind of assault can be carried out. The code even states that many current transactions may cause the smart contract to break at some time, stopping all future transactions. The function iteration in the loop is based on the current length of the address array. It can allow an attacker to exhaust resources by adding new return addresses to the array [70]. Additionally, since the refund function's transaction depends on the fixed sending function, which has a value of 2300 GAS, it is possible to artificially increase the GAS fee in this function.

Code snippet 10: DOS attack 2

```
address[] private refundAddresses;
mapping (address => uint) public refunds;
function refundAll() external onlyOwner {
// unknown length iteration based on how many addresses
participated
for(uint i; i < refundAddresses.length; i++) {
// doubly bad, now a single failure on send will hold up all
funds
require(refundAddresses[i].send(refunds[refundAddresses
[i]]))
}
}
```

4.5.2. A solution to Denial of Service attacks

The smart contract's codes should be securely examined, particularly in the control discussion, to defend against denial of service attacks (checking the actions of all departments and access permissions). Before deployment in the blockchain network, these codes must be rewritten if there are any issues. Additionally, the ability to interact with the smart contract should be limited, especially during procedures like the withdrawal of funds, and appropriate thresholds should be set to regulate the volume of requests. This procedure should be carried out within the block limit for the GAS Fee. Instead of using the transmit() method in certain circumstances, utilizing a transferFrom() function is preferable.

4.6. Failure to use appropriate random numbers

4.6.1. The challenge of generating random and pseudo-random numbers

In general, using actual random numbers is exceedingly challenging and impossible. Cyber attackers can also duplicate numbers and take advantage of the system, regardless of how far this process of creating random numbers is from reality. Despite the widespread use of random numbers (used in many programs such as lotteries, games, airdrops, etc.) and the direct link to financial issues, this issue is of utmost importance in smart contracts [7], [9]. Pay close attention to the lottery game example below to better understand the topic—a lottery based on random numbers where the winner receives cryptocurrency for their prize.

The DiceGame smart contract (Code snippet 11) serves as a prime illustration of this. The user must estimate the smart contract's random number in this code; if successful, he will be rewarded with one Ethereum. It is carried out using the random() function and the guess_the_dice() function. The random() function creates a random number by using the previous block's block number and the current block's timestamp [72].

Code snippet 11: Random Number Problem

```
contract DiceGame
{
constructor() payable{ }
```

```
function guess_the_dice(uint8 _guessDice) public {
uint8 dice = random();
if (dice == _guessDice) {
(bool sent, ) = msg.sender.call{value: 1 ether}("");
require(sent , "failed to transfer");
}
}
// source of randomness (1-6)
function random() private view returns (uint8) {
uint256 blockValue = uint256(blockhash(block.number-1
+ block.timestamp));
return uint8(blockValue % 5) + 1;
}
}
```

Looking at the smart contract's random function, it is obvious that a hacker could create this number using the following function and drain the balance of the account by guessing the winning lottery number [68].

Code snippet 12: Random Number Problem 2

```
function random() private view returns (uint8) {
uint256 blockValue = uint256(blockhash(block.number-1
+ block.timestamp));
return uint8(blockValue % 5) + 1;
}
```

4.6.2. The solution to the challenge of generating random and pseudo-random numbers

You should not use publicly accessible and predictable information, such as details about blocks (also known as Onchain) in the network, to defend against this attack. Additionally, it is preferable to generate random numbers using the provided libraries and special functions. One of the most crucial tools for this problem is libraries that offer VRF (Verifiable Random Function), offered by businesses like Chainlink. Smart contracts can produce statistically random numbers using these libraries. With one-way and asymmetric encryption, this method may verify the legitimacy of the random number's creator. It will cause anyone can identify the random numbers produced outside of the smart contract [73-74].

4.7. Forward transaction attacks

4.7.1. The challenge of forward transaction attacks

Transactions in the blockchain network are typically not immediately recorded. In fact, following the registration request, a collection of these transactions must be compiled into a block and come to a consensus in the network. Before choosing a block, all nodes in a network must be informed of transaction details due to the distributed nature of block creation. In other words, when a node in the blockchain creates a transaction, the pertinent information is broadcast to other nodes. After obtaining the indicated info by nodes, they place the transactional information in the "an unused pool" structure. The block creator node adds transactions to the block with the priority of the paid fee amount (the block creator seeks more profit) after adding a sufficient number of transactions (the size of writing a block) to this pool [64], [75-77]. Blockchain network users can prioritize their transactions by paying a more significant price. A cyber attacker can exploit this problem and elevates his transaction above other cases by charging a hefty fee for it. In smart contracts where timing is crucial (like NFT purchase competitions, first response competitions, etc.), it is possible to get around other users in this fashion [58-59]. The Findkey smart contract deserves your attention (Code snippet 13) [78].

Code snippet 13: Forward Transaction Attack

```
contract Findkey {
   bytes32 public constant key =

0x564ccaf7594d66b1eaaea24fe01f0585bf52ee70852af4ea
c0cc4b04711cd0e2;
   constructor() payable {}

   function guess(string memory solution) public {
      require(key ==
keccak256(abi.encodePacked(solution)), "Incorrect
answer");
      (bool sent, ) = msg.sender.call{value: 5 ether}("");
      require(sent, "Failed");
   }
}
```

Finding the value of a string will earn users 5 Ethereum in this smart contract. Let's say user number 1 is successful in figuring out this number and submits it in response to the guess function. The hacker in the blockchain network is currently looking at the transaction data in the "unused pool." He can immediately register it in the blockchain network with a greater registration charge after recognizing the amount in the transaction relating to User Number 1. By using the precedence of transactions in this way, it is possible to get around user number 1 and win 5 Ethereum in its place. You can see how to generally carry out the prior transaction assault in the Figure 3.

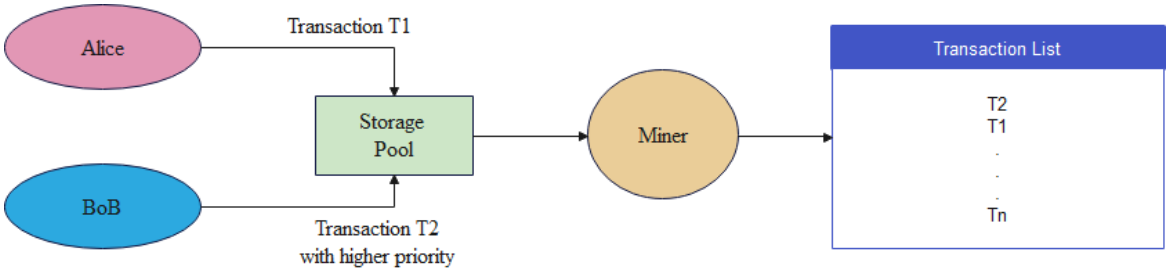


Figure 3. How to execute the front-end transaction attack [79]

4.7.2. Solution of forward transaction attacks

The use of the Commit-Reveal method is a viable answer to this issue. Using encryption techniques, user number 1 in the previous example sends his message locked and immutable over the network. After registering on the blockchain, he can unlock it. Because the attacker cannot see the transaction information before it is registered in the network, the forward transaction attack is no longer viable. The Submarine Send technique is another approach that experts have suggested to handle this predicament. This method involves sending the smart contract's related response to a third address in the blockchain network, encrypted using asymmetric encryption and viewable by the smart contract. The transaction details will then be made available to the smart contract by the sender, who will also disclose this transaction to it. Using this technique prevents the cyber attacker from seeing the transaction data in the collection of unused transactions [80-81].

4.8. Wrong time dependency vulnerability

4.8.1. The challenge of wrong time dependency vulnerability

The programmer may need to precisely record the time in smart contracts to use it in the program's logic. This value is called the time stamp. Any error in the definition of this topic could lead to a security flaw. This issue might arise, for instance, if the programmer utilizes a seal associated with the most recent block in the infrastructure blockchain network. One of the most difficult problems in this area is distributing an accurate clock among the nodes of a blockchain network. Therefore, the timing gap between nodes in a blockchain network is correct.

Additionally, this value may be modified by a node or smart contract [7], [9], [82-83]. Study the Code snippet for the EtherLotto smart contract to gain a better understanding of the problem [84]. This lottery application uses the smart contract code. Users must prepare a certain sum equal to the "competition ticket" and deposit it into the smart contract account to take part in this competition. Following (code snippet 14) the generation of a random number, if its value equals zero, the corresponding user will win the lottery. The matching award will then be sent to his address when this occurs.

Code snippet 14: Wrong Time Dependency

```
contract EtherLotto{
   uint constant TICKET_AMOUNT = 10;
   uint constant FEE_AMOUNT = 1;
   address public bank;
   uint public pot;
   function EtherLotto{ ()
      bank = msg.sender;
}
   function play() payable}
      assert(msg.value == TICKET_AMOUNT);
      pot += msg.value;
      var random = uint(sha3(block.timestamp)) % 2;
      if (random == 0){
         bank.transfer(FEE_AMOUNT);
         msg.sender.transfer(pot - FEE_AMOUNT);
         pot = 0;
}
}
```

The use of the block timestamp (block.timestamp) to produce a random value in the play function is the source of the issue in this smart contract. This timestamp is simple to change for the node to have the desired result. For instance, altering this number to 0 can result in the hacker winning the lottery.

Table 4. Changing the value of amount and address in the vulnerability of short addresses

4.8.2. Fixed the wrong time dependency vulnerability

Regarding how to address this vulnerability, there are two options. The first step is to adopt alternate techniques and eliminate vulnerable timestamps. Other secure techniques, such as Chainlink VRF, can be used to obtain the random number for the sample in the case above. Additionally, the timestamp can be obtained in another manner (for instance, by using JavaScript functions like the date() and converting it to the Unix timestamp standard or similar libraries) before comparing the two values. According to

the application's logic, the execution should be terminated if the difference exceeds a specific threshold.

4.9. Vulnerability of short addresses

4.9.1. The vulnerability challenge of short addresses

In general, the Ethereum virtual machine will append many zeros to the end of an address if it finds one shorter than the needed length. By altering the call function values and removing zeros from the end of the destination wallet address, a hacker can exploit this vulnerability [7,64,85] and make an unlawful credit withdrawal from the susceptible smart contract account. Consider the function Send(address, amount) to send a specific amount of cryptocurrency (amount) to the account associated with the address variable in the Ethereum blockchain network to better understand the topic. Below is a sample call to this function [64].
Send(0x1234...67890, 10).

The problem is how to deal with and manage the address and unit values in the smart contract before sending. Assume that these items are assigned values of 20 and 32 bytes in memory, respectively. Now, if a cyber attacker, instead of sending 20 bytes of the address, sends 19 bytes after removing the trailing zeros. With this, the API adds zeros to the amount section after coding. The binary equivalent of the address and the value are placed in the memory in a serialized form. A zero in the memory is added to the address part again (the zero corresponding to the beginning of the amount), but the amount part can be changed to multiples of the desired amount. Table 4 shows how this matter is. According to these values and the limitation in the number of zeros on the left side of the amount value in the memory, it is possible to add two of these zeros to the address value and turn the amount value into a multiple of the original request.

4.9.2. Short address vulnerability solution

Simply include a check function in the smart contract to address this problem. The size of the user's input is verified in this function. In the smart contract for the NonPayloadAttackableToken, in code snippet 15 an illustration of this review is provided. The transfer function's expectations for this checking are two 32-byte parameters and a 4-byte method signature [86].

Condition	Amount and address in memory
Normal Condition	0x3bdde1e9fbaef2579dd63e2abbf0be445ab93f000…001010
Execution of the attack	0x3bdde1e9fbaef2579dd63e2abbf0be445ab93f00…0101000

Code snippet 15: Short Address Attack

```
contract NonPayloadAttackableToken {
  modifier onlyPayloadSize(uint size) {
    assert(msg.data.length == size + 4);
    _;
  }
 function transfer(address _to, uint256 _value)
onlyPayloadSize(2 * 32) {
   // Do stuff
 }
}
```

4.10. Inventory lock vulnerability

4.10.1. Inventory lock vulnerability challenge

This vulnerability exists if it is feasible to deposit a balance of local blockchain network tokens to the smart contract address without being able to withdraw them. In fact, in the mentioned case, the deposit balance is locked to the contract, and the user can no longer withdraw it. The user may have forgotten to declare the withdrawal function, which can lead to this [9], [87-89]. Pay close attention to the code snippet 16 [90] to better understand the problem.

Code snippet16: Inventory lock vulnerability

```
contract Market {
   function deposit() payable {
   }

   function transfer() {
      uint y = msg.value;
   }
}
```

In this contract, there is only a deposit function, and if the token is deposited to the relevant address, withdrawing is no longer possible.

4.10.2. Inventory lock vulnerability solution

The answer to this vulnerability is simple—you must build a function that will withdraw the deposit balance using the program's logic.
The proposed vulnerabilities are organized into categories in Figure 4. The number of these vulnerabilities can undoubtedly rise. Additional categories, such as [91], are also offered. Unlike previous examples, the proposed taxonomy places the vulnerabilities in smaller groups and does not stop with a list of cases.

5. Statistical investigation of vulnerabilities in a dataset of Ethereum smart contracts

In this section, descriptive statistics are employed. Tables and graphs are examples of data visualization tools used in descriptive statistics, which facilitate analysis and understanding.

The findings of the analysis of a dataset of Ethereum smart contracts with public access to the source code are presented in this section concerning the pertinent vulnerabilities [92]. The prerequisites for the randomly chosen contracts were that they had an adequate number of transactions and the application had been active for at least the previous ten days. The average number of transactions, the time of the most recent transaction, and the total number of contracts are all displayed in Figure 5, the contracts in question. This figure indicates a lot of active transactions for the contracts indicated. The contracts, as mentioned earlier, also contain considerable cash resources.

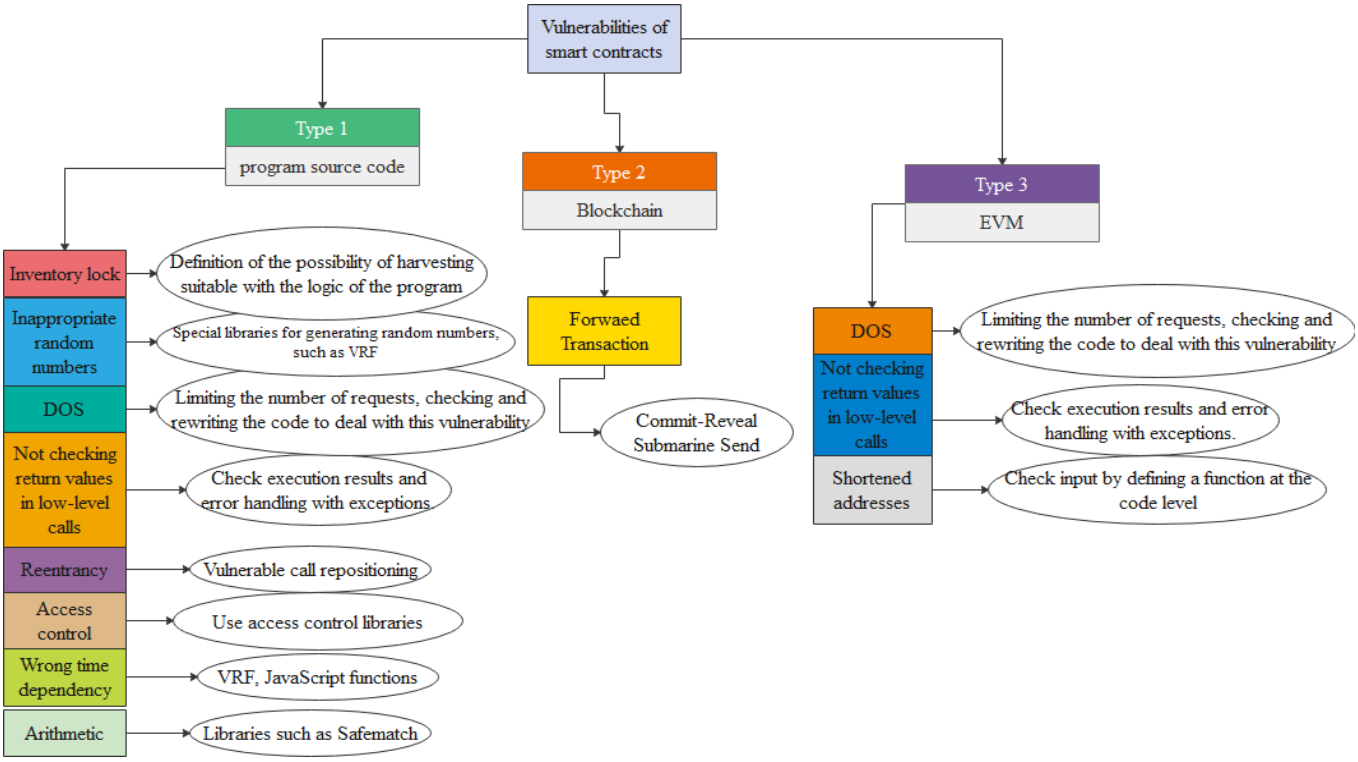


Figure 4. Classification of vulnerabilities related to smart contracts

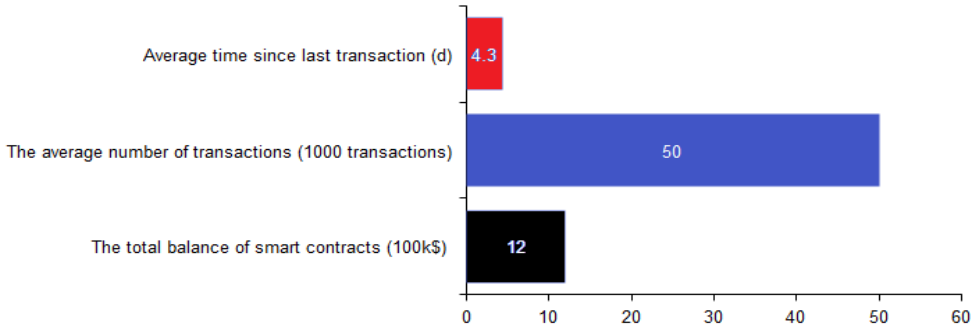


Figure 5. Information of smart contracts reviewed

Our review process includes static scanning at the code level with the help of four automatic vulnerability-checking tools: Slither, Smartcheck, Oyente, and Mythril [92-95]. In this review, a vulnerability warning in other tools is also considered with the help of observing the relevant outputs by each tool. In other words, some verification procedure has also been carried out for vulnerabilities with alerts in many tools. It should be recalled that a vulnerability requires three different circumstances, which were also discussed in sections 3-4. Only when employing the smart contract in the application, or, in other words, the final system, can all these things be investigated. As a result, some of these warnings might be inaccurate (the concept of a False Positive), or certain vulnerabilities might still be present in the system (the concept of a False Negative).

Slither: Slither is a smart contract static inspection tool that can sort the data in the code. This program creates an AST (Tree Syntax Abstract) structure to start, then pulls data from smart contracts like an inheritance graph, flow control graph, and list of Expressions to find the vulnerability. The vulnerability detection procedure is then performed using operations, including checking dependencies, variables, and accesses [93].

Smart check: a dynamic analysis tool for Solidity smart contracts that can be expanded. By transforming smart contract code into an intermediate XML representation and comparing it with patterns gathered from the actual world, this program finds weaknesses [94].

Oyente: A Symbolic review tool specifically made for analyzing smart contract code. This tool uses the CFG Builder (flow control graph maker) and Ethereum Blockchain Explorer modules to deliver the smart contract-related bytecodes in symbolic form to the central review module. This module's vulnerability detection procedure involves checking the smart contract's Symbolic output [95].

Mythril: An Ethereum virtual machine bytecode review tool that employs symbolic analysis, the SMT solver method, and the effect of different inputs on the program with the taint analysis approach to find smart contracts' weaknesses [96].

The reason for choosing these tools is to cover most of the introduced vulnerabilities. In addition, these tools must detect common vulnerabilities to be used in this investigation. Other tools, including Soda and Sfuzz [97-99], have also been published to check the security of smart contracts. However, these cases have been omitted due to the lack of direct coverage of the mentioned vulnerabilities or the lack of sharing in the list of detectable vulnerabilities.

In terms of percentage and diversity of vulnerability alerts, Figure 6 presents comparisons between smart contracts. As shown, at least one vulnerability warning exists at the code level for around a third of the extant smart contracts at the dataset level. This problem demonstrates both the significance of the issue and the lack of sufficient attention given by smart contract developers to the field's vulnerabilities. In Figure 7, you can also see the results of the vulnerability warning check in the data set by the vulnerability. According to the graph of arithmetic-type vulnerabilities, forward transactions and locked ether have the most repetitions among the examined data sets. Based on this, it can be said that inattention to the security of at least one calculation, as well as not considering (or not knowing) the definition of priorities in the infrastructure blockchain network or not defining a method for capital withdrawal from the smart contract are the most common vulnerabilities in smart contracts (at least in this data set). In this way, smart contract developers may employ an equal methodology.

This issue was validated by looking at the coding associated with the vulnerable smart contracts for the three vulnerabilities above. For instance, smart contracts that were flagged for having an arithmetic vulnerability employed unsecured routine methods to perform their computations. In addition, the outputs linked to the tools mentioned above still have many inaccuracies, as we discovered when manually reviewing the vulnerabilities. Put otherwise, and many pertinent reports were false positives since they could not be verified through manual examination.

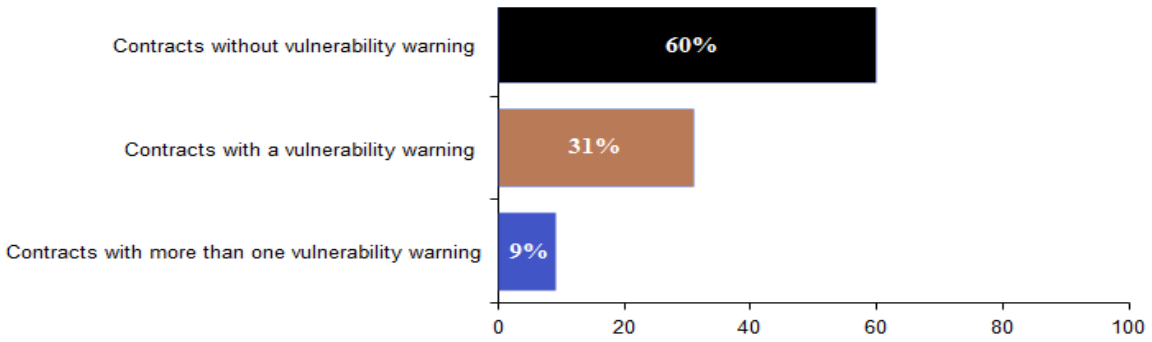


Figure 6. Comparison of smart contracts in terms of the number and type of vulnerability warning

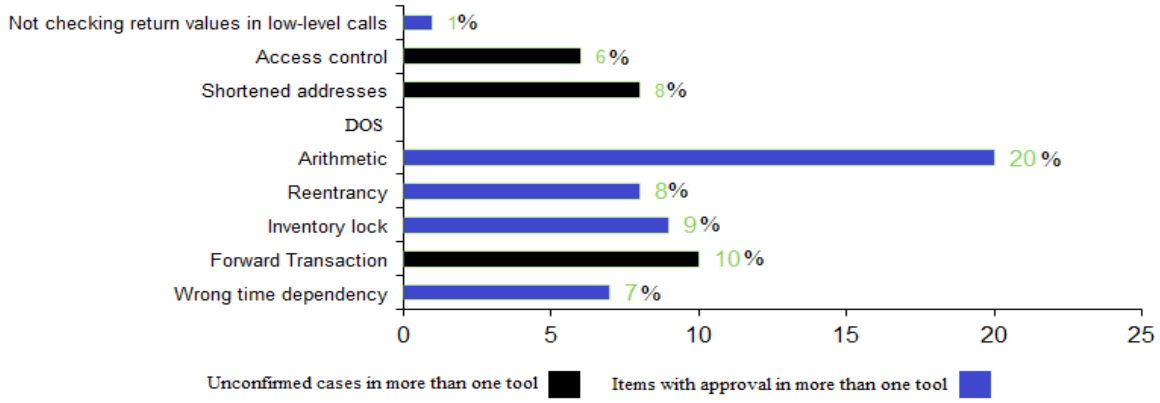


Figure 7. The results of checking the vulnerability alert in the dataset

In the chart below, you can see the mean, median and mode for alerts with confirmation in more than one tool.

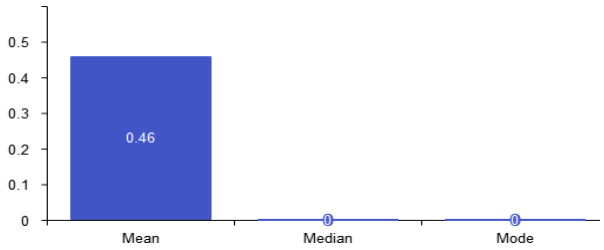


Figure 8. The mean, median and mode for alerts with confirmation in more than one tool

In Table 5, all tools that used in this section are compared.

Table 5. Comparison of automatic vulnerability detection tools

Tool name	Check method	The number of vulnerabilities to check from the list of 10 covered	Ability to access tool codes	Dynamic checking coverage
Slither	Intermediate Representation	8	yes (open source)	NO
Smartcheck	Intermediate Representation	6	yes (open source)	NO
Oyente	Symbolic Execution	4	yes (open source)	NO
Mythril	Symbolic Execution	8	yes (open source)	Yes

6. Conclusion and future work

Users can access the distributed infrastructure network through smart contracts in the blockchain sector. These agreements are crucial for information security since, in addition to their numerous uses, they directly impact the user's financial resources. Therefore, disclosing these contracts' vulnerabilities can be very helpful to blockchain application developers. In this study, the backdrop of the research and an explanation of a few technical words used in the field of blockchain technology were covered, along with an introduction to the issue and its significance.

In this study, the current vulnerabilities were thoroughly described, with real-world examples and practical solutions that were provided to fix them. Additionally, a statistical analysis of 100 active Ethereum smart contracts was done using the average results from four static security analysis tools. The Investigation's findings highlight the significance of the problem while also showing that several security weaknesses like arithmetic problem in the smart contracts under consideration recur. This represents a common response from developers to smart contract vulnerabilities. One of the primary causes of these problems in smart contracts may be the high complexity of large smart contracts and the developers' unfamiliarity with standard practices. With more contracts being evaluated and more (or more accurate and powerful) technologies being used, it is evident that the results will be more accurate. As a result, in further works based on this research, the number of vulnerabilities that have been identified would be raised, and testing can be carried out on a more considerable number of smart contracts. In addition, new automatic tools with various solutions are constantly being presented due to the spirit of active communities in this field. Future work should examine these tools' capabilities to finally identify why smart contract developers fail to pay attention to some widespread vulnerabilities. Better

training alternatives or ways to make smart contract programs less complicated for developers might be investigated further.

References

[1] Zou W, Lo D, Kochhar PS, Le XB, Xia X, Feng Y, Chen Z, Xu B. Smart contract development: Challenges and opportunities. IEEE Transactions on Software Engineering. 2019 Sep 24;47(10):2084-106.

[2] Wang S, Yuan Y, Wang X, Li J, Qin R, Wang FY. An overview of smart contract: architecture, applications, and future trends. In2018 IEEE Intelligent Vehicles Symposium (IV) 2018 Jun 26 (pp. 108-113). IEEE.

[3] Kushwaha, S. S., Joshi, S., Singh, D., Kaur, M., & Lee, H. N. (2022). Systematic review of security vulnerabilities in ethereum blockchain smart contract. *IEEE Access*, *10*, 6605-6621.

[4] Destefanis, G., Marchesi, M., Ortu, M., Tonelli, R., Bracciali, A., & Hierons, R. (2018, March). Smart contracts vulnerabilities: a call for blockchain software engineering?. In *2018 International Workshop on Blockchain Oriented Software Engineering (IWBOSE)* (pp. 19-25). IEEE.

[5] Sharma T, Zhou Z, Miller A, Wang Y. Exploring security practices of smart contract developers. arXiv preprint arXiv:2204.11193. 2022 Apr 24.

[6] Sifra EM. Security vulnerabilities and countermeasures of smart contracts: A survey. In2022 IEEE International Conference on Blockchain (Blockchain) 2022 Aug 22 (pp. 512-515). IEEE.

[7] Qian P, Liu Z, He Q, Huang B, Tian D, Wang X. Smart contract vulnerability detection technique: A survey. arXiv preprint arXiv:2209.05872. 2022 Sep 13.

[8] Singh, A., Parizi, R. M., Zhang, Q., Choo, K. K. R., & Dehghantanha, A. (2020). Blockchain smart contracts formalization: Approaches and challenges to address vulnerabilities. *Computers & Security*, *88*, 101654.

[9] Praitheeshan P, Pan L, Yu J, Liu J, Doss R. Security analysis methods on ethereum smart contract vulnerabilities: a survey. arXiv preprint arXiv:1908.08605. 2019 Aug 22.

[10] Wang Z, Jin H, Dai W, Choo KK, Zou D. Ethereum smart contract security research: survey and future research opportunities. Frontiers of Computer Science. 2021 Apr;15:1-8.

[11] Wohrer M, Zdun U. Smart contracts: security patterns in the ethereum ecosystem and solidity. In2018 International Workshop on Blockchain Oriented Software Engineering (IWBOSE) 2018 Mar 20 (pp. 2-8). IEEE.

[12] Zheng Z, Zhang N, Su J, Zhong Z, Ye M, Chen J. Turn the Rudder: A Beacon of Reentrancy Detection for Smart Contracts on Ethereum. arXiv preprint arXiv:2303.13770. 2023 Mar 24.

[13] Chen J, Huang M, Lin Z, Zheng P, Zheng Z. To healthier ethereum: A comprehensive and iterative smart contract weakness enumeration. arXiv preprint arXiv:2308.10227. 2023 Aug 20.

[14] Ray I. Security vulnerabilities in smart contracts as specifications in linear temporal logic (Master's thesis, University of Waterloo).

[15] He D, Deng Z, Zhang Y, Chan S, Cheng Y, Guizani N. Smart contract vulnerability analysis and security audit. IEEE Network. 2020 Jul 17;34(5):276-82.

[16] Yashavant CS, Kumar S, Karkare A. Scrawld: A dataset of real world ethereum smart contracts labelled with vulnerabilities. arXiv preprint arXiv:2202.11409. 2022 Feb 23.

[17] Ren M, Yin Z, Ma F, Xu Z, Jiang Y, Sun C, Li H, Cai Y. Empirical evaluation of smart contract testing: What is the best choice?. InProceedings of the 30th ACM SIGSOFT International Symposium on Software Testing and Analysis 2021 Jul 11 (pp. 566-579).

[18] Zhou H, Milani Fard A, Makanju A. The state of ethereum smart contracts security: Vulnerabilities, countermeasures, and tool support. Journal of Cybersecurity and Privacy. 2022 May 27;2(2):358-78.

[19] Prasad B. Vulnerabilities and attacks on smart contracts over blockChain. Turkish Journal of Computer and Mathematics Education (TURCOMAT). 2021 May 10;12(11):5436-49.

[20] Chen J, Xia X, Lo D, Grundy J, Luo X, Chen T. Defining smart contract defects on ethereum. IEEE Transactions on Software Engineering. 2020 Apr 20;48(1):327-45.

[21] Vani S, Doshi M, Nanavati A, Kundu A. Vulnerability Analysis of Smart Contracts. arXiv preprint arXiv:2212.07387. 2022 Dec 14.

[22] Durieux T, Ferreira JF, Abreu R, Cruz P. Empirical review of automated analysis tools on 47,587 ethereum smart contracts. InProceedings of the ACM/IEEE 42nd International conference on software engineering 2020 Jun 27 (pp. 530-541).

[23] Kushwaha SS, Joshi S, Singh D, Kaur M, Lee HN. Ethereum smart contract analysis tools: A systematic review. IEEE Access. 2022 Apr 22;10:57037-62.

[24] Wohrer M, Zdun U. Smart contracts: security patterns in the ethereum ecosystem and solidity. In2018 International Workshop on Blockchain Oriented Software Engineering (IWBOSE) 2018 Mar 20 (pp. 2-8). IEEE.

[25] Zheng Z, Xie S, Dai HN, Chen X, Wang H. Blockchain challenges and opportunities: A survey. International journal of web and grid services. 2018;14(4):352-75.

[26] Lashkari B, Musilek P. A comprehensive review of blockchain consensus mechanisms. IEEE Access. 2021 Mar 12;9:43620-52.

[27] Zheng Z, Xie S, Dai H, Chen X, Wang H. An overview of blockchain technology: Architecture, consensus, and future trends. In2017 IEEE international congress on big data (BigData congress) 2017 Jun 25 (pp. 557-564). Ieee.

[28] Gervais A, Karame GO, Wüst K, Glykantzis V, Ritzdorf H, Capkun S. On the security and performance of proof of work blockchains. InProceedings of the 2016 ACM SIGSAC conference on computer and communications security 2016 Oct 24 (pp. 3-16).

[29] Bentov I, Lee C, Mizrahi A, Rosenfeld M. Proof of activity: Extending bitcoin's proof of work via proof of stake [extended abstract] y. ACM SIGMETRICS Performance Evaluation Review. 2014 Dec 8;42(3):34-7.

[30] Joshi S. Feasibility of proof of authority as a consensus protocol model. arXiv preprint arXiv:2109.02480. 2021 Aug 30.

[31] Ekparinya P, Gramoli V, Jourjon G. The attack of the clones against proof-of-authority. arXiv preprint arXiv:1902.10244. 2019 Feb 26.

[32] Manolache MA, Manolache S, Tapus N. Decision making using the blockchain proof of authority consensus. Procedia Computer Science. 2022 Jan 1;199:580-8.

[33] Singh PK, Singh R, Nandi SK, Nandi S. Managing smart home appliances with proof of authority and blockchain. InInnovations for Community Services: 19th International

Conference, I4CS 2019, Wolfsburg, Germany, June 24-26, 2019, Proceedings 19 2019 (pp. 221-232). Springer International Publishing.
[34] Saad SM, Radzi RZ. Comparative review of the blockchain consensus algorithm between proof of stake (pos) and delegated proof of stake (dpos). International Journal of Innovative Computing. 2020 Nov 19;10(2).
[35] Yang F, Zhou W, Wu Q, Long R, Xiong NN, Zhou M. Delegated proof of stake with downgrade: A secure and efficient blockchain consensus algorithm with downgrade mechanism. IEEE Access. 2019 Aug 14;7:118541-55.
[36] Hu Q, Yan B, Han Y, Yu J. An improved delegated proof of stake consensus algorithm. Procedia Computer Science. 2021 Jan 1;187:341-6.
[37] Snider M, Samani K, Jain T. Delegated proof of stake: features & tradeoffs. Multicoin Cap. 2018 Mar 2;19:1-9.
[38] Castro M, Liskov B. Practical byzantine fault tolerance and proactive recovery. ACM Transactions on Computer Systems (TOCS). 2002 Nov 1;20(4):398-461.
[39] Abraham I, Gueta G, Malkhi D, Alvisi L, Kotla R, Martin JP. Revisiting fast practical byzantine fault tolerance. arXiv preprint arXiv:1712.01367. 2017 Dec 4.
[40] Gao S, Yu T, Zhu J, Cai W. T-PBFT: An EigenTrust-based practical Byzantine fault tolerance consensus algorithm. China Communications. 2019 Dec;16(12):111-23.
[41] Consensus Algorithms in Blockchain Systems [Internet]. DEV Community. 2020 [cited 2024 Feb 3]. Available from: https://dev.to/akroutihamza/consensus-algorithms-in-blockchain-systems-44ag
[42] Makhdoom I, Abolhasan M, Ni W. Blockchain for IoT: The challenges and a way forward. InICETE 2018-Proceedings of the 15th International Joint Conference on e-Business and Telecommunications 2018 Jan 1.
[43] Lang D, Friesen M, Ehrlich M, Wisniewski L, Jasperneite J. Pursuing the vision of Industrie 4.0: Secure plug-and-produce by means of the asset administration shell and blockchain technology. In2018 IEEE 16th International Conference on Industrial Informatics (INDIN) 2018 Jul 18 (pp. 1092-1097). IEEE.
[44] 1.The Blockchain Generations [Internet]. Ledger. Available from: https://www.ledger.com/academy/blockchain/web-3-the-three-blockchain-generations
[45] Anwar S, Anayat S, Butt S, Butt S, Saad M. Generation Analysis of Blockchain Technology: Bitcoin and Ethereum. International Journal of Information Engineering & Electronic Business. 2020 Aug 1;12(4).
[46] Efanov D, Roschin P. The all-pervasiveness of the blockchain technology. Procedia computer science. 2018 Jan 1;123:116-21.
[47] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. Decentralized business review. 2008 Oct 31.
[48] Rostami M, Bahaghighat M, Zanjireh MM. Bitcoin daily close price prediction using optimized grid search method. Acta Universitatis Sapientiae, Informatica. 2021;13(2):265-87.
[49] Brighente A, Conti M, Kumar S. Extorsionware: Exploiting smart contract vulnerabilities for fun and profit. arXiv preprint arXiv:2203.09843. 2022 Mar 18.
[50] Egbertsen W, Hardeman G, van den Hoven M, van der Kolk G, van Rijsewijk A. Replacing paper contracts with Ethereum smart contracts. Semantic Scholar. 2016 Jun 10;35:1-35.
[51] Top Smart Contract Applications and Use Cases - Scalable Solutions [Internet]. 2021. Available from: https://scalablesolutions.io/news/smart-contract-applications-and-use-cases/
[52] Dannen C. Introducing Ethereum and solidity. Berkeley: Apress; 2017.
[53] Zhang P, Xiao F, Luo X. A framework and dataset for bugs in ethereum smart contracts. In2020 IEEE International Conference on Software Maintenance and Evolution (ICSME) 2020 Sep 28 (pp. 139-150). IEEE.
[54] Krsul I, Spafford E, Tripunitara M. Computer vulnerability analysis. COAST Laboratory, Purdue University, West Lafayette, IN, Technical Report. 1998 May 6.
[55] Samreen NF, Alalfi MH. Reentrancy vulnerability identification in ethereum smart contracts. In2020 IEEE International Workshop on Blockchain Oriented Software Engineering (IWBOSE) 2020 Feb 18 (pp. 22-29). IEEE.
[56] Mehar MI, Shier CL, Giambattista A, Gong E, Fletcher G, Sanayhie R, Kim HM, Laskowski M. Understanding a revolutionary and flawed grand experiment in blockchain: the DAO attack. Journal of Cases on Information Technology (JCIT). 2019 Jan 1;21(1):19-32.
[57] Grossman S, Abraham I, Golan-Gueta G, Michalevsky Y, Rinetzky N, Sagiv M, Zohar Y. Online detection of effectively callback free objects with applications to smart contracts. Proceedings of the ACM on Programming Languages. 2017 Dec 27;2(POPL):1-28.
[58] Prechtel D, Groß T, Müller T. Evaluating spread of 'gasless send'in ethereum smart contracts. In2019 10th IFIP international conference on new technologies, mobility and security (NTMS) 2019 Jun 24 (pp. 1-6). IEEE.
[59] Oualid Z, Oualid Z. What is a reentrancy attack in Solidity? | Technical examples [Internet]. Get Secure World. 2022. Available from: https://www.getsecureworld.com/blog/what-is-a-reentrancy-attack-in-solidity-technical-examples/
[60] Samreen NF, Alalfi MH. A survey of security vulnerabilities in ethereum smart contracts. arXiv preprint arXiv:2105.06974. 2021 May 14.
[61] Samreen NF, Alalfi MH. A survey of security vulnerabilities in ethereum smart contracts. arXiv preprint arXiv:2105.06974. 2021 May 14.
[62] Palladino S. The parity wallet hack explained. July-2017.[Online]. Available: https://blog. zeppelin. solutions/on-the-parity-wallet-multisighack-405a8c12e8f7. 2017 Jul 20.
[63] Wöhrer M, Zdun U. Design patterns for smart contracts in the ethereum ecosystem. In2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) 2018 Jul 30 (pp. 1513-1520). IEEE.
[64] DASP - TOP 10 [Internet]. www.dasp.co. [cited 2024 Feb 3]. Available from: https://www.dasp.co
[65] Khan ZA, Namin AS. Ethereum smart contracts: Vulnerabilities and their classifications. In2020 IEEE International Conference on Big Data (Big Data) 2020 Dec 10 (pp. 1-10). IEEE.
[66] Thanh LY. Prevent Integer Overflow in Ethereum Smart Contracts [Internet]. Medium. 2018 [cited 2024 Feb 3]. Available from: https://yenthanh.medium.com/prevent-integer-overflow-in-ethereum-smart-contracts-a7c84c30de66
[67] Gao J, Liu H, Liu C, Li Q, Guan Z, Chen Z. Easyflow: Keep ethereum away from overflow. In2019 IEEE/ACM 41st International Conference on Software Engineering:

Companion Proceedings (ICSE-Companion) 2019 May 25 (pp. 23-26). IEEE.
[68] Scanning Live Ethereum Contracts for the "Unchecked-Send" Bug [Internet]. Hacking Distributed. Available from: https://hackingdistributed.com/2016/06/16/scanning-live-ethereum-contracts-for-bugs/
[69] Atzei N, Bartoletti M, Cimoli T. A survey of attacks on ethereum smart contracts (sok). InPrinciples of Security and Trust: 6th International Conference, POST 2017, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2017, Uppsala, Sweden, April 22-29, 2017, Proceedings 6 2017 (pp. 164-186). Springer Berlin Heidelberg.
[70] Kulkarni Y. Denial of Service (DoS) Attack on Smart Contracts [Internet]. Be on the Right Side of Change. 2022. Available from: https://blog.finxter.com/denial-of-service-dos-attack-on-smart-contracts/
[71] Bhardwaj A, Shah SB, Shankar A, Alazab M, Kumar M, Gadekallu TR. Penetration testing framework for smart contract blockchain. Peer-to-Peer Networking and Applications. 2021 Sep;14:2635-50.
[72] Smart Contract Randomness or ReplicatedLogic Attack – Be on the Right Side of Change [Internet]. 2023 [cited 2024 Feb 3]. Available from: https://blog.finxter.com/randomness-or-replicatedlogic-attack-on-smart-contracts/
[73] Yao S, Zhang D. An Anonymous Verifiable Random Function with Applications in Blockchain. Wireless Communications and Mobile Computing. 2022 Apr 19;2022.
[74] Verifiable Random Function (VRF) - Explained | Chainlink [Internet]. chain.link. [cited 2024 Feb 3]. Available from: https://blog.chain.link/verifiable-random-function-vrf/
[75] Behnke R. What Is a Front-Running Attack? [Internet]. www.halborn.com. 2021 [cited 2024 Feb 3]. Available from: https://halborn.com/what-is-a-front-running-attack/
[76] Frontrunning - Ethereum Smart Contract Best Practices [Internet]. consensys.github.io. Available from: https://consensys.github.io/smart-contract-best-practices/attacks/frontrunning/
[77] Mense A, Flatscher M. Security vulnerabilities in ethereum smart contracts. InProceedings of the 20th international conference on information integration and web-based applications & services 2018 Nov 19 (pp. 375-380).
[78] ImmuneBytes. A Techno-Manual on the Front Running Attack - ImmuneBytes [Internet]. 2022 [cited 2024 Feb 3]. Available from: https://www.immunebytes.com/blog/front-running-attack/
[79] Front-running attack in DeFi applications - how to deal with it? [Internet]. Securing. 2022. Available from: https://www.securing.pl/en/front-running-attack-in-defi-applications-how-to-deal-with-it/
[80] Libsubmarine.org. 2022. Available from: https://libsubmarine.org/
[81] Arulprakash M, Jebakumar R. Commit-reveal strategy to increase the transaction confidentiality in order to counter the issue of front running in blockchain. InAIP Conference Proceedings 2022 Aug 26 (Vol. 2460, No. 1). AIP Publishing.
[82] Dika A, Nowostawski M. Security vulnerabilities in ethereum smart contracts. In2018 IEEE international conference on Internet of Things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE Smart Data (SmartData) 2018 Jul 30 (pp. 955-962). IEEE.
[83] Tang X, Zhou K, Cheng J, Li H, Yuan Y. The vulnerabilities in smart contracts: A survey. InAdvances in Artificial Intelligence and Security: 7th International Conference, ICAIS 2021, Dublin, Ireland, July 19-23, 2021, Proceedings, Part III 7 2021 (pp. 177-190). Springer International Publishing.
[84] Ethereum Contract Diff Checker [Internet]. etherscan.io. [cited 2024 Feb 3]. Available from: https://etherscan.io/contractdiffchecker?a1=0xa11e4ed59dc94e69612f3111942626ed513cb172
[85] Zhu H, Niu W, Liao X, Zhang X, Wang X, Li B, He Z. Attacker Traceability on Ethereum through Graph Analysis. Security and Communication Networks. 2022 Jan 27;2022.
[86] CoinFabrik. Smart Contract Short Address Attack Mitigation Failure [Internet]. CoinFabrik. 2017 [cited 2024 Feb 3]. Available from: https://blog.coinfabrik.com/smart-contract-short-address-attack-mitigation-failure/
[87] Perez D, Livshits B. Smart contract vulnerabilities: Vulnerable does not imply exploited. In30th USENIX Security Symposium (USENIX Security 21) 2021 (pp. 1325-1341).
[88] Perez D, Livshits B. Smart contract vulnerabilities: Vulnerable does not imply exploited. In30th USENIX Security Symposium (USENIX Security 21) 2021 (pp. 1325-1341).
[89] Sayeed S, Marco-Gisbert H, Caira T. Smart contract: Attacks and protections. IEEE Access. 2020 Jan 30;8:24416-27.
[90] Bug Security : Locked Ether · Issue #19930 · ethereum/go-ethereum [Internet]. GitHub. [cited 2024 Feb 3]. Available from: https://github.com/ethereum/go-ethereum/issues/19930
[91] Smart Contract Weakness Classification (SWC) [Internet]. swcregistry.io. [cited 2024 Feb 3]. Available from: https://swcregistry.io
[92] SmartCDS/Addresses.txt at main · Csreasercher/SmartCDS [Internet]. GitHub. [cited 2024 Feb 3]. Available from: https://github.com/Csreasercher/SmartCDS/blob/main/Addresses.txt
[93] Feist J, Grieco G, Groce A. Slither: a static analysis framework for smart contracts. In2019 IEEE/ACM 2nd International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB) 2019 May 27 (pp. 8-15). IEEE.
[94] Tikhomirov S, Voskresenskaya E, Ivanitskiy I, Takhaviev R, Marchenko E, Alexandrov Y. Smartcheck: Static analysis of ethereum smart contracts. InProceedings of the 1st international workshop on emerging trends in software engineering for blockchain 2018 May 27 (pp. 9-16).
[95] Luu L, Chu DH, Olickel H, Saxena P, Hobor A. Making smart contracts smarter. InProceedings of the 2016 ACM SIGSAC conference on computer and communications security 2016 Oct 24 (pp. 254-269).
[96] MythX: Preparing for a smart contract audit [Internet]. mythx.io. [cited 2024 Feb 3]. Available from: https://mythx.io/about
[97] Chen T, Cao R, Li T, Luo X, Gu G, Zhang Y, Liao Z, Zhu H, Chen G, He Z, Tang Y. SODA: A Generic Online Detection Framework for Smart Contracts. InNDSS 2020 Feb 23.
[98] Nguyen TD, Pham LH, Sun J, Lin Y, Minh QT. sfuzz: An efficient adaptive fuzzer for solidity smart contracts. InProceedings of the ACM/IEEE 42nd International Conference on Software Engineering 2020 Jun 27 (pp. 778-788).

[99] Chen J, Xia X, Lo D, Grundy J, Luo X, Chen T. Defectchecker: Automated smart contract defect detection by analyzing evm bytecode. IEEE Transactions on Software Engineering. 2021 Jan 27;48(7):2189-207.

Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
(Print copies up to 2016)
e-ISSN: 2395-3454

Vol. 23 | No. 4 | pp. 2143-2155 | 2024

Review Research Paper

https://doi.org/10.46488/NEPT.2024.v23i04.019

Open Access Journal

Microplastics in Agricultural Soil and Their Impact: A Review

P. Solanki[1†], S. Jain[1], R. Mehrotra[2], P. Mago[2] and S. Dagar[3]

[1]IGNOU, Dyal Singh College, Delhi, India
[2]Shaheed Rajguru College of Applied Science for Women, University of Delhi, Delhi, India
[3]Department of Environmental Engineering, Delhi Technological University, Delhi, India
†Corresponding author: Pooja Solanki; solankipooja8@gmail.com

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 19-03-2024
Revised: 07-05-2024
Accepted: 18-05-2024

Key Words:
Microplastics
Agricultural soil
Sludge
Sewage
Microbes
Soil health

ABSTRACT

The rapid global plastic production of 348 million tonnes in 2018 has led to widespread environmental pollution, especially in terrestrial ecosystems. This study examines microplastics in agricultural soils, coming alarmingly. Particles ≤5 mm, which are defined as microplastics, have detrimental effects on the earth's environment. Because of its ecological importance, soil acts as an important microplastic sink, affecting soil and plant health and microbial activity. A variety of factors contribute to microplastic pollution in agricultural soils, including plastic mulching, manure, agricultural products (silage nets, twine), sewage sludge, weathering, and other indirect processes. These microplastics migrate, threatening soil integrity and biodiversity. Soil microplastics are analyzed for size, volume fraction, and polymer. Common materials include polyethylene, polypropylene, polyamide, polystyrene, polyvinyl chloride, and polyesters. Techniques, including optical microscopy and spectroscopy, extract and analyze microplastics. This comprehensive review calls for increased concern about the ecological effects of microplastics in agricultural soils. It emphasizes the importance of managing plastics to solve environmental challenges. The integrated environmental assessment highlights the complex relationship between microplastics and soil ecosystems, providing insights into potential risks and suggesting strategies to combat this looming environmental threat.

INTRODUCTION

Plastic is a flexible, long-lasting, and cost-effective material used in a variety of important industries such as packaging, electronics, agricultural production, etc. (Plastic Europe 2018). The widespread use of these synthetic materials increased manufacturing, resulting in a huge amount of plastic litter in the environment (Geyer et al. 2017). In 2018, the anticipated global plastic production was 348 million tonnes (Plastic Europe 2018). Around ≥6000 Mt of plastic garbage was generated in 2015, with around 80% of the material ending up being dumped directly into the environment that finds its way to landfills (Geyer et al. 2017). It is also observed that around 8078 Mt of Plastic has been generated in the last 50 years between 1950 and 2020 (Plastic Europe 2021). Plastic waste is ubiquitous and has been discovered in a variety of environmental compartments (de Souza Machado et al. 2018a), where it is subjected to increasing fragmentation caused mostly by thermo-oxidation, photo-oxidation, UV light, and mechanical abrasion (Wang et al. 2019, Hayes 2019 & Da Costa et al. 2019). Unfortunately, the fragmentation process does not entirely disintegrate the plastic waste but instead transforms it into a plethora of fine-sized plastic particles, encompassing Microplastics defined as 5 mm in diameter (Arthur et al. 2009, Thompson et al. 2004).

CHARACTERISTIC FEATURES OF MICROPLASTICS

The sink of Microplastics discovered in the soil can indicate localized usage, adjacent artificial activities, or atmospheric deposition, but the soil is a potential sink for Microplastics from different sources (Forster et al.2020). The capacity to understand the formation of Microplastics in soil, where the distribution distance on both temporal and spatial scales, among the point of origin and the sampling site, may be shorter than that in atmospheric transfer, is made possible by correlations among different kinds and occurrences of Microplastics discovered and regional human activities (Eerkes-Medrano et al. 2018, Luo et al. 2020). Additionally, the effects of different types of Microplastics on terrestrial systems vary (de Souza Machado et al. 2019, Lozano et al. 2020).

Shape

According to researchers (Xu et al. 2020, Yang et al. 2021), the Microplastics seen may be divided into numerous shapes, including pellets, fragments, foam, fiber, and film. According to a laboratory weathering experiment, PP pellets may create 6084±1061 particles after being subjected to UV light for 12 months and mechanical abrasion for 2 months (Yang et al. 2021).

In addition to being the master batch for products like industrial pellets, pellet particles are also connected to personal care items like cosmetics and cleaning supplies (Xu et al. 2020). Additionally, a new class called "fiber balls" is created since fibers frequently tangle and form balls. Such a fiber ball often comes in bundles and is made up of fibers in various colors (Weber & Opp 2020).

Size Fractions

The definition of Microplastics by researchers, as well as the sensitivity of the extraction and analytical techniques utilized, makes determining the size of Microplastics fairly challenging. The concept of Microplastics has become clearer as a result of an early study that focused on their presence in aquatic ecosystems. Microplastics are known to have a size of less than 5 mm. However, the sampling, pretreatment, and identification procedures directly affect the minimum particle size (Yang et al. 2021, Weber & Opp. 2020). In soil environments, Microplastics with minute particle sizes (≤ 1 mm) have been commonly reported.

Polymer Identification

According to the study of Yang et al. (2021), the presence of Microplastics is crucial for determining the source of contamination. Polyethylene, polypropylene, polyvinyl chloride, polyamide, polystyrene, and polyester are the most prevalent polymers found in soil. Polyethylene and polypropylene are the most commonly found in soil.

Microplastic pollution has recently received a lot of attention both from the general public as well as scientific communities all over the world, with a focus on aquatic settings, particularly the marine environment (Hidalgo-Ruz et al. 2012, Auta et al. 2017). The availability of Microplastics in the oceans has been linked mostly to ongoing inputs and also the degradation and fragmentation of large plastic litter (Hidalgo-Ruz et al. 2012), the vast majority of which are emitted from land (Rezania et al. 2018).

Microplastics in aquatic conditions may be consumed by Oligochaeta, crustacea, mollusks, nematodes, and vertebrates (Jambeck et al. 2015, Desforges et al. 2015, Hurley et al. 2007, Lei et al. 2018b). Microplastics and plastic-derivative compounds, such as plastic additives and adhered contaminants, have been linked to diversified toxicological effects that include inflammatory responses, metabolic disorders, stunted growth and reproduction issues, and other lethal issues (Hurley et al. 2007). Such circumstances are also likely for soil biota. Microplastics have been a research-intensive subject in aquatic environments for over a decade (Van Cauwenberghe et al.2015, Akdogan & Guven et al. 2019, Rillig 2018, Cole et al. 2014, Riilig & Bonkowski 2018, Cozar et al. 2014, Ivleva et al. 2017).

Microplastics accumulate more in terrestrial soil than in aquatic habitats (Zhao et al. 2018). As per UNEP reports, substantial amounts of particle plastics observed in the marine environment worldwide originate from land-based sources (UNEP, 2016). According to Rezania (2018), around 4.8 - 12.7 Mt of terrestrial plastic garbage is found entering the ocean every year, accounting for approximately 1.7-4.6% of total plastic waste generated globally. Sediment transmission during soil erosion is an event that permits particle plastics to be transported from terrestrial to aquatic habitats. Despite this connection to terrestrial resources, numerous scientific studies on the particles of plastic have overlooked the consequences of these synthetic materials (Bolan & Bradney 2019). Given that the majority of plastic waste is generated and discharged on land, it is surprising that Microplastics research has only recently begun to focus on terrestrial systems, where soil appears as a long-term sink for Microplastics debris (Kumar et al. 2020, Moller et al. 2020, Rochman 2018). Terrestrial domains, such as soils, are more vulnerable to plastic contamination than the oceans. According to Nizzetto et al. (2016), the annual input of Microplastics from sewage and wastewater treatment sludge on agricultural fields could exceed the total amount of Microplastics currently floating in the global oceans. Although the underlying mechanisms are unknown, preliminary data suggest that the presence of Microplastics in soils may impact the soil properties, plant performance, and microbial activity (de Souza Machado et al. 2019).

Recent research has discovered a considerable amount of filamentous and fragmented Microplastics in soils all around the world (Zhao et al. 2018). For example, recently discovered fragmented-dominated Microplastics in agricultural soils, where sewage sludge application promotes a Microplastics buildup (van den Berg et al. 2020). Microplastics accumulated in the soil can be easily taken up by plants and transferred through the food chain (Guo et al. 2020).

Although the genesis and potential translocation pathways of Microplastics in soil are varied, including the use of sewage-generated sludge and organic compost (Huerta

This publication is licensed under a Creative Commons Attribution 4.0 International License

Lwanga et al. 2017), irrigation (Blasing & Amelung 2018), plastic mulching (Yang et al. 2021), littering (Akdogan & Given 2019), and atmospheric deposition (Allen et al. 2019), urban soil used for agriculture is more prone to Microplastics pollution since they are often exposed to Microplastics (Moller et al. 2020, Chase et al. 2018).

Microplastics have been shown to harm soil health and function (de Souza Machado et al. 2018b, Liu et al. 2018), as well as in marine environments. Contamination will certainly result in inadvertent Microplastics ingestion by soil fauna. Worms, including Earthworms and ringworms, have been shown to consume Microplastics, with the rate increasing substantially as the amount of Microplastics rises; for example, Huerta Lwanga et al. (2017) detected around 14.8-28.8 Microplastics particles/gram of earthworm casts and 129.8- 82.3 particles/grams of chicken feces in home garden soils. Panebianco et al. (2019) discovered Microplastics in the majority of the snails (a total of 425 specimens), with an average of 0.92-1.21 particles/5 snails.

Plastic trash, particularly biodegradable plastics, is more prone to physical fragmentation than decay by mineralization, resulting in smaller plastic sizes. Natural disintegration and degradation of Microplastics can produce plastic particles as small as 0.1 mm in diameter, known as nano-plastics (NPs)(de Souza Machado et al. 2018a, 2018b). The plastics gradually weather and accumulate in the soil, contaminating the soil with Microplastics fragments. Furthermore, the recent increase in the number of waste sites has made soil huge Microplastics sink. Microplastics migration in soil happens both vertically as well as horizontally, i.e., they are conveyed to people and animals horizontally via the terrestrial food chain and drain down vertically into groundwater with run-off. Microplastics have also been identified in sheep faces, which were most likely biomagnified by feed and the surrounding environment (Beriot et al. 2021).

In agricultural fields, plastic mulching is widely used across the world to boost yields, improve fruit quality, and improve water usage efficiency (Ashrafuzzamann et al. 2011). Furthermore, due to their specific optical and material properties, plastic mulches are employed globally (Chalker-Scott et al. 2007). Besides this, the use of organic fertilizers and films made up of plastic in agricultural operations is the primary cause of Microplastics (MP) buildup in farms (Weithmann et al. 2018).

A study by Kumar et al. (2021) shows that plastic mulching has been detected in four locations for producing vegetables viz tomatoes and beans in different regions of Tamil Naidu, India. The soil from the Sulur region had more plastic residues than the soil compared to the other areas. At a depth of 0-10 cm, the plastic content of the soil taken from Sulur varied from 0.092 ± 0.02 to 4.96 ±0.08 g kg^{-1} soils. The plastic content varied from 0.075 ± 0.01 to 3.45 ± 0.01 at a depth of 11-20 cm. At a depth of around 21-30 cm, the plastic content was determined to be in a range of 0.01 ± 0.02 to 2.81± 0.01 g kg^{-1} soils.

SOURCES OF MICROPLASTICS IN AGRICULTURAL SOIL

Microplastic sources are primarily characterized as either primary or secondary (Cole et al. 2011, Thompson & Richard 2015). Primary Microplastics are designed with specific applications in mind, such as cosmetic harsh chemicals, drug vectors, and engineering-related uses such as air blasting (Auta et al. 2017, Hays et al. 1974). Microplastics are difficult to eliminate utilizing sewage disposal systems, and once they enter the wastewater, they eventually accumulate in the environment (Van Cauwenberghe et al. 2015).

Secondary Microplastics are formed when bigger plastics are gradually shattered into smaller bits by a variety of complicated environmental factors, including wave action, temperature of wind, and UV radiation (Andrady 2011, Rocha-Santos & Duarte 2015). Repeated usage of products made of plastics can induce fragmentation and the development of additional Microplastics (Hartline et al. 2016). Furthermore, plastic emissions from vehicle transportation, such as wearing and tearing of tires, brakes, and road markings, are major contributors to Microplastics in the natural environment (Gieré et al. 2018). The World average of Microplastics emissions from road vehicle tire abrasion was calculated to be 0.81 $kg.year^{-1}.capita^{-1}$ (Kole et al. 2016). Aside from Vehicles, abrasion from aircraft tires accounts for around 2% of total tire disintegration emissions in the Netherlands (Kole et al. 2016).

Furthermore, artificial grass is a major secondary source of Microplastics, with estimates ranging from 760-4500 $t.y^{-1}$ (Kole et al. 2016, Lassen et al. 2015, Magnusson et al. 2016) result, many types of Microplastics are discharged into natural environments and ecosystems. Compared to the sources of Microplastics in the ocean, which primarily include land-based sources (80%), tourism on the coast, recreational and commercial fishing gear contributing 18%, shipping and marine industries (e.g., aquaculture, oil rigs, etc.) (Cole et al. 2011, Hays & Cormons 1974, Kole et al. 2016). Microplastics enter soil through a variety of pathways, including landfills (He et al. 2019), amendments of the soil (Zubris et al. 2005), sewage sludge land application (Corradini et al. 2019, Mintenig et al. 2017, Ziajahromi et al. 2017), wastewater irrigation. Furthermore, plastic debris in soil also fragments into Microplastics by the actions of soil fauna, such as feeding, digesting, and excretion (Chae et al. 2018).

This publication is licensed under a Creative Commons Attribution 4.0 International License

Direct Sources/Primary Sources of Microplastics in Agricultural Oils

Plastic Mulching

Massive but undetectable sources of plastics are continually flooding the soil, causing Microplastics deposition, but with no obvious ability to stop or disrupt them. Mulch film, which is made of polyvinyl chloride and polyethylene, has become a popular technique in worldwide agriculture because of its numerous economic benefits, including increased harvest, enhanced quality of fruits, and water use efficiency (Yang et al. 2021). In 2016, the worldwide market for agricultural films made of plastic was 4 million tonnes, and it is predicted to rise at a pace of 5.6 percent per year by 2030 (Jambeck 2015).

Large chunks of plastic mulch debris lying on the farm land's surface are naturally exposed to UV light, producing photo-degradation and becoming brittle (Astner et al. 2019). Shear pressures on plastic trash in farmland will also be felt when agricultural soils are plowed and cultivated, potentially fragmenting already fragile polymers (Piehl et al. 2018).

According to the study by Kumar et al. (2021), Microplastics are found in the 3 regions of India in the amount of 37.97%, 35.07%, and 36.99% of the tomato fields that use the process of Plastic Mulching in Fig. 1.

Littering

The excessive growth of plastics and unplanned or poor management procedures have increased the presence of a large variety and range of Microplastics waste in the soil (Akdogan & Given 2019, Kumar et al. 2020). Microplastics in aquatic environments can emerge from a substantial volume of plastic trash (Blasing et al. 2018). From the 1950s till 2015, roughly 6.3 Bt of plastic rubbish was created globally, with 4.97 Bt ending up in landfills and natural habitats (Geyer et al. 2017).

Silage and Bale Nets

In the first research addressing the issue of the presence of Microplastics in soils at an agricultural scale, (Piehl et al. 2018) indicated the pollution of livestock feeds due to the use of plastics in the form of wrapped grass bales and silage after intake, may enter the soil through excretion.

Twine

The twine composed of Polypropylene (PP) is used for various agricultural uses (Guerrini et al. 2017). Twine helps to secure plants to stakes for significant crops, including tomatoes, crucifers, sweet peppers, etc. It is utilized in the cultivation of bananas to connect plants and keep them from toppling over (Hernandez & Witter 2016). During harvesting, the twine is cut and frequently dumped carelessly in the fields only, where it ends up in the form of Microplastics in the soil. There are initiatives to promote the use of biodegradable twine, which may be gathered along with plant leftovers and composted (Guerrini et al. 2017, Biothop 2019).

Plastics Used for Plant Protection

Plastic films, along with non-woven textiles made of plastic, such as those utilized in greenhouses, polytunnels, shade

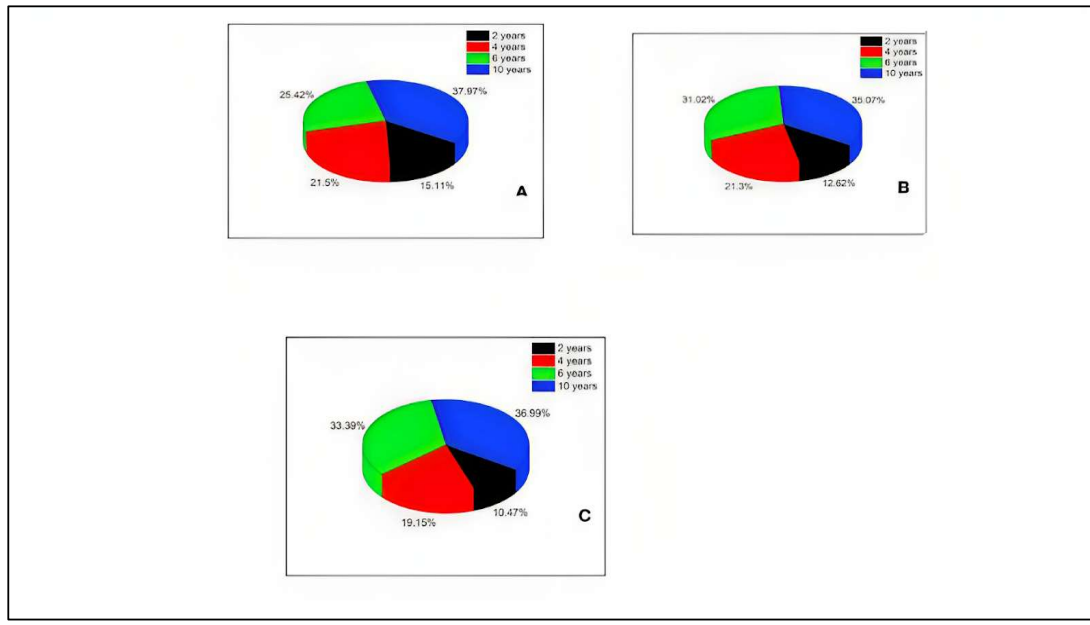


Fig. 1: Effect of the duration of plastic film mulching on percent distribution of plastic residues at different soil depths (A) 0-10 cm (B) 11-20 cm (C) 21-30 cm by Kumar et al. (2021).

This publication is licensed under a Creative Commons Attribution 4.0 International License

nets, and also as wind barriers, enhance the presence of Microplastics in soil samples. Liu (2021) reported that for an identical productive area, samples of soil collected from farms without greenhouses contained fewer Microplastics compared to soil samples collected from farms with greenhouses. At fields where greenhouses were initially utilized in the 1980s, the authors recorded a range of 1000 - 3786 particles of MP kg^{-1}. Liu (2021) also observed an average of 2110 MP kg^{-1} in soil samples collected within greenhouses compared to 310/kg Microplastics in soil samples collected outdoors.

Improper Storage

According to Svensk Ensilageplast Retur, Plastic may be found in abundance on an agricultural farm. Farmers have to put additional efforts into collecting and storing old plastics since they frequently lack the time or technical competence to securely clean and preserve the used ones. To keep plastic garbage clean and avoid it from blowing away, it must be deposited in a dry spot that is shielded from the wind. Plastic waste management in agricultural regions is a major problem, which can be ascribed in part to inappropriate plastic storage on farms.

Indirect Sources/Secondary Sources of Microplastics

Sewage Sludge

Sludge from sewage and wastewater treatment plants causes Microplastics contamination, and Microplastics can build in soil with repeated sludge use (Xu et al. 2021). Microplastics enter wastewater treatment facilities via a variety of routes (Gao 2018). Micro beads from personal cleaning and care products, polymer fibers released from washing textiles, plastic masterbatches seeped from the plastic production facility, and Microplastics from automobile tires are all transmitted to sewage. These tiny particles flow and settle throughout the sewage treatment process. A portion of them is released from the sewage system, while the majority is segregated during the sewage treatment sedimentation process and eventually enters the wastewater sludge (Gao 2018).

Compost

Compost soil addition can potentially provide a conduit for Microplastics to enter the soil. Organic waste is often placed in fields as nutrients after it has been composted and fermented for reuse nutrients, minerals, trace elements, and humus. Composts made from biological waste, for example, have been shown to include plastics as a result of improper disposal and insufficient waste categorization (Blasing & Amelung 2018). The concentration of plastic pieces detectable to the human eye in a composting factory in Bonn is 2.38 to 180 mg kg^{-1}, confirming the presence of Microplastics in organic compost (Blasing & Amelung 2018). Weithmann et al. (2018) discovered that the compost from municipal organic waste and green clipping in German contained 24MP particles/kg ranging in size from 1mm to 5mm. Furthermore, according to (Crossman 2020), despite compliance with current regulations, biosolid applications may result in significant rates of Microplastics export. Liu (2021) discovered that total Microplastics concentrations in soils are 545.9 and 87.6 items/kg after yearly launch with 30 and 15 t ha^{-1} of sludge composts, respectively, which is considerably higher than soil lacking compost application.

Irrigation

The presence of Microplastics in agricultural irrigation water resources has been widely verified (Jian et al. 2020). Rivers, lakes, reservoirs, and aquifers are the primary irrigation water sources worldwide. Sewage is also utilized for irrigation in some locations when water resources are restricted (Blasing & Amelung 2018). Despite the fact that a substantial number of Microplastics may be eliminated during the sewage treatment process, high quantities of Microplastics remain in the purified wastewater (Ziajahromi 2017). Several studies have shown significant amounts of Microplastics in rivers, lakes, reservoirs, and aquifers (Koelmans et al. 2019). Microplastics contained within water reservoirs will be transported to the soil by irrigation, creating a pedigree of Microplastics in the soil.

Flooding and Street Runoff

Street runoff and floods, in addition to purposeful irrigation, are key channels for the transfer and accumulation of Microplastics into the soil (Blasing & Amelung 2018). Street runoff and floods can introduce unmanaged rubbish dumping near roadways, as well as rubber tire abrasion into soils. Some of them already constitute Microplastics, while others are progressively changed into Microplastics as a result of numerous environmental encounters.

Input From the Atmosphere

Atmospheric transmission is a significant mode of Microplastics deposition to land. A study (Liu et al. 2019) quantified and recorded the first to record and quantify the accumulation of fibrous Microplastics both indoors and outdoors, with the settled flux of atmospheric Microplastics outside reaching 0.3-1.5 fibers m^{-3} (Liu et al. 2019). Every day, relative averages of around 249 fragments, 74 pieces of film, and 43 fibers might be deposited in a distant mountainous catchment region (Allen et al. 2019).

TRANSPORTATION OF MICROPLASTICS TO AGRICULTURAL SOILS

Migration, which includes horizontal and vertical movement as well as biological and non-biological transportation, is

This publication is licensed under a Creative Commons Attribution 4.0 International License

a critical link for extending the effect of Microplastics in soil (Xu et al. 2021). Surface runoff or wind can transport Microplastics in surface soil (Koelmans et al. 2019, Qi et al. 2018). The presence of Microplastics in deep soil indicates that Microplastics migrate downhill (Liu et al. 2018). Since Soil is porous, Microplastics in the micrometer (μm) range can be percolated through soil pores via leaching. External pressures such as biological disturbance by the fauna and flora and agricultural operations cause bigger Microplastics to move in the soil.

Also, the bioturbation of plant roots in soil may influence Microplastics translocation through root growth and movement, root water extraction, and furthermore. Soil fauna may help to move Microplastics vertically and horizontally in the soil (Xu et al. 2020). Microplastics have been discovered to be transferred and dispersed by earthworms and collembola species, either by adhesion or excretion (de Souza Machado et al. 2019, Maaß et al. 2017). Furthermore, the formation of fractures in the soil produced by the dry environment allows Microplastics to enter deep soil (Koelmans et al. 2019, Qi et al. 2018).

METHODS TO EXTRACT MICROPLASTICS FROM THE SOIL

Density Separation

The technique of density separation is one of the most often used techniques for separating soil Microplastics. To ensure every particle in the bulk sample sinks or floats, the soil is first treated with ultrasonics (Liu et al. 2018). Sodium chloride is a low-cost, ecologically friendly salt that is commonly used in suspension solutions (Zhou et al. 2018). Yet, the density of a saturated solution of NaCl is 1.2 g cm^{-3}, implying that high-density polymers such as polyvinyl chloride (PVC, 1.35 g cm^{-3}), polyethylene terephthalate (PET, 1.38 g cm^{-3}), and others cannot be separated in this method (Ruggero et al. 2020). As a result, for the suspension medium, several researchers employ a saturated solution of zinc chloride ($ZnCl_2$), sodium iodide (NaI), sodium bromide (NaBr), calcium chloride ($CaCl_2$), and zinc bromide ($ZnBr_2$) (Imhof et al. 2017, Scheurer & Bigalke 2018).

Electrostatic Separation

Plastics are not electrically conductive, unlike soil minerals and other particles. An external electric field can be used to separate the two because of the difference in electrostatic characteristics. Electrostatic separation is a dry processing technology that uses electric forces working on charged particles to separate main and secondary raw materials (Deotterl et al. 2000). It was investigated that plastic particle's electrostatic behavior can be improved before separation from sediment samples using a tiny electrostatic separation apparatus. Without sacrificing Microplastics, up to 99% of the original sample mass might be eliminated (Felsing et al. 2018).

Oil Separation

The lipophilic characteristics of microplastics are used as an alternative to density-based oil recovery technologies in a new, cost-effective oil extraction process (OEP)(Crichton et al. 2017). For seven polymers, the OEP exhibited a recovery ratio of 90-100%, showing a better efficiency than density separation in a salt solution. OEP is less complicated, easier, and less expensive than salt solution separation. However, oil interferes with Fourier Transform Infrared Spectroscopy (FTIR) for identification; thus, a wash with 90% ethanol is required following extraction. Using castor oil, Mani et al. (2019) separated MPs from fluvial suspended surface solids, marine suspended surface solids, marine beach sediments, and agricultural soil substrates. In this investigation, 0.3-1 mm MP particles were extracted utilizing four virgin polymers [polypropylene (PP), polystyrene (PS), polymethyl-methacrylate (PMMA), and glycol-modified polyethylene terephthalate (PETG)]. The average SD MPs spike recovery percentage was 99 4%, with a 95 4% matrix decrease (dry weight, n = 16). This process is less expensive, less risky, and more rapid than salt solution isolation.

Froth Flotation

Froth flotation takes the use of the material's density and the hydrophilic nature of its surface. It is widely employed in the recycling sector. Froth preferentially binds to hydrophobic particles and lifts them upward, segregating them from hydrophilic molecules. To remove plastics from dirt, this approach employs various hydrophilic properties. Plastic flotation techniques include gamma flotation, reagent adsorption, and surface modification (Fraunholcz 2014, Huang et al. 2017) attained a 95% recovery rate for PVC and PMMA by using pinacol (97.71% pure) as the foaming agent and potassium permanganate as the surface modification. In study conducted by Imhof et al. (2017) extracted MPs from sediments using froth flotation but observed low observed found low efficiency and substantial variance across various polymers.

Magnetic Extraction

Grbic et al. (2019) created a magnetic plastic extraction technique in which Magnetized hydrophobic iron nanoparticles are bonded to plastic particles. Iron nanoparticles bond to the surface of Microplastics after being treated with hydrophobic hydrocarbons using cetyl

This publication is licensed under a Creative Commons Attribution 4.0 International License

trimethyl silane (HDTMS) and may be retrieved using a magnetic field. Microplastics (polyethylene, polystyrene, polyurethane, PVC, and polypropylene) recovered from fresh water and sediment were 84% and 78%, respectively, spanning particle sizes ranging from 200 m to 1 mm.

ANALYTICAL TECHNIQUES FOR ANALYSIS OF MICROPLASTICS

Identification and quantification of Microplastics from the ambient matrix are required after separation and purification (Kumar et al. 2020). The common strategy is to first identify obvious/possible Microplastics using a microscope, followed by confirmation using spectroscopy and thermodynamic methods such as Fourier Transform infrared spectroscopy (FTIR) or Raman spectroscopies, and Pyrolysis gas chromatography-mass spectrometry (Yang et al. 2021). Optical microscopes, particularly stereomicroscopes, are essential tools for documenting the physical features of Microplastics (Wang et al. 2018a, 2018b)

By using the microscope, the micron (μm) range of the particles can be analyzed. This approach can swiftly identify Microplastics and record their physical properties and abundance. However, Microplastics with diameters of 1 mm are difficult to detect. As a result, only visible Microplastics are easily identified. Furthermore, in order to avoid inaccurate results and other misinterpretations, the technique necessitates labor-intensive pre-concentration and laboratory hygiene. Integrating high-resolution digital cameras into microscopes, on the other hand, allows for the identification of smaller particles as well as the determination of particle size. This approach has the advantage of being non-destructive (Zhang et al. 2019). Unfortunately, visual detection of Microplastics can be occasionally inaccurate (David et al. 2018, David et al. 2019). Additionally, without the assistance of FTIR and Raman spectroscopies, which use their distinctive absorption spectrum to identify the associated functional groups, the microscope can't determine the detailed chemical composition of Microplastics.

Furthermore, without the support of FTIR and Raman spectroscopies, which employ distinctive absorption spectra to identify the relevant functional groups, the microscope cannot detect the exact chemical composition of Microplastics. Optical microscopes, particularly stereomicroscopes, are essential tools for documenting the physical features of Microplastics ((Wang et al. 2018a, Wang et al. 2018b). However, visual assessment alone might result in a large number of false positives, particularly for small fibers. FTIR and its optimization technologies, such as micro-FTIR (- FTIR), attenuated total reflectance FTIR (ATR-FTIR), and focal plane arrays FTIR (FPA-FTIR), thereby demonstrate significantly increased Microplastics characterization capabilities (Wang et al. 2018a, 2018b). These infrared spectroscopic sensors have a detection limit of 5-10 m for Microplastics (Mintenig et al. 2017, Yang et al. 2021). FTIR can identify Microplastics with a particular thickness and has a detection limit of 10 m. The use of ATR-FTIR can benefit from the high signal-to-noise ratio and the extensive literature spectrum (Yang et al. 2021). Focal plane array FTIR can be employed in precision equipment to automatically identify the Microplastics in the sample filter of the preliminary polymer types allocated (Mintenig et al. 2017).

Raman spectroscopy is a further technique for identifying Microplastics. When used in conjunction with a microscope, it can identify Microplastics as tiny as 1 m in size, with spatial resolution reaching as low as 500 nm in some situations (Elert et al. 2017). Another significant benefit of Raman spectroscopy is its ability to analyze wet materials while also identifying fillers or pigments (Dumichen et al. 2017).

Three mass spectrometry analysis methods offer novel approaches for identifying Microplastics (Pual et al. 2019). Pyrolysis-gas chromatography-mass spectrometry (Pyr-GC-MS) (Nuelle et al. 2014), thermo-gravimetric analysis-mass spectrometry (TGA-MS) (Majewsky et al. 2016), as well as thermal extraction desorption-gas chromatography-mass spectrometry (TED-GC-MS) (Pual et al. 2019) have also been demonstrated to be in useful identifying and quantifying Microplastics (Wang et al. 2018). Thermo analytical methods do not need sample pretreatment, and the processed particle size that may be analyzed is limited only by the ability to manually place them into the pyrolysis tube (Zhao et al. 2018). Unfortunately, these procedures remove the Microplastics' color, size, and shape information, which is critical for assessing the potential hazards of Microplastics.

A new approach for size-independent Microplastics analysis has developed (Piehl et al. 2018). Polycarbonate (PC) and polyethylene terephthalate (PET) Microplastics were effectively quantified using the alkali-assisted heating depolymerization technique (Wang et al. 2018). This technology represents a significant advancement in the painstaking separation, recognition, and counting of Microplastics. However, for polymers containing a wide range of important structural components, this technique e requires further validation.

Several technologies, including NIR spectroscopy (Du et al. 2020a), quantitative H-NMR spectroscopy, and hyperspectral imaging technology, offer alternative options for high-throughput Microplastics investigation (Shan et al. 2018). Although these methods need minimum sample preparation, they do have certain intrinsic limitations that

This publication is licensed under a Creative Commons Attribution 4.0 International License

may limit their use (Wang et al. 2018a, 2018b). So far, the sensitivity and specificity trials of the integrated NIR spectroscopic chemometric method have failed (Du et al. 2020a). To reduce signal variations, the best setting for the H^-NMR approach is to remove any organic debris from the sample (Moller et al. 2020). It is now hard to entirely remove organic materials from environmental samples without destroying the Microplastics, and the application of H^-NMR for the examination of soil Microplastics samples is dubious (Moller et al. 2020).

(TOF-SIMS) Time-of-flight secondary ion mass spectrometry might offer information on the Microplastic's size distribution and chemical components (Rillig et al. 2017a). This approach, however, can only be utilized to analyze Microplastics with known composition.

IMPACT AND POTENTIAL RISK OF MICROPLASTICS IN AGRICULTURAL SOIL ECOSYSTEM

Soil nature influences Microplastics migration, and Microplastics change soil properties such as soil structure and functioning as well as the diversity of microbes (Rillig et al. 2017b, Zhang et al. 2018), which may have implications for plant and animal health and pose possible risks for the safety and quality of food, ultimately jeopardizing human health (Rillig et al. 2018). The presence of substantial residual plastic films in soil has been found to reduce soil-saturated hydraulic conductivity and influence soil microbial activity and abundance, thereby influencing soil fertility (Wan et al. 2018).

Impact on Soil Structure

Because Microplastics can interact with different soil features, soil nature may be the fundamental metric for assessing the dangers presented to terrestrial ecosystems by Microplastics (de Souza Machado et al. 2018a). To variable degrees, Microplastics particles may penetrate soil aggregations and clumps: loosely in fragment types and more firmly in linear types (Liu et al. 2018). Furthermore, de Souza Machado et al. (2018b) discovered that polyester fibers may significantly boost capacity while decreasing bulk density and water-stable aggregation; yet, the impacts of polyethylene (PE) and polyacrylic acid on water-holding capacity show no clear trends. As a result, Microplastics of various materials have varying impacts on soil. Microplastics have also been demonstrated to influence soil permeability and water retention, which affects the evaporation of water (de Souza Machado et al. 2018a, Wang et al. 2018). Wan evaluated how the addition of Microplastics affects the evaporation of water and desiccation cracking in two clay soils and found that both are significant and rise with increasing Microplastics concentration (Allison & Jastrow et al. 2006). According to these findings, Microplastics can change the water cycle in soils, increase soil water shortages, and influence pollutant migration into deep soil layers through fissures (Rillig 2018).

Soil Fertility and Nutrient

Soil enzymes with high catalytic capacity are closely associated with a variety of soil biochemical processes; these enzymes serve as an indicator for assessing soil fertility and play an important role in regulating the process of soil nutrient cycling for nutrients like C, N, and P (Trasar-Cepeda et al. 2008, Arthur et al. 2012). Since Microplastics include polymer chains, Microplastics -Carbon may be disguised as a significant caused by human components of the soil organic carbon pool (Rillig 2018). Therefore, according to de Souza Machado et al. (2018b), the impacts of Microplastics on soil are largely dependent on Microplastics content as well as the exposure period.

Soil Microorganisms

According to researchers ((Rillig 2018, Girvan et al. 2003, Naveed et al. 2016, and Rubol et al. 2013), soil characteristics and nutrients are highly linked with soil microbial activity. Changes in the physical environment of the soil, particularly soil aggregation, which has been shown to include linear microfibers (de Souza Machado et al. 2018a, Zhang et al. 2019), are likely to affect microbial development more significantly than non-microfiber-structured soils (Rillig et al. 2017b, Zhang et al. 2018). Furthermore, Microplastics-induced changes in soil porosity and wetness may affect the flow of oxygen in the soil, altering the proportions of anaerobic and aerobic microbes (Veresoglou et al. 2015). Changes in pore spaces induced by Microplastics may also result in the extinction of indigenous microorganisms (Judy et al. 2019). Furthermore, DeForest et al. (2004a) discovered that the addition of Microplastics considerably interacted with the microbial community composition, and the substrate-induced respiration (SIR) levels dramatically dropped, showing that Microplastics generated alterations in soil microbial function. Because Dissolved organic matter (DOM) serves as a substrate and an important source of carbon for microorganisms, it has been linked to both water eutrophication and the greenhouse effect (DeForest et al. 2004b, Marschner & Kalbitz 2003, Alimi et al. 2018). Thus, changes in DOM caused by Microplastics could impact soil function and microbial communities (Judy et al. 2019).

Soil Contamination

The growing prominence of Microplastics as an ecosystem stressor impacts not only soil health and function but also soil

This publication is licensed under a Creative Commons Attribution 4.0 International License

biophysical characteristics, resulting in complicated changes in the environmental behavior of other soil contaminants (Wang et al. 2018a, 2018b, Yang et al. 2021, Hahladakis et al. 2018). Due to their high specific surface area and elevated adsorption capacity, Microplastics not only contain additives like diethylhexyl phthalate (DEHP), a common organic pollutant used during the manufacture of plastic (Groh et al. 2019, Brennecke et al. 2016), but they also adsorb dangerous contaminants, such as heavy metals like zinc, copper, and lead, antibiotics, toxic organic chemicals like polybrominated diphenyl ether (PBDE) and perfluorochemicals (PFOS), and PFOS (Gaylor et al. 2013, Li et al. 2018, Lagana et al.2018, He et al. 2018).

Transfer Along Food Chains

The most alarming findings in the research on Microplastics in soil come from the ecological and health concerns posed by Microplastics exposure (Guo et al. 2020, Sarker et al. 2020, Kumar et al. 2020, Schwabl et al. 2019). The concept that Microplastics can be transported from prey (at a lower nutritional level) to a predator (at a higher nutritional level) in the food chain is supported by food chain modeling and field experiments (Guo et al. 2020). Evidence of macro- and Microplastics transmission from soils to chickens in traditional Mayan household gardens in Southeast Mexico is well documented by Huerta Lwanga et al. (2017). Microplastics have recently been found in human feces and adult colectomy tissues, demonstrating their presence in the human digestive system (Ibrahim et al. 2021, Jiang et al. 2019).

Uptake of Microplastics by Plants

According to Zhou et al. (2018), the presence of Microplastics would alter the physical and chemical properties of the soil, which will alter the root system and the vegetative phase and thus impair plant growth. Certain Present studies showed significant effects of Microplastics on plants, including wheat (Triticum aestivum) (Qi et al. 2018), perennial ryegrass (Lolium perenne), Vicia faba (Khalid et al. 2020), Polystyrene Microplastics(PS-MPs) were shown to cause evident growth suppression, genotoxic and oxidative damage to hydroponic Vicia faba, and a substantial number of 100 nm PS-MPs were found to collect in root tips using laser confocal scanning microscopy (Khalid et al. 2020).

Agricultural Production

Soil Microplastics can cause direct crop harm in the early stages by physically blocking the seed capsule openings or roots (Pignattelli et al. 2020). Indeed, with Microplastics exposure, extremely short-term unfavorable impacts on edible plant development may show as early as 6 days after sowing (Napper & Thompson 2019). Furthermore, the extreme durability of polymeric plastic particles hinders breakdown processes; even biodegradable plastic bags stay unchanged after 27 months in soil (Kumari et al. 2022, Silva et al. 2021). The direct absorption of Microplastics from soils through apoplastic and symplastic routes, and by distribution to the plant as a whole through the vascular system, is well documented to affect the development of agriculturally important plants. According to Bouaicha et al. (2022), from the perspective of agricultural output, Microplastics generally have a detrimental influence on crop productivity.

FUTURE CHALLENGES

There are several future challenges associated with Microplastics contamination in the agricultural soil -

- **Widespread Soil Contamination:** The proliferation of microplastics in soil poses an ongoing challenge, as current levels continue to rise due to persistent plastic use and inadequate waste management.
- **Ecosystem Disruption:** Microplastics impact soil ecosystems, potentially altering microbial communities, nutrient cycling, and overall soil health, leading to cascading effects on plant and animal life.
- **Agricultural Concerns:** The use of plastic mulching in agriculture, a primary source of microplastics in soil, presents a dilemma, as alternatives must be developed and adopted to reduce environmental harm without compromising crop productivity.
- **Human Health Risks:** The potential transfer of microplastics through the food chain raises concerns regarding human health impacts. Research is needed to figure out the extent of these risks and implement strategies to minimize exposure.
- **Lack of Comprehensive Regulation:** The absence of strict regulations governing plastic production, use, and disposal contributes to the persistence of microplastic pollution. A coordinated global effort is necessary to address these gaps.
- **Limited Biodegradation:** The slow degradation of plastics exacerbates the persistence of microplastics in the environment. Developing and adopting more biodegradable alternatives is essential for reducing long-term environmental impacts.

REFERENCES

Akdogan, Z. and Guven, B., 2019. Microplastics in the environment: A critical review of current understanding and identification of future research needs. *Environmental Pollution*, 254, p.113011.

This publication is licensed under a Creative Commons Attribution 4.0 International License

Alimi, O.S., Farner Budarz, J., Hernandez, L.M. and Tufenkji, N., 2018. Microplastics and nanoplastics in aquatic environments: aggregation, deposition, and enhanced contaminant transport. *Environmental Science & Technology*, 52(4), pp.1704-1724.

Allen, S., Allen, D., Phoenix, V.R., Le Roux, G., Jimenez, P.D., Simonneau, A., Binet, S. and Galop, D., 2019. Author Correction: Atmospheric transport and deposition of microplastics in a remote mountain catchment. *Nature Geoscience*, 12, pp.339-344.

Allison, S.D. and Jastrow, J.D., 2006. Activities of extracellular enzymes in physically isolated fractions of restored grassland soils. *Soil Biology & Biochemistry*, 38, pp.3245-3256.

Andrady, A.L., 2011. Microplastics in the marine environment. *Marine Pollution Bulletin*, 62, pp.1596-1605.

Arthur, C., Baker, J. and Bamford, H., 2009. The occurrence, effects, and fate of microplastic marine debris. *NOAA,* 11, p.30.

Arthur, E., Moldrup, P., Holmstrup, M., Schjonning, P., Winding, A., Mayer, P. and de Jonge, L.W., 2012. Soil microbial and physical properties and their relations along a steep copper gradient. *Agriculture, Ecosystems & Environment*, 159, pp.9-18.

Ashrafuzzaman, M., Halim, M.A., Ismail, M.R., Shahidullah, S.M. and Hossain, M.A., 2011. Effect of plastic mulch on growth and yield of chili (Capsicum annuum L.). *Brazilian Archives of Biology and Technology*, 54(2), pp.321-330.

Astner, A.F., Hayes, D.G., O'Neill, H., Evans, B.R., Pingali, S.V., Urban, V.S. and Young, T.M., 2019. Mechanical formation of micro-and nano-plastic materials for environmental studies in agricultural ecosystems. *Science of the Total Environment*, 685, pp.1097-1106.

Auta, H.S., Emenike, C.U. and Fauziah, S.H., 2017. Distribution and ecological impact of marine microplastics: A review of scientific evidence. *Environment International*, 102, pp.165-176.

Beriot, N., Peek, J., Zornoza, R., Geissen, V. and Lwanga, E.H., 2021. Low-density microplastics detected in sheep feces and soil: A case study from intensive vegetable farming in southeast Spain. *Science of the Total Environment*, 755, p.142653.

Biothop, 2019. *Biotwine Hop Waste Transformation Into Novel Product Assortments for Packaging and Horticulture Sector.* Biothop.

Blasing, M. and Amelung, W., 2018. Plastics in soil: Analytical methods and possible sources. *Science of the Total Environment*, 612, pp.422-435.

Bolan, N. and Bradney, L., 2019. Particulate plastics in soils. *Chemistry in Australia*, 18.

Bouaicha, O., Mimmo, T., Tiziani, R., Praeg, N., Polidori, C., Lucini, L., Vigani, G., Terzano, R., Sanchez-Hernandez, J.C., Illmer, P. et al., 2022. Microplastics make their way into the soil and rhizosphere: A review of the ecological consequences. *Rhizosphere*, 22, p.100542.

Brennecke, D., Duarte, B., Paiva, F., Cacador, I. and Canning-Clode, J., 2016. Microplastics as vectors for heavy metal contamination from the marine environment. *Estuarine, Coastal and Shelf Science*, 175, pp.189-195.

Chae, Y. and An, Y.J., 2018. Current research trends on plastic pollution and ecological impacts on the soil ecosystem: A review. *Environmental Pollution*, 240, pp.387-395.

Chalker-Scott, L., 2007. Impacts of mulches on landscape plants and the environment - A review. *Journal of Horticulture and Forestry*, 25, p.239.

Cole, M., Lindeque, P., Halsband, C. and Galloway, T.S., 2011. Microplastics as contaminants in the marine environment: A review. *Marine Pollution Bulletin*, 62, pp.2588-2597.

Cole, M., Webb, H., Lindeque, P.K., Fileman, E.S., Halsband, C. and Galloway, T.S., 2014. Isolation of microplastic in biota-rich seawater samples and marine organisms. *Scientific Reports*, 4, p.4528.

Corradini, F., Meza, P., Eguiluz, R., Casado, F., Huerta-Lwanga, E. and Geissen, V., 2019. Evidence of microplastic accumulation in agricultural soils from sewage sludge disposal. *Science of the Total Environment*, 671, pp.411-420.

Cozar, A., Echevarria, F., Gonzalez-Gordillo, J.I., Irigoien, X., Ubeda, B., Hernandez-Leon, S., Palma, A.T., Navarro, S., Garcia-de-Lomas, J., Ruiz, A., Fernandez-de-Puelles, M.L. and Duarte, C.M., 2014. Plastic debris in the open ocean. *Proceedings of the National Academy of Sciences of the United States of America*, 111, pp.10239-10244.

Crichton, E.M., Noel, M., Gies, E.A. and Ross, P.S., 2017. A novel, density-independent, and FTIR-compatible approach for the rapid extraction of microplastics from aquatic sediments. *Analytical Methods*, 9(9), pp.1419-1428.

Crossman, J., Hurley, R.R., Futter, M. and Nizzetto, L., 2020. Transfer and transport of microplastics from biosolids to agricultural soils and the wider environment. *Science of the Total Environment*, 724, p.138334.

Da Costa, J.P., Paço, A., Santos, P., Duarte, A.C. and Rocha-Santos, T., 2019. Microplastics in the environment: Challenges in analytical chemistry – A review. *Environmental Chemistry*, 16, pp.145-156.

David, J., Steinmetz, Z., Kučerík, J. and Schaumann, G.E., 2018. Quantitative analysis of poly(ethylene terephthalate) microplastics in soil via thermogravimetry-mass spectrometry. *Analytical Chemistry*, 90, pp.8793-8799.

David, J., Weissmannová, H.D., Steinmetz, Z., Kabelíková, L., Demyan, M.S. and Šimečková, J., 2019. Introducing a soil universal model method (SUMM) and its application for the qualitative and quantitative determination of poly(ethylene), poly(styrene), poly(vinyl chloride), and poly(ethylene terephthalate) microplastics in a model soil. *Chemosphere*, 225, pp.810-819.

de Souza Machado, A.A., Kloas, W., Zarfl, C., Hempel, S. and Rillig, M.C., 2018a. Microplastics as an emerging threat to terrestrial ecosystems. *Global Change Biology*, 24(4), pp.1405-1416.

de Souza Machado, A.A., Lau, C.W., Kloas, W., Bergmann, J., Bachelier, J.B., Faltin, E., Becker, R., Gorlich, A.S. and Rillig, M.C., 2019. Microplastics can change soil properties and affect plant performance. *Environmental Science & Technology*, 53(10), pp.6044-6052.

de Souza Machado, A.A., Lau, C.W., Till, J., Kloas, W., Lehmann, A., Becker, R. and Rillig, M.C., 2018b. Impacts of microplastics on the soil biophysical environment. *Biology*, 24, pp.1405-1416.

DeForest, J.L., Zak, D.R., Pregitzer, K.S. and Burton, A.J., 2004a. Atmospheric nitrate deposition and the microbial degradation of cellobiose and vanillin in a northern hardwood forest. *Soil Biology and Biochemistry*, 36, pp.965-971.

DeForest, J.L., Zak, D.R., Pregitzer, K.S. and Burton, A.J., 2004b. Atmospheric nitrate deposition, microbial community composition, and enzyme activity in northern hardwood forests. *Soil Science Society of America Journal*, 68, pp.132-138.

Deotterl, M., Wachsmuth, U., Waldmann, L., Flachberger, H., Mirkowska, M., Brands, L., Beier, P.-M. and Stahl, I., 2000. Extraction and identification methods of microplastics and nanoplastics in agricultural soil: A review. *Journal of Environmental Management*, 294.

Desforges, J.P., Galbraith, M. and Ross, P.S., 2015. Ingestion of microplastics by zooplankton in the Northeast Pacific Ocean. *Environmental Contamination and Toxicology*, 69, pp.320-330.

Doyle, M.J., Watson, W., Bowlin, N.M. and Sheavly, S.B., 2011. Plastic particles in coastal aquatic ecosystems of the Northeast Pacific Ocean. *Marine Environmental Research*, 71(1), pp.41-52.

Du, C., Liang, H., Li, Z. and Gong, J., 2020a. Pollution characteristics of microplastics in soils in southeastern suburbs of Baoding City, China. *International Journal of Environmental Research and Public Health*, 17, p.845.

Dumichen, E., Eisentraut, P., Bannick, C.G., Barthel, A.K., Senz, R. and Braun, U., 2017. Fast identification of microplastics in complex environmental samples by a thermal degradation method. *Chemosphere*, 174, pp.572-584.

Eerkes-Medrano, D. and Thompson, R., 2018. Occurrence, fate, and effect of microplastics in freshwater systems. *Environmental Pollution*, pp.95-132.

This publication is licensed under a Creative Commons Attribution 4.0 International License

Elert, A.M., Becker, R., Duemichen, E., Eisentraut, P., Falkenhagen, J., Sturm, H. and Braun, U., 2017. Comparison of different methods for microplastic detection: What can we learn from them, and why asking the right question before measurements matters? *Environmental Pollution*, 231, pp.1256-1264.

Felsing, S., Kochleus, C., Buchinger, S., Brennholt, N., Stock, F. and Reifferscheid, G., 2018. A new approach in separating microplastics from environmental samples based on their electrostatic behavior. *Environmental Pollution*, 234, pp.20-28.

Forster, N.A., Tighe, M.K. and Wilson, S.C., 2020. Microplastics in soils of wilderness areas: What is the significance of outdoor clothing and footwear? *Geoderma*, 378, p.114612.

Fraunholcz, N., 2014. Separation of waste plastics by froth flotation - A review, Part 1. *Minerals Engineering*, 17(2), pp.261-268.

Gao, H., Yan, C., Liu, Q., Ding, W., Chen, B. and Li, Z., 2018. Effects of plastic mulching and plastic residue on agricultural production: A meta-analysis. *Science of the Total Environment*, 651, pp.484-492.

Gaylor, M.O., Harvey, E. and Hale, R.C., 2013. Polybrominated diphenyl ether (PBDE) accumulation by earthworms (Eisenia fetida) exposed to biosolids-, polyurethane foam microparticle-, and penta-BDE-amended soils. *Environmental Science and Technology*, 47, pp.13831-13839.

Geyer, R., Jambeck, J.R. and Law, K.L., 2017. Production, use, and fate of all plastics ever made. *Science Advances*, 3, pp.1-5.

Gieré, R., Sommer, F., Dietze, V., Baum, A., Gilge, S., Sauer, J. and Maschowski, C., 2018. Tire abrasion is a major source of microplastics in the environment. *Geological Society of America*, 50, p.6.

Girvan, M.S., Bullimore, J., Pretty, J.N., Osborn, A.M. and Ball, A.S., 2003. Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. *Applied and Environmental Microbiology*, 69, pp.1800-1809.

Grbic, J., Nguyen, B., Guo, E., You, J.B., Sinton, D. and Rochman, C.M., 2019. Magnetic extraction of microplastics from environmental samples. *Environmental Science & Technology Letters*, 6(2), pp.68.

Groh, K.J., Backhaus, T., Carney-Almroth, B., Geueke, B., Inostroza, P.A., Lennquist, A., Leslie, H.A., Maffini, M., Slunge, D., Trasande, L., Warhurst, A.M. and Muncke, J., 2019. Overview of known plastic packaging-associated chemicals and their hazards. *Science of the Total Environment*, 651, pp.3253-3268.

Guerrini, S., Borreani, G. and Voojis, H., 2017. Biodegradable materials in agriculture: Case histories and perspectives. In: Malinconico, M. (Ed.), *Soil Degradable Bioplastics for Sustainable Agriculture*. Springer, Berlin Heidelberg, pp.35-65.

Guo, J.J., Huang, X.P., Xiang, L., Wang, Y.Z., Li, Y.W., Li, H., Cai, Q.Y., Mo, C.H. and Wong, M.H., 2020. Source, migration, and toxicology of microplastic in soil. *Environmental International*, 137, p.105263.

Hahladakis, J.N., Velis, C.A., Weber, R., Iacovidou, E. and Purnell, P., 2018. An overview of chemical additives present in plastics: Migration, release, fate and environmental impact during their use, disposal, and recycling. *Journal of Hazardous Materials*, 344, pp.179-199.

Hartline, N.L., Bruce, N.J., Karba, S.N., Ruff, E.O., Sonar, S.U. and Holden, P.A., 2016. Microfiber masses recovered from conventional machine washing of new or aged garments. *Environmental Science & Technology*, 50(21), pp.11532-11538.

Hayes, D., 2019. Biodegradable mulch performance and adoption in sustainable agriculture: A review. *Biodegradable Mulch Performance and Adoption*, 124, pp.1-6.

Hays, H. and Cormons, G., 1974. Plastic particles are found in tern pellets on coastal beaches and at factory sites. *Marine Pollution Bulletin*, 5, pp.44-46.

He, L., Wu, D., Rong, H.F., Li, M., Tong, M.P. and Kim, H., 2018. Influence of nano- and microplastic particles on the transport and deposition behaviors of bacteria in quartz sand. *Environmental Science and Technology*, 52, pp.11555-11563.

Hernandez, C.E. and Witter, S., 2016. Evaluating and managing the environmental impact of banana production in Costa Rica: A systems approach. *Ambio*, 25(3), pp.171-178.

Hidalgo-Ruz, V., Gutow, L., Thompson, R.C. and Thiel, M., 2012. Microplastics in the marine environment: A review of methods, occurrence, and effects. *Environmental Science & Technology*, 46, pp.3060-3075.

Huang, L., Wang, H., Wang, C., Zhao, J. and Zhang, B., 2017. Microwave-assisted surface modification for the separation of polycarbonate from polymethylmethacrylate and polyvinyl chloride waste plastics by flotation. *Waste Management & Research*, 35(3), pp.294-300.

Huerta Lwanga, E., Mendoza Vega, J., Ku Quej, V., Chi, J.L.A., Sanchez Del Cid, L., Chi, C., Escalona Segura, G., Gertsen, H., Salanki, T., van der Ploeg, M., Koelmans, A.A. and Geissen, V., 2017. Field evidence for transfer of plastic debris along a terrestrial food chain. *Scientific Reports*, 7, p.14071.

Hurley, R.R., Woodward, J.C. and Rothwell, J.J., 2007. Ingestion of microplastics by freshwater tubifex worms. *Environmental Science & Technology*, 51, pp.12844-12851.

Ibrahim, Y.S., Tuan Anuar, S., Azmi, A.A., Wan Mohd Khalik, W.M.A., Lehata, S., Hamzah, S.R., Ismail, D., Ma, Z.F., Dzulkarnaen, A., Zakaria, Z., Mustaffa, N., Tuan Sharif, S.E. and Lee, Y.Y., 2021. Detection of microplastics in human colectomy specimens. *JGH Open*, 5, pp.116-121.

Imhof, H.K., Sigl, R., Brauer, E., Feyl, S., Giesemann, P., Klink, S., Leupolz, K., Löder, M.G., Löschel, L.A., Missun, J. and Muszynski, S., 2017. Spatial and temporal variation of macro, meso, and microplastic abundance on a remote coral island of the Maldives, Indian Ocean. *Marine Pollution Bulletin*, 116(1-2), pp.340-347.

Ivleva, N.P., Wiesheu, A.C. and Niessner, R., 2017. Microplastics in aquatic environments. *Angewandte Chemie International Edition*, 56, pp.1720-1739.

Jambeck, J.R., Geyer, R., Wilcox, C., Siegler, T.R., Perryman, M. Andrade, A., Narayan, R. and Law, K.L., 2015. Plastic waste inputs from land into the ocean. *Science*, 347, pp.768-771.

Jian, M., Zhang, Y., Yang, W., Zhou, L., Liu, S. and Xu, E.G., 2020. Occurrence and distribution of microplastic in China's largest freshwater lake system. *Chemosphere*, 261, p.128186.

Jiang, X., Chen, H., Liao, Y., Ye, Z., Li, M. and Klobučar, G., 2019. Ecotoxicity and genotoxicity of polystyrene microplastics on higher plant Vicia faba. *Environmental Pollution*, 250, pp.831-838.

Judy, J.D., Williams, M., Gregg, A., Oliver, D., Kumar, A., Kookana, R. and Kirby, J.K., 2019. Microplastics in soil: Microplastics in municipal mixed-waste organic outputs induce minimal short to long-term toxicity in key terrestrial biota. *Environmental Pollution*, 252, pp.522-531.

Khalid, N., Aqeel, M. and Noman, A., 2020. Microplastics could be a threat to plants in terrestrial systems directly or indirectly. *Environmental Pollution*, 267, p.115653.

Koelmans, A.A., Mohamed Nor, N.H., Hermsen, E., Kooi, M., Mintenig, S.M. and De France, J., 2019. Microplastics in fresh and drinking water: Critical review and assessment of data quality. *Water Research*, 155, pp.410-422.

Kole, P.J., Löhr, A.J., Van Belleghem, F. and Ragas, A., 2017. Wear and tear of tires: A stealthy source of microplastic in the environment. *International Journal of Environmental Research and Public Health*, 14(10), p.1265.

Kumar, M., Xiong, X., He, M., Tsang, D.C.W., Gupta, J. and Khan, E., 2020. Microplastics as pollutants in agricultural soils. *Environmental Pollution*, 265, p.114980.

Kumar, M., Xiong, X., He, M., Tsang, D.C.W., Gupta, J., Khan, E., Harrad, S., Hou, D., Ok, Y.S. and Bolan, N.S., 2020. Microplastics as pollutants in agricultural soils. *Environmental Pollution*, 265, p.114980.

Kumar, M.V. and Sheela, A.M., 2021. Effect of plastic film mulching on the distribution of plastic residues in agricultural fields. *Chemosphere*, 273, pp.128590.

This publication is licensed under a Creative Commons Attribution 4.0 International License

Kumari, A., Rajput, V.D., Mandzhieva, S.S., Rajput, S., Minkina, T., Kaur, R., Sushkova, S., Kumari, P., Ranjan, A. and Kalinitchenko, V.P., 2022. Microplastic pollution: An emerging threat to terrestrial plants and insights into its remediation strategies. *Plants*, 11, p.340.

Laganà, P., Caruso, G., Corsi, I., Bergami, E., Venuti, V., Majolino, D., Ferla, R.L., Azzaro, M. and Cappello, S., 2018. Do plastics serve as a possible vector for the spread of antibiotic resistance? First insights from bacteria associated with a polystyrene piece from King George Island (Antarctica). *International Journal of Hygiene and Environmental Health*, 222, pp.89-100.

Lassen, C., Hansen, S.F., Magnusson, K., Hartmann, N.B., Rehne Jensen, P., Nielsen, T.G. and Brinch, A., 2015. Microplastics: Occurrence, effects, and sources of releases to the environment in Denmark. *Danish Environmental Protection Agency*, 1793, p.206.

Lei, L., Wu, S., Lu, S., Liu, M., Song, Y., Fu, Z., Shi, H., Raley-Susman, K.M. and He, D., 2018b. Microplastic particles cause intestinal damage and other adverse effects in zebrafish *Danio rerio* and nematode Caenorhabditis elegans. *Science of the Total Environment*, 619, pp.1-8.

Li, J., Zhang, K. and Zhang, H., 2018. Adsorption of antibiotics on microplastics. *Environmental Pollution*, 237, pp.460-467.

Liu, M., Lu, S., Song, Y., Lei, L., Hu, J., Lv, W., Zhou, W., Cao, C., Shi, H., Yang, X. and He, D., 2018. Microplastic and mesoplastic pollution in farmland soils in suburbs of Shanghai, China. *Environmental Pollution*, 242, pp.855-862.

Liu, M., Song, Y., Lu, S., Qiu, R., Hu, J., Li, X., Bigalke, M., Shi, H. and He, D., 2019. A method for extracting soil microplastics through the circulation of sodium bromide solutions. *Science of the Total Environment*, 691, pp.341-347.

Liu, Y., Zhang, J., Tang, Y., He, Y., Li, Y., You, J., Breider, F., Tao, S. and Liu, W., 2021. Effects of anthropogenic discharge and hydraulic deposition on the distribution and accumulation of microplastics in surface sediments of a typical seagoing river: The Haihe River. *Journal of Hazardous Materials*, 404, p.124180.

Lozano, Y.M. and Rillig, M.C., 2020. Effects of microplastic fibers and drought on plant communities. *Environmental Science & Technology*, 54, pp.6166-6173. https://doi.org/10.1021/acs.est.0c01051.

Luo, Y., Zhang, Y., Xu, Y., Guo, X. and Zhu, L., 2020. Distribution characteristics and mechanism of microplastics mediated by soil physicochemical properties. *Science of the Total Environment*, 726, p.138389.

Maaß, S., Daphi, D., Lehmann, A. and Rillig, M.C., 2017. Transport of microplastics by two collembolan species. *Environmental Pollution*, 225, pp.456-459.

Magnusson, K., Eliasson, K., Fråne, A., Haikonen, K., Hultén, J., Olshammar, M., Stadmark, J. and Voisin, A., 2016. *Swedish sources and pathways for microplastics to the marine environment*. IVL Svenska Miljöinstitutet

Mani, T., Frehland, S., Kalberer, A. and Burkhardt-Holm, P., 2019. Using castor oil to separate microplastics from four different environmental matrices. *Analytical Methods*, 11 (13), pp.1788-1794.

Marschner, B. and Kalbitz, K., 2003. Controls of bioavailability and biodegradability of dissolved organic matter in soils. *Geoderma*, 113, pp.211-235.

Mintenig, S.M., Int-Veen, I., Loder, M.G.J., Primpke, S. and Gerdts, G., 2017. Identification of microplastic in effluents of wastewater treatment plants using focal plane array-based micro-Fourier-transform infrared imaging. *Water Research*, 108, pp.365-372.

Moller, J.N., Loder, M.G.J. and Laforsch, C., 2020. Finding microplastic in soils: A review of analytical methods. *Environmental Science & Technology*, 54, pp.2078-2090.

Napper, I.E. and Thompson, R.C., 2019. Environmental deterioration of biodegradable, oxo-biodegradable, compostable, and conventional plastic carrier bags in the sea, soil, and open air over a 3-year period. *Environmental Science & Technology*, 53, pp.4775-4783.

Naveed, M., Herath, L., Moldrup, P., Arthur, E., Nicolaisen, M., Norgaard, T., Ferre, T.P.A. and de Jonge, L.W., 2016. Spatial variability of microbial richness and diversity and relationships with soil organic carbon, texture, and structure across an agricultural field. *Applied Soil Ecology*, 103, pp.44-55.

Nizzetto, L., Futter, M. and Langaas, S., 2016. Are agricultural soils dumps for microplastics of urban origins? *Environmental Science & Technology*, 50, pp.10777-10779.

Panebianco, A., Nalbone, L., Giarratana, F. and Ziino, G., 2019. First discoveries of microplastics in terrestrial snails. *Science of the Total Environment*, 106722, pp.1-7.

Paul, A., Wander, L., Becker, R., Goedecke, C. and Braun, U., 2019. High-throughput NIR spectroscopic (NIRS) detection of microplastics in soil. *Environmental Science and Pollution Research International*, 26, pp.7364-7374.

Piehl, S., Leibner, A., Löder, M.G.J., Dris, R., Bogner, C. and Laforsch, C., 2018. Identification and quantification of macro-and microplastic on agricultural farmland. *Scientific Reports*, 8, p.17950.

Pignattelli, S., Broccoli, A. and Renzi, M., 2020. Physiological responses of garden cress (Lepidium sativum) to different types of microplastics. *Science of the Total Environment*, 727, p.138609.

Plastics Europe, 2018. *Plastics - The Facts 2018: An analysis of European plastic production, demand, and waste data for 2018*. EUA

PlasticsEurope, 2021. *Plastics - The Facts 2021: An analysis of European plastics production, demand and waste data*. Retrieved from: https://plasticseurope.org/knowledge-hub/plastics-the-facts-2021/.

Qi, Y., Yang, X., Pelaez, A.M., Lwanga, E.H., Beriot, N., Gertsen, H., Garbeva, P. and Geissen, V., 2018. Macro-and microplastics in soil-plant system: Effects of plastic mulch film residues on wheat (Triticum aestivum) growth. *Science of the Total Environment*, 645, pp.1048-1056.

Rezania, S., Park, J., Md Din, M.F., Mat Taib, S., Talaiekhozani, A., Kumar Yadav, K. and Kamyab, H., 2018. Microplastics pollution in different aquatic environments and biota: A review of recent studies. *Marine Pollution Bulletin*, 133, pp.191-208.

Rillig, M.C., 2018. Microplastics disguising as soil carbon storage. *Environmental Science & Technology*, 52, pp.6079-6080.

Rillig, M.C. and Bonkowski, M., 2018. Microplastic and soil protists: A call for research. *Environmental Pollution*, 241, pp.1128-1131.

Rillig, M.C., Ingraffia, R., de Souza, M. and Anderson, A., 2017. Microplastic incorporation into soil in agroecosystems. *Frontiers in Plant Science*, 8, pp.305719.

Rillig, M.C., Muller, L.A.H. and Lehmann, A., 2017. Soil aggregates as massively concurrent evolutionary incubators. *ISME Journal*, 11, pp.1943-1948.

Rocha-Santos, T. and Duarte, A.C., 2015. A critical overview of the analytical approaches to the occurrence, the fate, and the behavior of microplastics in the environment. *Trends in Analytical Chemistry*, 65, pp.47-53.

Rochman, C.M., 2018. Microplastic research: From sink to source. *Science*, 360, pp.28-29.

Rubol, S., Manzoni, S., Bellin, A. and Porporato, A., 2013. Modeling soil moisture and oxygen effects on soil biogeochemical cycles, including dissimilatory nitrate reduction to ammonium (DNRA). *Advances in Water Resources*, 62, pp.106-124.

Sarker, A., Deepo, D.M., Nandi, R., Rana, J., Islam, S., Rahman, S., Hossain, M.N., Islam, M.S., Baroi, A. and Kim, J.E., 2020. A review of microplastic pollution in the soil and terrestrial ecosystems: A global and Bangladesh perspective. *Science of the Total Environment*, 733, p.139296.

Scheurer, M. and Bigalke, M., 2018. Microplastics in Swiss floodplain soils. *Environmental Science & Technology*, 52, pp.3591-3598.

Schwabl, P., Koppel, S., Konigshofer, P., Bucsics, T., Trauner, M., Reiberger, T. and Liebmann, B., 2019. Detection of various

This publication is licensed under a Creative Commons Attribution 4.0 International License

microplastics in human stool: A prospective case series. *Annals of Internal Medicine*, 171, pp.453-457.

Silva, G., Madrid, F.G., Hernández, D., Pincheira, G., Peralta, A., Gavilán, M.U., Vergara-Carmona, V. and Fuentes-Peñailillo, F., 2021. Microplastics and their effect on horticultural crops: Food safety and plant stress. *Agronomy*, 11, pp.1528.

Thompson, R.C., Olsen, Y., Mitchell, R.P., Davis, A., Rowland, S.J., John, A.W.G., McGonigle, D. and Russell, A.E., 2004. Lost at sea: Where is all the plastic? *Science*, 304, pp.838.

Thompson, Richard C., 2015. Microplastics in the Marine Environment: Sources, Consequences, and Solutions. In: Bergmann, M., Gutow, L. and Klages, M. (Eds.), *Marine Anthropogenic Litter*. Springer International Publishing, Cham, pp.185-200.

Trasar-Cepeda, C., Leiros, M.C. and Gil-Sotres, F., 2008. Hydrolytic enzyme activities in agricultural and forest soils: Some implications for their use as indicators of soil quality. *Soil Biology and Biochemistry*, 40, pp.2146-2155.

Van Cauwenberghe, L., Claessens, M., Vandegehuchte, M.B. and Janssen, C.R., 2015. Microplastics are taken up by mussels (Mytilus edulis) and lugworms (Arenicola marina) living in natural habitats. *Environmental Pollution*, 199, pp.10-17.

van den Berg, P., Huerta-Lwanga, E., Corradini, F. and Geissen, V., 2020. Sewage sludge application as a vehicle for microplastics in eastern Spanish agricultural soils. *Environmental Pollution*, 261, p.114198.

Veresoglou, S.D., Halley, J.M. and Rillig, M.C., 2015. Extinction risk of soil biota. *Nature Communications*, 6, p.8862.

Wan, Y., Wu, C., Xue, Q. and Hui, X., 2018. Effects of plastic contamination on water evaporation and desiccation cracking in soil. *Science of the Total Environment*, 654, 576-582.

Wang, F., Wong, C.S., Chen, D., Lu, X., Wang, F. and Zeng, E.Y., 2018. Interaction of toxic chemicals and microplastics: A critical review. *Water Research*, 139, pp.208-219.

Wang, J., Liu, X., Li, Y., Powell, T., Wang, X., Wang, G. and Zhang, H., 2019. Microplastics as contaminants in the environment: A review focusing on effects to organisms. *Science of the Total Environment*, 691, pp.848-857.

Wang, J., Taylor, A., Xu, C.Y., Schlenk, D. and Gan, J., 2018. Evaluation of different methods for assessing bioavailability of DDT residues during soil remediation. *Environmental Pollution*, 238, pp.462-470.

Weber, C.J. and Opp, C., 2020. An overview of the potential risks, sources, and analytical methods for microplastics in soil. *Environmental Pollution*, 267, p.115390.

Weithmann, N., Möller, J.N., Löder, M.G.J., Piehl, S., Laforsch, C. and Freitag, R., 2018. Organic fertilizer is a vehicle for the entry of microplastic into the environment. *Science Advances*, 4, p.8060.

Weithmann, N., Möller, U.N., Löder, M.G.J., Piehl, S., Laforsch, C. and Freitag, R., 2018. Organic fertilizer as a vehicle for the entry of microplastic into the environment: A critical review. *Science Advances*, 4, eaap8060.

Xu, C., Zhang, B., Gu, C., Shen, C., Yin, S., Aamir, M. and Li, F., 2020. Are we underestimating the sources of microplastic pollution in terrestrial environments? *Journal of Hazardous Materials*, 400, p.123228.

Xu, C., Zhang, B., Gu, C., Shen, C., Yin, S., Aamir, M. and Li, F., 2020. Environmental impacts of microplastics: A critical review of the literature. *Journal of Hazardous Materials*, 400, p.123228.

Xu, G., Liu, Y. and Yu, Y., 2021. Effects of polystyrene microplastics on uptake and toxicity of phenanthrene in soybean. *Science of the Total Environment*, 783, 147016.

Yang, L., Zhang, Y., Kang, S., Wang, Z. and Wu, C., 2021. Microplastics in freshwater sediment: A review on methods, occurrence, and sources. *Science of the Total Environment*, 780, p.146546.

Yang, L., Zhang, Y., Kang, S., Wang, Z. and Wu, C., 2021. Corrigendum to Microplastics in soil: A review on methods, occurrence, sources, and potential risk. *Science of the Total Environment*, 780, pp.146546.

Zhang, G.S., Zhang, F.X. and Li, X.T., 2019. Effects of polyester microfibers on soil physical properties: Perception from a field and a pot experiment. *Science of the Total Environment*, 670, pp.1-7.

Zhang, S., Yang, X., Gertsen, H.N., Peters, P., Salánki, Y. and Geissen, V., 2018. A simple method for the extraction and identification of light-density microplastics from soil. *Science of the Total Environment*, 616-617, pp.1056-1065.

Zhao, S., Zhu, L., Gao, L. and Li, D., 2018. Limitations for microplastic quantification in the ocean and recommendations for improvement and standardization. *Marine Pollution Bulletin*, 127, pp.27-49.

Zhou, Q., Zhang, H., Fu, C., Zhou, Y., Dai, Z., Li, Y., Tu, C. and Luo, Y., 2018. The distribution and morphology of microplastics in coastal soils adjacent to the Bohai Sea and Yellow Sea. *Geoderma*, 322, pp.201-208.

Ziajahromi, S., Neale, P.A., Rintoul, L. and Leusch, F.D., 2017. Plastic pollutants in the water environment. *Water Research*, 112, pp.93-99.

Zubris, K.A.V. and Richards, B.K., 2005. Synthetic fibers as an indicator of land application of sludge. *Environmental Pollution*, 138(2), pp.201-211.

This publication is licensed under a Creative Commons Attribution 4.0 International License

Arabian Journal for Science and Engineering
https://doi.org/10.1007/s13369-024-09782-0

RESEARCH ARTICLE-ELECTRICAL ENGINEERING

Mitigation of Residential EV Charging Effects on Power Distribution Networks with Optimal Allocation of DGs Using Coati Optimization

Shefali Painuli[1] · **Suman Bhowmick**[1] · **Radheshyam Saha**[1]

Received: 9 May 2024 / Accepted: 12 November 2024
© King Fahd University of Petroleum & Minerals 2024

Abstract
Electric vehicle (EV) charging deteriorates the performance of power distribution networks. This may lead to issues such as distortion of the daily demand profile, deterioration in bus voltage magnitudes, increased system losses, phase imbalances and power quality concerns. This paper examines the effects of slow, residential EV charging on a 240-bus power distribution network in Iowa, U.S.A. The investigation estimates the charging demand of approximately 1,500 EVs across four different segments, representing over 1,000 U.S. residents, using data from the 2017 National Household Travel Survey. The findings reveal that slow, residential EV charging significantly increases network load, particularly during or after evening hours. To counteract these adverse effects, wind power-based distributed generation (DG) resources were optimally allocated in the 240-bus network using the coati optimization algorithm (COA), which is a recently developed metaheuristic algorithm. A multi-objective optimization problem was formulated to minimize both active power losses and the voltage magnitude deviation of the network. Wind speed patterns for the site were sourced from Iowa Environmental Mesonet data. The COA simplifies implementation, reduces complexity and scales efficiently, making it ideal for real-world power systems. It outperforms traditional methods by finding optimal solutions faster while minimizing computational time and resources for DG sizing. The results demonstrate that the proposed technique effectively reduces network active power losses by 38.45% and network voltage deviation by 51.61%, validating its ability to mitigate the negative impacts of EV charging demand on the power distribution network performance.

Keywords Coati · Distributed generation · Distribution network · Electric vehicle · NHTS · Optimization · Residential charging

1 Introduction

Electrification of transportation has garnered significant interest from the government, industry and the academia, as concerns about environmental pollution and the shortage of fossil fuels grow globally. EVs show promise as a future replacement for conventional automobiles [1]. The transportation and power sectors will undergo a massive transformation as a result of the EVs' growing trend of adoption. EV charging impairs the performance of power distribution networks by distorting the daily load dynamics, deteriorating the bus voltage magnitude profile, increasing network losses, increasing the phase unbalancing and creating power quality issues [2, 3]. Large-scale EV charging demand will put additional strain on power distribution networks, making their operation more difficult. This is a matter of serious concern and needs to be addressed.

1.1 Literature Review

To assess the impact of multiple EV charging on the power distribution network, the primary step is to model and estimate the EV charging demand profile. Modeling EV charging demand is crucial for a variety of applications, including the planning and operation of power systems, designing EV charging infrastructure and managing grid loads effectively. In this regard, authors in [4] have modeled the EV charging profile on the basis of EV charger, battery and network capacity. To estimate the power consumption, charging of three different types of EVs with capacities of 9.6 kW, 11.5 kW

✉ Shefali Painuli
shefalipainuli_2k21phdee01@dtu.ac.in

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

Published online: 05 December 2024

and 15.4 kW were considered. However, [4] fails to consider any transportation survey data which provides insights into the driver behavior, the distance traveled by a vehicle, the number of vehicles charging at a time, their arrival and departure times, etc., to accurately calculate the EV charging demand profile. In [5], authors have simulated the load of EVs for residential charging. The EV demand estimation in [5] is confined only to plug-in hybrid EVs (PHEVs) without considering any real-time driving data and is done under the assumption that the driving habits of EV owners will not affect the load demand calculation which is practically not true. In [6], the stochastic modeling of EV load is done for slow charging, fast charging and battery swapping. However, the model inputs random values for EV charging power, starting time of charging and daily mileage driven [6]. Therefore, the model fails to consider realistic data for estimating the EV demand efficiently. Moreover, the model considered here is restricted to PHEVs only and doesn't apply to all EVs in general.

Several smart charging strategies and control methods have been extensively studied in the literature and integrated into distribution system planning to mitigate the adverse effects of EV charging. For example, in [7], smart scheduling of EV charging is recommended to smooth out the load curve. Additionally, [8] has proposed distributed algorithms that empower EV owners participating in demand response programs to smooth the demand curve by decreasing the peak demand through a game theory-based algorithm. To flatten the daily load profile of the grid and lessen system load variability, distributed EV charging techniques are introduced in [9]. [10] has suggested control strategies for charging and discharging to alleviate power fluctuations resulting from EV charging, utilizing a predictive control methodology. [7–10] have examined various network demand side management strategies such as load shifting, valley filling and peak shaving. However, these methods may not be optimal for addressing the challenges posed by EV charging, given that charging decisions are solely at the discretion of the vehicle owner. Also, it would be impractical to impose restrictions on EV owners regarding the time and duration of charge. Moreover, implementing such strategies could inadvertently increase demand during off-peak hours, potentially creating new peak periods that pose technical challenges like degraded voltage profile issues and feeder overloading. These issues are particularly concerning during winter months when heating device usage is high. Therefore, the proposed methods may compromise the convenience of EV owners and raise concerns about grid reliability.

To address the aforementioned issues in [1–3], additional renewable energy sources can be incorporated at selected buses in the power distribution network as an effective solution to function as DGs for augmenting the network capacity. The most recent studies carried out in this area, as listed in Table 1, are reviewed in this section. In Table 1, entries with dashed lines indicate that specific information is not provided in the referenced paper. It is noted that most of the modeling of EVs relies on probability distribution functions; however, real-time travel data for accurately estimating the EV demand in the distribution network are generally lacking in [11–14, 16, 18–20]. Also, the algorithms used in these studies are either based on traditional methods [12, 14, 17], or do not involve recent metaheuristic techniques [11, 19]. Furthermore, the distribution networks analyzed are not large-scale, which limits the ability to capture a comprehensive view of real-time EV charging demand in large residential networks and its potential impacts. The shortcomings and their contributions of the abovementioned works are presented in Sect. 1.2.

1.2 Article Shortcomings and Contributions

Based on the literature presented in the aforementioned section, the shortcomings (S) and the contributions (C) of this paper are summed up as follows:

- **S-1:** To the best of knowledge of the authors, a generalized model to estimate the charging demand of both PHEVs and other categories of EVs is not abundant in the literature. In addition to this, a very few studies have considered practical transportation survey or travel data for the modeling process.

C-1: In order to address S-1, a generalized model to estimate the hourly EV charging demand due to slow, residential charging is presented in this paper. This modeling considers the latest data of EVs present in market which includes different EV categories and their battery capacities, ranging from small city cars to sports utility vehicles (SUVs) as well as realistic transportation survey data from NHTS like number and type of EVs in a house, their distance traveled, start of charging times, etc., to extract various parameters and compute the 24-hour, residential EV charging demand.

- **S-2:** The literature is more focused on the demand side management approach to confront the additional demand and peaks due to the charging process of EVs. Such an approach often imposes restrictions on EV owners regarding the time and duration of charging. Moreover, it may inadvertently increase demand during off-peak hours, potentially creating new peaks. Therefore, such methods may compromise the convenience of EV owners along with raising various concerns about grid reliability.

C-2: It is observed that slow, residential EV charging distorts the daily demand profile, deteriorates the bus voltage magnitude profile and increases the network power losses.

Table 1 Summary of few recent works [11–20]

References no.	Year	Modeling of EV charging demand	Contribution	Objective	Real-time travel survey data	Distribution network studied	Algorithm used
[11]	2016	Randomly distributed based on normal distribution	Random allocation of DG to residential customers in a distribution network	Normalize the voltage profile even for high level of EV uptakes	Not used	UK low-voltage generic	– (Probabilistic approach)
[12]	2017	–	Proposed joint optimization of control strategy and component size of the home battery energy storage with the participation of EV and PV arrays	Minimize a summation of the total electric energy cost and the home battery energy storage system cost	Not used	–	Convex programming
[13]	2018	Randomly distributed based on lognormal distribution	Capacity optimization methodology for a stand-alone microgrid employing DGs	Minimize the cost, reduce greenhouse gases emissions and curtail dump energy	Not used	Stand-alone microgrid system supplying a typical residential demand of Dammam, located in the eastern region of Saudi Arabia	–
[14]	2020	Scenario-based models are used for the EV charging demand in the network	Optimal siting and sizing of soft opening points to control the line power flows in the distribution network	Minimize network losses and total uncharged EV network demand levels	Not used	Eight-node	Interior-point
[15]	2021	Mathematical model	Developed a framework for modeling and controlling residential smart EV chargers	Reduce the peak demand of the home	Used	–	–

Table 1 (continued)

References no.	Year	Modeling of EV charging demand	Contribution	Objective	Real-time travel survey data	Distribution network studied	Algorithm used
[16]	2022	Gaussian mixture model	Assessed the interaction of EVs with the residential electricity distribution grid in terms of deterioration of the quality of the grid voltage and overloading of the grid assets	The performance of the charging strategies from EV owner's perspective is studied using metrics such as delay in charging and average cost of charging	Not used	Modified version of a suburban-type Kerber network	–
[17]	2022	Probability distribution	To forecast the orderly charging load of EVs	Minimize the variance of the overall load in the residential area and the orderly charging load curves	Used	Residential area in a city in Henan Province	Nonlinear programming
[18]	2022	Gaussian mixture model	Aggregated charging management solution is developed, to keep the overall demand below the maximum limit at the medium voltage transformer level	Reduce the overload occurring in medium voltage networks by managing the electricity consumption with smart grid	Not used	–	–
[19]	2023	Mathematical model	Proposed charging scheduling based on intermittent PV	To reduce peak loads during intermittent PV events	Not used	–	Genetic algorithm
[20]	2023	Gaussian mixture model	Constructed a commercial demand analysis model for orderly charging of EVs in residential areas	Implementation of orderly charging in residential areas	Not used	– (Load power data of residents in a residential area in Hangzhou City, Zhejiang Province, China, is used)	–

These adverse effects can be addressed by the placement of appropriately sized DG resources in the distribution network. Another benefit of using DG units is that they lessen the carbon footprint associated with conventional generation (as an alternative to coal or diesel). Therefore, in this paper, a latest metaheuristic technique, COA is applied to the optimal allocation of DG resources in the 240-bus distribution network. A multi-objective problem is formulated to minimize the network power losses (P_{Tloss}) and voltage deviation (VD) with residential EV charging. Moreover, the effectiveness and superiority of the COA algorithm is verified over other popularly existing metaheuristic algorithms like gray wolf optimization (GWO), particle swarm optimization (PSO) and genetic algorithm (GA) for the proposed problem. It is discovered that COA performs significantly better in terms of accuracy and speed of convergence over the other optimization methods.

- **S-3:** As far as optimization for distribution networks is concerned, focus is concentrated either on the optimal allocation of fast charging stations or several DG units in the distribution network individually. However, very limited research has been conducted on the mitigation of the detrimental effects of slow, residential EV charging on distribution networks by allocation of DG units of appropriate size at suitable locations within the network.

C-3: This paper uses recently developed COA to optimally allocate DGs in the 240-bus distribution network to mitigate the adverse effects of slow, residential charging of about 1500 EVs pertaining to more than 1000 U.S residents in the state of Iowa, U.S.A. It is observed that with residential EV charging, the network load escalates mainly during or after the evening hours. Hence, only wind-based DGs are considered in this work. Case studies demonstrate that after optimal DG allocation using COA method, the network active power losses and the voltage deviation reduce by 38.45 and 51.61 percent, respectively. This validates the effectiveness of the proposed technique.

1.3 Organization of the Paper

Figure 1 presents the layout of the proposed work. The rest of the paper is arranged as follows: Sect. 2 explains the charging demand model for EVs which includes the mathematical equations and filtering of the NHTS 2017 data. Network modeling and DG operation are detailed in Sect. 3 which includes the description of 240-bus distribution network, direct power flow method, DG operation, selection of network buses for DG allocation and overview of COA. Section 4 discusses the case studies and results along with comments on the proposed problem which includes the formulation of the optimization problem and the effectiveness of the proposed technique vis-à-vis daily load dynamics, network power losses, voltage deviation and the bus voltage magnitude profiles, followed by the practical implications of the study and its scope for further work. Finally, Sect. 5 concludes the paper.

2 Charging Demand Model for EVs

Generally, there exist two types of EV charging methods, onboard charging and off-board charging, as shown in Fig. 2. The onboard charger for an AC grid connection can be either single-phase AC charging as defined by SAE J1772, Level 1 and Level 2 or three-phase AC charging as described by SAE J3068, whereas the off-board DC fast charging takes place according to SAE J1772 combo or CHAdeMO standard [21]. Onboard charging is a slow but cost-effective method of charging EVs in locations like residences, parking lots and public charging stations where EVs can be kept idle for extended periods of time without the need for emergency refilling. On the other hand, certain EV refueling stations offer off-board charging, which can facilitate rapid driving range extension in a time frame comparable to that of refilling a conventional gasoline vehicle. Although some industry leaders have suggested supercharging and battery swapping methods that can drastically reduce refueling times, the widespread implementation of such methods requires time for both completion and reaching of sufficient market demand [22]. Therefore, it is still anticipated that the majority of individual EV owners will choose to charge their vehicles conventionally overnight in residential areas in the foreseeable future [22]. Thus, in this paper, modeling of EVs is done for slow, residential charging using mathematical equations, NHTS 2017 data and EV specifications.

2.1 Modeling of EV Charging Demand

This study focuses exclusively on slow, residential EV charging within the U.S., utilizing the Level-1 SAEJ1772 standard which employs a single-phase, 120 V AC, 12A, 1.44 kW grounded outlet [23]. Since Level-2 equipment offers charging from 208 V, it does not pertain to a single-phase residential charging in the U.S., and is not considered in this paper. To understand the impact of EV charging on the electrical distribution system, it is crucial to take into account several parameters of EVs, including battery capacity, the state of charge (SOC), the energy required for charging and the charging level.

According to data from source [24], EVs are classified into four categories ranging from small city cars to SUVs, as detailed in Table 2.

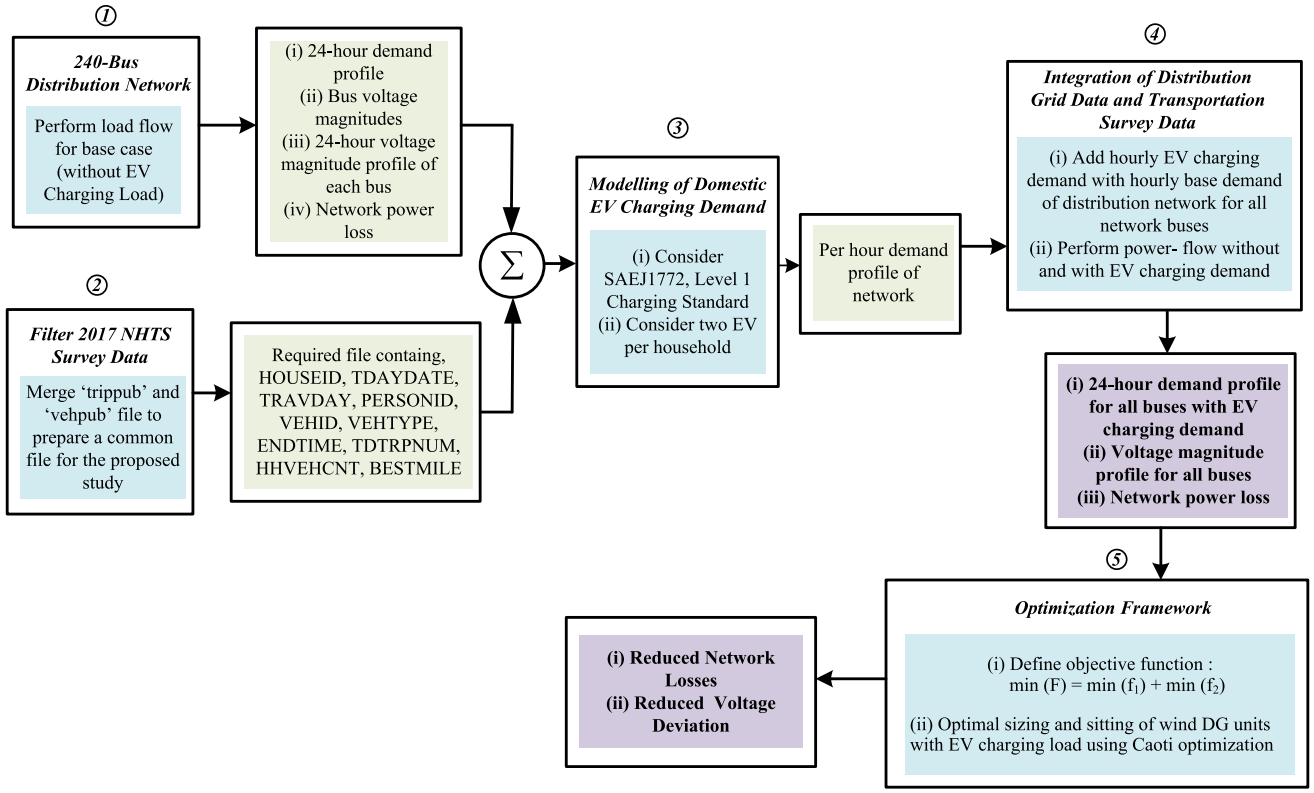
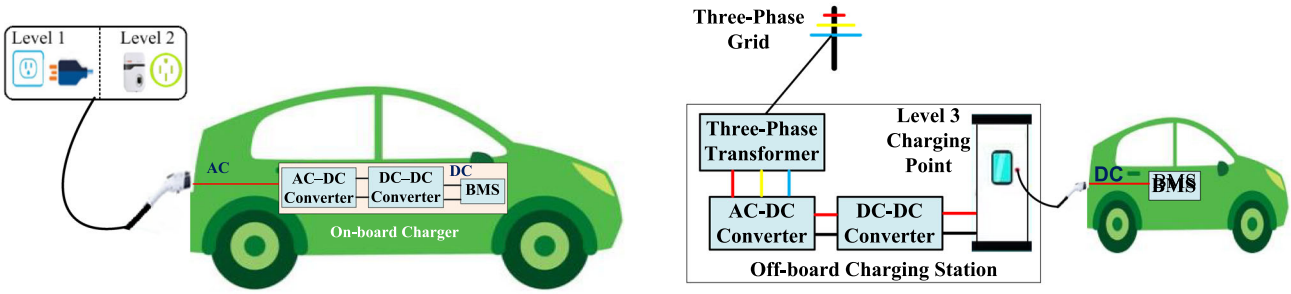


Fig. 1 Layout of the Proposed Work



(a) On-board Charging (b) Off-board Charging

Fig. 2 Architecture for EV Charging System

Table 2 ECPM and battery capacities for different EV types

S. no.	Vehicle Type	ECPM (kWh/mile)	EV Battery Capacity (kWh)
1	Small	0.24955	42
2	Medium	0.25599	58
3	Large	0.29946	77
4	SUV	0.38157	70

The SOC of an EV can be calculated as follows:

$$\mathrm{SOC} = \left(1 - \frac{\mathrm{ECPM} * d}{C_{\mathrm{batt}}}\right)100 \quad (1)$$

where, ECPM represents the electrical energy consumption per mile in kWh/mile, d is the total distance traveled in a day by the EV in miles and C_{batt} defines the EV battery capacity in kWh.

The SOC calculation identifies the amount of energy required to fully recharge the EV battery following the vehicle's most recent trip. The electrical energy required to achieve a full charge is drawn from the power grid and can be expressed mathematically as follows,

$$E_g = \frac{\left(1 - \frac{\mathrm{SOC}}{100}\right)C_{\mathrm{batt}}}{\eta} \quad (2)$$

where, η is the efficiency of charging the battery of an EV, which is taken as 0.9 [14] in this article. Finally, the time needed to charge the EV, T_{ch} is calculated as,

$$T_{ch} = \frac{E_g}{ch_\text{level}} \tag{3}$$

where, *ch*_level is the EV charging power rate in kW as per the charging standard.

Since, SAEJ1772 standard for Level-1 charging has been adopted for residential EV charging; therefore, the ch_level is set at 1.44 kW. In this paper, the type of charging considered is constant power-constant voltage (CP-CV) according to which the EV battery draws constant power till 80% SOC and as the battery SOC exceeds 80%, it starts to draw very small amount of current to protect the EV battery and maximize its lifespan [25].

2.2 Filtered NHTS 2017 Data

Transportation survey reports are the premier sources for information on vehicle and trip characteristics. Transportation survey data, when integrated with grid operation data, can enhance the understanding of the broader impacts of EV charging on the electrical power distribution network. This includes analyzing how transportation trends influence local and regional grid loads and identifying potential issues like voltage irregularities, daily load dynamics, network power loss, etc. The study in this paper utilizes NHTS 2017 data to extract vehicle trip characteristics. NHTS provides comprehensive travel pattern data for the USA [26], detailing EV counts per household, type of EV (dictates battery capacity), distances traveled, trip end times and more. Using data from the NHTS 2017, two Microsoft Excel files, 'trippub.csv' and 'vehpub.csv,' are utilized to analyze vehicle owners' behavior and extract important vehicle characteristics. This section thoroughly examines the NHTS dataset, extracting and explaining the relevant data from the report in detail.

2.3 A. trippub.csv

In this paper, nine attributes from this file have been utilized: household ID number (HOUSEID), person ID number (PERSONID), travel date (TDAYDATE), travel day of the week (TRAVDAY), vehicle ID number (VEHID), type of vehicle (VEHTYPE), travel day trip end time (ENDTIME), travel day trip number (TDTRPNUM) and the number of household vehicles (HHVEHCNT). PERSONID and TDTRPNUM specify the trip characteristics for a particular household member. VEHID and VEHTYPE identify the specific vehicle used in a trip and its type. ENDTIME denotes when the trip concludes, while TRAVDAY and TDAYDATE specify the trip's day of the week and date, respectively. In this paper, the trip end time (ENDTIME) is considered as the charging start time of EVs.

2.4 B. vehpub.csv

Eight attributes from this file are used in this paper: HOUSEID, TDAYDATE, TRAVDAY, PERSONID, VEHID, VEHTYPE, HHVEHCNT and BESTMILE. The first seven attributes are identical to those in the previous file. BESTMILE represents the best estimation of the annual miles driven by each vehicle [26]. To determine the daily mileage, BESTMILE is divided by 365.

To summarize, 'trippub.csv' provides trip characteristics, while 'vehpub.csv' details vehicle characteristics. To accurately determine EV characteristics, it is essential to know the total miles driven by the vehicle daily and the end time of each trip when it arrives home.

Finally, these two files are merged to generate useful data for use in this paper. The common attributes HOUSEID, TDAYDATE, TRAVDAY, PERSONID, VEHID, VEHTYPE and HHVEHCNT are used to combine the files, and the sample of which is presented in Table 3.

Figure 3 illustrates the frequency distribution of vehicle arrivals during a typical May weekday. To generate the 24-h network demand curve in the baseline scenario (excluding EV charging), smart meter data of the distribution network is utilized from all customers. Then, we calculate the additional dynamic demand from EV charging using Eqs. (1)–(3) and incorporate it into the baseline demand, resulting in the 24-h demand curve including EV charging as shown in Fig. 4.

3 Network Modeling and DG Operation

3.1 240-Bus Distribution Network

Figure 5 illustrates the 240-bus distribution network, which is part of a municipal utility in Iowa, U.S.A [27]. This network serves a residential area and includes smart meters at every customer location. There are 1079 customers in the network. Each customer is considered as a household receiving single-phase supply at 120 V. The customers are appropriately distributed [27] among 139 single-phase and 55 three-phase load buses. To ensure phase balance, the allocation of customers to each three-phase bus is done by dividing them evenly among the three phases ('A,' 'B' and 'C'). The network structure consists of three feeders with bus numbers starting from '1,' '2' and '3' representing feeder '1,' '2' and '3' buses respectively as shown in Fig. 5. It is noted in the study that all customers are assumed to charge their EVs at home after their final trips. The following methodology is used to allocate the EV charging demand in the distribution network:

Table 3 Sample of filtered NHTS 2017 data used

HOUSEID	TDAYDATE	TRAVDAY	PERSONID	VEHID	VEHTYPE	ENDTIME	TDTRPNUM	HHVEHCNT	BESTMILE
30000007	201608	2	3	1	1	900	1	5	14611.92
30000008	201608	5	1	4	4	2340	2	4	9192.44
300000012	201607	5	1	2	3	605	1	2	4002.55
300000019	201605	5	1	1	1	1515	1	2	7027.68
300000029	201610	3	1	1	3	845	1	2	14029.07

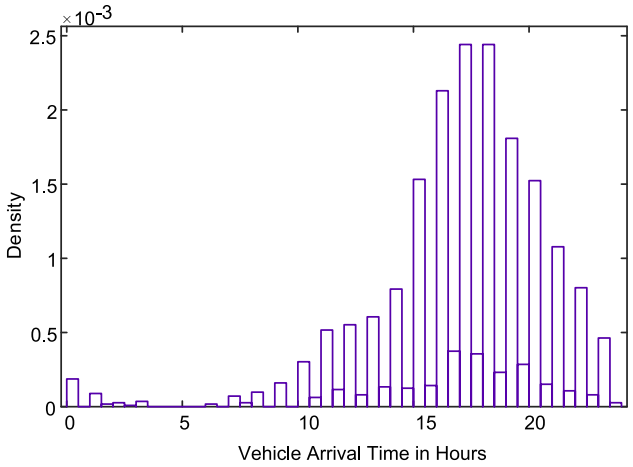


Fig. 3 Vehicle Home Arrival Time Pattern from NHTS 2017

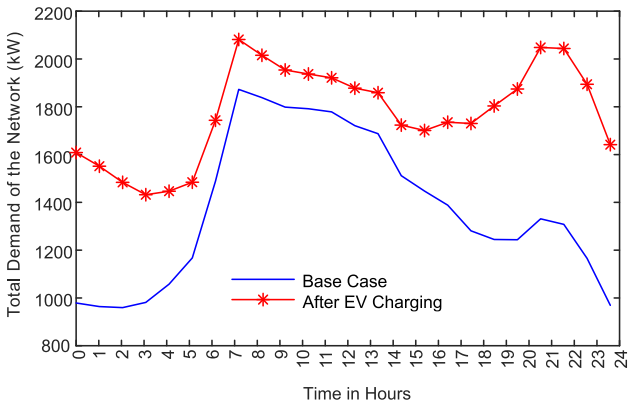


Fig. 4 Total Network Demand for a Typical Weekday of May

(i) Each household (from NHTS 2017 data) represents a single-phase customer.
(ii) The EV charging demand of each household (customer) is computed using (1)—(3) as explained in Sect. 2.
(iii) For three-phase buses of the network, the customers are distributed evenly among the three phases ('A,' 'B' and 'C'), whereas for single-phase bus, the customers are allocated to the phase in which that single-phase bus is connected. ('A'/'B'/'C').
(iv) Finally for all buses of the distribution network, the demand due to EV charging load of each phase is calculated by summing the demands of all the customers pertaining to that phase.
(v) Following the aforementioned points (i–iv), a 24-h demand profile for the complete 240-bus distribution network is computed (shown in Fig. 4).

3.2 Direct Power Flow Method

Most distribution networks are built in a radial structure with a low ratio of reactance to resistance. Traditional load flow methods like Gauss–Seidel and Newton–Raphson are ineffective in these cases. Therefore, this article uses a direct search-based approach for power flow analysis [33]. The suggested power flow approach involves creating two matrices: one for bus injection to branch current (BIBC) and another for branch current to bus voltage (BCBV). By multiplying these matrices together, the load flow solutions can be determined efficiently. Figure 6 illustrates a radial 6-bus distribution network, with P_1 to P_5 denoting branch currents and Z_{12} to Z_{56}, representing the impedance of the respective branches.

The current injected at node 'i' at 'k^{th}' iteration is determined as,

$$I_i^k = \left(\frac{P_i + jQ_i}{V_i^{k-1}}\right)^* \tag{4}$$

where, I_i, P_i, Q_i and V_i^{k-1} represents the injected current, active power, reactive power and voltage at thei^{th} bus, respectively.

The current described in (4) can be decomposed into its active and reactive components as,

$$\text{Real}\ (I_i) = \frac{P_i \cos\theta_i + Q_i sin\theta_i}{|V_i|} \tag{5}$$

$$\text{Imag}\ (I_i) = \frac{P_i \sin\theta_i - Q_i \cos\theta_i}{|V_i|} \tag{6}$$

where, Θ_i is the voltage angle ati^{th} bus.

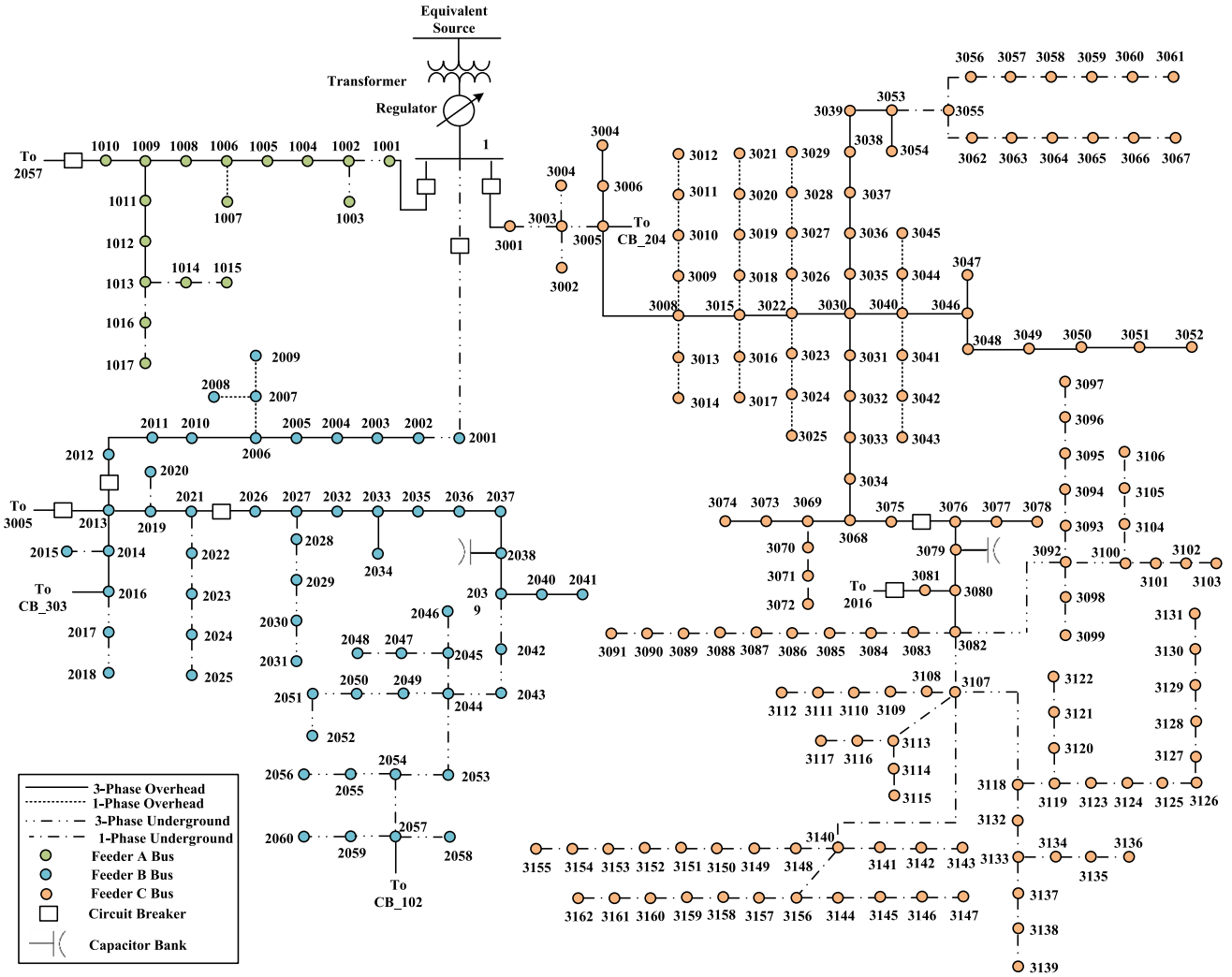


Fig. 5 240-Bus Distribution Network

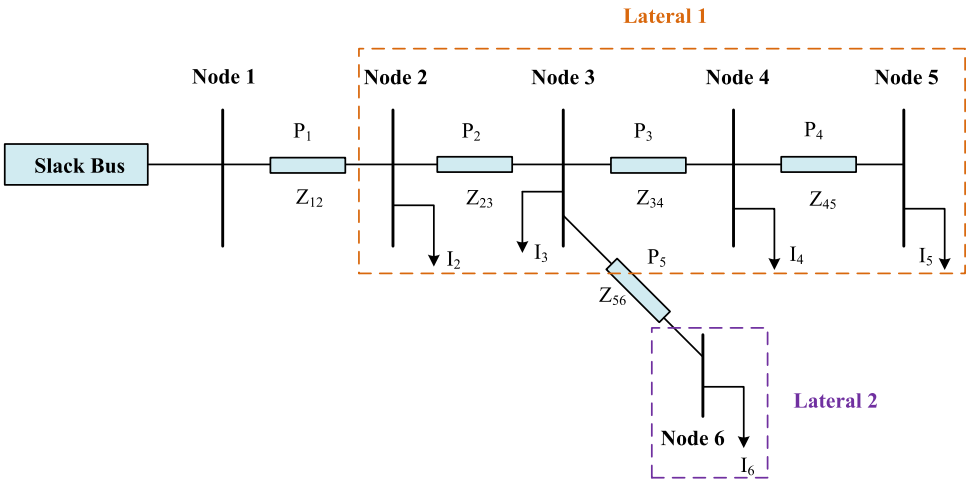


Fig. 6 A Simplified 6-Bus Distribution Network

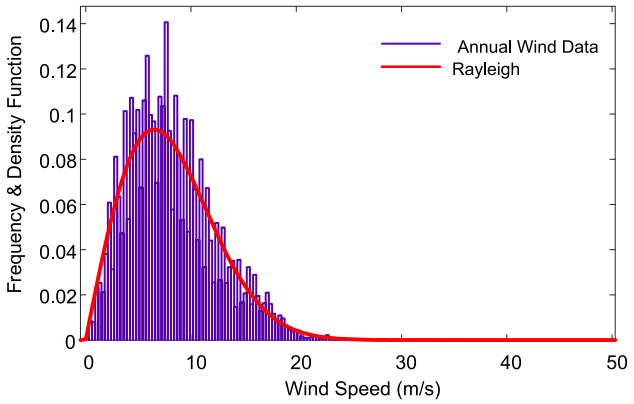


Fig. 7 Annualized Wind Speed Pattern in Iowa [28]

The branch current matrix B and matrix P is computed using Kirchhoff's current law (KCL) in accordance with (6).

$$\begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{bmatrix} = \begin{bmatrix} 1\,1\,1\,1\,1 \\ 0\,1\,1\,1\,1 \\ 0\,0\,1\,1\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \end{bmatrix} \begin{bmatrix} I_2 \\ I_3 \\ I_4 \\ I_5 \\ I_6 \end{bmatrix} \tag{7}$$

$$[P] = [BIBC][I] \tag{8}$$

Conversely, the voltage drops at each bus relative to the reference bus is calculated utilizing Kirchhoff's voltage law (KVL) in the subsequent manner as,

$$\begin{bmatrix} V_1 \\ V_1 \\ V_1 \\ V_1 \\ V_1 \end{bmatrix} - \begin{bmatrix} V_2 \\ V_3 \\ V_4 \\ V_5 \\ v_6 \end{bmatrix} = \begin{bmatrix} Z_{12} & 0 & 0 & 0 & 0 \\ Z_{12} & Z_{23} & 0 & 0 & 0 \\ Z_{12} & Z_{23} & Z_{34} & 0 & 0 \\ Z_{12} & Z_{23} & Z_{34} & Z_{45} & 0 \\ Z_{12} & Z_{23} & 0 & 0 & Z_{36} \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{bmatrix} \tag{9}$$

$$[\Delta V] = [BCBV][P] \tag{10}$$

Substituting the matrix P value from (7) into (9), we obtain

$$\begin{bmatrix} V_1 \\ V_1 \\ V_1 \\ V_1 \\ V_1 \end{bmatrix} - \begin{bmatrix} V_2 \\ V_3 \\ V_4 \\ V_5 \\ v_6 \end{bmatrix} = \begin{bmatrix} Z_{12} & 0 & 0 & 0 & 0 \\ Z_{12} & Z_{23} & 0 & 0 & 0 \\ Z_{12} & Z_{23} & Z_{34} & 0 & 0 \\ Z_{12} & Z_{23} & Z_{34} & Z_{45} & 0 \\ Z_{12} & Z_{23} & 0 & 0 & Z_{36} \end{bmatrix} * \begin{bmatrix} 1\,1\,1\,1\,1 \\ 0\,1\,1\,1\,1 \\ 0\,0\,1\,1\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \end{bmatrix} * \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \end{bmatrix} \tag{11}$$

$$[\Delta V] = [BCBV]\,[BIBC]\,[I] \tag{12}$$

$$[\Delta V] = [DLF][I] \tag{13}$$

where, DLF represents distribution load flow matrix which is employed to assess the voltage drop at each bus concerning the reference bus. Two steps are involved in deriving the DLF as given below:

(1) Set the branch impedance vector Z_b using the branch data.
(2) Transform the branch impedance vector into a diagonal matrix, and set all off-diagonal elements to zero, leaving only the diagonal elements unchanged. Then, multiply this diagonal matrix by the current vector I to calculate ΔV

3.3 DG Operation

To mitigate the negative impact of EV charging on the network performance, DGs of suitable sizes are strategically located at appropriate bus locations within the network. The increase in the magnitude of the network demand caused by residential EV charging typically occurs during or after the 17th hour (as shown in Fig. 4), particularly in the evening hours which is obvious as the EV owners return back to their home from their respective workplaces (as shown in Fig. 3) and plug-in their EVs for charge. Since solar-based DGs are not effective at these hours, only wind-based DGs are considered to tackle this problem.

The performance of a wind turbine is influenced by the wind speed at its location. Studies in this paper are conducted with the goal of meeting 20% of the overall network demand using wind turbine-based DGs. Figure 7 illustrates the annual wind speed profile in Iowa, located in the U.S. Midwest [28]. The wind speed is modeled assuming a Rayleigh distribution [29, 30], represented by:

$$f(\upsilon) = \frac{2}{c^2} e^{-(\frac{\upsilon}{c})^2} \quad 0 \le \upsilon \le \infty \tag{14}$$

where, c and v are scale index of the Rayleigh distribution and wind speed in m/s, respectively.

The rated power delivered by the rotor is governed by the following equation [37]:

$$P_{w,r} = 0.5 * C_p * \rho * \pi * R^2 * v_r^3 \tag{15}$$

where, C_p is the turbine performance coefficient, ρ is the air density in kg/m^3 and R is the radius of the turbine in m.

Table 4 presents the parameters utilized for the wind DG in this study. The wind speed data for the location is sourced

Table 4 Specifications of the wind turbine

Parameter	Value
C_p	0.3326
R	12.5 m
$P_{w,r}$	100 kW
v_r	10 m/s
v_{ci}	2.5 m/s
v_{co}	20 m/s
ρ	1.225 kg/m^3

from the IEM [28]. The turbine's output power, across varying wind speeds, adheres to the formula described in [29] as follows,

$$P = \begin{cases} 0 & v_w \le v_{ci} \text{ OR } v_w \ge v_{co} \\ P_{w,r}\left(\frac{v_w - v_{ci}}{v_r - v_{ci}}\right)100 & v_{ci} \le v_w \le v_r \\ P_{w,r} & v_r \le v_w \le v_{co} \end{cases} \quad (16)$$

where, $P_{w,r}$ is the rated power output in kW.

v_w is the wind speed in m/s for which power output is to be determined.

v_r is wind speed in m/s.

v_{ci} is cut in wind speed in m/s.

v_{co} is cut out wind speed in m/s.

Based on the analysis depicted in Fig. 4, the network's peak demand of 2048.31 kW is observed to occur at 22 h with EV charging. Consequently, optimal allocation of DG resources is carried out to address this specific peak demand as the planning of DG capacity should be done considering the peak demand. To meet this peak demand with EV charging load and assuming a wind power penetration rate of 20%, four 100 kW wind turbines (detailed specifications in Table 4) are optimally allocated in the distribution network using COA [31].

3.4 Selection of Network Buses for DG Allocation

To select the suitable nodes (buses) for placing DGs, the initial step involves choosing three-phase nodes (buses) in the distribution network. The final determination of the most suitable nodes among these is based on the loss sensitivity factor (LSF) [32], calculated as follows:

$$\frac{\partial P_{T\text{loss}}}{\partial P_{\text{eff}}} = \frac{2 * P_{\text{eff}}[q] * r_j}{(V[q])^2} \quad (17)$$

where, the total effective reactive power supplied beyond node *'q'* is denoted as $P_{eff}[q]$, and r_j represents the resistance of the j^{th} line. The computed values of LSF are subsequently arranged in the descending order for all the lines of the network. This ordering of LSF the descending order dictates the sequence for considering buses in the allocation of DG. Additionally, at these buses, the normalized voltage magnitudes are calculated as,

$$\text{norm}\ (i) = \frac{V[i]}{0.97} \quad (18)$$

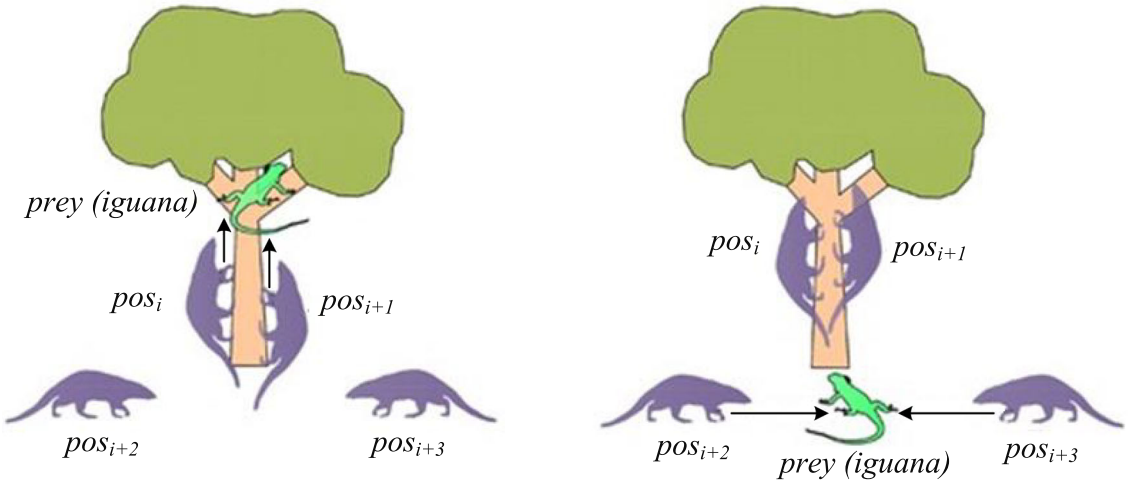
For buses, whose *norm(i)* value is less than 1.01 is considered as candidate buses for DG allocation. These candidate buses are stored in a vector and are considered as buses which requires DG placement. In this way, the LSF determines whether a bus needs DG placement or not, i.e., 'DG siting,' whereas the optimal size of DG unit, i.e., 'DG sizing' for these candidate buses is calculated using COA optimization in this paper, which is detailed in the next subsection (Fig. 8).

3.5 Overview of COA

3.5.1 Initialization

COA is a metaheuristic optimization algorithm whose fundamental principles are detailed in [31], where researchers have been intrigued by coatis' intricate behaviors like hunting iguanas (prey) and evading predators. The metaheuristic COA optimizes populations by treating them as active contributors, with variable values determined by the positional data of each coati within the search space. Figure 9 presents the flowchart of the algorithm, while Algorithm 1 depicts its pseudo-code. Initially, in COA implementation, coatis' positions within the search space are randomly set between the lower and upper boundary as shown in (19).

$$\text{pos} = \begin{bmatrix} \text{pos}_1 \\ . \\ . \\ \text{pos}_i \\ . \\ . \\ \text{pos}_N \end{bmatrix}_{N\text{X dim}} = \begin{bmatrix} \text{pos}_{1,1} & \text{pos}_{1,j} & \text{pos}_{1,dim} \\ . & . & . \\ . & . & . \\ \text{pos}_{i,1} & \text{pos}_{i,j} & \text{pos}_{i,dim} \\ . & . & . \\ . & . & . \\ \text{pos}_{N,1} & \text{pos}_{N,j} & \text{pos}_{N,dim} \end{bmatrix}_{N\text{ X dim}} \quad (19)$$



(a) Exploration

(b) Exploitation

Fig. 8 Exploration and Exploitation Phase of COA

where, *pos i*s coati's position, *N* represents the number of coatis and *dim* represents the dimension.

In the optimization process, when evaluating a potential location, if its fitness exceeds that of the current location, the current fitness value is updated to this higher value. Conversely, if the next evaluation does not yield a superior fitness, the current fitness value remains unchanged. This method ensures that the best fitness value discovered thus far is maintained throughout the optimization process. By consistently adjusting the fitness value to reflect improved solutions, the algorithm works toward converging to the optimal solution. This approach of managing and updating fitness values enables the COA to monitor progress and preserve the highest fitness value achieved during the process.

3.5.2 Exploration

Initially, coatis are generated randomly within a search space as explained in the aforementioned section. However, this phase is the exploration phase of the algorithm where half of these coatis ascend a tree to startle iguanas, while the remaining half waits beneath the tree to ambush the falling iguanas. The climbing coatis adjust their positions based on (20). When the coatis frighten an iguana, the iguana relocates randomly within the search space. Each coati's position, denoted as *pos*, is updated in accordance with (21), which compares the fitness of their current position to the optimal fitness in the mathematical model as shown in (21).

$$\text{prey}_j = lb_j + \text{rand}\,(0,\ 1)\,.\,(ub_j - lb_j), \quad j = 1,\ 2,\ \ldots\ldots,\ \dim \tag{20}$$

$$\text{pos}_{i,\,j}^{p1} = \begin{cases} pos_{i,\,j} + \text{rand}(0,\ 1).\left(\text{prey}_j - I.\text{pos}_{i,\,j}\right), & F_{\text{prey}} < F_i \\ pos_{i,\,j} + \text{rand}(0,\ 1).\left(\text{pos}_{i,\,j} - \text{prey}_j\right), & \text{else} \end{cases}$$
$$\text{for } i = \frac{N}{2} + 1,\ \frac{N}{2} + 2,\ \ldots\ldots,\ N \quad \text{and} \quad j = 1,\ 2,\ \ldots\ldots,\ \dim \tag{21}$$

In this context, *prey* refers to the location of the iguanas on the ground,*j* represents the dimension, pos_i denotes the current coati location, *p1* represents the first phase and *I* is a randomly generated integer from the set [1, 2], while F_{prey} involves evaluating the fitness of the current position. Further, (22) is used to compute the appropriate value for the current position. If the current fitness is lower than the optimal fitness, it indicates that the current position of the coatis is preferable, and their position is updated accordingly. Otherwise, the coatis remain in their original position.

$$\text{pos}_i = \begin{cases} \text{pos}_i^{p1}, & F_i^{p1} < F \\ \text{pos}_i, & \text{else} \end{cases} \tag{22}$$

where, $F_i{}^{p1}$ represents the fitness of the current position, $pos_i{}^{p1}$ signifies the calculation of the updated position and pos_i refers to the updated position.

3.5.3 Exploitation

The next phase discusses the exploitation phase, which is mathematically represented through a simulation of coatis responding to predators. The coatis' positions are adjusted based on (23) and (24) as they evade predators. When threatened by a predator, the coatis swiftly move away from their

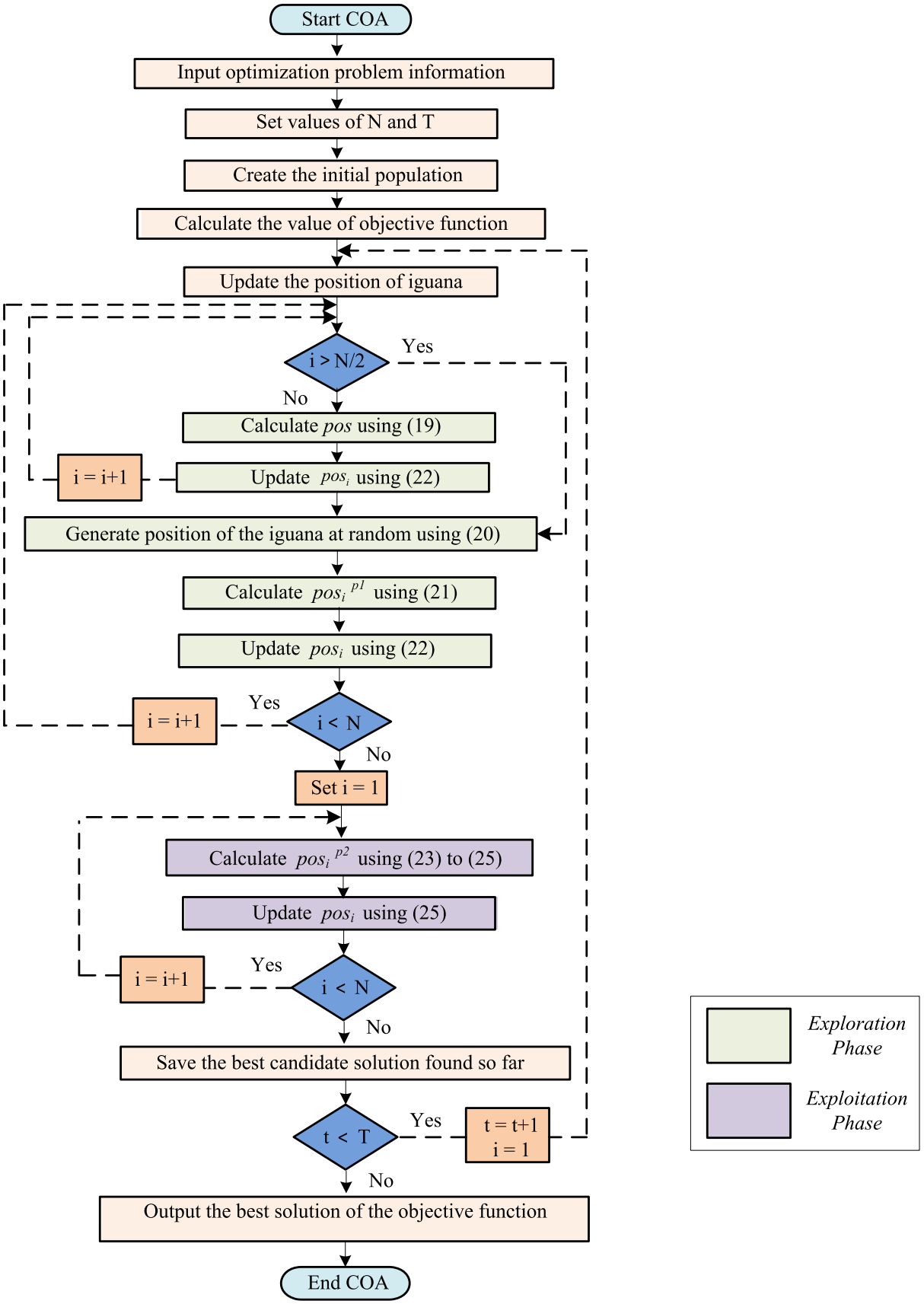


Fig. 9 Flowchart of COA

current location using a predator avoidance tactic to maintain a safe proximity. This behavior showcases the coatis' adeptness at conducting local searches.

$$lb_j^L = \frac{lb_j}{t},\ ub_j^L = \frac{ub_j}{t},\ t = 1, 2, ..., T \tag{23}$$

$$\text{pos}_{i,j}^{p2} = \left\{\text{pos}_{i,j} + (1 - 2\text{Xrand}(0, 1)).\left(lb_j^L + \text{rand}(0, 1).\left(ub_j^L - lb_j^L\right)\right)\right. \quad \text{for } i = 1, 2, ..., N \quad \text{and} \quad j = 1, 2, ..., \text{dim} \tag{24}$$

the current position of the coati is considered superior; otherwise, the coatis maintain their original positions. Here, $F_i{}^{p2}$ stands for the fitness of the calculated current position, $pos_i{}^{p2}$ denotes the current position and pos_i represents the updated position.

$$\text{pos}_i = \begin{cases} \text{pos}_i^{p2}, & F_i^{p2} < F \\ \text{pos}_i, & \text{else} \end{cases} \tag{25}$$

The pseudo-code for the COA algorithm is given in Algorithm 1.

Algorithm 1 Pseudo-code for COA

Input: N, T
Output: The global optimum
Begin
pos: Initialize the random positions of *N* coatis
for t = 1: *T*
F_i *:* Calculate the fitness of individual pos_i member
end
Update the position of the iguana based on the position of the best member of the population

Exploration Phase:
for i = 1 : (*N*/2)
$Pos_i{}^{p1}$*:* Update coati's position according to (19)
F_i : The greedy strategy to select the optimal fitness is used as per (22)
end
for i = 1+(*N*/2) : *N*
$Pos_i{}^{p1}$*:* Update coati's position according to (21)
F_i : The greedy strategy to select the optimal fitness is used as per (22)
end

Exploitation Phase:
Calculate the local bounds for variables using (23)
for i : *N*
$Pos_i{}^{p2}$*:* Update coati's position according to (23) and (24)
F_i : The greedy strategy to select the optimal fitness is used as per (25)
end
Maintain the current optimal position and optimal solution found so far
End
Return the global optimum solution for the proposed problem by COA
End

In the given scenario, *p2* denotes the second stage; $lb_j{}^L$ and $ub_j{}^L$ signify the local lower and upper bounds in dimension *j*, respectively; lb_j and ub_j denote the lower and upper bounds of dimension *j*; *T* represents the maximum number of evaluations; $pos_j{}^{p2}$ indicates the new position of the coatis calculated in the second stage; $pos_{i,j}{}^{p2}$ specifies the position of dimension *j*. Then, the current position's fitness is determined using (25). If this fitness is less than the optimal fitness,

4 Results and Comments

4.1 Formulation of Optimization Problem for DG allocation Using COA

As presented in Fig. 4, when residential EV charging is included, the distribution network's peak load happens at 22 h, causing a significant increase in demand. Therefore, the voltage magnitude profiles for the 240-bus network at

22 h are analyzed both with and without EV charging (the latter being the base case). Following this, the optimal allocation of wind-based DGs using COA method is determined for the 240-bus network pertaining to 22-h peak load, which represents a worst-case scenario with EV charging included. In this paper, the objective function formulated for the optimal placement of wind-based DG units with EV charging load is formulated as a multi-objective problem with the aim to reduce the total network losses and voltage deviation, which are given in (26)-(28). Equal weights are assigned to both the objectives in (26).

$$\text{minimize} \quad F = w_1 * P_{T\text{loss}} + w_2 * VD \tag{26}$$

where, $w_1 = w_2 = 0.5$

$$P_{T\text{loss}} = \frac{\sum_{j=1}^{b} |I_j|^2 r_j}{P_{T\text{losss}}^{\text{base}}} \tag{27}$$

$$VD = \sum_{k=1}^{n} (|V_k| - V_{\text{nominal}})^2 \quad p.u. \tag{28}$$

The aim is to maximize the effectiveness of the objective function within the boundaries of the established constraints. The constraints considered here are:

(1) DG power capacity:

$$P_{DG,\min} \leq P_{DG,n} \leq P_{DG,\max}$$

(2) Node Voltage:

$$V_{\min} \leq V_n \leq V_{\max}$$

(3) Thermal limit of each branch:

$$I_j \leq I_{\max,j}$$

where, V_{min} and V_{max} are the minimum and maximum value of V_n.

$P_{DG,min}$ and $P_{DG,max}$ are the minimum and maximum value of P_{DG}.

$I_{max,j}$ is the maximum current capacity of thej^{th} branch.

4.2 Performance Evaluation

With the objective to reduce network ohmic losses and voltage deviation, a multi-objective optimization problem to optimally allocate wind-based DGs is formulated in this paper to cater to the increased demand due to EV charging, which is already discussed in the previous section. With this, the analysis is done for the power system parameters like daily load dynamics, network active power loss, network voltage deviation and bus voltage magnitude profiles, which is discussed in detail in the following subsections.

4.2.1 Daily Load Dynamics

The 24-h demand profiles for a typical weekday of May pertaining to both base case (with only network base load and without EV charging load) and with EV charging (with network base load and EV charging load) are shown in Fig. 10. The plots for the phases 'A,' 'B' and 'C' are shown in red, yellow and blue, respectively. It is observed that in general, for all three phases, the network load escalates mainly during evening hours starting from the 16th hour onwards. This is because for slow, residential charging, the EV owner plug-in their EVs to charge after returning home from their respective workplaces. In addition to this, it is observed that for the.

240-bus distribution network considered, phases 'A' and 'C' show slightly more erratic nature as compared to phase 'B.' This is not unexpected as the (i) type (dictates battery capacity), (ii) distance traveled (dictates battery SOC) and (iii) trip end time (taken as charging start times) of the EV(s) are random variables. Hence, two customers having the same number and type of EVs may have different charging demands. The increase in the magnitude of the demand with EV charging for the 240-bus distribution network (with peaks at 22 h) is detailed in Table 5.

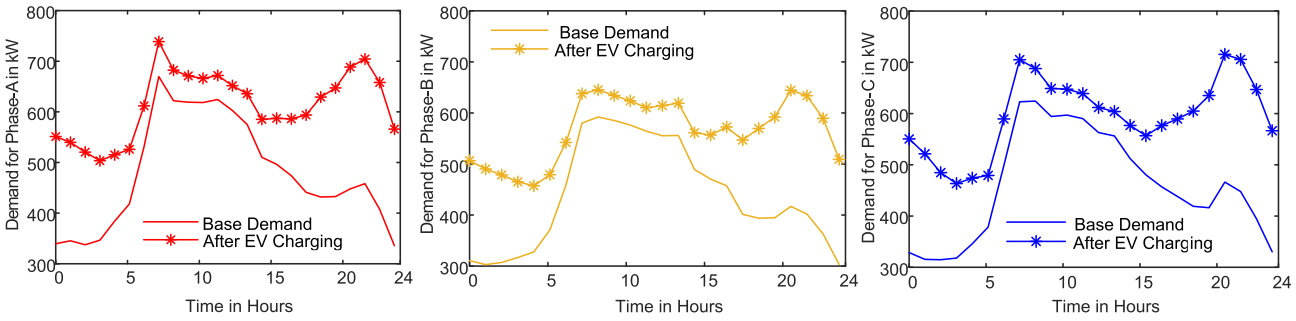


Fig. 10 Individual Phase Demand Profiles for a Typical Weekday of May

Table 5 Per phase load dynamics for different scenarios

Scenario	At 22 Hours		
Phase	'A'	'B'	'C'
Base case demand (kW)	458.17	401.968	447.866
Total demand with EV charging (kW)	704.41	633.808	705.626
Increase in demand (kW)	246.24	231.84	257.76

4.2.2 Optimization Results (Network Active Power Loss and Bus Voltage Magnitude Deviation)

The results for the optimal allocation of wind-based DG units in the 240-distribution network using COA method based on (26), (27), (28) and the aforementioned constraints are presented in Tables 6 and 7. The optimal sites (buses) for DG allocation are found out to be at buses 3081, 3078, 3077 and 3074 with capacities of 97.1276 kW, 98.2244 kW, 94.4234 kW and 98.7513 kW, respectively. It is observed that using the proposed COA method for finding out the appropriate 'DG sizing' and using LSF for finding out the candidate buses for 'DG siting,' the network active power loss reduces by 38.45 percent, whereas the percentage reduction for the network bus voltage magnitude deviation is found out to be 51.61 percent. This validates the effectiveness of the proposed method.

Moreover, in order to verify the efficacy of the proposed COA method, the convergence curve of COA is presented for hundred iterations with other well-known and commonly used metaheuristic optimization algorithms like GWO, PSO and GA for the proposed optimization problem is shown in Fig. 11a, while Fig. 11b provides the magnified picture of the obtained results. As the proposed multi-objective optimization problem in (26) is of minimization (minimization of network active power loss and bus voltage magnitude deviation), therefore, the algorithm which results in the least minimum value of the objective function will be considered as superior in finding out the best optimal solution subject to the network constraints.

Table 7 DG placement results: multi-objective COA (bus voltage magnitude deviation)

Parameters	VD before DG placement (p.u.)	VD after DG Placement (p.u.)	% reduction in VD
Base case (without EV, without DG)	0.07	–	–
After EV charging	0.31	0.15	51.61

From Fig. 11b, it is observed that COA yields the minimum value of the objective function in the least number of iterations as compared to the other existing algorithms. On comparing the performance of the algorithms in Fig. 11 b, it is observed that the COA, GWO, GA and PSO achieves the objective value of 16.8629, 16.8639, 16.8662 and 16.872 respectively in 23, 92, 77 and 79 iterations, respectively. Therefore, it can be concluded that among all the four methods, COA has better accuracy and has the fastest convergence, thus making it more efficient in reaching optimal or near-optimal solutions within a fewer number of iterations as compared to the other existing algorithms. The detailed results of the comparative performance analysis of COA with other methods is provided in Table 8 which illustrates that COA maintains a good balance between the exploration of the global search space and the exploitation of local optima which helps in avoiding local minima and finding the global optimum solution for DG placement and sizing. This validates the proficiency of the COA algorithm used in this study to minimize the network active power losses and bus voltage magnitude deviation with residential EV charging.

4.2.3 Bus Voltage Profiles

Figure 12 presents the bus voltage magnitude profiles of the 240-bus distribution network for phases 'A,' 'B' and 'C' considering three different scenarios: (a) base case, (b) with EV charging but without any DGs and (c) with EV charging and

Table 6 DG placement results: multi-objective COA (network active power loss)

Parameters	Bus no	Wind turbine output (kW)	Losses before DG placement (kW)	Losses after DG placement (kW)	% reduction in losses
Base case (without EV, without DG)	–	–	31.06	–	–
After EV charging	3081 3078 3077 3074	97.1276 98.2244 94.4234 98.7513	54.6468	33.6368	38.45

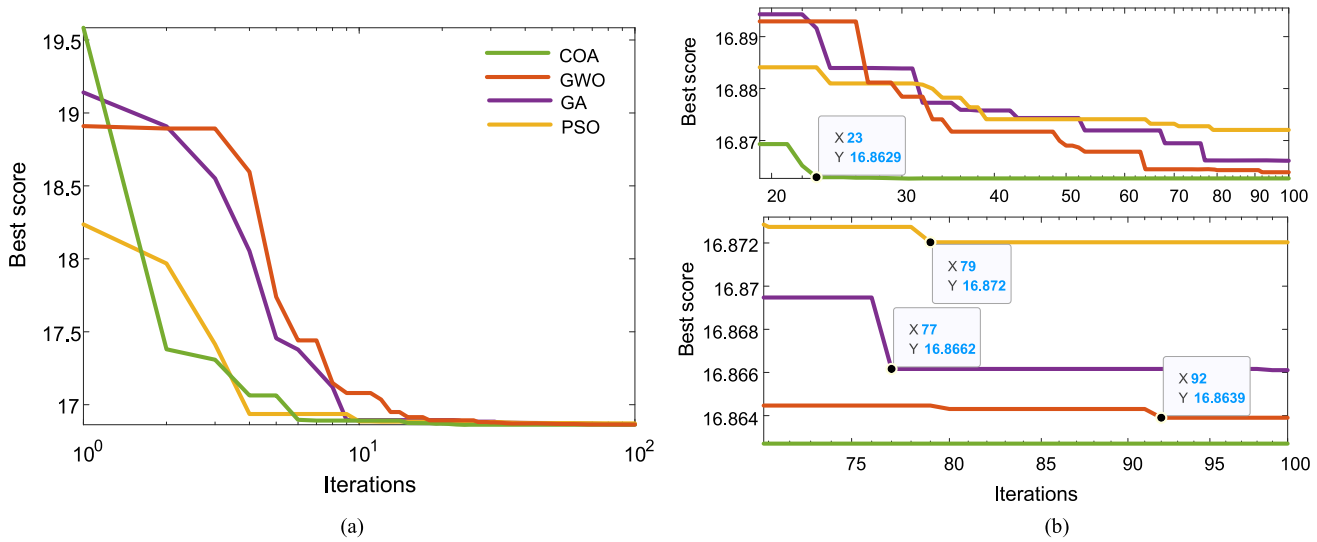


Fig. 11 Comparative Plot for Convergence Curves of COA with Other Algorithms

Table 8 Comparative performance evaluation of COA with other existing algorithms

Parameters	COA	GWO	GA	PSO
Best score	16.8629	16.8639	16.8662	16.872
Convergence iteration count	23	92	77	79
CPU time (sec)	162.98	715.29	406.44	433.57

with optimal DG allocation. From Fig. 12, it is observed that in general, as the bus number increases, their voltage magnitude decreases due to the radial topology of the network. However, the network consists of three different feeders, with the first bus of each feeder connected to the substation. Hence, from Fig. 12, the first bus of each feeder is also observed to have a voltage magnitude of 1 p.u.

From Fig. 12, it is further observed that after considering EV charging, the bus voltage magnitudes for all three phases have reduced from their base case values and the decrease in the bus voltage magnitude is more for phases 'A' and 'C' than phase 'B.' This is due to the difference in the EV charging demands across the different phases, as already discussed at the end of Sect. 4.2.1. To address this deterioration of the bus voltage magnitudes, DGs of appropriate sizes are placed at the optimal locations using the COA method, with the results shown in the green plot (Fig. 12). Subsequent to optimal DG allocation, from Fig. 12, it is also observed that for phases 'A' and 'C' which are highly affected, the bus voltage magnitudes are almost matching the base case bus voltage magnitude profile. However, for Phase 'B,' with DG allocation, the bus voltage magnitudes display much improvement over the base case bus voltage magnitude profile. This is due to the fact that unlike phases 'A' and 'C,' the lower EV charging demand in phase 'B' is overcompensated by the three-phase DG(s). Therefore, with optimal DG allocation in the 240-bus distribution network using the COA method, the bus voltage magnitude profiles for all the phases have improved significantly which confirms the effectiveness of the proposed method in countering the deteriorating effects of the EV charging demand.

4.2.4 Practical Implications of the Results

This paper focuses on slow, residential EV charging, as it is the most basic and essential mode of charging for EV owners due to its convenience and cost-effectiveness. This paper addresses the mitigation of detrimental effects of slow, residential EV charging on distribution networks like higher network active power losses and bus voltage magnitude deviations. To achieve this, optimal allocation of wind-based DG resources within a 240-bus power distribution network is carried out using COA technique. Detailed results of this approach are explained in the subSects. 4.2.1 to 4.2.3 of this paper. The key findings and practical implications of this work in improving the distribution network performance with high EV adoption are summarized as follows:

- Power distribution networks are complex and large, featuring numerous buses. COA technique is used for the allocation of appropriately sized DGs at suitable network bus locations to mitigate the detrimental effects of EV charging on the network performance. The core principles of the COA facilitate easier implementation and reduced

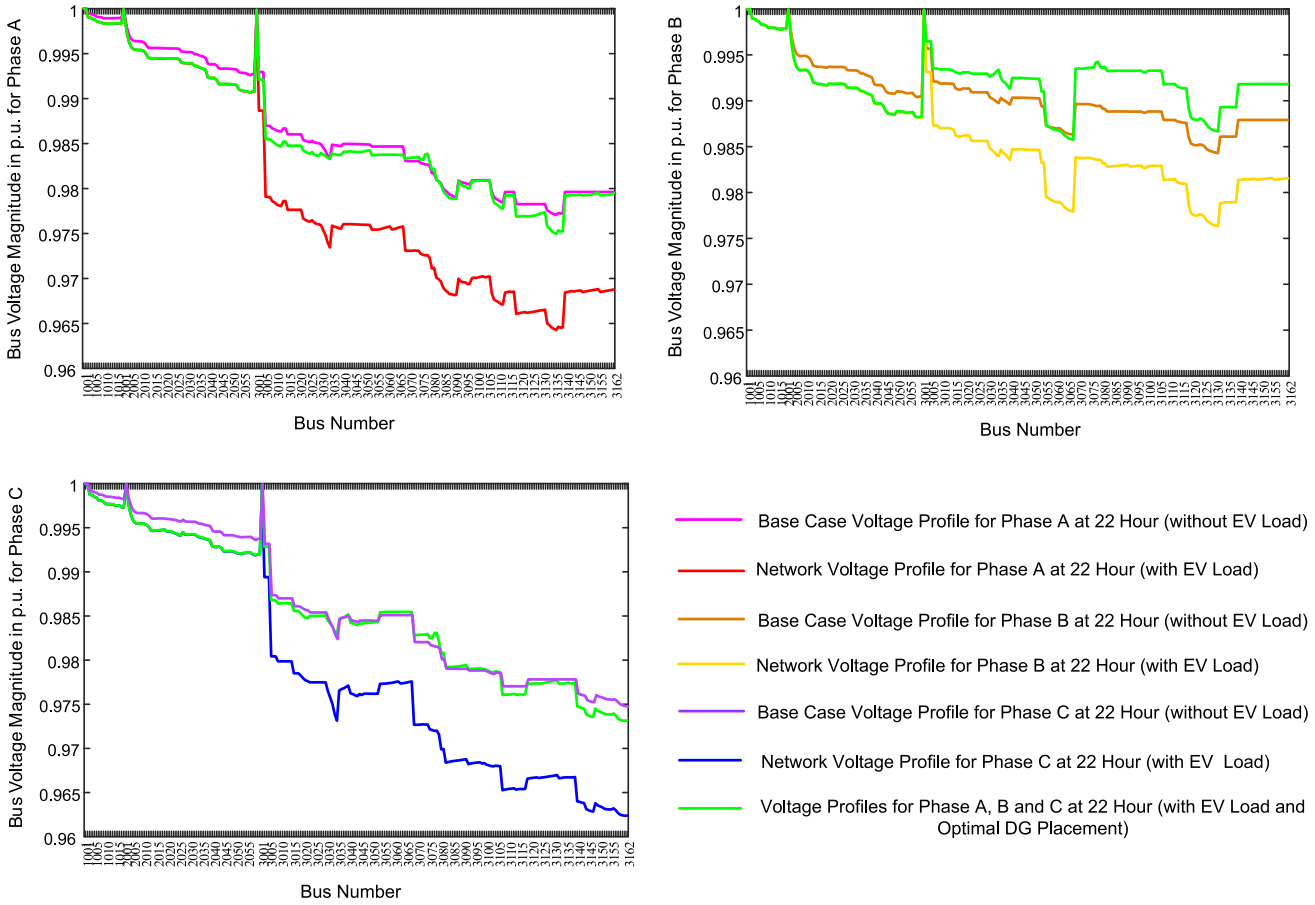


Fig. 12 Bus Voltage Profiles for Individual Phases at Peak Load Period (22 Hours)

computational complexity which enhances its appeal for practical applications. Additionally, the COA is scalable and adept at managing large-scale optimization problems efficiently, making it well-suited for real-world power system applications.

- Optimal DG allocation often encompasses multiple goals, such as minimizing network power losses, enhancing bus voltage magnitude profiles etc. The COA effectively addresses these multi-objective, high-dimensional challenges in optimization by balancing the exploration of the global search space with the exploitation of the local optima. Thus, COA avoids local minima and identifies the global optimum solution for DG placement and sizing. This capability makes it well-suited for handling practical large-scale problems efficiently.
- Power distribution networks face several uncertainties, including load variations and fluctuations of outputs of renewable energy resources. The COA method is robust in managing these uncertainties effectively, resulting in more reliable and resilient DG sizing solutions.
- The COA can handle a range of constraints, including power balance, voltage limits and line capacity limits, to ensure that DG sizing solutions are feasible and practical for implementation. Additionally, the algorithm can be readily combined with other optimization techniques or incorporated into more comprehensive decision-making frameworks, further increasing its effectiveness in addressing the DG sizing problem.
- Compared to some traditional optimization techniques, COA converges to optimal solutions faster, thereby reducing the computational time and resources needed to solve the DG sizing problem in power systems.

In summary, considering all the points above, COA can be leveraged by utilities or policymakers to improve the network performance in areas with significant EV adoption.

4.2.5 Limitations and Scope for Further Work

The work in this paper is entirely focused on the detrimental effects of slow, residential level-1 AC charging of EVs on the network performance indices like distortion of daily demand profile, network active power losses and bus voltage magnitude deviation along with their mitigation using

wind-based DG resources. Level-2 AC charging and DC fast charging of EVs are not considered in this work. However, as the slow residential EV charging demand peaks after the evening hours, solar-PV based DGs are not considered in this work. This work also does not consider DG harmonics. Thus, the future directions of study should include the investigations on the impact of level-2 AC charging and DC fast charging of EVs on power distribution networks, along with power quality issues. The optimal allocation of solar-PV based DGs should also be considered as the EV charging demand with DC fast charging may peak during the morning hours or daytime. Also, the approach used in this paper can be further scaled to larger networks and varied geographical locations. Additionally, according to the no free lunch theorem, no algorithm can be considered as universally superior. Therefore, future research should also focus on developing new methodologies and more robust algorithms to effectively address the problems discussed in this paper.

5 Conclusions

This paper introduces a recent metaheuristic optimization technique, COA, for the optimal placement of wind-based DG units in a realistic 240-bus distribution network, to mitigate the adverse effects of slow, residential EV charging. A multi-objective problem is formulated to minimize the overall network active power losses and bus voltage magnitude deviations caused by EV charging. Compared to other well-known algorithms, COA delivers superior optimal solutions in terms of both accuracy and speed of convergence. Thus, the conclusions of this work can be summarized as:

- A generalized model is presented to estimate the dynamic, residential EV charging demand using data from the NHTS 2017 transportation survey. This model extracts parameters such as distance traveled, trip start and end times, type of EV and its count per house, etc. It incorporates the latest EV data, categorizing them into four segments from small city cars to SUVs. Based on this, the 24-h varying charging demand of around 1,500 EVs from 1,000 U.S. residents is calculated, and its impact on a realistic 240-bus distribution network based in Iowa, U.S.A., is analyzed.
- The increased demand on the distribution network during and after evening hours, primarily due to slow, residential EV charging deteriorates the network performance indices like distortion of daily load dynamics, increased network losses and bus voltage magnitude deviation. To counter these effects, wind-based DG resources are optimally allocated in the 240-bus network using COA technique. This achieved a reduction in the network overall active power losses by 38.45% and bus voltage magnitude deviation by 51.61%.
- The COA is efficient and scalable, making it ideal for large-scale optimization problems in real-world power systems. It balances global exploration with local exploitation, effectively managing uncertainties and network constraints like power balance, voltage and line capacity, resulting in practical and reliable DG sizing solutions for power system studies. Moreover, compared to other existing algorithms, COA is observed to have better accuracy and faster convergence.

References

1. Jin, W.; Li, C.; Zheng, M.Y.: Sustainable energy management in electric vehicle secure monitoring and blockchain machine learning model. Comput. Electr. Eng. **115**, 109093 (2024). https://doi.org/10.1016/j.compeleceng.2024.109093
2. Gholami, K.; Karimi, S.; Rastgou, A.; Nazari, A.; Moghaddam, V.: Voltage stability improvement of distribution networks using reactive power capability of electric vehicle charging stations. Comput. Electr. Eng. **116**, 109160 (2024). https://doi.org/10.1016/j.compeleceng.2024.109160
3. Ucer, E.; Koyuncu, I.; Kisacikoglu, M.C.; Yavuz, M.; Meintz, A.; Rames, C.: Modeling and analysis of a fast-charging station and evaluation of service quality for electric vehicles. IEEE Trans. Transp. Electrif. **5**(1), 215–225 (2019). https://doi.org/10.1109/TTE.2019.2897088
4. Islam, M.R.; Lu, H.; Hossain, J.; Islam, M.R.; Li, L.: Multiobjective optimization technique for mitigating unbalance and improving voltage considering higher penetration of electric vehicles and distributed generation. IEEE Syst. J. **14**(3), 3676–3686 (2020). https://doi.org/10.1109/JSYST.2020.2967752
5. Sortomme, E.; Hindi, M.M.; MacPherson, S.D.J.; Venkata, S.S.: Coordinated charging of plug-in hybrid electric vehicles to minimize distribution system losses. IEEE Trans. Smart Grid **2**(1), 198–205 (2011). https://doi.org/10.1109/TSG.2010.2090913
6. Wang, S.; Dong, Z.Y.; Luo, F.; Meng, K.; Zhang, Y.: Stochastic collaborative planning of electric vehicle charging stations and power distribution system. IEEE Trans. Industr. Inf. **14**(1), 321–331 (2018). https://doi.org/10.1109/TII.2017.2662711
7. Hua, L.; Wang, J.; Zhou, C.: Adaptive electric vehicle charging coordination on distribution network. IEEE Trans. Smart Grid **5**(6), 2666–2675 (2014). https://doi.org/10.1109/TSG.2014.2336623
8. Rassaei, F.; Soh, W.-S.; Chua, K.-C.: Demand response for residential electric vehicles with random usage patterns in smart grids. IEEE Trans. Sustain. Energy **6**(4), 1367–1376 (2015). https://doi.org/10.1109/TSTE.2015.2438037
9. Karfopoulos, E.L.; Panourgias, K.A.; Hatziargyriou, N.D.: Distributed coordination of electric vehicles providing V2G regulation services. IEEE Trans. Power Syst. **31**(4), 2834–2846 (2016). https://doi.org/10.1109/TPWRS.2015.2472957
10. Zhang, W.; Zhang, D.; Mu, B.; Wang, L.Y.; Bao, Y.; Jiang, J.; Morais, H.: Decentralized electric vehicle charging strategies for reduced load variation and guaranteed charge completion in regional distribution grids. Energies **10**, 147 (2017). https://doi.org/10.3390/en10020147
11. Mohd Shariff, N.B.; Al Essa, M.; and Cipcigan, L.: Probabilistic analysis of electric vehicles charging load impact on residential distributions networks. In: 2016 IEEE International Energy Conference (ENERGYCON), Leuven, Belgium, 2016, pp 1-6. https://doi.org/10.1109/ENERGYCON.2016.7513943

12. Wu, X.; Hu, X.; Teng, Y.; Qian, S.; Cheng, R.: Optimal integration of a hybrid solar-battery power source into smart home nanogrid with plug-in electric vehicle. J. Power. Sources **363**, 277–283 (2017). https://doi.org/10.1016/j.jpowsour.2017.07.086
13. Akram, U.; Khalid, M.; Shafiq, S.: An improved optimal sizing methodology for future autonomous residential smart power systems. IEEE Access **6**, 5986–6000 (2018). https://doi.org/10.1109/ACCESS.2018.2792451
14. Li, M.; Niu, Y.; Hu, H.; and Wang, K.: Optimal siting and sizing of SOPs in distribution networks with DG and EV charging. In: 2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe), The Hague, Netherlands, 2020, pp 675-679. https://doi.org/10.1109/ISGT-Europe47291.2020.9248981
15. Blonsky, M.; Munankarmi, P.; and Balamurugan, S.P.: Incorporating residential smart electric vehicle charging in home energy management systems. In: 2021 IEEE Green Technologies Conference (GreenTech), Denver, CO, USA, 2021, pp 187-194. https://doi.org/10.1109/GreenTech48523.2021.00039
16. Singh, R.S.; Mier, G.; Bosma, T.; Eijgelaar, M.; Bloemhof, G.; and Sauba, G.: Assessment of EV charging strategies and their effect on residential grids using co-simulation. In: 2022 International Conference on Smart Energy Systems and Technologies (SEST), Eindhoven, Netherlands, 2022, pp 1–6. https://doi.org/10.1109/SEST53650.2022.9898417.
17. Liao, X.; Hua, Y.; Wang, S.; Han, D.; Bai, H.; and Wang, Y.: A medium and long term orderly charging load planning method for electric vehicles in residential areas. In: 2022 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia), Shanghai, China, 2022, pp 1867-1872. https://doi.org/10.1109/ICPSAsia55496.2022.9949656
18. Dumlu, F.; and Zehir, M.A.: Technical coordination of aggregated electric vehicle charging and residential loads at the medium voltage level. In: 2022 4th Global Power, Energy and Communication Conference (GPECOM), Nevsehir, Turkey, 2022, pp 546-551. https://doi.org/10.1109/GPECOM55404.2022.9815665
19. Putri, S.M.; Ashari, M.; and Suryoatmojo, H.: EV charging scheduling with genetic algorithm as intermittent PV mitigation in centralized residential charging stations. In: 2023 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, Indonesia, 2023, pp 286-291. https://doi.org/10.1109/ISITIA59021.2023.10221086
20. Xiang, X.; Huo, Y.; He, Y.; Chen, Y.; Xu, J.; and Geng, G.: Business scheme analysis of orderly electric vehicle charging in residential areas. In: 2023 IEEE 7th Conference on Energy Internet and Energy System Integration (EI2), Hangzhou, China, 2023, pp 3160–3165. https://doi.org/10.1109/EI259745.2023.10513039
21. Das, H.S.; Rahman, M.M.; Li, S.; Tan, C.W.: Electric vehicles standards, charging infrastructure, and impact on grid integration: a technological review. Renew. Sustain. Energy Rev. **120**, 109618 (2020). https://doi.org/10.1016/j.rser.2019.109618
22. Lepolesa, L.J.; Adetunji, K.E.; Ouahada, K.; Liu, Z.; Cheng, L.: Optimal EV charging strategy for distribution networks load balancing in a smart grid using dynamic charging price. IEEE Access **12**, 47421–47432 (2024). https://doi.org/10.1109/ACCESS.2024.3382124
23. Acharige, S.S.G.; Haque, M.E.; Arif, M.T.; Hosseinzadeh, N.; Hasan, K.N.; Oo, A.M.T.: Review of electric vehicle charging technologies, standards, architectures, and converter configurations. IEEE Access **11**, 41218–41255 (2023). https://doi.org/10.1109/ACCESS.2023.3267164
24. Koch, L. et al.: Accurate physics-based modeling of electric vehicle energy consumption in the SUMO traffic micro simulator. In: 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), Indianapolis, IN, USA, 2021, pp 1650-165. https://doi.org/10.1109/ITSC48978.2021.9564463
25. Jeon, S.U.; Park, J.-W.; Kang, B.-K.; Lee, H.-J.: Study on battery charging strategy of electric vehicles considering battery capacity. IEEE Access **9**, 89757–89767 (2021). https://doi.org/10.1109/ACCESS.2021.3090763
26. National Household Travel Survey [Online]. Available: http://nhts.ornl.gov
27. Bu, F.; Yuan, Y.; Wang, Z.; Dehghanpour, K.; and Kimber, A.: A time-series distribution test system based on real utility data. In: 2019 North American Power Symposium (NAPS), Wichita, KS, USA, 2019, pp 1-6. https://doi.org/10.1109/NAPS46351.2019.8999982
28. https://mesonet.agron.iastate.edu/request/download.phtml
29. Liu, Z.; Wen, F.; Ledwich, G.: Optimal siting and sizing of distributed generators in distribution systems considering uncertainties. IEEE Trans. Power Delivery **26**(4), 2541–2551 (2011). https://doi.org/10.1109/TPWRD.2011.2165972
30. Feijoo, A.E.; Cidras, J.; Dornelas, J.L.G.: Wind speed simulation in wind farms for steady-state security assessment of electrical power systems. IEEE Trans. Energy Convers. **14**(4), 1582–1588 (1999). https://doi.org/10.1109/60.815109
31. Dehghani, M.; Montazeri, Z.; Trojovská, E.; Trojovský, P.: Coati optimization algorithm: a new bio-inspired metaheuristic algorithm for solving optimization problems. Knowl. Based Syst. **259**, 110011 (2023)
32. Rao, R.S.; Ravindra, K.; Satish, K.; Narasimham, S.V.L.: Power loss minimization in distribution system using network reconfiguration in the presence of distributed generation. IEEE Trans. Power Syst. **28**(1), 317–325 (2013). https://doi.org/10.1109/TPWRS.2012.2197227
33. Teng, J.-H.: A direct approach for distribution system load flow solutions. IEEE Trans. Power Del. **18**(3), 882887 (2003). https://doi.org/10.1109/TPWRD.2003.813818

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

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3509692

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Mobile secured IoT sensors-driven network using efficient QoS management

MOHAMMAD SIRAJ[1] (SENIOR MEMBER IEEE), MAJID ALTAMIMI[2] (MEMBER IEEE), ZEESHAN ABBASI[3]

[12]Department of Electrical Engineering, College of Engineering, King Saud University, Riyadh 11543, Saudi Arabia; (email: siraj@ksu.edu.sa, mtamimi@ksu.edu.sa)

[3]University Polytechnic, Jamia Millia Islamia, New Delhi, India; (email: zaabbasi@jmi.ac.in)

Corresponding author: Mohammad Siraj (email: siraj@ksu.edu.sa)

The authors extend their appreciation to the Deputyship for Research and Innovation, 433 Ministry of Education, Saudi Arabia for funding this research (IFKSUOR3-459).

ABSTRACT Autonomous services use future networks with many consumer electronic devices for the growth of real-time systems. Internet of Things (IoT) networks collect data from unpredictable environments using wireless technologies, process and forward it to requested users. Next-generation networks pose significant research challenges in load balancing and mobility maintenance due to the uncertain behavior of devices. Most solutions increase data flow using centralized computing, but distributed channels and heterogeneous devices still cause communication and reliability issues. With the use of mobile agents, this research offers quality-based services and a cooperative protocol for unpredictable IoT networks. By examining device mobility patterns, it effectively manages massive amounts of data across heterogeneous connections and lowers communication faults in more diverse services. Moreover, the data routes are optimized by considering the energy and resource considerations. To increase the reliability of wireless networks, mobile devices are connected with intelligence and quality-aware lightweight processing strategies. Secondly, different performance metrics are incorporated to evaluate the robustness of security in next-generation IoT applications. The significance of the proposed protocol is evaluated and performance results revealed its noteworthy performance other than state-of-the-art schemes.

INDEX TERMS Internet of Things, mobile networks, reliability, sensors management, future technologies

I. INTRODUCTION

IoT is a cutting-edge technology that supports the observation and analysis of real-time applications to assist network users and communication devices [1, 2]. Sensors-enabled technologies and Quality of Service (QoS) are interconnected to provide a significant role in the development and implementation of next-generation services [3, 4]. The integration of IoT devices and next-generation technologies offered many real-time developments in consumer applications for healthcare, transportation, smart home, etc. They provide crucial services in unpredictable environments and facilitate communication services with timely responses [5, 6]. Moreover, edge computing provides local computing for the constraint sensors and improves the QoS performance in critical communication services [7, 8]. In this context, data forwarding and collaboration among consumer technologies describe the procedure to facilitate communication with external networks, such as cloud systems. However, achieving QoS-aware routing presents an innovative challenge in ensuring secure data management. With the incorporation of trusted and secured methods, critical operations require high-quality and timely delivery of collected data. As a result, many researchers today are concentrating on developing reliable, performance-oriented techniques for mobile networks with the support of data protection and security measurement [9, 10]. Moreover, due to unpredictable network environments, communication between real-time devices should be able to adopt the decision by exploring various realistic factors [11, 12]. In recent decades, most of the exising solutions are unable to handle the security of big data security under high-traffic distribution and provide ineffective services to consumer applications. Such inefficiency not only has an impact on data processing and transmission in real time but also exposes sensitive data to potential security breaches [13, 14]. Therefore, the incorporation of mobile networks with IoT devices has made demanded opportunities for the development of intelligent systems. These networks offer higher flexibility and coverage making them more appropriate for the growth of smart applications. The challenges, however, lie in ensuring an effective data management system with reliability and timely performance[15, 16]. Also, sensors with constrained

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3509692

processing, memory, and power resources require an effective and lightweight communication system to optimize the usage of resources [17, 18]. Thus, mobile networks improve the process of data collection in dynamic environments like disaster management, wildlife tracking, or smart cities by increasing coverage and flexibility. Mobility, however, also introduces difficulties in preserving network stability, efficient routing, and energy consumption. As a result, we require an autonomous, predictive communication system that can make effective decisions with a high degree of dependability [19, 20]. In the proposed work, sustainable communication is achieved by recording the mobility patterns and enhancing network survivability. In addition, the load balancing is examined in conjunction with QoS considerations to identify robust routes and reduce extra overheads for significant applications. Furthermore, a major security improvement is made to ensure the persistence of the network while the adaptation of high-level edge computing, results in securing data forwarding points and ultimately eliminating disruptive events. The proposed work makes the following primary contributions.

i. Distributed computing provides support for dynamic attributes and mobility patterns of communication devices. It offers congestion-free multi-path routing for prolonged network survivability.
ii. For a diverse and constrained network environment, QoS-driven decisions are combined to maintain routing trust while providing effective and intelligent resource management.
iii. It addresses various security concerns to attain privacy and authentication for real-time data.

The following subsections comprise the organization of this research work. Section II presents the related work. Section III gives the problem background of the issue. Section IV discusses the proposed protocol and system model. The simulation and a discussion of the experimental findings are covered in Section V. In Section VI, the conclusion is provided.

II. LITERATURE WORK

In smart cities, an ecosystem of interconnected devices, including mobile devices, actuators, sensors, and many intermediate devices [21-23]. Integrating consumer electronics into sensors-enabled networks provides the development of adaptive communication systems with the support of a variety of services, including smart home, industrial IoT, and healthcare monitoring [24, 25]. The next-generation network offers an integrated and more efficient communication infrastructure to support a variety of services in the digital age. Such a system improves communication services by integrating various functionalities of IoT devices [26-28]. By utilizing IoT technologies, there is a lot of potential for creating a more inventive, effective, and connected digital ecosystem. Sensor-equipped smart devices help consumer networks to collect data from different resources. However, due to the sensitive nature of the gathered information, security is also an essential factor for next-generation networks [29, 30]. In [31], authors proposed a novel protocol for WSN that utilizes sender-based responsive techniques for energy, mobility, and efficient routing. It tackles a variety of packet routing issues, particularly those related to node mobility, energy optimization, and energy balancing in communication. The proposed protocol predicts the optimal path for an effective transmission system and enhances the fundamental QoS metrics for each connection. It proposed an energy-efficient and secure routing protocol using node trust values and fuzzy rules. However, under highly dynamic environments where energy efficiency and mobility management require more advanced, context-aware mechanisms, it may face limitations in scalability and adaptability to diverse network conditions. A WSN model for forest fire detection was proposed in [32] using an effective clustering and routing technique. The proposed model is referred to as Energy Efficient Routing Protocol (EERP). It reduces cluster head idle listening, saving sensor node energy. By limiting the number of sensor nodes that can report an event to nearby nodes, EERP reduces redundant data. Data is sent from source nodes to the Base Station (BS) via multi-hop routes by EERP. The proposed model is compared with the MAC protocol to evaluate its performance under various scenarios. However, it causes uneven energy consumption between nodes, potentially due to early node failure in critical areas. Moreover, clustering may not adapt well to rapidly changing environmental conditions, affecting real-time fire detection reliability. The development of a novel energy-efficient two-stage routing protocol (EETSP) [33] aims to decrease sensor network energy consumption and extend network lifetime. It reduces network energy consumption by the cluster heads (CHs) and secondary cluster heads (SCHs) and greatly increases the amount of packets communicated to the BS. SCH serves as a backup for primary CH. There are two stages to the EETSP. In the first step, CH and SCH are chosen based on the input parameters and the second stage involves both intra-cluster and inter-cluster routing. Though, due to the two-stage process of the proposed protocol, route management incurs additional complexity and communication costs. In [34], authors proposed a secure routing model to perform optimal path selection and encryption. First, nodes or optimal paths are selected for secure transmission in optimal link-state multipath routing. An algorithm called Crossover Mutated Marriage in Honey Bee (CM-MH) is developed and proposed for optimal path selection destination and source. To ensure secure transmission, encryption is applied. The Better Blowfish Algorithm (IBFA) is proposed as a secure authentication method. Lastly, there is monitoring of the updates. Finally, the comparison tests are performed for the proposed approach and existing work. It overlooked the factor of scalability in large networks, which results in more computational overhead as the number of devices increases. Furthermore, although the CM-MH algorithm prioritizes optimal path selection, however, the proposed model may

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3509692

fail to account for the dynamic nature of real-time traffic, and leads to a sub-optimal routing under different network scenarios. Relying solely on parameters like response time, distance, and queue length may not be sufficient to address security issues or dynamic traffic patterns, potentially resulting in imbalances or congestion under changing network conditions. In mobiles sensors, authors[35] examined the effects of version number attacks in RPL networks. To identify the malicious nodes carrying out version number attacks, they proposed a solution that utilizes the Q-Learning strategy. While significantly reducing the overhead on the nodes of low power and lossy networks, the proposed method accurately identifies malicious network nodes. The main drawback of the proposed solution include higher overhead in data processing that could impair scalability on wireless devices and decrease the network adaption in the context of IoT unpredictability. Authors [36] introduced a new DAG-Blockchain architecture for MANET-IoT security. The network is secured by the Multi-Factor PUF identification technique. All allowed nodes are divided into clusters according to the network topology. In addition, based on trust, Dijikstra's method was introduced to secure data transfer with multiple criteria. A Bi-Directional GRU for deep packet analysis was also presented. Intrusions are found through deep packet inspection, and later blocking mechanism stops them. Using a combination of algorithms, the proposed method increased packet delivery ratio, productivity, time analysis, detection accuracy, and security level. However, the architecture's reliance on consensus mechanism the r s may also introduce latency, which would impair eal-time performance required in dynamic MANET environments. For MANET-IoT networks, the authors proposed a gateway selection mechanism [37]. The proposed mechanism provides load balancing within the network as well as between gateways by taking into account the gateways' response time, distance, and queue length (QL). The proposed method outperforms current schemes in terms of load balancing, delay, and packet delivery ratio (PDR) according to simulation results. In the dynamic MANET-IoT scenarios, the selection of adaptive gateways may impose additional overhead and latency in the network. In addition, the unpredictable environment of MANET lead to sub-optimal selection of gateways.

III. PROPOSED METHODOLOGY

This section includes a network model and a detailed analysis of the proposed methodology.

A. NETWORK MODEL

Our proposed system incorporates geographical sensors to sense and process the environmental data. Consumer devices such as gateways are deployed to perform a vital role between sensors and edges. We consider the sensors as mobile with predefined transmission power. Sensors have limited transmission power and can communicate with neighbors either using single or multi-hop routing. Moreover, they can perform limited computing and are not allowed to communicate with edges directly. Their positions are frequently altered with time, so they are called mobile nodes. Nodes are uniformly distributed, thus each node has an approximate number of neighbors in its radius. Network operations are performed in uncertain and unreliable environments, thus leading to security issues for privacy and data authentication. The following are assumed network considerations in the development of a proposed protocol.

i. Nodes have a restricted communication range and can only communicate with nearby nodes.
ii. The deployment area has a uniform distribution of nodes.
iii. Nodes are mobile and their initial positions are known.
iv. Malicious entities may attempt to intercept the network communication.
v. Edges and sinks are enough resource devices with significant computing and processing capabilities.

B. PROPOSED PROTOCOL

In the proposed protocol, the mobile sensors initially distribute their local information and create routing tables. The routing tables are dynamically updated using the latest information and network conditions whenever any event is triggered in the field. The extracted information makes it easy to adapt reliable and more efficient paths from the sensing field toward edge devices. Each node has to maintain unique identification ID, hop count HC, and velocity VC and timestamp TS factors in its local table and share it with neighbors in the proximity, as defined in Equation 1.

$$N_i(TS) = [ID, HC, VC] \quad (1)$$

ID is a unique identifier that helps in identifying the data of sensors and involvement in any routing path. HC denotes the distance in terms of hop count from the source node towards the edge and should be the least for optimal energy consumption. Let us consider that $S_i, S_{i+1}, S_{i+2}, \ldots, S_n$ are series of neighbors from N_i to edge ED_i, then HC can be defined in Equation 2.

$$HC(N_i \rightarrow ED_i) = HC(S_i) + HC(S_i, S_{i+1}) + \ldots\ldots + HC(S_n, ED_i) \quad (2)$$

VC determines the mobility speed of the sensor in its proximity and the node whose velocity rate is not highest in a particular time interval, selected for the optimal decision. We consider the distance parameter to compute the velocity of nodes as defined in Equation 3. Let us consider mobile sensor-covered distance d_j at time interval t_j from distance d_i at time interval t_i.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3509692

$$VC_i = (d_j - d_i)/(t_j - t_i) \quad (3)$$

The proposed protocol determines the distance level D of the computed node velocity within the preset distance threshold DT. If the VC_i value is higher than a threshold, then its entry is recorded separate list indicates not suitable for considering the node for the routing scheme, as defined in Equation 4.

$$\begin{cases} D(VC_i) \leq DT, & \text{True} \\ D(VC_i) \geq DT, & False \end{cases} \quad (4)$$

The proposed protocol explores the aggregation function Z with composite computing based on the weights (w_1, w_2) and criterion parameters (HC, VC) using Equation 5. Afterward, the proposed protocol chooses the mobile node for routing the data packets whose aggregated function is minimum.

$$min\ (Z) = w_1 \text{ x } HC + w_2 \text{ x } VC \quad (5)$$

where the weighted parameters w_1 and w_2 uniformly contribute to the hop count and velocity factors, so that the sum of these values equals 1. Edge nodes store the values of the aggregated functions in their table and keep track of the energy resources of the selected mobile node. In case, the energy level EL falls than a certain threshold it sends beacon messages in its proximity to recompute the Z value for the competitive nodes. The next phase of the proposed protocol is to provide the security analysis for authentication, privacy, and data availability. Edge nodes group the mobile nodes in various groups based on proximity computing. Those nodes that are closest to a certain distance limit are arranged into a particular group. We assume that G_i denotes the set of groups that are comprised of various mobile sensors as denoted by Equation 6.

$$G_i = \sum_{i=0}^{k} S_n, \quad (6)$$

where S_n is mobile nodes fall in a particular distance limit. Later, the edge node generates particular keys K_i and share among each group G_i as defined in Equation 7.

$$K_i = [K_1, K_2, \ldots\ldots, K_n] \quad (7)$$

Moreover, each key K_i is encrypted using a public key of the edge node E_{ed} and stamp time T' is incorporated as given in Equation 8.

$$ED \rightarrow G_i : E_{ed}\ (K_i) + T' \quad (8)$$

Whenever any mobile node needs to transfer the network data, it needs to authenticate using the grouped key and accordingly, it is allowed to be a part of the routing scheme. Furthermore, group keys are updated by the proposed protocol based on fixed time intervals. Such methods reduce the probability of data compromise and increase the trust level among the nodes. The new keys K_i' are generated by exploring the random number generator (RNG) to produce a unique random value R_i , as given in Equation 9.

$$K_i' = R_i + K_i \quad (9)$$

Figure 1 depicts the three main stages of the proposed protocol in a consumer-based networked system. In the beginning, consumer devices and sensors interact with each other based on proximity and record the routing information in the form of tables. The tables are refined by exploring the new information and network conditions. In the middle layer, edges and gateways are explored to perform the role of central hubs between consumer networks and cloud systems. The various criterion parameters are identified to compute the aggregated function and the least weighted value node is selected among sensors as a data forwarder of consumer applications. Moreover, threshold-based analysis is performed to identify another round of data forwarder selection. Later, group-based keys are generated and distributed to individual groups. Keys are updated using random numbers which reduces the chances of data leakage and increases the degree of authentication. In the end, complex data security and availability are attained between central hubs and cloud systems.

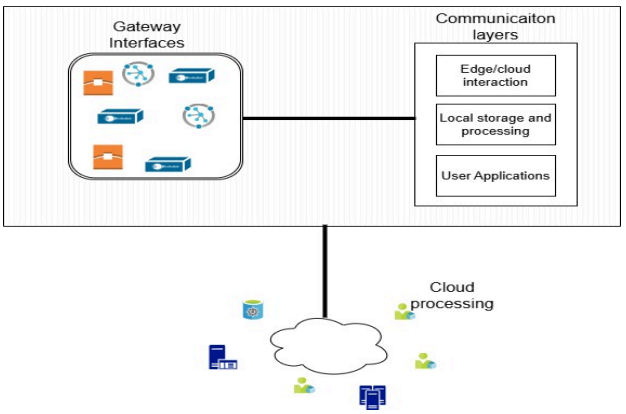


Figure 1. **Stages of proposed mobility aware secured QoS aware routing for mobile IoT applications**

In the proposed protocol, several key strategies improve security mechanisms and address previous limitations over most of the existing work such as key compromise, replay attack, man in the middle (MitM) attack and data temparing. By using group-based key distribution and frequent key updates, the approach mitigates key compromise risk. Each group has a unique key, and periodic updates ensure that even if a group is compromised, its lifespan is short, minimizing potential damage. Moreover, by exploring secure key distribution, even if an attacker intercepts the communication, then frequent key updates with integrity checks make MitM and data eavesdropping attacks difficult. The integration of

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3509692

distributed edge processing mitigates Denial of Service (DoS) attacks by ensuring that other network segments can function even if one is overloaded. In addition, to the time-based methods, the proposed protocol also provides an effective mechanism to avoid the replay attack using the unique key pattern. Figure 2 illustrates the flowchart of the developed routing protocol in the proposed protocol. The routing strategies are based on the analysis of the multi-criterion which is comprised of distance in terms of hop count and velocity of mobile sensors. Both factors are aggregated in a composite manner and offer the optimal data forwarders. Based on the preset thresholds, the evaluation criteria are re-accessed and the network is flooded with the newly selected forwarders. Accordingly, routing tables are also updated and neighbors are notified of the latest identifications of data forwarders. Figure 3 shows the flowchart for the proposed security methods. Edge devices initiate to determine the nearest mobile sensors using distance parameters. Once they are found, then groups are created and each group is assigned a particular key. The keys are securely distributed between the groups and after the time interval, the keys are updated. This provides less chances of data breaches and offers a high degree of authentication for mobile sensors. Moreover, to attain privacy and availability, the proposed protocol computes security functions on the incoming data and key patterns. The pseudocodes of the proposed protocols are depicted in Algorithm 1 and Algorithm 2. The main procedures of the developed algorithms are mobile routes detection MOB_SRoutes ( ) and device authentication DEVs_Authentication ( ). In MOB_SRoutes ( ), by exploring multiple parameters aggregated function is executed and determines the weighted value. The computed value is dynamically updated based on the energy level threshold and readjusts the mobile routes. To enable adaptive route management, the computed weighted value is dynamically modified based on the nodes' energy level threshold. The uniform weighted values contribute to each factor in a way that balances load distribution across sensor nodes and communication links, optimizing network performance. Using DEVs_Authentication ( ), the proposed protocol is made suitable for resource-constrained environments, resulting in providing an authentication mechanism for node authenticity with minimal processing overhead. Moreover, edges confirm the incoming request for the specific group, and keys are safely generated to distribute throughout the groups. To further increase the unpredictability of intrusions and reduce the likelihood of data breaches, keys are updated at random. However, as the number of nodes increases, the proposed MOB_SRoutes ( ) procedure leads to high computational complexity for constraint devices, because the recomputing of weighted values imposes additional overhead for the management of routes. Moreover, as the network grows, the generation and sharing of cryptographic keys using the DEVs_Authentication ( ) procedure also may lead to increased processing time.

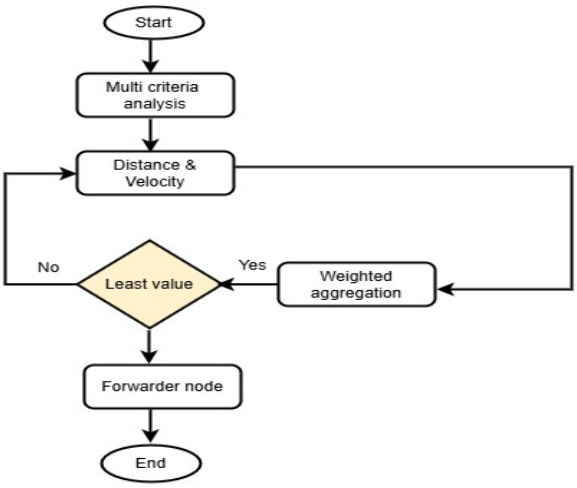


Figure 2. **Flowchart of the proposed routing in the sensors-driven IoT network**

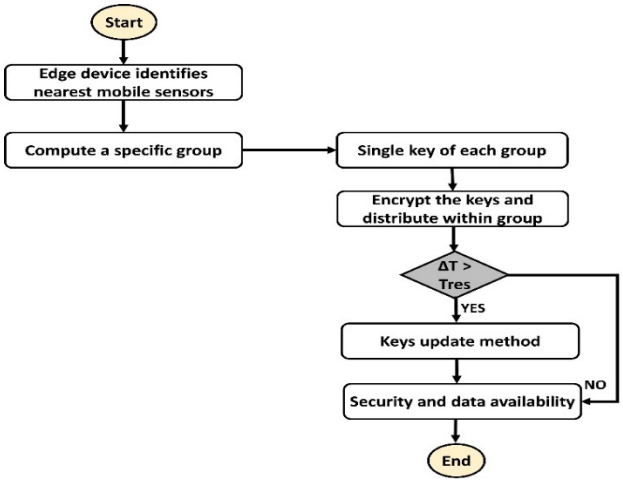


Figure 3. **Security against authentication and privacy for the proposed protocol.**

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3509692

Algorithm 1: Optimized mobility-aware efficient sensors data routing

```
Procedure MOB_SRoutes ( )
  collect local data
  update routing tables
  for (i=1; i<=N; i++)
    do
      multi-criterion analysis
      determine aggregated function
      min (Z) = w1 x HC + w2 x VC
      announced to neighbors
  end for
  compute the energy level threshold
  if forwarder.energy< threshold then
    send RREQ
    recall multi-criterion analysis
    update the entries
  end if
end procedure
```

Algorithm 2: Group-based device authentication and security for sensor data

```
Procedure DEVs_Authentication ( )
  establish a group of mobile sensors
  generate keys for each group
  securely distribute the set of keys  K_i
  if the timer is expired then
    regenerate the new keys K_i' using RGN
    ED → G_i: E_ed (K_i) + T'
  end if
  perform data security using XoR methods for peer nodes
end procedure
```

B. SECURITY METHODS WITH ANALYSIS

The proposed mobile IoT protocol for edge cloud environment provides a robust mechanism to attain data security in the following ways.

i. The use of separate keys for each group of mobile sensors ensures that data is encrypted and kept private among sensors. Unauthorized entities outside of the group cannot access the data because they do not possess the keys.

ii. In the proposed protocol, key-based authentication ensures that only authorized devices can send and receive group data. It minimizes the risk of unauthorized access caused by leaked or compromised keys by periodically updation policy.

iii. Security functions verify data integrity by exploring digital signature during transmission, ensuring it hasn't been tampered. Thus results in prevents data injection or attack by malicious entities.

iv. Distributed security functions at the edge reduce the chances of cloud service availability. Even during interruptions in cloud connectivity, it ensures that the system remains operational and responsive.

IV. RESULTS DISCUSSION

This section evaluates the proposed protocol against relevant studies. To evaluate the experimental results, we utilized network simulator NS-2.35. The scripting files are used to compute statistical analysis, and the simulations are recorded in trace files. We used mobile sensors along with edges that perform the role of gateways in the network environment. The mobile sensors are rotated in the surroundings with some mobility speed and gather the data. The size of the network field is set to 1000m X 1000m. Sensor nodes are varied from 100-400 with an initial energy of 2j. The number of sinks is set to 2. Each node is equipped with GPS and its transmission power is set to 3m. We ran 25 simulations to evaluate the average improvement of the proposed protocol with EERP and EETSP by exploring scenarios: varying iterations (1000s – 6000s) and varying numbers of nodes (100 - 400). In Table I, the simulation parameters are provided.

TABLE 1. **Network parameters**

Factor	Value
Mobile sensors	100-400
Number of packets	30-150
Malicious nodes	8
Transmission radius	10m
Initial energy	5j
Sink nodes	2
Mode of data flow	CBR
Bandwidth	2Mbps
w_1 and w_2	0.5, 0.5

The evaluation of network throughput for the proposed protocols with different numbers of iterations is depicted in Figure 4. The experimental results illustrate that the proposed protocol outperforms existing approaches by an average of 17% and 21%. The proposed protocol employs the concept of parallel route identification and attains reliable communication channels until sensor data reaches the destination. Moreover, it optimizes the use of network resources by re-computing parameters that balance the load on transmission links, ensuring efficient traffic distribution. Also, agents are mobile and crucial components of the proposed protocol, as they serve as intermediary devices in closest proximity to the low-powered devices. They intelligently collected the IoT data and decreased the transmission distance with effective data delivery performance. To reduce risks and secure the data transmission process, the proposed protocol has a robust security stage that examines communication channels and retains the desired throughput for critical network applications.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3509692

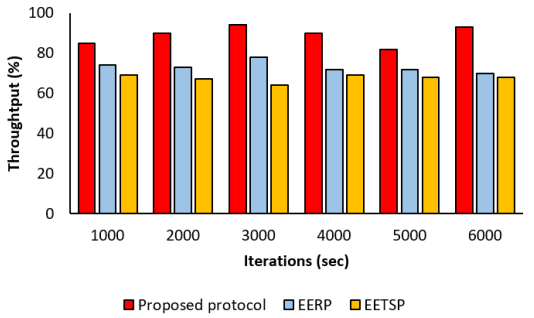


Figure 4. **Performance of network throughput with varying iterations (1000s – 6000s)**

The experimental outcomes of the proposed protocol in comparison to the existing solution are illustrated in Figure 5. The proposed protocol balances resource consumption and node congestion to optimize sensor performance through the use of distributed computing. It was noticed that the proposed protocol increases the network throughout by an average of 18% and 23.4% under different numbers of sensors, owing to the selection of several criteria-based parameters. The network edges are mobile that not only limits the transmission power of the constraint devices but also eliminates the interrupted links for transmissions of data. The dynamic mobility of network edges, which not only reduces the transmission power requirements of resource-constrained devices but also proactively removes unstable or interrupted links, is a significant advancement in the protocol. The proposed protocol significantly improves the performance and reliability of IoT networks under a variety of network topology changes, node mobility, and traffic load conditions by utilizing adaptive routing and load-balancing techniques.

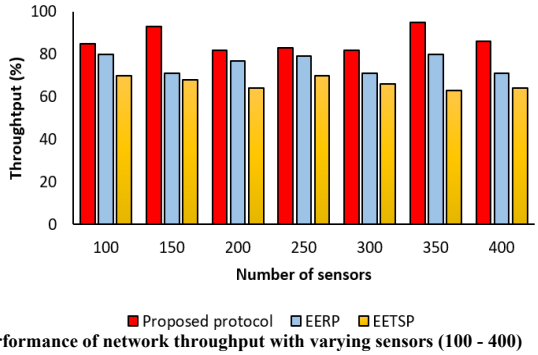


Figure 5. **Performance of network throughput with varying sensors (100 - 400)**

In comparison to the existing solutions, the proposed protocol reduces network latency for varying iterations by an average of 15.6% and 19.8%, as illustrated in Figure 6. This is because forwarding routes are utilized to assess node traffic and processing costs. Unlike most of the existing approaches, the proposed protocol temporarily marked a route as invalid when data forwarding is congested, and the source node is notified to find an alternative path. By identifying and mitigating faulty or malicious channels, mobile agents play a crucial role in enhancing network security and reliability. When a channel is compromised, mobile agents send alert messages to all nearby routes, preventing data transmission. The protocol also optimizes route selection for data transmission between constrained devices and mobile agents by evaluating multiple candidate paths and choosing the most efficient one. This optimized route selection significantly reduces communication overhead and uses the most efficient paths.

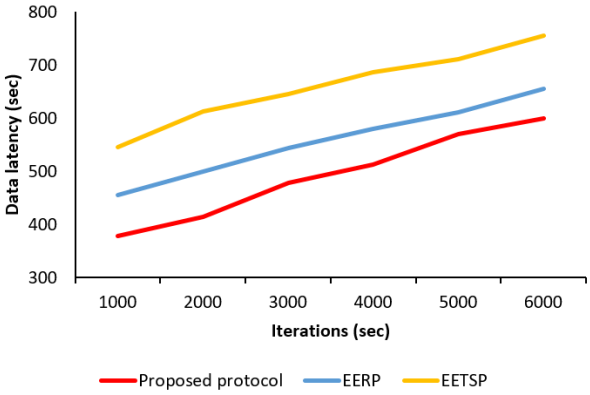


Figure 6. **Performance of data latency with varying iterations (1000s – 6000s)**

Under varying numbers of sensors, the performance of the proposed protocol in terms of data latency is shown in Figure 7 as compared to existing solutions. Based on the statistical results, it was seen that the proposed protocol improved the latency performance in terms of varying sensors by an average of 18% and 24.6% respectively. This results from the cooperative methods among sensors and mobile agents and storing the neighboring information by exploring the quality-aware assessment. Moreover, by generating route requests only in response to actual data forwarding events, extraneous routing overhead is reduced and latency is further decreased. Using distinct route IDs, the protocol can identify redundant links within already-existing routes, helping to simplify routing paths and reduce traffic. By moving around their proximity, mobile agents dynamically refine their neighbor tables, marking nodes as ineffective if their cost value is below a threshold, ensuring that only the best nodes are used for routing. In addition, with effective load balancing among the devices, the rotation of edges around the network boundary reduces energy holes while simultaneously enhancing route stability and decreasing data latency.

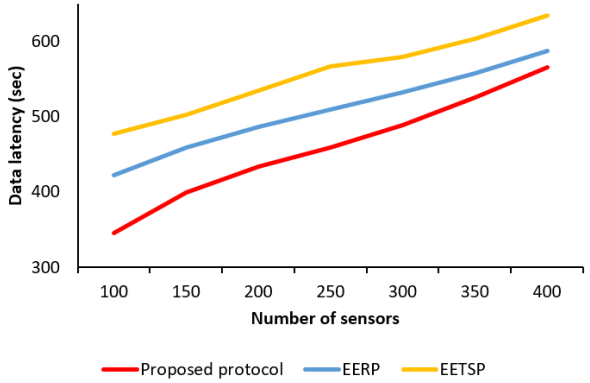


Figure 7. **Performance of data latency with varying sensors (100 - 400)**

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3509692

As shown in Figure 8, the performance of the proposed protocol and the existing solution is simulated for the node residual energy. Experimental results indicate that as the number of iterations increases, the proposed protocol enhances performance by an average of 20.5% to 27.3%. This improves the efficiency of communication bandwidth utilization by optimizing routing table information through the examination of energy, distance, and processing error parameters. The proposed protocol measures the mobile agent's distance from the source node before forwarding IoT data; if this distance is less than a predetermined threshold, data is forwarded in directly. If not, routing tables are examined to retrieve the most recent data and identify the suitable subsequent hop. If, after a certain amount of time, the designated route is still unavailable, its entry is eliminated from the routing table and the neighbors are notified. Furthermore, the verification of nodes is established because time-oriented secret keys and encrypted timestamps are used. In this way, malicious devices are excluded from routing call, which ultimately lengthens the stability period for the network.

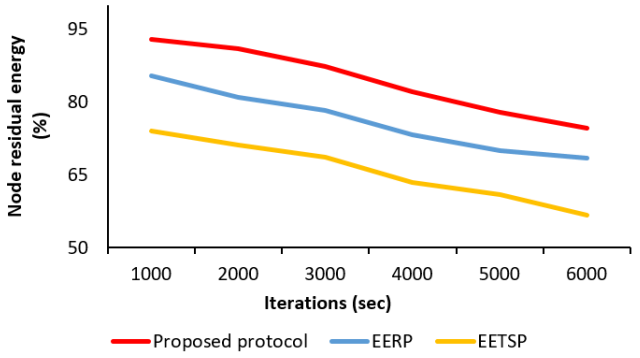


Figure 8. **Performance of network lifetime with varying iterations (1000s - 6000s)**

In Figure 9, the performance evaluation of the proposed protocol is compared with existing solutions in terms of node residual energy. It was improved for varying sensors by an average of 21.4% and 26.8% using the proposed protocol by exploring the adaptive approaches. Unlike the existing methods, the proposed protocol computes the routing criteria between the nodes using an intelligent method with balanced energy consumption across the network. Moreover, routing tables remove longer paths and only keep updated data. Unlike other systems, it sends data to wireless systems while continuously assessing device parameters and detecting the specifics of a running state. It also increases network stability by offering the most reliable neighbors for establishing node-to-node connections. Finally, mobile edges improves the nodes' consumption level of energy by collecting the sensor data with limited transmission power of the devices.

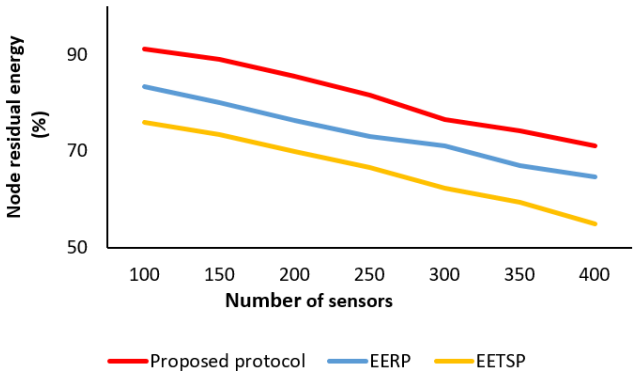


Figure 9. **Performance of network lifetime with varying sensors (100 - 400)**

Figure 10 shows the performance outcomes in terms of transmission load for the proposed protocol and existing solutions. It was observed that the transmission load increases with the deployment of malicious nodes, however, the findings indicate that the retransmission load of the proposed protocol is improved by an average of 9.4% and 12.6% as compared to earlier work for varying iterations. This is due to the involvement of mobile agents in gathering the sensed data intelligence from the low-powered sensors, and adopting the multi-hop model for data transmission. The routing decision is based on multiple parameters, and routing calls are announced in the proximity of source nodes rather than in the whole transmission area. In addition, the invalid routes are discarded from the routing tables, and up-to-date information is stored to improve the response time for end users while receiving the network data.

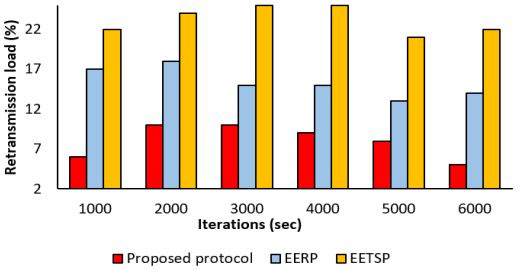


Figure 10. **Performance of retransmission load with varying iterations (1000s - 6000s)**

The performance evaluation of the retransmission load under varying sensors is demonstrated in Figure 11. The statistical analysis shows that the proposed protocol improved the transmission load of the proposed protocol by an average of 16.7% and 26.8% respectively. The proposed protocol reduces the cost of processing and forwarding data requests by using lightweight computing techniques, ensuring a cost-effective method of resource management. To improve adaptability to changing conditions, the protocol analyzes weighted edges and takes into account multiple factors in a dynamic environment when determining the next hop. The protocol efficiently distributes the load among the sensors during the routing phase, lessening the burden on individual sensors and enhancing the network's overall reliability. Additionally, the lightweight mechanism is supported and the more trustworthy neighbors are identified for data

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3509692

transmission by the security process, which employs centralized administration with the support of mobile edges. The mobile edge authenticated the incoming data from various sources and limited the overheads in case of data retransmissions.

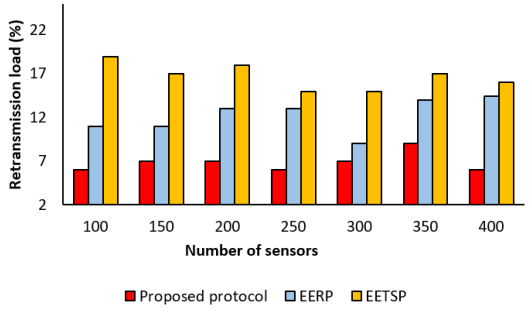


Figure 11. **Performance of retransmission load with varying sensors (100 - 400)**

V. CONCLUSION

Wireless networks provide seamless and secure connectivity throughout urban infrastructures, they are the driving system behind the integration of IoT devices for smart cities. To ensure protected and effective network operations, they are crucial for facilitating secure data exchange systems with the integrity of vital services. One of the main research challenges in these networks under unpredictable environments is reliable and efficient network controlling of the constraint system, despite the fact that many approaches have been developed to improve data collection and transmission operations of IoT networks. The formation of a resilient communication system critically depends on the security of data transmission over global channels. This paper presents a protocol that investigates the use of mobile sensors and intelligent computing to improve the routing between IoT-based applications, offer networked scalability, and implement security for next-generation systems. The edges implement distributed computation for communication devices and guarantee the security of the decentralized environment. It was noted, however, that the proposed protocol is unable to learn the detection strategy against threats as the number of malicious devices and false messages increases. Future work could focus on exploring machine learning algorithms to enhance QoS in mobile IoT networks by predicting patterns of network traffic and dynamic allocation of network resources. Moreover, enhanced network resilience against emerging threats can be achieved by the integration of automated vulnerability assessments with lightweight computing and regular routing updates.

REFERENCES

1. Malik, P.K., R. Sharma, R. Singh, A. Gehlot, S.C. Satapathy, W.S. Alnumay, D. Pelusi, U. Ghosh, and J. Nayak, *Industrial Internet of Things and its applications in industry 4.0: State of the art.* Computer Communications, 2021. **166**: p. 125-139.
2. Shafique, K., B.A. Khawaja, F. Sabir, S. Qazi, and M. Mustaqim, *Internet of things (IoT) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5G-IoT scenarios.* Ieee Access, 2020. **8**: p. 23022-23040.
3. Wen, M., Q. Li, K.J. Kim, D. López-Pérez, O.A. Dobre, H.V. Poor, P. Popovski, and T.A. Tsiftsis, *Private 5G networks: Concepts, architectures, and research landscape.* IEEE Journal of Selected Topics in Signal Processing, 2021. **16**(1): p. 7-25.
4. Aski, V.J., V.S. Dhaka, A. Parashar, and I. Rida, *Internet of Things in healthcare: A survey on protocol standards, enabling technologies, WBAN architectures and open issues.* Physical Communication, 2023: p. 102103.
5. Mishra, A., A.V. Jha, B. Appasani, A.K. Ray, D.K. Gupta, and A.N. Ghazali, *Emerging technologies and design aspects of next generation cyber physical system with a smart city application perspective.* International Journal of System Assurance Engineering and Management, 2023. **14**(Suppl 3): p. 699-721.
6. Esenogho, E., K. Djouani, and A.M. Kurien, *Integrating artificial intelligence Internet of Things and 5G for next-generation smartgrid: A survey of trends challenges and prospect.* IEEE Access, 2022. **10**: p. 4794-4831.
7. Aldossari, M.Q. and A. Sidorova, *Consumer acceptance of Internet of Things (IoT): Smart home context.* Journal of Computer Information Systems, 2020. **60**(6): p. 507-517.
8. Ram, S.K., B.B. Das, K. Mahapatra, S.P. Mohanty, and U. Choppali, *Energy perspectives in IoT driven smart villages and smart cities.* IEEE Consumer Electronics Magazine, 2020. **10**(3): p. 19-28.
9. Alfandi, O., S. Khanji, L. Ahmad, and A. Khattak, *A survey on boosting IoT security and privacy through blockchain: Exploration, requirements, and open issues.* Cluster Computing, 2021. **24**: p. 37-55.
10. Latif, S., Z. Idrees, J. Ahmad, L. Zheng, and Z. Zou, *A blockchain-based architecture for secure and trustworthy operations in the industrial Internet of Things.* Journal of Industrial Information Integration, 2021. **21**: p. 100190.
11. Iftikhar, S., S.S. Gill, C. Song, M. Xu, M.S. Aslanpour, A.N. Toosi, J. Du, H. Wu, S. Ghosh, and D. Chowdhury, *AI-based fog and edge computing: A systematic review, taxonomy and future directions.* Internet of Things, 2023. **21**: p. 100674.
12. Walia, G.K., M. Kumar, and S.S. Gill, *AI-empowered fog/edge resource management for IoT applications: A comprehensive review, research challenges and future perspectives.* IEEE Communications Surveys & Tutorials, 2023.
13. Amanullah, M.A., R.A.A. Habeeb, F.H.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3509692

Nasaruddin, A. Gani, E. Ahmed, A.S.M. Nainar, N.M. Akim, and M. Imran, *Deep learning and big data technologies for IoT security.* Computer Communications, 2020. **151**: p. 495-517.

14. Deepa, N., Q.-V. Pham, D.C. Nguyen, S. Bhattacharya, B. Prabadevi, T.R. Gadekallu, P.K.R. Maddikunta, F. Fang, and P.N. Pathirana, *A survey on blockchain for big data: Approaches, opportunities, and future directions.* Future Generation Computer Systems, 2022. **131**: p. 209-226.
15. Khan, M.M.I. and G. Nencioni, *Resource Allocation in Networking and Computing Systems: A Security and Dependability Perspective.* IEEE Access, 2023.
16. Khan, N., R. bin Salleh, Z. Khan, A. Koubaa, M. Hamdan, and A.M. Abdelmoniem, *Ensuring reliable network operations and maintenance: The role of PMRF for switch maintenance and upgrades in SDN.* Journal of King Saud University-Computer and Information Sciences, 2023. **35**(10): p. 101809.
17. Kumar, S., A. Singh, A. Benslimane, P. Chithaluru, M.A. Albahar, R.S. Rathore, and R.M. Álvarez, *An optimized intelligent computational security model for interconnected blockchain-IoT system & cities.* Ad Hoc Networks, 2023. **151**: p. 103299.
18. El-Fouly, F.H. and R.A. Ramadan, *Real-time energy-efficient reliable traffic aware routing for industrial wireless sensor networks.* IEEE Access, 2020. **8**: p. 58130-58145.
19. Heidari, A., N. Jafari Navimipour, M. Unal, and G. Zhang, *Machine learning applications in internet-of-drones: systematic review, recent deployments, and open issues.* ACM Computing Surveys, 2023. **55**(12): p. 1-45.
20. Rehman, A., T. Saba, K. Haseeb, R. Singh, and G. Jeon, *Smart health analysis system using regression analysis with iterative hashing for IoT communication networks.* Computers and Electrical Engineering, 2022. **104**: p. 108456.
21. El-Garoui, L., S. Pierre, and S. Chamberland, *A new SDN-based routing protocol for improving delay in smart city environments.* Smart Cities, 2020. **3**(3): p. 1004-1021.
22. Ahad, M.A., S. Paiva, G. Tripathi, and N. Feroz, *Enabling technologies and sustainable smart cities.* Sustainable cities and society, 2020. **61**: p. 102301.
23. Meslin, A., N. Rodriguez, and M. Endler, *Scalable mobile sensing for smart cities: The MUSANet experience.* IEEE Internet of Things Journal, 2020. **7**(6): p. 5202-5209.
24. Fan, Y.-C., *Emerging Sensing Technologies in Consumer Electronics*. 2021, MDPI. p. 7689.
25. Ha, Y.-G., *Dynamic integration of zigbee home networks into home gateways using OSGI service registry.* IEEE Transactions on Consumer Electronics, 2009. **55**(2): p. 470-476.
26. Bansal, S. and D. Kumar, *IoT ecosystem: A survey on devices, gateways, operating systems, middleware and communication.* International Journal of Wireless Information Networks, 2020. **27**: p. 340-364.
27. Gupta, N., S. Sharma, P.K. Juneja, and U. Garg. *Sdnfv 5g-iot: A framework for the next generation 5g enabled iot*. in *2020 International Conference on Advances in Computing, Communication & Materials (ICACCM)*. 2020. IEEE.
28. Haseeb, K., M. Siraj, F.A. Alzahrani, and G. Jeon, *Sensor-based optimization multi-decision model for sustainable smart cities.* Sustainable Energy Technologies and Assessments, 2023. **60**: p. 103452.
29. Cook, J., S.U. Rehman, and M.A. Khan, *Security and privacy for low power iot devices on 5g and beyond networks: Challenges and future directions.* IEEE Access, 2023.
30. Rachakonda, L.P., M. Siddula, and V. Sathya, *A comprehensive study on IoT privacy and security challenges with focus on spectrum sharing in Next-Generation networks (5G/6G/beyond).* High-Confidence Computing, 2024: p. 100220.
31. Dhanalakshmi, B., L. SaiRamesh, and K. Selvakumar, *Intelligent energy-aware and secured QoS routing protocol with dynamic mobility estimation for wireless sensor networks.* Wireless Networks, 2021. **27**: p. 1503-1514.
32. Pedditi, R.B. and K. Debasis, *Energy Efficient Routing Protocol for an IoT-Based WSN System to Detect Forest Fires.* Applied Sciences, 2023. **13**(5): p. 3026.
33. Dwivedi, A.K., P.S. Mehra, O. Pal, M.N. Doja, and B. Alam, *EETSP: Energy‐efficient two‐stage routing protocol for wireless sensor network‐assisted Internet of Things.* International Journal of Communication Systems, 2021. **34**(17): p. e4965.
34. Alotaibi, M., *Improved blowfish algorithm-based secure routing technique in IoT-based WSN.* IEEE Access, 2021. **9**: p. 159187-159197.
35. Sharma, G., J. Grover, and A. Verma, *QSec-RPL: detection of version number attacks in RPL based mobile IoT using Q-learning.* Ad Hoc Networks, 2023. **142**: p. 103118.
36. Ilakkiya, N. and A. Rajaram, *A secured trusted routing using the structure of a novel directed acyclic graph-blockchain in mobile ad hoc network internet of things environment.* Multimedia Tools and Applications, 2024: p. 1-26.
37. Quy, V.K., N.T. Ban, D. Van Anh, N.M. Quy, and D.C. Nguyen, *An adaptive gateway selection mechanism for MANET-IoT applications in 5G networks.* IEEE Sensors Journal, 2023.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3509692

MOHAMMAD SIRAJ received his Ph.D. from Universiti Technologi Malaysia. He holds Master of Engineering degree in Computer Technology and Applications from Delhi College of Engineering, Delhi, India and Bachelor of Engineering degree in Electronics and Communication Engineering from Jamia Millia Islamia, New Delhi. Currently he is working as an Assistant Professor in Electrical Engineering, King Saud University. His research interests are Cognitive Wireless Networks, Wireless Mesh Networks, Sensor Networks, Internet of Things, Cloud Computing and Telecom Optical Networks.

He has numerous peered publications in well-known international journals and Conferences. He is a Senior Member IEEE and reviewer of many well-known International Journals and Conferences.

MAJID AL TAMIMI (Member, IEEE) received the B.Sc.Eng. degree (Hons.) in electrical engineering from King Saud University, Saudi Arabia, in 2004, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Canada, in 2010 and 2014, respectively. In 2015, he joined the Department of Electrical Engineering, King Saud University, as an Assistant Professor. He served as a TA with the Department of Electrical and Computer Engineering, Univer-

sity of Waterloo, in 2013, and the Department of Electrical Engineering, King Saud University, from 2004 to 2006. His current research interests include analyzing the energy cost for wireless handheld devices and cloud comput- ing architecture, integrating mobile computing with cloud computing, and studying and designing green ICT solutions.

ZEESHAN AHMAD ABBASI received his Ph.D. from Jamia Millia Islamia, Delhi.He did his master from University of Delhi and Bachelor degree from Faculty of Engineering Jamia Delhi in E&C Engineering.He has worked In Police Wireless and Police Research and Development in Ministry of Home Affairs, India.He taught many papers at Delhi College of Engineering,Hamdard University,IGNOU and Mewat Engineering College at UG and PG level. Presently he is working as Assistant Professor University Polytechnic JMI Delhi. His area of Interest is Wireless Communication, Wireless mesh networks and Internet of Things. He has numerous peered publications in national and international journals and conferences. He has served as reviewer and session chair of many reputed national and international conferences.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

Eur. Phys. J. D (2024)78:145
https://doi.org/10.1140/epjd/s10053-024-00934-3

THE EUROPEAN PHYSICAL JOURNAL D

Regular Article – Quantum Information

Modified six-state cryptographic protocol with entangled ancilla states

Rashi Jain[a] and Satyabrata Adhikari[b]

Department of Applied Mathematics, Delhi Technological University, Delhi 110042, India

Received 31 August 2024 / Accepted 8 November 2024
© The Author(s), under exclusive licence to EDP Sciences, SIF and Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract. In a realistic situation, it is very difficult to communicate securely between two distant parties without introducing any disturbances. These disturbances might occur either due to external noise or may be due to the interference of an eavesdropper. In this work, we consider and modify the six-state Quantum Key Distribution (QKD) protocol in which Eve can construct the unitary transformation that make all ancilla states entangled at the output, which is not considered in Bruss's work [29]. Using the above proposed modification, we would like to study the effect of entangled ancilla states on the mutual information between Alice and Eve. To achieve this task, we calculate the mutual information between Alice and Bob and Alice and Eve and identify the region where the secret key is generated even in the presence of Eve. We find that, in general, the mutual information of Alice and Eve depends not only on the disturbance D but also on the concurrence of the ancilla states. We show that the entanglement of the ancilla states helps in generating the secret key in the region where Bruss's six-state QKD protocol failed to do so. We have further shown that it is possible to derive the disturbance-free mutual information of Alice and Eve, if Eve manipulates her entangled ancilla state in a particular manner. Thus, in this way, we can show that a secret key can be generated between Alice and Bob even if the disturbance is large enough.

1 Introduction

One of the most important challenges of the information era is the ability to communicate secretly. QKD [1,2] is the most advanced application of quantum mechanics [3] and information theory [4,5], wherein a secret key is established between two trusted parties using which information can be transmitted secretly. The distribution of the secret key is the most challenging task in the field of quantum cryptography [6], as there might be an eavesdropper present between the sender and receiver, assumed to possess every possible operation acceptable by the laws of quantum mechanics [7]. The security of encrypted messages depends directly on the security of key distribution protocols. QKD uses the most powerful features and resources of quantum mechanics, such as quantum entanglement [8,9] and quantum teleportation [10–12]. QKD protocols motivated the development of quantum cryptographic protocol variants, including quantum secret sharing [13] and quantum-secured blockchain [14].

Bennett and Brassard [15] were the first to introduce a revolutionary idea of quantum key distribution protocol, which exploits the properties of the quantum mechanics [16]. They designed a four-state quantum cryptographic protocol for secure key distribution between two distant parties. In this scheme, the sender (Alice) transmits four single qubits in a random manner to the receiver (Bob). The qubit may be prepared either in the computational bases $B_1 = \{|0\rangle, |1\rangle\}$ or in the conjugate bases $B_2 = \{|\bar{0}\rangle, |\bar{1}\rangle\}$, [17] defined as

$$\begin{aligned} |\bar{0}\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ |\bar{1}\rangle &= \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \end{aligned} \tag{1}$$

Bob performs measurements on one of the two bases mentioned above at random. Then, in the next step, the bases used by both Alice and Bob are disclosed in the public channel where they compare the bit values for a randomly chosen sample of few bits from the string prepared by both sender and receiver. A secret key is said to be established using the cases in which Alice's and Bob's bases coincide. It is then important for Alice and Bob to know whether anyone in the middle may steal the information about the key. Alice and Bob can reveal the existence of the third party, i.e., the eavesdropper in the middle, by calculating the quantum bit error rate (QBER). QBER may be calculated

[a] e-mail: rashijain_23phdam07@dtu.ac.in
[b] e-mail: satyabrata@dtu.ac.in (corresponding author)

Published online: 04 December 2024

as the ratio of the error rate to the key rate. Therefore, if the QBER is too high, Alice and Bob decide to abort the protocol. Otherwise, using either one-way [18–20] or two-way [21,22] classical communication, they apply a classical post-processing protocol to distill a secret key.

The security of the standard BB84 scheme has been provided in [23], which is closer to a realistic experimental situation. Later, simple proof of security of BB84 was presented in [24–26]. A general framework of optimal eavesdropping on BB84 protocol was studied in [27], and they derived an upper bound on mutual information and described a specific type of interaction and the corresponding measurement that achieves the bound. Extending their work, in 2017, the uniqueness of optimal interaction up to rotation has been established in [28].

In 1998, D. Bruß [29] introduced a generalization of the BB84 scheme in the sense that the author uses three bases, i.e., six states in her QKD protocol instead of four states used in BB84 protocol. The third set of bases may be denoted as B_3 in which the states can be defined as

$$\begin{aligned} |\bar{\bar{0}}\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle) \\ |\bar{\bar{1}}\rangle &= \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle) \end{aligned} \tag{2}$$

In the Bruß scheme [29], Alice transmits one of these six states to Bob, with equal probability, and in this scheme, it is assumed that all six possible states occurred at Bob's site with the same disturbance. Using the optimal mutual information between Alice and Eve, it was shown that this scheme [29] is more secure against eavesdropping in comparison to BB84 scheme [15]. In 2008, an optimal six-state eavesdropping QKD scheme [32] was proposed wherein Alice deliberately adds equal noise to the signal states before transmitting it to Bob, and they found that the six-state protocol with mixed states is more robust than the six-state protocol with pure states. The proof of security of the six-state QKD protocol has been proposed by [33,34].

Moreover, we can also find that the security increases not only by increasing the number of non-orthogonal bases but also by increasing the level of the system. This can be easily verified from the study, where Bechmann and Peres [30] suggested the use of three-level systems rather than two-level systems for establishing a secure quantum key. Later, in 2002, Bruß and Macchiavello [31] introduced the optimal eavesdropping scheme with three-dimensional systems and showed that it would be more secure against symmetric attacks than two-dimensional systems. One can also note that, as the number of bases increases, the security also increases, but the efficiency of key formation decreases [31]. Further, in 1999, Gisin and Wolf [35] studied a protocol in which as long as Alice and Bob share some entanglement, they can use either quantum or classical key-agreement protocols to extract a secret key.

Despite all these studies related to QKD protocols, we observe that there is a gap in which we can think of a situation, where Eve can execute her unitary operation in such a way that she may extract the information as much as possible successfully from the intercepted qubit by allowing the small disturbance and thus omitting her presence in the protocol. Hence, this gap motivates us to study about Eve, a person in the middle, who believed to be stronger in the sense that she can use quantum resource efficiently.

The aim of this paper is two-fold. Firstly, we observe that the ancilla states included in the unitary transformation introduced by Bruss [29] are not all entangled. The question here arises that if all the ancilla states are entangled, then what effect will it have on the mutual information between Alice and Eve. Thus, our aim is to show that if Eve constructs the unitary transformation in such a way that it may transform her input ancilla state to the entangled ancilla states at the output. With the help of this construction, our purpose is to show that the mutual information of Alice and Eve may depend not only on the disturbance but also on the concurrence of the entangled ancilla states. Secondly, our aim is to show that the generation of the secret key also depends on the entanglement of ancilla states.

The rest of the paper is organized as follows: In Sect. 2, we will discuss in detail the unitary transformation constructed by Eve that may generate the entangled ancilla states at the output of the transformation. In Section 3, we derive the mutual information of Alice and Eve and show that it depends on the concurrence of the entangled ancilla states. Moreover, we also derive the mutual information of Alice and Bob, which solely depends on the disturbance D. In Section 4, we discuss the generation of the secret key in the following two cases: (i) when the mutual information of Alice and Eve depends on the disturbance D caused by Eve's interference and (ii) when it is independent of the same. Finally, we conclude in Section 5.

2 Construction of a Unitary transformation

In a six-state QKD protocol, the sender (Alice) has to send her qubit prepared in any one of the basis chosen from $\{B_1, B_2, B_3\}$ to the receiver (Bob). But before reaching to Bob, the eavesdropper (Eve) intercepts the qubit with the following motivations: (i) Eve would like to extract as much information as possible from the qubit sent by Alice so that she would know about the secret key shared between Alice and Bob. (ii) While extracting the information from the intercepting qubit, she also has to keep in mind that the disturbance will be the same irrespective of the basis $\{B_1, B_2, B_3\}$ chosen for the preparation of Alice's qubit. This ensures the fact that Bob may not discover the presence of an Eavesdropper.

Thus, to achieve her motivations, Eve has to apply a suitable unitary transformation on the intercepted qubit. Therefore, it is crucial to consider all the above points while constructing the unitary transformation.

The unitary transformation U constructed by Eve may be defined as

$$\begin{aligned} U|0\rangle|X\rangle &= \sqrt{F}|0\rangle|\phi_0\rangle + \sqrt{D}|1\rangle|\psi_0\rangle \\ U|1\rangle|X\rangle &= \sqrt{F}|1\rangle|\phi_1\rangle + \sqrt{D}|0\rangle|\psi_1\rangle \end{aligned} \tag{3}$$

where $|X\rangle$ denote the initial normalized ancilla state before interaction with the input qubit while $|\phi_i\rangle$ and $|\psi_i\rangle$ $(i = 0, 1)$ refers to normalized ancilla states after interaction. D denotes the disturbance at the output when the ancillary state of Eve interacted with the input qubit. The fidelity between the input and output qubit is described by F, and it is related with D as given below

$$F + D = 1 \tag{4}$$

2.1 Constraints to be satisfied by the ancilla states at the output

The constraints that can be satisfied by the ancilla states at the output might come from two assumptions: (a) the unitarity of the transformation and (b) the transformation should keep the disturbance D same for all input qubits that come from any basis under consideration.

(a) Unitarity of the transformation gives

$$\langle\phi_0|\psi_1\rangle + \langle\psi_0|\phi_1\rangle = 0 \tag{5}$$

(b) If the transformation is treating all the input states in the same way, i.e., keeping the disturbance same for all input qubits generated by sender from any basis, then the following conditions hold

$$\langle\psi_0|\psi_1\rangle = 0 \tag{6}$$

$$\mathrm{Re}\langle\phi_1|\phi_0\rangle = \frac{1-2D}{1-D} \tag{7}$$

$$\langle\phi_0|\psi_0\rangle + \overline{\langle\phi_1|\psi_1\rangle} = 0 \tag{8}$$

2.2 Structure of the two-qubit ancilla states

We are now in a position to determine the structure of the two-qubit ancilla states generated at the output when the initial ancilla states of Eve interact with the qubit sent by the sender, Alice. The specific form of the ancilla states may be obtained by using the constraints given in (6-8).

2.2.1 Specific form of $|\psi_0\rangle$ and $|\psi_1\rangle$

To start with, let us first choose the ancilla state $|\psi_0\rangle$ and $|\psi_1\rangle$ in such a way that (6) holds. To fulfill the requirement, we may choose

$$|\psi_0\rangle = \frac{|00\rangle - \tau_1|11\rangle}{\sqrt{1+\tau_1^2}} \tag{9}$$

and

$$|\psi_1\rangle = \frac{\tau_1|00\rangle + |11\rangle}{\sqrt{1+\tau_1^2}} \tag{10}$$

where $0 \le \tau_1 \le 1$.

Defining in this way (9-10), we can say that the ancilla states $|\psi_0\rangle$ and $|\psi_1\rangle$ represent non-maximally entangled states. They are product states when $\tau_1 = 0$, while maximally entangled states when $\tau_1 = 1$. To quantify the amount of entanglement in the states $|\psi_0\rangle$ and $|\psi_1\rangle$, we can make use of concurrence [37]. The concurrence C_{ψ_0} and C_{ψ_1} of the state $|\psi_0\rangle$ and $|\psi_1\rangle$, respectively, may be given as

$$C_{\psi_0} = C_{\psi_1} = \frac{2\tau_1}{1+\tau_1^2} \tag{11}$$

From (11), we can express τ_1^2 in terms of C_{ψ_0} as

$$\tau_1^2 = \frac{2 - C_{\psi_0}^2 - 2\sqrt{1-C_{\psi_0}^2}}{C_{\psi_0}^2} \tag{12}$$

The proof of (12) is given in Appendix A.

We should note here that when $\tau_1 = 0$, the transformation (3) reduces to the transformation prescribed by Bruss [29]. Here, we may also stress that when $0 < \tau_1 \le 1$, the transformation may shed new light on the distribution of the secret key between the sender and the receiver. We will show that the formation of a secret key is possible, although the modified transformation may increase the efficiency of Eve with respect to the stealing of information.

2.2.2 Specific form of $|\phi_0\rangle$ and $|\phi_1\rangle$

Let us now move on to the ancilla states $|\phi_0\rangle$ and $|\phi_1\rangle$, which can be expressed initially in the general form of two-qubit pure states in the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ as

$$\begin{aligned} |\phi_0\rangle &= \alpha_{\phi_0}|00\rangle + \beta_{\phi_0}|01\rangle + \gamma_{\phi_0}|10\rangle + \delta_{\phi_0}|11\rangle \\ \text{with } &|\alpha_{\phi_0}|^2 + |\beta_{\phi_0}|^2 + |\gamma_{\phi_0}|^2 + |\delta_{\phi_0}|^2 = 1 \end{aligned} \tag{13}$$

and

$$\begin{aligned} |\phi_1\rangle &= \alpha_{\phi_1}|00\rangle + \beta_{\phi_1}|01\rangle + \gamma_{\phi_1}|10\rangle + \delta_{\phi_1}|11\rangle \\ \text{with } &|\alpha_{\phi_1}|^2 + |\beta_{\phi_1}|^2 + |\gamma_{\phi_1}|^2 + |\delta_{\phi_1}|^2 = 1 \end{aligned} \tag{14}$$

Now, our task is to determine the parameters $\alpha_{\phi_i}, \beta_{\phi_i}, \gamma_{\phi_i}$ and δ_{ϕ_i} $(i = 0, 1)$, for which the constraints (5-8) are satisfied. Simplifying the constraints (5-8) for the states $|\phi_0\rangle$ and $|\phi_1\rangle$, we get (Appendix B)

$$\alpha_{\phi_0} = \alpha_{\phi_1} = 0 \tag{15}$$

$$\delta_{\phi_0} = \delta_{\phi_1} = 0 \tag{16}$$

Therefore, the ancilla states $|\phi_0\rangle$ and $|\phi_1\rangle$ given in (13) and (14) reduce to

$$|\chi_0\rangle = \beta_{\phi_0}|01\rangle + \gamma_{\phi_0}|10\rangle,\ |\beta_{\phi_0}|^2 + |\gamma_{\phi_0}|^2 = 1 \quad (17)$$
$$|\chi_1\rangle = \beta_{\phi_1}|01\rangle + \gamma_{\phi_1}|10\rangle,\ |\beta_{\phi_1}|^2 + |\gamma_{\phi_1}|^2 = 1 \quad (18)$$

It may be easily seen that the ancilla states $|\chi_0\rangle$ and $|\chi_1\rangle$ are entangled unless the parameters $|\beta_{\phi_0}|$ or $|\gamma_{\phi_0}|$ and $|\beta_{\phi_1}|$ or $|\gamma_{\phi_1}|$ vanishes. If they are entangled, then the amount of entanglement can be calculated by concurrence, and it is given by

$$\begin{aligned} C_{\chi_0} &= 2|\mathrm{Re}(\beta_{\phi_0}\gamma^*_{\phi_0})| \\ &= 2|b_1 g_1 + b_2 g_2| \end{aligned} \quad (19)$$

where $\beta_{\phi_0} = b_1 + ib_2$ and $\gamma_{\phi_0} = g_1 + ig_2$.
If we now choose the parameters b_1, b_2, g_1 and g_2 in such a way that $b_1 g_2 = b_2 g_1$, then the equation (19) reduces to

$$C^2_{\chi_0} = 4\Big(|\beta_{\phi_0}|^2(1 - |\beta_{\phi_0}|^2)\Big) \quad (20)$$

Solving (20), we can obtain the value of $|\beta_{\phi_0}|^2$ in terms of the concurrence C_{χ_0} as

$$|\beta_{\phi_0}|^2 = \frac{1 \pm \sqrt{1 - C^2_{\chi_0}}}{2}, 0 \le C_{\chi_0} \le 1 \quad (21)$$

One of the two values of $|\beta_{\phi_0}|^2$ can be considered from (21) and then the ranges of $|\beta_{\phi_0}|^2$ can be found out, which are discussed below:

(i) If

$$|\beta_{\phi_0}|^2 = \frac{1 - \sqrt{1 - C^2_{\chi_0}}}{2} \quad (22)$$

then we can find that $|\beta_{\phi_0}|^2$ is lying in $[0, \frac{1}{2}]$. We may also observe that $|\beta_{\phi_0}|^2$ is an increasing function of C_{χ_0}.

(ii) If

$$|\beta_{\phi_0}|^2 = \frac{1 + \sqrt{1 - C^2_{\chi_0}}}{2} \quad (23)$$

then $|\beta_{\phi_0}|^2 \in [\frac{1}{2}, 1]$. In this case, we find that $|\beta_{\phi_0}|^2$ is a decreasing function of C_{χ_0}.

In a similar fashion, the concurrence of the ancilla state $|\chi_1\rangle$ can be denoted by C_{χ_1}, which is given by

$$C^2_{\chi_1} = 4\Big(|\beta_{\phi_1}|^2(1 - |\beta_{\phi_1}|^2)\Big) \quad (24)$$

Therefore, $|\beta_{\phi_1}|^2$ can be expressed in terms of the concurrence C_{χ_1} as

$$|\beta_{\phi_1}|^2 = \frac{1 \pm \sqrt{1 - C^2_{\chi_1}}}{2}, 0 \le C_{\chi_1} \le 1 \quad (25)$$

Following the similar discussion above, we can have

$$\begin{aligned} &(i)\ \text{If}\ |\beta_{\phi_1}|^2 = \frac{1 - \sqrt{1 - C^2_{\chi_1}}}{2}\ \text{then}\ |\beta_{\phi_1}|^2 \in \left[0, \frac{1}{2}\right] \\ &(ii)\ \text{If}\ |\beta_{\phi_1}|^2 = \frac{1 + \sqrt{1 - C^2_{\chi_1}}}{2}\ \text{then}\ |\beta_{\phi_1}|^2 \in \left[\frac{1}{2}, 1\right] \end{aligned} \quad (26)$$

3 Mutual Information

Mutual information can be defined as the relative entropy between the joint distribution and the product of the marginal distributions. If $p(x, y)$ denote the joint probability mass function of the random variables X and Y and if $p(x)$ and $p(y)$ denote the marginal probability mass functions, then the mutual information may be defined mathematically as [4,38]

$$I^{XY} = \sum_x \sum_y p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \quad (27)$$

Mutual information I^{XY} may also be interpreted as how much information we can extract about the random variable X by performing a measurement on other random variable Y.

Therefore, our interest in this section would be to know how well Eve will use her entangled ancilla states to extract information from the intercepted qubit, which was sent by the sender, Alice.

3.1 Mutual information of Alice and Eve

Let Alice prepares a two-qubit maximally entangled state of the form

$$|\psi^{(B_1)}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \quad (28)$$

Here, we should note that it is not necessary that Alice prepare the qubit in only computational basis $B_1 = \{|0\rangle, |1\rangle\}$ but she may use the basis $B_2 = \{\frac{|0\rangle+|1\rangle}{\sqrt{2}}, \frac{|0\rangle-|1\rangle}{\sqrt{2}}\}$ or $B_3 = \{\frac{|0\rangle+i|1\rangle}{\sqrt{2}}, \frac{|0\rangle-i|1\rangle}{\sqrt{2}}\}$ also. Using the bases B_2 and B_3, the two-qubit maximally entangled state can be re-expressed as

$|\psi^{(B_2)}\rangle$

$$= \frac{1}{\sqrt{2}}\Big[(|0\rangle + |1\rangle)(|0\rangle + |1\rangle) + (|0\rangle - |1\rangle)(|0\rangle - |1\rangle)\Big] \tag{29}$$

$$|\psi^{(B_3)}\rangle = \frac{1}{\sqrt{2}}\Big[(|0\rangle + i|1\rangle)(|0\rangle + i|1\rangle) + (|0\rangle - i|1\rangle)(|0\rangle - i|1\rangle)\Big] \tag{30}$$

She then transmits one of the qubits to Bob, but in the midway, Eve intercepts the qubit, and the intercepted qubit undergoes the transformation (3). Eve does not know the fact that in which basis, Alice has prepared her qubits. But Eve's transformation will act on the qubit in the same way irrespective of the basis chosen by the sender, Alice, to prepare the state. This means that the disturbance created by Eve's transformation will be same irrespective of the chosen basis. When the intercepted qubit interacts with Eve's ancilla state, the four-qubit state at Eve's site is given by

$$|\chi\rangle_{AEE_1E_2} = \frac{1}{\sqrt{2}}\Big(\sqrt{F}|00\rangle_{AE}|\phi_0\rangle_{E_1E_2} + \sqrt{D}|01\rangle_{AE}|\psi_0\rangle_{E_1E_2} + \sqrt{F}|11\rangle_{AE}|\phi_1\rangle_{E_1E_2} + \sqrt{D}|10\rangle_{AE}|\psi_1\rangle_{E_1E_2}\Big) \tag{31}$$

where the first qubit possessed by Alice, the second qubit is in Eve's possession and the remaining two-qubits are from Eve's ancilla state.
Now, tracing out the two ancilla qubits E_1 and E_2, the reduced state shared between Alice and Eve is given by

$$\begin{aligned}\rho_{AE} = \frac{1}{2}[&\left(F|\beta_{\phi_0}|^2 + \eta\right)|00\rangle\langle 00| + \sqrt{F\eta}\left(\beta_{\phi_0} + \gamma^*_{\phi_1}\right)|00\rangle\langle 11| \\ &+ \left(F|\gamma_{\phi_0}|^2 + \eta\tau_1^2\right)|01\rangle\langle 01| + \sqrt{F\eta}\tau_1^2\left(\gamma_{\phi_0} - \beta^*_{\phi_1}\right)|01\rangle\langle 10| \\ &+ \sqrt{F\eta}\tau_1^2\left(\beta_{\phi_1} + \gamma^*_{\phi_1}\right)|10\rangle\langle 01| + \left(F|\beta_{\phi_1}|^2 + \eta\tau_1^2\right)|10\rangle\langle 10| \\ &+ \sqrt{F\eta}\left(\gamma_{\phi_1} + \beta^*_{\phi_0}\right)|11\rangle\langle 00| + \left(F|\gamma_{\phi_1}|^2 + \eta\right)|11\rangle\langle 11|]\end{aligned} \tag{32}$$

where $\eta = \frac{D}{1+\tau_1^2}$.
The mutual information of Alice and Eve is given by

$$\begin{aligned}I^{AE}(|\beta_{\phi_0}|^2, |\beta_{\phi_1}|^2, \tau_1, D) = {} & F(|\beta_{\phi_0}|^2 + |\beta_{\phi_1}|^2) + D - \\ & \left(F(|\beta_{\phi_0}|^2 + |\beta_{\phi_1}|^2) + D\right)\log_2\left(F(|\beta_{\phi_0}|^2 + |\beta_{\phi_1}|^2) + D\right) \\ & + \epsilon\left[F|\beta_{\phi_0}|^2 + \eta, F|\beta_{\phi_1}|^2 + \eta\tau_1^2\right]\end{aligned} \tag{33}$$

where the parameters $|\beta_{\phi_0}|^2$, $|\beta_{\phi_1}|^2$ are defined in (22), (23) and (26) and the function $\epsilon[a, b]$ may be defined as

$$\epsilon[a, b] = a\log a + b\log b - (a + b)\log(a + b) \tag{34}$$

Therefore, the mutual information between Alice and Eve may be re-expressed in terms of the concurrences C_{χ_0} and C_{χ_1} as

(i) $I_1^{AE}(C_{\chi_0}, C_{\chi_1}, \tau_1, D)$

$$\begin{aligned}&= F\left(\frac{1 - \sqrt{1 - C_{\chi_0}^2}}{2} + \frac{1 - \sqrt{1 - C_{\chi_1}^2}}{2}\right) + D + \epsilon\Bigg[F\left(\frac{1 - \sqrt{1 - C_{\chi_0}^2}}{2}\right) \\ &\quad + \eta, F\left(\frac{1 - \sqrt{1 - C_{\chi_1}^2}}{2}\right) + \eta\tau_1^2\Bigg],\end{aligned}$$

$$|\beta_{\phi_0}|^2 \in \left[0, \frac{1}{2}\right], |\beta_{\phi_1}|^2 \in \left[0, \frac{1}{2}\right] \tag{35}$$

(ii) $I_2^{AE}(C_{\chi_0}, C_{\chi_1}, \tau_1, D)$

$$\begin{aligned}&= F\left(\frac{1 - \sqrt{1 - C_{\chi_0}^2}}{2} + \frac{1 + \sqrt{1 - C_{\chi_1}^2}}{2}\right) + D + \epsilon\Bigg[F\left(\frac{1 - \sqrt{1 - C_{\chi_0}^2}}{2}\right) \\ &\quad + \eta, F\left(\frac{1 + \sqrt{1 - C_{\chi_1}^2}}{2}\right) + \eta\tau_1^2\Bigg],\end{aligned}$$

$$|\beta_{\phi_0}|^2 \in \left[0, \frac{1}{2}\right], |\beta_{\phi_1}|^2 \in \left[\frac{1}{2}, 1\right] \tag{36}$$

(iii) $I_3^{AE}(C_{\chi_0}, C_{\chi_1}, \tau_1, D)$

$$\begin{aligned}&= F\left(\frac{1 + \sqrt{1 - C_{\chi_0}^2}}{2} + \frac{1 - \sqrt{1 - C_{\chi_1}^2}}{2}\right) + D + \epsilon\Bigg[F\left(\frac{1 + \sqrt{1 - C_{\chi_0}^2}}{2}\right) \\ &\quad + \eta, F\left(\frac{1 - \sqrt{1 - C_{\chi_1}^2}}{2}\right) + \eta\tau_1^2\Bigg],\end{aligned}$$

$$|\beta_{\phi_0}|^2 \in \left[\frac{1}{2}, 1\right], |\beta_{\phi_1}|^2 \in \left[0, \frac{1}{2}\right] \tag{37}$$

(iv) $I_4^{AE}(C_{\chi_0}, C_{\chi_1}, \tau_1, D)$

$$\begin{aligned}&= F\left(\frac{1 + \sqrt{1 - C_{\chi_0}^2}}{2} + \frac{1 + \sqrt{1 + C_{\chi_1}^2}}{2}\right) + D + \epsilon\Bigg[F\left(\frac{1 + \sqrt{1 - C_{\chi_0}^2}}{2}\right) \\ &\quad + \eta, F\left(\frac{1 + \sqrt{1 - C_{\chi_1}^2}}{2}\right) + \eta\tau_1^2\Bigg]\end{aligned}$$

$$|\beta_{\phi_0}|^2 \in \left[\frac{1}{2}, 1\right], |\beta_{\phi_1}|^2 \in \left[\frac{1}{2}, 1\right] \tag{38}$$

3.2 Mutual information of Alice and Bob

Eve extracts the information from the intercepted qubit by using her ancilla states in such a way that it keeps the disturbance same for all the intercepted qubit irrespective of the bases in which Alice has prepared the qubit. After eavesdropping, Eve sent the intercepted qubit to Bob so that Alice and Bob does not reveal her presence in the intermediate place. After receiving the qubit, the shared state between Alice and Bob is given

by

$$\rho_{AB} = \begin{pmatrix} \frac{1-D}{2} & 0 & 0 & \frac{1-2D}{4} \\ 0 & \frac{D}{2} & 0 & 0 \\ 0 & 0 & \frac{D}{2} & 0 \\ \frac{1-2D}{4} & 0 & 0 & \frac{1-D}{2} \end{pmatrix} \tag{39}$$

Therefore, the mutual information of Alice and Bob is given by

$$I^{AB}(D) = 1 + D\log_2 D + (1-D)\log_2(1-D) \tag{40}$$

Therefore, using (4) and (41), the mutual information $I^{AE}(|\beta_{\phi_0}|^2, |\beta_{\phi_1}|^2, \tau_1, D)$ given in (33) reduces to

$$\begin{aligned} I^{AE}(|\beta_{\phi_0}|^2, \tau_1, D) = 1 &+ \left[(1-D)|\beta_{\phi_0}|^2 + \eta\right] \\ &\times \log_2\left[(1-D)|\beta_{\phi_0}|^2 + \eta\right] \\ &+ \left[(1-D)(1-|\beta_{\phi_0}|^2) + \eta\tau_1^2\right] \\ &\times \log_2\left[(1-D)(1-|\beta_{\phi_0}|^2) + \eta\tau_1^2\right] \end{aligned} \tag{42}$$

The analysis of the generation of secret key can be studied in the following two cases:

Case-I: When $|\beta_{\phi_0}|^2 \in [0, \frac{1}{2}]$.

In this case, the relationship between $|\beta_{\phi_0}|^2$ and the concurrence $C^2_{\chi_0}$ of the ancilla state $|\chi_0\rangle$ is given by using (12) and (22), the expression of $I^{AE}(|\beta_{\phi_0}|^2, \tau_1^2, D)$ given in (42) reduces to

$$\begin{aligned} I^{AE}(C_{\psi_0}, C_{\chi_0}, D) = 1 &+ \frac{1}{2}\left((1-\sqrt{1-C^2_{\chi_0}}) + \frac{D\left(C^2_{\psi_0} - (1-\sqrt{1-C^2_{\chi_0}})(1-\sqrt{1-C^2_{\psi_0}})\right)}{1-\sqrt{1-C^2_{\psi_0}}}\right) \\ &\log_2\left[\frac{1}{2}\left((1-\sqrt{1-C^2_{\chi_0}}) + \frac{D\left(C^2_{\psi_0} - (1-\sqrt{1-C^2_{\chi_0}})(1-\sqrt{1-C^2_{\psi_0}})\right)}{1-\sqrt{1-C^2_{\psi_0}}}\right)\right] \\ &+ \frac{1}{2}\left((1+\sqrt{1-C^2_{\chi_0}}) + \frac{D\left(2-C^2_{\psi_0} - 2\sqrt{1-C^2_{\psi_0}} - \left(1+\sqrt{1-C^2_{\chi_0}}\right)\left(1-\sqrt{1-C^2_{\psi_0}}\right)\right)}{1-\sqrt{1-C^2_{\psi_0}}}\right) \\ &\log_2\left[\frac{1}{2}\left((1+\sqrt{1-C^2_{\chi_0}}) + \frac{D\left(2-C^2_{\psi_0} - 2\sqrt{1-C^2_{\psi_0}} - \left(1+\sqrt{1-C^2_{\chi_0}}\right)\left(1-\sqrt{1-C^2_{\psi_0}}\right)\right)}{1-\sqrt{1-C^2_{\psi_0}}}\right)\right] \end{aligned} \tag{43}$$

4 Generation of Secret Key in six-state modified QKD protocol

In this section, we will provide a detailed analysis of when the secret key will be generated in the quantum key distribution protocol in spite of the presence of an eavesdropper with entangled ancilla states. It is known that the two parties can share a secret key if the mutual information of A and B is strictly greater than the mutual information of A and E [36]. To analyze it in detail, we divide the whole fact into two parts: In the first part, we will investigate the generation of the secret key when the mutual information of Alice and Eve depends on the disturbance D, and in the second part, we probe when mutual information of Alice and Eve is independent of D.

4.1 Relationship between $I^{AE}(C_{\psi_0}, C_{\chi_0}, D)$ and the concurrences C_{ψ_0} and C_{χ_0} of the ancilla states

Eve can enhance the amount of extracted information from the intercepted qubit if she chooses the parameter $|\beta_{\phi_0}|^2$ and $|\beta_{\phi_1}|^2$ in (33) in the following way:

$$|\beta_{\phi_0}|^2 + |\beta_{\phi_1}|^2 = 1 \tag{41}$$

It may be observed that if we vary the parameters C_{χ_0} and C_{ψ_0} in between zero and unity and if D is lying between zero and a value which is very close to 0.5, then we get

$$I^{AB}(D) - I^{AE}(C_{\psi_0}, C_{\chi_0}, D) > 0, \tag{44}$$

Thus, equation (44) represents the fact that, in this case, a secret key can be generated. But, there exists a critical value of D, which is lying in [0.4999999.0.5) and for the concurrence $C_{\psi_0} \in (0, 0.009]$, for which a secret key cannot be generated. This can be verified if we choose the values of the concurrences C_{χ_0} and C_{ψ_0} in such a way that $C_{\chi_0} = C_{\psi_0}$. In this particular case, $I^{AE}(C_{\psi_0}, C_{\chi_0}, D)$ given in (43) can be re-expressed as

$$\begin{aligned} I^{AE}(C_{\psi_0}, D) = 1 &+ \frac{1}{2}\left(1-(1-2D)\sqrt{1-C^2_{\psi_0}}\right) \\ &\log_2\left[\frac{1}{2}\left(1-(1-2D)\sqrt{1-C^2_{\psi_0}}\right)\right] \\ &+ \frac{1}{2}\left(1+(1-2D)\sqrt{1-C^2_{\psi_0}}\right) \\ &\log_2\left[\frac{1}{2}\left(1+(1-2D)\sqrt{1-C^2_{\psi_0}}\right)\right] \end{aligned} \tag{45}$$

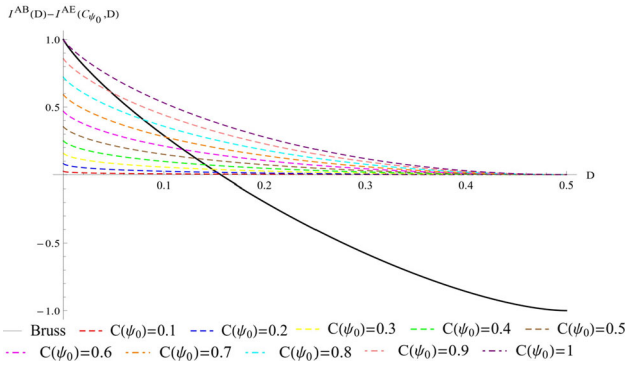


Fig. 1 Secret key generation for the six-state modified QKD protocol with each ancilla state as entangled states compared to the six-state protocol adopted by [29]. The D-axis represents the amount of disturbance $D \in (0, 0.5)$ in the system as a result of Eve's interference, and the $I^{AB}(D) - I^{AE}(C_{\psi_0}, D)$ axis represents the possibility for generation of the secret key between Alice and Bob in Eve's presence

Now, in the above expression (45), if we choose the values of the parameter C_{ψ_0} from 0.1 to 0.9 with a step length of 0.1 and varying D in the interval (0,0.5), then the plot of $I^{AB}(D) - I^{AE}(C_{\psi_0}, D)$ versus D may be shown in Fig. 1.

The following insights can be drawn from the graph (Fig.1) obtained above:

(i) As the concurrence C_{ψ_0} increases from 0 to 1, the value of $I^{AB}(D) - I^{AE}(C_{\psi_0}, D)$ increases as the disturbance D increases from 0 to 0.5.
(ii) After a certain value of disturbance D, the key is not generated for the model proposed by [29]. However, it can be easily seen that almost all the curves in our case lie in the upper half of the D-axis, implying that the secret key can be generated as $I^{AB}(D) - I^{AE}(C_{\psi_0}, D) > 0$ for all values of $C_{\psi_0} \in (0.009, 1]$ and for the values of D lying in $(0, 0.4999999)$. The secret key cannot be generated when $D \in [0.4999999, 0.5)$ and $C_{\psi_0} \in (0, 0.009]$.

Case-II: When $|\beta_{\phi_0}|^2 \in [\frac{1}{2}, 1]$.
Using (12) and (23), we find that the expression given in (42) will be independent of D, and it will be studied in detail in the next subsection.

4.2 Relationship between the concurrence C_{ψ_0} of the ancilla states and D independent mutual information

Here, we discuss the case when there is no effect of D on the mutual information of Alice and Eve. That is, mutual information remains the same for whatever disturbance induced by Eve on the intercepted qubit while extracting information from it. In this scenario, mutual information depends only on the entanglement of the ancilla states.

Recalling (42), let's re-express it in the following way

$$\begin{aligned} I^{AE}(|\beta_{\phi_0}|^2, \tau_1^2, D) = 1 + \left[(1-D)|\beta_{\phi_0}|^2 + \eta\right] \times \\ \log_2\left[(1-D)|\beta_{\phi_0}|^2 + \eta\right] \\ + \left[(1-D)|\beta_{\phi_1}|^2 + \eta\tau_1^2\right] \times \\ \log_2\left[(1-D)|\beta_{\phi_1}|^2 + \eta\tau_1^2\right] \end{aligned} \tag{46}$$

If we now choose the parameters $|\beta_{\phi_0}|^2$ and $|\beta_{\phi_1}|^2$ in such a way that

$$|\beta_{\phi_0}|^2 = \frac{1}{1+\tau_1^2}, |\beta_{\phi_1}|^2 = \frac{\tau_1^2}{1+\tau_1^2} \tag{47}$$

then (41) holds and the mutual information of Alice and Eve will now become independent of D, and it may be given as

$$\begin{aligned} I^{AE}(|\beta_{\phi_0}|^2, |\beta_{\phi_1}|^2) =& 1 + |\beta_{\phi_0}|^2 \log_2 |\beta_{\phi_0}|^2 \\ &+ |\beta_{\phi_1}|^2 \log_2 |\beta_{\phi_1}|^2 \end{aligned} \tag{48}$$

Now, our task is to express $I^{AE}(|\beta_{\phi_0}|^2, |\beta_{\phi_1}|^2)$ in terms of the concurrences C_{ψ_0} and C_{χ_0}. Since $|\beta_{\phi_0}|^2$ and $|\beta_{\phi_1}|^2$ can take two values in terms of concurrences, we need to analyze the relations given in (47).

If possible let $|\beta_{\phi_0}|^2 = \frac{1-\sqrt{1-C_{\chi_0}^2}}{2}$. Then (12) and (47) give

$$\begin{aligned} &\left(1-\sqrt{1-C_{\chi_0}^2}\right)\left(1-\sqrt{1-C_{\psi_0}^2}\right) = C_{\psi_0}^2 \\ &\implies \left(\sqrt{1-C_{\psi_0}^2}-1\right)\left(\sqrt{1-C_{\psi_0}^2}+\sqrt{1-C_{\chi_0}^2}\right) = 0 \end{aligned} \tag{49}$$

Since $0 < C_{\psi_0} \leq 1$ and $0 \leq C_{\chi_0} < 1$, (49) is not possible and we arrive at a contradiction. Thus, $|\beta_{\phi_0}|^2 \neq \frac{1-\sqrt{1-C_{\chi_0}^2}}{2}$, Hence, $|\beta_{\phi_0}|^2$ does not belong to $[0, 0.5)$.

It can be shown that if $|\beta_{\phi_0}|^2 = \frac{1+\sqrt{1-C_{\chi_0}^2}}{2}$ then

$$C_{\psi_0} = C_{\chi_0} \tag{50}$$

holds.
If $C_{\chi_0} = 1$, then $I^{AE}(\frac{1}{2}, \frac{1}{2}) = 0$. So $C_{\chi_0} \neq 1$. Therefore, alternatively, we can say that the equation (50) holds when $|\beta_{\phi_0}|^2 \in (0.5, 1]$.
Similarly, we can also prove that if $|\beta_{\phi_1}|^2 \in [0, 0.5)$ then

$$C_{\psi_0} = C_{\chi_1} \tag{51}$$

holds.
Moreover, one can easily verify that $|\beta_{\phi_1}|^2$ cannot belong to $(0.5, 1]$.

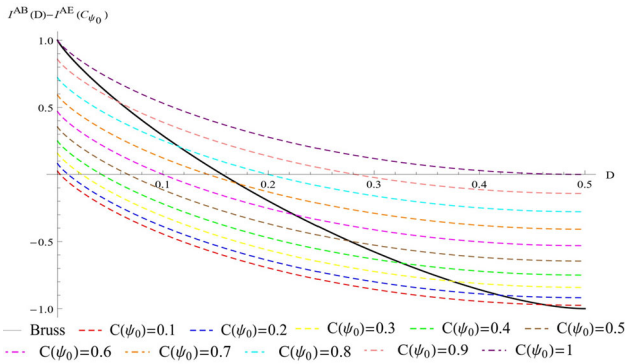


Fig. 2 The figure represents that in the modified six-state QKD protocol, the secret key is generated when the curves are lying above $D-$ axis in $(D, I^{AB}(D) - I^{AE}(C_{\psi_0}))$-plane and the secret key is not generated when the curves lying below the $D-$ axis. The different curves in the graph are drawn for different values of the concurrence of the ancilla states. The black curve represents when the secret key is generated or is not generated in Bruss's six-state QKD protocol

Therefore, if $|\beta_{\phi_0}|^2 \in (0.5, 1]$ and $|\beta_{\phi_1}|^2 \in [0, 0.5)$, then we have

$$C_{\psi_0} = C_{\chi_0} = C_{\chi_1} \tag{52}$$

Hence, using (48) and (52), the expression for $I^{AE}(|\beta_{\phi_0}|^2, |\beta_{\phi_1}|^2)$ can be re-expressed as

$$\begin{aligned} I^{AE}(C_{\psi_0}) =& 1 + \frac{1+\sqrt{1-C_{\psi_0}^2}}{2} \log_2 \frac{1+\sqrt{1-C_{\psi_0}^2}}{2} \\ &+ \frac{1-\sqrt{1-C_{\psi_0}^2}}{2} \log_2 \frac{1-\sqrt{1-C_{\psi_0}^2}}{2} \end{aligned} \tag{53}$$

The following insights may be drawn from the graph (Fig.2) obtained above:

(i) As the value of concurrence C_{ψ_0} increases from 0 to 1, the value of $I^{AB}(D) - I^{AE}(C_{\psi_0})$ increases for increasing value of the disturbance D from 0 to 0.5.
(ii) Unlike the model proposed by [29], the secret key may not be generated after a certain value of disturbance D, depending upon $C_{\psi_0} \in (0, 1]$. However, one may find that the secret key may only be generated in the full range of D when the ancilla state $|\psi_0\rangle$ is maximally entangled state.

5 Conclusion

To summarize, we have studied a six-state QKD protocol by modifying the unitary transformation prescribed in Bruss's work [29]. The modified unitary operation might be able to generate entangled ancilla states at the output, and it might give extra power to the eavesdropper, Eve, to gain knowledge about the key that would be generated between two distant partners, Alice and Bob. The amount of entanglement in ancilla states is measured by the concurrence. We have shown that the mutual information of Alice and Eve not only depends on the disturbance produced while extracting the information but also on the concurrence of the ancilla state and we denote it by $I^{AE}(C_{\chi_0}, C_{\chi_1}, \tau_1, D)$. Moreover, we have identified the region where the inequality $I^{AB}(D) > I^{AE}(C_{\chi_0}, C_{\chi_1}, \tau_1, D)$ holds, i.e., we have identified the region where the secret key may be generated in spite of the presence of Eve. To analyze this study, we divided the modified protocol into two parts. Firstly, we investigate the case where mutual information of Alice and Eve depends upon both disturbance D and concurrence of the entangled ancilla states. We find that as concurrence increases, the value of $I^{AB}(D) - I^{AE}(C_{\psi_0}, C_{\chi_0}, D)$ will be positive and increases as the disturbance increases. Further, we noticed in the proposal proposed in [29] that the secret key is not generated after a certain value of disturbance D, i.e., when $D > 0.1565$, but in our case, the secret key is generated up to a critical value D_c of D, i.e. the secret key is generated when $0 < D_c < 0.4999999$; otherwise, the secret key is not generated, i.e., when $0.4999999 \leq D_c < 0.5$. In the second case, we find that there exists a relationship between the concurrences of the ancilla states for which the mutual information of Alice and Eve is independent of D. In this scenario, the value of $I^{AB}(D) - I^{AE}(C_{\psi_0})$ increases for increasing values of concurrence and the disturbance D, respectively. One may also find that in this case, the secret key is not generated by following Bruss's six-state QKD protocol when the disturbance $D > 0.1565$ but in our modified six-state QKD protocol, the secret key is generated even for $D > 0.1565$ but the concurrence of the ancilla state is adjusted between $(0.74, 1)$. The key is generated in the whole domain of D if the ancilla component $|\psi_0\rangle$ is a maximally entangled state. To conclude, we feel that it would be interesting to study the same when Alice prepares her state as a mixed state instead of a pure state.

Author contributions

RJ and SA contributed to conceptualization, methodology, software, writing—original draft preparation, visualization, and writing—reviewing and editing.

Data Availability Statement Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Appendix A: Derivation of the expression of τ_1^2

Recalling (11) and squaring both sides, we get

$$(1+\tau_1^2)^2 C_{\psi_0}^2 = 4\tau_1^2 \tag{A1}$$

Simplifying and expressing τ_1^2 in terms of the concurrence C_{ψ_0}, we get

$$\tau_1^2 = \frac{2 - C_{\psi_0}^2 \pm 2\sqrt{1-C_{\psi_0}^2}}{C_{\psi_0}^2} \tag{A2}$$

Now, we will prove that

$$\tau_1^2 = \frac{2 - C_{\psi_0}^2 - 2\sqrt{1-C_{\psi_0}^2}}{C_{\psi_0}^2} \tag{A3}$$

holds.

Proof If possible, let

$$\tau_1^2 = \frac{2 - C_{\psi_0}^2 + 2\sqrt{1-C_{\psi_0}^2}}{C_{\psi_0}^2} \tag{A4}$$

Since $0 < \tau_1^2 \leq 1$, we have

$$0 < \frac{2 - C_{\psi_0}^2 + 2\sqrt{1-C_{\psi_0}^2}}{C_{\psi_0}^2} \leq 1 \tag{A5}$$

We may observe that LHS inequality is always true, but simplifying the RHS inequality, we get

$$\sqrt{1-C_{\psi_0}^2}(\sqrt{1-C_{\psi_0}^2}+1) \leq 0 \tag{A6}$$

We arrive at a contradiction because $0 \leq C_{\psi_0} \leq 1$. Thus, we have

$$\tau_1^2 = \frac{2 - C_{\psi_0}^2 - 2\sqrt{1-C_{\psi_0}^2}}{C_{\psi_0}^2} \tag{A7}$$

Hence proved.

Appendix B: Detailed calculation to derive the specific form of $|\phi_0\rangle$ and $|\phi_1\rangle$

To start with, let us consider the general form of initially considered pure two-qubit ancilla state $|\phi_0\rangle$ and $|\phi_1\rangle$ in the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ as

$$|\phi_0\rangle = \alpha_{\phi_0}|00\rangle + \beta_{\phi_0}|01\rangle + \gamma_{\phi_0}|10\rangle + \delta_{\phi_0}|11\rangle,$$
$$\text{with } |\alpha_{\phi_0}|^2 + |\beta_{\phi_0}|^2 + |\gamma_{\phi_0}|^2 + |\delta_{\phi_0}|^2 = 1 \tag{B1}$$

and

$$|\phi_1\rangle = \alpha_{\phi_1}|00\rangle + \beta_{\phi_1}|01\rangle + \gamma_{\phi_1}|10\rangle + \delta_{\phi_1}|11\rangle$$
$$\text{with } |\alpha_{\phi_1}|^2 + |\beta_{\phi_1}|^2 + |\gamma_{\phi_1}|^2 + |\delta_{\phi_1}|^2 = 1 \tag{B2}$$

Further, we choose the states $|\psi_0\rangle$ and $|\psi_1\rangle$ in such a way that (6), i.e., $\langle\psi_0|\psi_1\rangle = 0$ is satisfied. Therefore, we can consider the following form of the states

$$|\psi_0\rangle = \frac{|00\rangle - \tau_1|11\rangle}{\sqrt{1+\tau_1^2}} \tag{B3}$$

and

$$|\psi_1\rangle = \frac{\tau_1|00\rangle + |11\rangle}{\sqrt{1+\tau_1^2}} \tag{B4}$$

where $0 < \tau_1 \leq 1$.

Let us now calculate the disturbance made by Eve if Alice prepares the state in the basis B_1:

$$\langle 0|\rho_B|0\rangle = 1 - D \tag{B5}$$
$$\langle 1|\rho_B|1\rangle = 1 - D \tag{B6}$$

where $\rho_B = \mathrm{Tr}_{AE_1E_2}(\rho_{ABE_1E_2})$ represents the state sent by Alice to Bob which is intercepted by Eve.

In the same way, if Alice prepares the state in the bases B_2 or B_3 and keeping in mind the restrictions that make the disturbance same for all the states in the three non-orthogonal bases $\{B_1, B_2, B_3\}$, then for $0 < \tau_1 \leq 1$, we have

$$(\alpha_{\phi_0} + \alpha_{\phi_1}^*)(1-\tau_1) + (\delta_{\phi_0} + \delta_{\phi_1}^*)(1+\tau_1) = 0 \tag{B7}$$
$$(\alpha_{\phi_0} + \alpha_{\phi_1}^*)(1+\tau_1) + (\delta_{\phi_0} + \delta_{\phi_1}^*)(1-\tau_1) = 0 \tag{B8}$$

Solving (B7-B8), we obtain

$$\delta_{\phi_0} + \delta_{\phi_1}^* = 0 \tag{B9}$$
$$\alpha_{\phi_0} + \alpha_{\phi_1}^* = 0 \tag{B10}$$

Recalling (5) and (8), and using (B1-B4), we get

$$\begin{aligned} \tau_1(\alpha_{\phi_0}^* - \delta_{\phi_1}) + \delta_{\phi_0}^* + \alpha_{\phi_1} = 0 \\ \tau_1(\alpha_{\phi_1} - \delta_{\phi_0}^*) + \delta_{\phi_1} + \alpha_{\phi_0^*} = 0 \end{aligned} \tag{B11}$$

To make (B11) independent of τ_1, we have

$$\alpha_{\phi_0}^* - \delta_{\phi_1} = 0 \tag{B12}$$
$$\alpha_{\phi_1} - \delta_{\phi_0}^* = 0 \tag{B13}$$

Using (B12-B13), equation (B11) reduces to

$$\delta_{\phi_0}^* + \alpha_{\phi_1} = 0 \tag{B14}$$

$$\delta_{\phi_1} + \alpha^*_{\phi_0} = 0 \tag{B15}$$

Solving (B12) & (B15) and (B13) & (B14), we get

$$\alpha_{\phi_0} = \delta_{\phi_0} = \alpha_{\phi_1} = \delta_{\phi_1} = 0 \tag{B16}$$

Using (B16), equation (B1) and (B2) reduces to

$$|\chi_0\rangle = \beta_{\phi_0}|01\rangle + \gamma_{\phi_0}|10\rangle,\ |\beta_{\phi_0}|^2 + |\gamma_{\phi_0}|^2 = 1$$
$$|\chi_1\rangle = \beta_{\phi_1}|01\rangle + \gamma_{\phi_1}|10\rangle,\ |\beta_{\phi_1}|^2 + |\gamma_{\phi_1}|^2 = 1$$

References

1. N. Gisin, G. Ribordy, W. Tittel, H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002)
2. C. Portmann, R. Renner, Rev. Mod. Phys. **94**, 025008 (2022)
3. L.E. Ballentine, *Quantum mechanics: a modern development* (World Scientific Publishing Company, 2014)
4. T.M. Cover, *Elements of information theory* (Wiley, 1999)
5. C. Arndt, *Information measures: information and its description in science and engineering* (Springer, 2001)
6. H. K. Lo, T. Spiller, & S. Popescu, Introduction to quantum computation and information. World Scientific, (1998)
7. D.J. Griffiths, *Introduction to quantum mechanics*, 1st edn. (Prentice Hall, New Jersey, 1995)
8. R. Horodecki, P. Horodecki, M. Horodecki, K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009)
9. S. Pirandola, U.L. Andersen, L. Banchi, M. Berta, Advances in quantum cryptography. Adv. Opt. Photonics **12**, 1012–1236 (2020)
10. C.H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W.K. Wootters, Phys. Rev. Lett. **70** (1895)
11. F. Verstraete, H. Verschelde, Phys. Rev. Lett. **90**, 097901 (2003)
12. G. Alber, T. Beth, M. Horodecki, P. Horodecki, R. Horodecki, M. Rötteler, H. Weinfurter, R. Werner, A. Zeilinger, *Quantum information: An introduction to basic theoretical concepts and experiments, 173* (Springer, 2003)
13. M. Hillery, V. Buzek, Berthiaume. Phys. Rev. A **59**, 1829–1834 (1999)
14. E.O. Kiktenko, N.O. Pozhar, M.N. Anufriev, A.S. Rushechkin, R.R. Yunusov, Y.V. Kurochkin, A.I. Lvovsky, A.K. Fedorov, Quantum Sci. Technol. **3**, 035004 (2018)
15. C. H. Bennett, G. Brassard, In: Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India (IEEE, New York, 1984), pp. 175-179
16. A. Das, *Lectures on quantum mechanics* (World Scientific Publishing Company, 2012)
17. M.A. Nielsen, I.L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, New Delhi, 2008)
18. D. Stucki, N. Brunner, N. Gisin, V. Scarani, H. Zbinden, Appl. Phys. Lett. **87**, 194108 (2005)
19. B. Kraus, N. Gisin, R. Renner, Phys. Rev. Lett. **95**, 080501 (2005)
20. R. Renner, N. Gisin, B. Kraus, Phys. Rev. A **72**, 012332 (2005)
21. D. Gottesman, H.K. Lo, IEEE Trans. Inform. Theory **49**, 457–475 (2003)
22. N.J. Beaudry, M. Lucamarini, S. Mancini, R. Renner, Phys. Rev. A **88**, 062302 (2013)
23. D. Mayers, J. ACM. **48**, 351–406 (2001)
24. H.K. Lo, H.F. Chau, Science **283**, 2050 (1999)
25. H.K. Lo, Quantum Inf. Comput. **1**, 2 (2001)
26. P.W. Shor, J. Preskill, Phys. Rev. Lett. **85**, 441 (2000)
27. C.A. Fuchs, N. Gisin, R.B. Griffiths, C.S. Niu, A. Peres, Phys. Rev. A **56**, 1163 (1997)
28. A. Acharyya, G. Paul, Phys. Rev. A **95**, 022326 (2017)
29. D. Bruß, Phys. Rev. Lett. **81**, 3018 (1998)
30. H. Bechmann-Pasquinucci, A. Peres, Phys. Rev. Lett. **85**, 3313 (2000)
31. D. Bruß, C. Macchiavello, Phys. Rev. Lett. **88**, 127901 (2002)
32. Z. Shadman, H. Kampermann, T. Meyer, D. Bruß, Int. J. Quantum Inf. **7**, 297–306 (2009)
33. H. Inamori, Security of EPR-based Quantum Key Distribution using three bases, preprint (2000), arXiV:0008076 [quant-ph]
34. V. Scarani, H. Bechmann-Pasquinucci, N.J. Cerf, M. Dušek, N. Lütkenhaus, M. Peev, Rev. Mod. Phys. **81**, 1301 (2009)
35. N. Gisin, S. Wolf, Phys. Rev. Lett. **83**, 4200 (1999)
36. I. Csiszar, J. Korner, IEEE Trans. Inf. Theory IT- **24**, 339 (1978)
37. W. K. Wootters, Quantum Inf. Comput. **1**(1) (2001)
38. R. McEliece, *The theory of information and coding* (Cambridge University Press, Cambridge, 2002)

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

J Mater Sci

Chemical routes to materials

Molecularly imprinted polypyrrole decorated $Ti_3C_2T_x$ electrochemical sensor for highly selective and sensitive detection of levofloxacin

Divya Hudda[1] and Devendra Kumar[1,*]

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

Received: 18 April 2024
Accepted: 31 October 2024

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

ABSTRACT

Herein, a rapid and efficient molecular imprinted polymer (MIP) modified $Ti_3C_2T_x$-based electrochemical sensing platform was developed for levofloxacin detection (LEV). The MIP sensor was designed by depositing the synthesized $Ti_3C_2T_x$ electrophoretically onto the indium tin oxide (ITO) glass surface, followed by electropolymerization of pyrrole as a monomer and LEV as a template molecule. $Ti_3C_2T_x$ significantly enhances the electron transfer rate and offers a larger surface area to increase the number of imprinted sites for precise recognition of LEV. The fabricated sensor (MIP/$Ti_3C_2T_x$/ITO), under all optimized parameters, demonstrated superior selectivity with a low detection limit (LOD) of 0.76 pM and a linear relationship with LEV concentration ranging (1 pM–100 nM). Additionally, this newly developed sensor exhibited robust stability and sensitivity and successfully detected LEV in actual samples with satisfactory outcomes.

Introduction

Levofloxacin (LEV) belongs to the third generation of fluoroquinolones that exhibits high antibacterial properties and is widely utilized in both human and veterinary medicine to treat pneumonia, gastroenteritis, infections in the urinary and respiratory tract, skin, and soft tissues [1–3]. However, long-term use of LEV often increases the risk of certain health problems, including tendon damage, heart diseases, immune suppression, muscle atrophy, and abdominal discomfort. Hence, there is a need to develop a highly sensitive and reliable method for monitoring the presence of LEV in medical, food, and environmental samples [4–6].

At present, various analytical techniques are used for LEV detection, including fluorimetry [7], flow-injection spectrophotometry [8], capillary electrophoresis [9], nuclear magnetic resonance [10], chemiluminescence [11] and high-performance liquid chromatography [12, 13]. Although these methods possess satisfactory results, even some demerits are associated with these techniques, such as costly equipment, extended analysis periods, low sensitivity, and complex sample preparation [14, 15]. In contrast, electrochemical sensors (ECS) have gained much interest due to their high sensitivity, fast response, easy operation, and low

Handling Editor: Pedro Camargo.

Address correspondence to E-mail: dkumar@dce.ac.in

https://doi.org/10.1007/s10853-024-10403-2
Published online: 27 November 2024

cost [16]. However, the accurate and reliable detection of LEV, among other fluoroquinolone antibiotics, remains a significant challenge.

To improve selectivity, ECS often utilizes natural receptors (antibodies, enzymes, and hormones) as specific recognition components. Nonetheless, the practical applications are impeded by their high cost, short shelf-life, deactivation susceptibility, and limited environmental tolerance. Therefore, artificial receptors are also incorporated into the electrochemical sensors [17–19]. Recently, Molecular imprinting polymers (MIPs) have emerged as favorable recognition components in ECSs due to their numerous imprinted cavities that possess shape, size, and functionalities resemblance to template molecules. This alignment facilitates the selective binding to the template molecules, ultimately improving overall selectivity [20–22]. Despite their selectivity capabilities, MIPs exhibit low sensitivity, incomplete template removal, and high diffusion barriers. These limitations restrict the number of available recognition sites [23]. Therefore, the supporting material with high surface area and excellent conductivity can extend the imprinted film, resulting in the generation of more imprinted cavities and thereby effectively enhancing the sensitivity [20].

The two-dimensional layered nanomaterials with transition metal carbides, nitrides, and carbonitrides ($Ti_3C_2T_x$) have recently attracted increasing attention due to its distinct advantages, including excellent electrical conductivity (2400 Scm^{-1}), high surface area (98 m^2g^{-1}), remarkable chemical stability, good hydrophilic surface and mechanical properties [24, 25]. These characteristics have led to the extensive applications of MXene in the fields of energy storage [26], electrochemical sensing [27], electromagnetic shielding [28], and electrocatalysis [29]. Conducting polymers have been utilized for the synthesis of MIPs for selective recognition and detection of different molecules [30]. Among the conducting polymers, polypyrrole (Ppy) is a suitable candidate for molecular imprinting sensors because of its unique properties of high conductivity, good stability, redox reversibility, environmental-friendly, and superb biocompatibility [31–33].

In this study, hydrothermally synthesized $Ti_3C_2T_x$ was initially deposited onto indium tin oxide (ITO) coated glass substrate using an electrophoretic deposition (EPD) technique, which was thus coupled with MIP through one-step electropolymerization for the fabrication of novel MIP electrochemical sensor (MIP/$Ti_3C_2T_x$/ITO) to detect the LEV. The prepared MIP layer offers high specificity for the detection of LEV, and $Ti_3C_2T_x$ was utilized to enhance the sensitivity of the electrochemical sensor. The constructed MIP electrochemical sensor exhibits excellent selectivity and reproducibility, and it was successfully employed for detecting the LEV in real samples with a satisfactory recovery range.

Experimental section

Materials and instrumentation

All chemicals and instrumentation details are discussed in the supporting information (S1).

Synthesis of $Ti_3C_2T_x$

$Ti_3C_2T_x$ was prepared using the hydrothermal method as per the previous report with little modification [34]. In brief, LiF (1 g) was dissolved in 20 mL HCl (6 M) with constant stirring for 30 min. Then, Ti_3AlC_2 (1 g) powder was added slowly to the above suspension and continuously stirred at room temperature for half an hour. The resultant suspension was poured into a 100 mL Teflon-lined autoclave for hydrothermal treatment at 180 °C for 16 h. The obtained precipitate was washed through DI water with centrifugation at 3500 rpm until the supernatant pH reached 6. Finally, the obtained MXene powder was dried in a vacuum oven at 60 °C for 12 h.

Electrophoretic deposition (EPD) of $Ti_3C_2T_x$ on ITO

To fabricate the electrode, $Ti_3C_2T_x$ was deposited electrophoretically on the pre-hydrolyzed ITO. Before deposition, $Ti_3C_2T_x$ (0.5 mg) was dispersed in DI water (15 mL) through ultrasonication to form a colloidal suspension. For EPD, a two-electrode system was utilized where platinum wire served as a reference electrode, placed 1 cm apart from the ITO (working electrode). For smooth deposition of the film, a DC potential of 10 V was applied for 8 s.

Fabrication of MIP/$Ti_3C_2T_x$/ITO

To prepare the MIP/$Ti_3C_2T_x$/ITO sensor, the modified electrode ($Ti_3C_2T_x$/ITO) was immersed in polymerization solution (0.1 M PBS, pH = 7) containing 0.4 mM LEV and 2 mM pyrrole monomer. Then, CV electropolymerization was carried out over the potential range of − 0.6–1.6 V at 50 mVs^{-1} with 10 consecutive cycles. Subsequently, the modified MIP electrode was stirred in the eluent solution comprising NaOH (0.2 M) and ethanol (V/V: 1:1) for 20 min to remove the embedded template molecules (LEV), then washed with DI water and dried at room temperature. Figure 1 represents the schematic approach used for the fabrication of MIP/$Ti_3C_2T_x$/ITO. Furthermore, the non-imprinted polymer-modified electrode (NIP/$Ti_3C_2T_x$/ITO) was prepared by using the same procedure as discussed above without adding LEV in the polymerization solution.

Results and discussion

Design of proposed MIP/$Ti_3C_2T_x$ /ITO sensor

In this work, a novel 2D-layered material $Ti_3C_2T_x$ is prepared by hydrothermal method (Fig. 1A). The surface of $Ti_3C_2T_x$ is terminated with different functional groups (—F, —OH, —O), resulting in high hydrophilicity and a large electrochemically active surface, which further enhances the analytical performance of the proposed sensor. For the fabrication of the MIP sensor, the synthesized $Ti_3C_2T_x$ is deposited electrophoretically on the ITO glass substrate (Fig. 1B). After that, pyrrole is electropolymerized onto the fabricated transducer surface, while LEV is trapped within the polymer by non-covalent interactions such as hydrogen bonds and electrostatic forces (Fig. 1C). Removal of the template molecule resulted in the formation of imprinting cavities that could specifically identify the LEV molecule (Fig. 1D). After adsorption, the template molecules fill up the more imprinted sites, leading to a decrease in current (Fig. 1E).

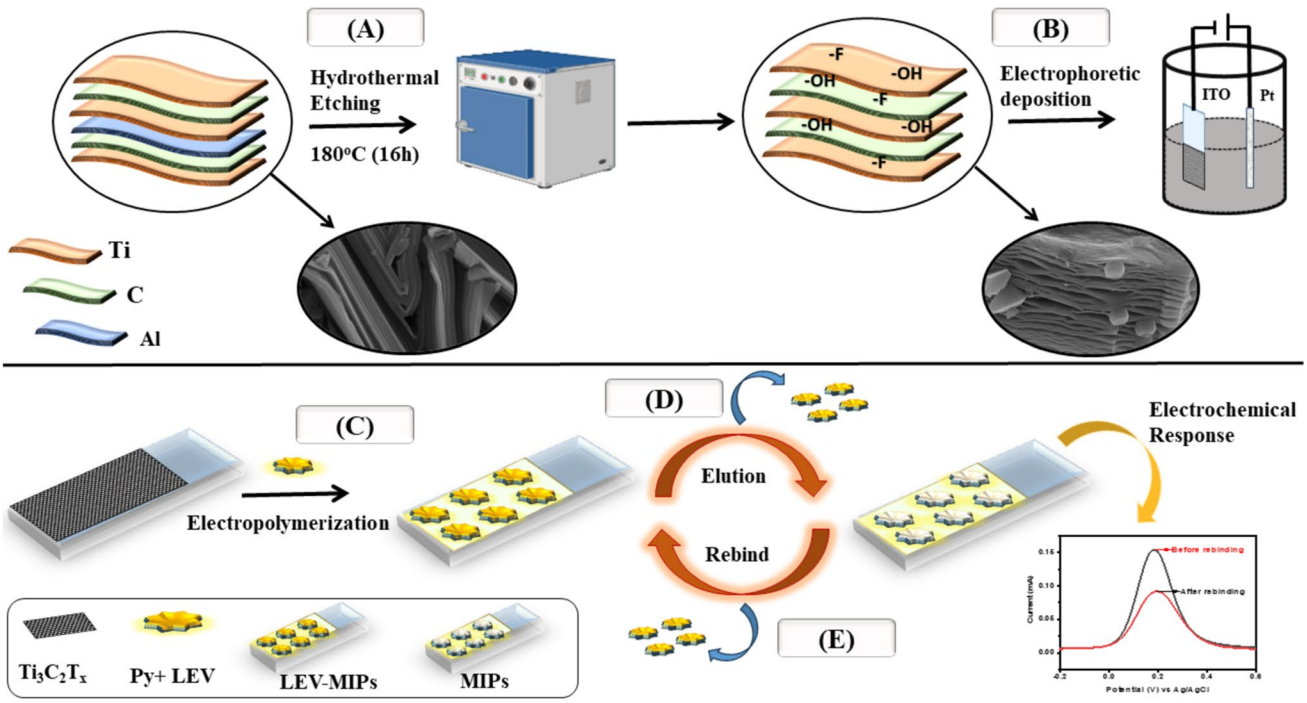


Figure 1 Schematic representation of LEV detection using a fabricated MIP/$Ti_3C_2T_x$/ITO electrode.

Structural and morphological characterization

Figure 2a represents the XRD spectra of Ti_3AlC_2 (black) and $Ti_3C_2T_x$ (red). For Ti_3AlC_2, the major diffraction peaks were observed at 9.5°, 34.1°, 39.0°, 41.9°, 48.5°, 56.4°, 60.3° and 65.3°. After hydrothermal treatment, the diffraction peak of Ti_3AlC_2 at 9.5° is broadened and shifted towards a lower 2θ angle of 6.1°, indicating the interlayer spacing of $Ti_3C_2T_x$ (13.9 Å) has been increased as compared to Ti_3AlC_2 (9.2 Å) [35]. Moreover, the diffraction peak of Ti_3AlC_2 at 39.0° has completely vanished in the spectra of $Ti_3C_2T_x$, suggesting that Al has been completely etched from Ti_3AlC_2. Figure S2 illustrates the FTIR spectra of $Ti_3C_2T_x$, which is used to examine the functional groups present on the surface of $Ti_3C_2T_x$. The bands observed at 1740, 1360, 1215, and 670 cm^{-1} can be ascribed to the stretching vibrations of C = O, O—H, C—F, and Ti—O, respectively [36].

The surface morphologies of Ti_3AlC_2, $Ti_3C_2T_x$, and polymerized electrodes are examined by FESEM. The SEM image of Ti_3AlC_2 illustrates a packed layered structure, as shown in Fig. 2b. After hydrothermal treatment, the morphology emerges an accordion-like 2D-multilayered structure as depicted in (Fig. 2c), illustrating 2D sheet-like MXene synthesized successfully. A smooth and uniform film of pyrrole on $Ti_3C_2T_x$/ITO was observed after electropolymerization in Fig. S3a. Further, on template extraction, the MIP/$Ti_3C_2T_x$/ITO exhibits a rough structure, as shown in Fig. S3b, suggesting the complete elution of the template from the polymer. Figure 2d and e represent the EDX spectra of Ti_3AlC_2 and $Ti_3C_2T_x$, respectively. In Fig. 2e, the absence of Al in $Ti_3C_2T_x$ in contrast to that of Ti_3AlC_2 from Fig. 2d confirms the successful synthesis of MXene.

Electrochemical studies of MIP/ $Ti_3C_2T_x$/ ITO

All the modified electrodes (Bare ITO, $Ti_3C_2T_x$/ITO, MIP/ $Ti_3C_2T_x$/ITO before elution, MIP/ $Ti_3C_2T_x$/ITO after elution, MIP/$Ti_3C_2T_x$/ITO after adsorption and NIP/ $Ti_3C_2T_x$/ITO) are examined electrochemically through cyclic voltammetry (CV) using $[Fe(CN)_6]^{3-/4-}$ as redox probe. As seen in Fig. 3a, bare ITO (curve i) exhibits a pair of reversible redox characteristic peaks with an electrochemical current of 0.28 mA. After being modified by $Ti_3C_2T_x$ (curve ii), the redox peak current increases

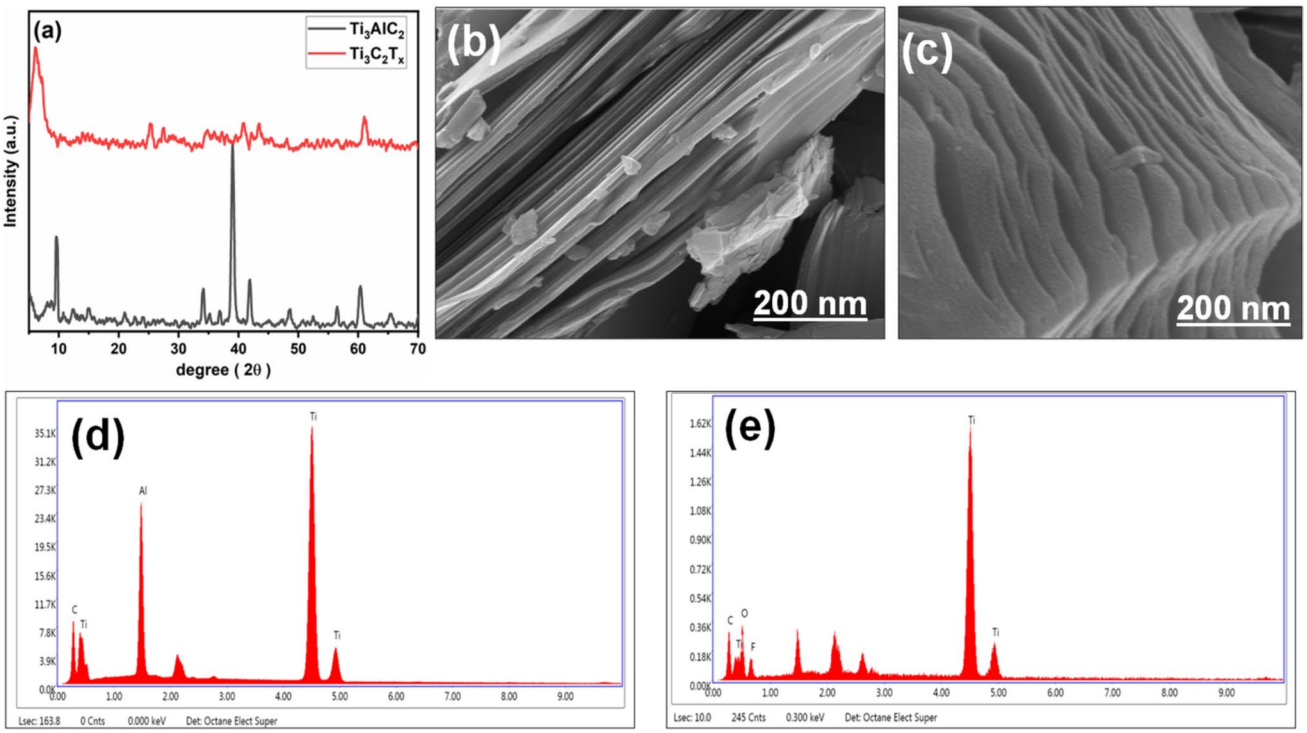


Figure 2 **a** XRD pattern; **b, c** morphology structure; and **d, e** EDX analysis of $TiAlC_2$ and $Ti_3C_2T_x$.

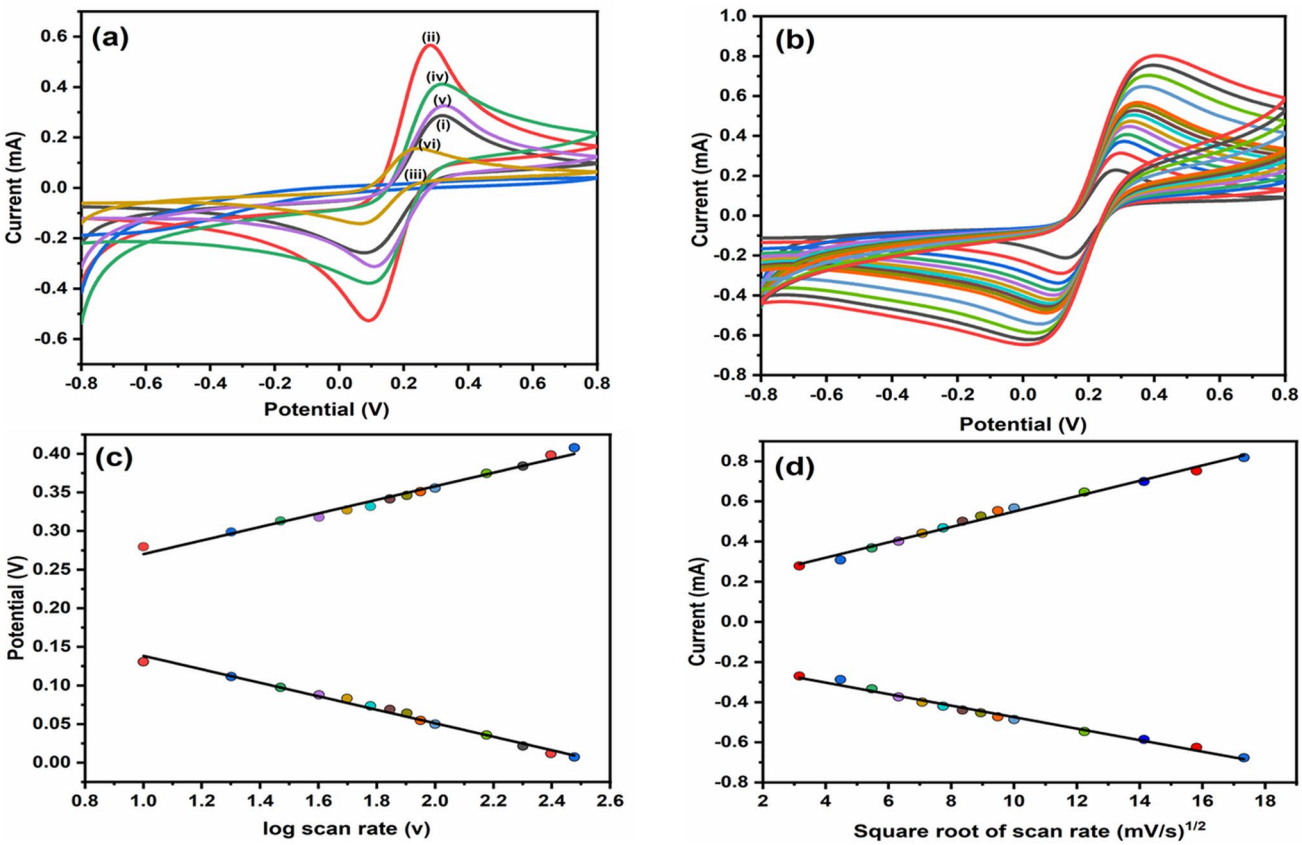


Figure 3 CV analysis of **a** different fabricated electrodes: (i) bare ITO; (ii) $Ti_3C_2T_x$/ITO; (iii) MIP/$Ti_3C_2T_x$/ITO before elution; (iv) MIP/$Ti_3C_2T_x$/ITO after elution; (v) MIP/$Ti_3C_2T_x$/ITO after adsorption; (vi) NIP/ $Ti_3C_2T_x$/ITO; **b** MIP/$Ti_3C_2T_x$/ITO electrode at varying scan rates (10–300 mV s^{-1}). The plots of **c** peak potential vs. log scan rate for MIP/$Ti_3C_2T_x$/ITO electrode and **d** peak current vs. square root of scan rate for MIP/$Ti_3C_2T_x$/ITO in a solution of 0.1 M PBS containing 5 mM $[Fe(CN)_6]^{3-/4-}$.

remarkably (0.56 mA) due to the excellent conductivity and larger surface area of $Ti_3C_2T_x$. After electropolymerization (curve iii), the redox peak almost disappears due to the non-conductive MIP layer, which inhibits the redox probe from entering the electrode surface. When the template molecules LEV are eluted (curve iv), the redox peak current is enhanced significantly (0.41 mA) again. This may be due to the formation of imprinting holes, which provide a pathway for the redox probe to access the electrode surface within the MIP film. After elution, the electrode was immersed in a solution containing template molecules (LEV), resulting in a decreased redox peak current (0.32 mA) shown by (curve v). This could be due to the blockage of some holes by template molecules LEV, thus hindering the electron transfer onto the electrode surface. In comparison, NIP/$Ti_3C_2T_x$/ITO exhibited no significant change in peak current after elution (0.15 mA) since there is no doping of LEV molecules (curve vi).

Figure 3b illustrates the CV analysis of the MIP/$Ti_3C_2T_x$/ITO electrode by varying the scan rates (υ) (10–300 mVs^{-1}). The figure clearly shows that oxidation peak current shifts towards a more positive direction while reduction peak current shifts towards a more negative direction with an increase in the scan rate. This clearly indicates the linear relationship between the anodic and cathodic peak potentials (E_{pa} & E_{pc}) vs. natural logarithm of scan rate (log υ) as shown in Fig. 3c, which can be represented by Eqs. (1) and (2):

$$E_{pa}\left[MIP/Ti_3C_2T_x/ITO\right] = 0.08783V \log(\upsilon) + 0.1823V; R^2 = 0.9825 \quad (1)$$

$$E_{pc}\left[MIP/Ti_3C_2T_x/ITO\right] = -0.08729\,V\,\log\,(\upsilon) + 0.2256\,V; R^2 = 0.9894 \quad (2)$$

Using Laviron's equation [37], the charge transfer coefficient (α) and charge transfer rate constant (K_s) for MIP/$Ti_3C_2T_x$/ITO are found to be 0.9216 and 0.0938 s^{-1}, respectively. Figure 3d shows a good linearity between the anodic and cathodic peak currents (I_{pa} and I_{pc}) with the square root of the scan rates ($\upsilon^{1/2}$), which indicates that the transfer of electrons through MIP/$Ti_3C_2T_x$/ITO electrode is surface controlled process as shown by Eqs. (3) and (4):

$$I_{pa}\left[MIP/Ti_3C_2T_x/ITO\right] = 38.30\,\mu A\,(mV\,s^{-1}) \times \upsilon^{1/2} mV\,s^{1/2} + 166.58\,\mu A; R^2 = 0.9906 \quad (3)$$

$$I_{pc}\left[MIP/Ti_3C_2T_x/ITO\right] = -28.64\,\mu A\,(mV\,s^{-1}) \times \upsilon^{1/2} mV\,s^{1/2} - 188.03\,\mu A; R^2 = 0.9887 \quad (4)$$

The electroactive surface area of $Ti_3C_2T_x$/ITO and MIP/$Ti_3C_2T_x$/ITO electrodes could be determined by the Randle-Sevcik equation as follows:

$I_p = (2.69 \times 10^5)\, n^{3/2} A D^{1/2} C \upsilon.^{1/2}$

where I_p, n, and A denote the peak current (A), the number of electrons oxidized or reduced, and the electroactive surface area (cm^2), respectively. D, C, and υ ½ represent $[Fe(CN)_6]^{3-/4-}$ diffusion coefficient (cm^2s^{-1}), molar concentration (mol cm^{-3}), and the square root of the scan rate (V s^{-1}), respectively. The calculated electroactive surface area for $Ti_3C_2T_x$/ITO and MIP/$Ti_3C_2T_x$/ITO are 0.27 cm^2 and 0.24 cm^2, respectively, consistent with the CV results.

Optimization of analytical conditions

The sensitivity and selectivity of the prepared MIP sensor were enhanced by optimizing some parameters, including the pH of PBS solution, the impact of template: monomer molar ratio, the number of electropolymerization cycles, different scan rates for the generation of MIP film, and elution as well as adsorption time.

Effect of pH

In Fig. 4a, the impact of different pH of PBS solution on the formation of polymerization film was tested ranging from 4.0 to 9.0. As seen in the figure, the peak current increases with an increase in pH up to 7 and then decreases gradually. This is because too acidic and alkaline environments might affect the imprinted film's structural integrity and reduce the sensitivity of the sensor [38]. Therefore, pH 7 was considered as the ideal pH for the electropolymerization solution.

Effect of ratio of template: monomer

The ratio of the template: monomer (LEV: py) was another parameter that affects the number of imprinted cavities within the polymer matrix during the electropolymerization process. Figure 4b represents the effect of different ratios of the template to monomer (1:1 to 1:9) on the current response. As seen clearly in the figure, the current increases with an increase in monomer concentration from 1:1 to 1:5 and then decreases gradually from 1:5 to 1:9. This is because insufficient monomer cannot bind with enough template molecules which leads to fewer recognition sites while an excess of monomer results into self-polymerization. Hence, 1:5 was selected as the optimal ratio for further experiment.

Effect of polymerization cycles

The number of polymerization cycles determines the thickness of the MIP film, which impacts the formation of imprinted cavities and the sensitivity of the sensor. In Fig. 4c, as the scan cycles increase from 4 to 10, the current response increases due to the formation of more imprinting sites. However, after exceeding 10 cycles, the current response decreased as the MIP layers were too thick to remove template (LEV) molecules. Thus, the optimal scan cycles for electropolymerization were found to be 10.

Effect of scan rate

The scan rate represents another parameter that plays a crucial role in the formation of MIP film. As shown in Fig. 4d, the impact of different scan rates on sensor performance was examined from 30 to 80 mV s^{-1}. The current response increases with an increase in scan rate from 30 to 50 mV s^{-1} which could be attributed to the slow binding of template molecules to the functional monomers, resulting in the increased density of MIP film. When the scan rate was too fast, the current

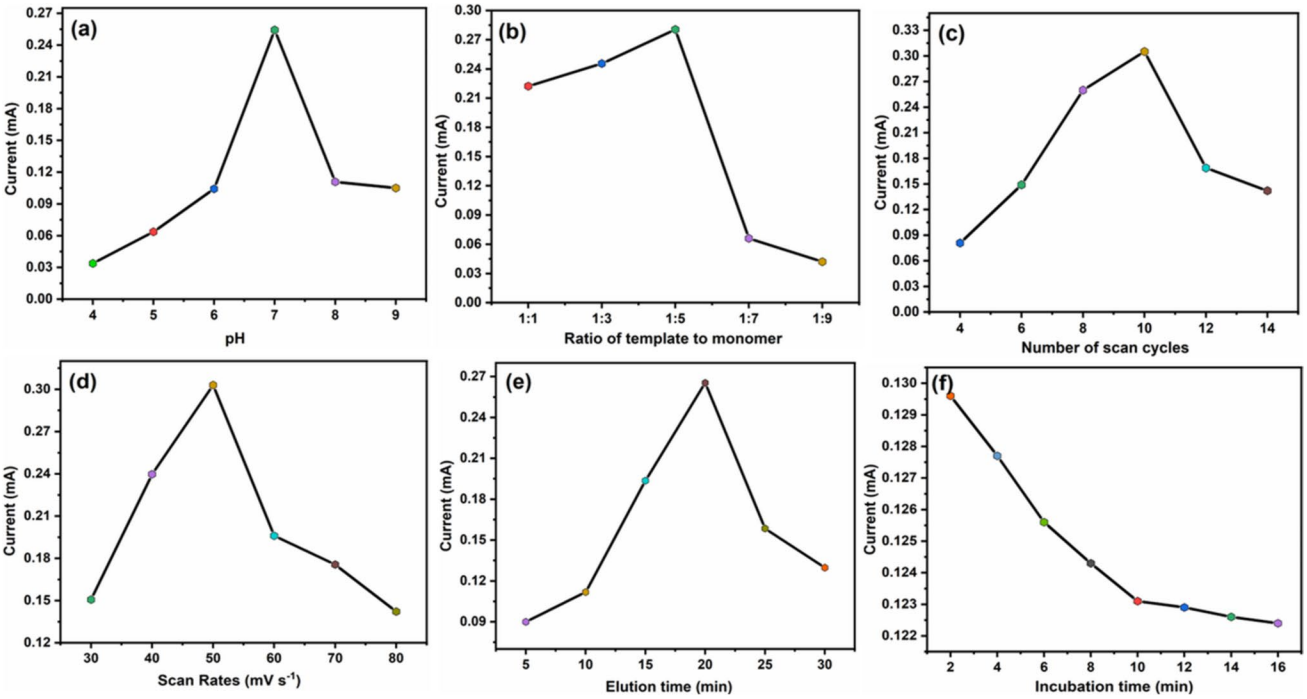


Figure 4 Optimization of analytical conditions: Effect of **a** pH of electrolyte; **b** ratio of template: monomer; **c** polymerization cycles; **d** scan rate; **e** elution time; **f** incubation time on the current response.

response decreased as the MIP film formed might be too loose, and the interaction between the template molecules and monomer was not sufficiently effective [38]. Therefore, 50 mV s^{-1} was the optimal scan rate.

Effect of elution time

Figure 4e illustrates the influence of elution time on the sensor's performance. As seen, the current response increases continuously with an increase in elution time from 5 to 20 min, and the maximum current response was observed when the elution time was 20 min. This indicates that many template molecules were removed from the polymerized film and the corresponding cavities are formed, further accelerating the electron transfer. As observed, the peak current decreases significantly after the elution duration extends from 20 to 30 min, indicating that too long elution time may damage the polymer film's surface. Therefore, the optimized elution time is found to be 20 min.

Effect of incubation time

The incubation time is another crucial parameter that impacts the performance of the developed sensor. In Fig. 4f, the current response decreased with an increase in incubation time and reached a plateau at 10 min, indicating LEV occupied the cavities exposed on the MIP/$Ti_3C_2T_x$/ITO surface. Thus, the optimal incubation time is 10 min.

Electrochemical sensing study of MIP/$Ti_3C_2T_x$/ITO sensor

Differential pulse voltammetry (DPV) technique was utilized to measure the electrochemical response of the fabricated MIP/$Ti_3C_2T_x$/ITO electrode towards LEV. As depicted in Fig. 5a, the peak current continuously decreases with an increase in LEV concentration from 1 pM to 100 nM. The peak current reduction occurs as more imprinted cavities are occupied by LEV molecules, inhibiting the redox probe's electron transfer process. The calibration curve (Fig. 5b) exhibits a linear behaviour between peak current

($\Delta I = I_{elution} - I_{adsorption}$) vs logarithm of LEV concentration (log (C_{LEV})) and obeys the following equation:

$$I = 9.649\,\mu A\,pM^{-1} \times \log C_{LEV} + 42.23\,\mu A; R^2 = 0.9799 \quad (5)$$

The sensitivity (S = slope/effective surface area) of the constructed MIP sensor was found to be 40.20 μA $(pM)^{-1}$ cm^{-2}. In contrast, the detection limit (LOD) was determined to be 0.76 pM by using the equation (LOD = 3σ/S), where σ and S represent the standard deviation of the electrode and slope of the calibration plot, respectively (Eq. 5). The efficiency of the fabricated MIP/$Ti_3C_2T_x$/ITO sensor was also compared with other reported techniques for LEV detection as outlined in Table 1. The proposed MIP sensor in this study exhibits a lower LOD value and wider detection range. This could be due to the large surface area, good electrical conductivity, the outstanding electrochemical performance of $Ti_3C_2T_x$, and specific MIP recognition.

Selectivity, reproducibility, and stability of MIP/$Ti_3C_2T_x$/ITO sensor

In order to examine the selectivity of the proposed MIP sensor, several potential structural analogs (NOR, CIP, and OME), organic matter (GLU), and ions (Na^+, CO_3^{2-}) were chosen as interfering agents. As seen in Fig. 5c, the current response of MIP/$Ti_3C_2T_x$/ITO to LEV is significantly higher even with 10 times more concentration of interfering species. These results demonstrate that the imprinted cavities from MIPs correspond well to the size, shape, and functional groups of LEV molecules. The imprinting factor (IF = $\Delta I_{MIP}/\Delta I_{NIP}$) is another essential aspect for assessing the selectivity of the MIP sensor (Fig. S4). The IFs for LEV, NOR, CIP, and OME are 9.41, 2.03, 3.47, and 1.14, respectively, indicating the high recognition ability of the fabricated sensor toward LEV.

To investigate reproducibility, five independent MIP/$Ti_3C_2T_x$/ITO electrodes were prepared under optimized conditions for LEV detection. As illustrated in Fig. 5d the relative standard deviation (RSD) of 10 nM LEV was

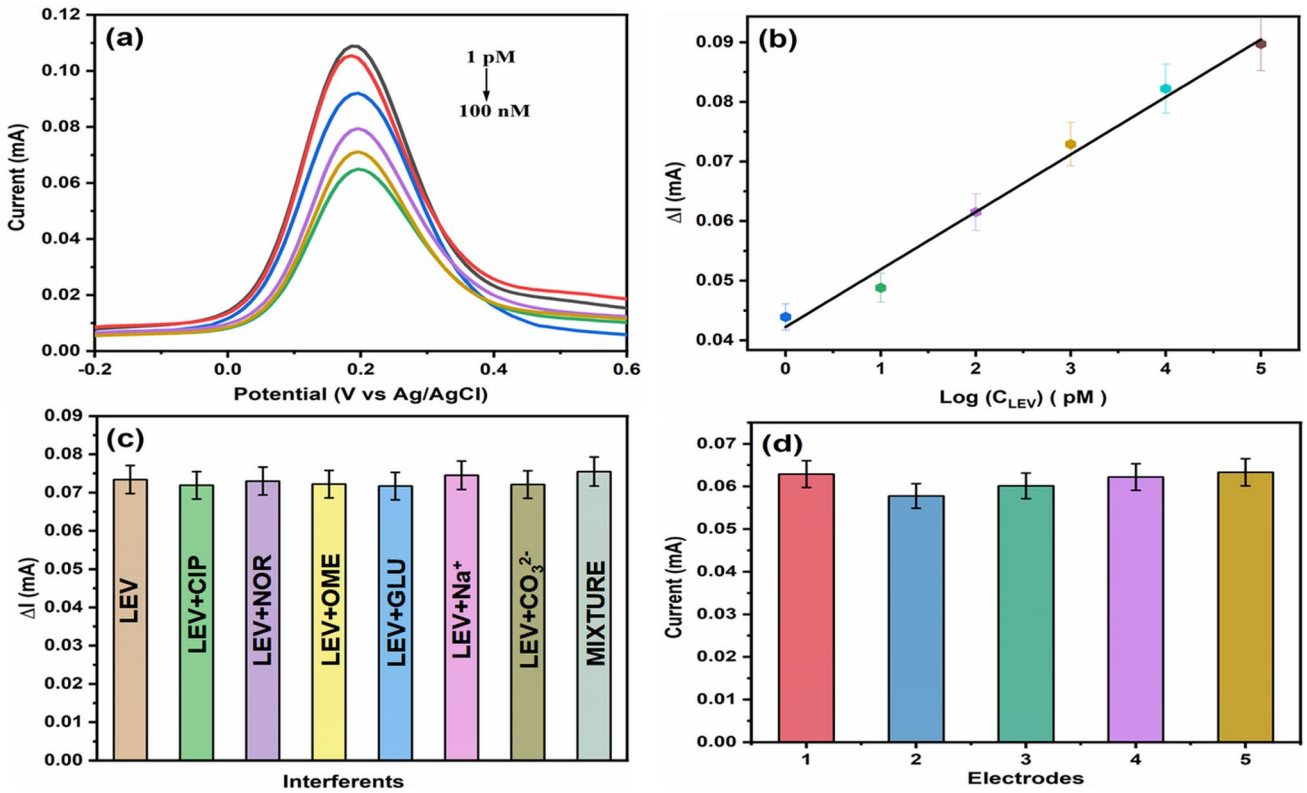


Figure 5 **a** DPV response of MIP/$Ti_3C_2T_x$/ITO electrode in various concentrations of LEV (1 pM to 100 nM); **b** Linearity plot between the ΔI and logarithm of different concentrations of LEV; **c** Selectivity study of MIP/$Ti_3C_2T_x$/ITO for LEV (10 nM) with other inteferents (100 nM of CIP, NOR, OME, GLU, Na^+, CO_3^{2-}), mixture; and **d** Reproducibility study of developed MIP/$Ti_3C_2T_x$/ITO.

3.75%, suggesting that the prepared MIP sensor exhibits good reproducibility.

Moreover, the stability of the MIP/$Ti_3C_2T_x$/ITO sensor was tested weekly in the presence of 10 nM LEV over four weeks (Fig. S5). The current response of the sensor retains 75.4% of its initial value for 4 weeks, thus exhibiting acceptable long-term stability.

Real sample analysis

The practicability and validity of the fabricated sensor to determine LEV in soil and tap water were evaluated using the standard addition method. Firstly, the soil sample was mixed with PBS buffer (7 pH), followed by centrifugation for 15 min at RT. Similarly, the tap water was filtered and centrifuged for 20 min at room temperature. After that, these real samples were spiked with various concentrations of LEV (0.1, 1, and 10 nM) and used for a sensing study for the MIP/$Ti_3C_2T_x$/ITO sensor. The recoveries and their corresponding RSD values for the detection of LEV from soil and tap water are found to be 93.6–96.9%, 93.6–104.5%, and 2.19–4.65%, 2.78–4.61%, respectively, as shown in Table 2. Moreover, the accuracy of the constructed sensor was evaluated using the HPLC–MS method. The recovery level of the electrochemical detection was found to be in good agreement with the HPLC–MS method, indicating that the proposed sensor has satisfactory reliability and accuracy for its application in real samples.

Conclusion

In summary, a novel and simple molecular imprinted electrochemical sensor for highly specific, sensitive, and rapid detection of LEV has been reported. The $Ti_3C_2T_x$-modified electrode provides a large surface area and abundant recognition sites, enhancing the electrochemical response of the MIP sensor with a wide detection linear range (1 pM—100 nM) and low LOD (0.76 pM). The sensor also possesses good selectivity, satisfactory stability, and reproducibility for LEV detection. Moreover, the precision and applicability of this MIP sensor for practical applications were validated with two

Table 1 Comparison of biosensing characteristics of the MIP/$Ti_3C_2T_x$/ITO electrode for LEV detection

S.No	Immobilization matrix	Linear range	LOD	Detection Technique	Sample	Ref
1	Au/PDDA/rGO	10–200 µM 200–800 µM	3.9 µM	LSV	LV tablets	[39]
2	MIP/G-AuNPs	1–100 µM	0.53 µM	DPV	LEV capsule	[1]
3	PoAP/MWCNTs	1–200 µM	1 µM	DPV	LV capsules	[40]
4	BDD electrode	10–80.9 µM	2.88 µM	SWV	Human urine Human serum	[41]
5	Poly(p-ABSA)/rGO	2–30 µM 30–500 µM	0.12 µM	LSV	Human urine	[4]
6	AgNs/NiO	0.25–100 µM	27 nM	SWV	Human serum	[42]
7	β-CD/rGO	100 pM–100 nM	30 pM	DPV	Spiked human serum	[43]
8	PGE/Au-NPs/poly (oPD-co-L-Dopa)	1–100 µM and 100–10000 µM	0.46 µM	DPV	Tablets Spiked human plasma	[44]
9	Ag/$AgVO_3$/N-rGO	0.09–671 µM	7.9 pM	DPV	Human blood serum Human urine Wastewater	[45]
10	MIP/$Ti_3C_2T_x$	1 pM –100 nM	0.76 pM	DPV	Soil Tap water	Present work

Abbreviations: AuNPs, gold nanoparticles; PDDA, poly(diallydimethylammonium chloride); rGO, reduced graphene oxide; G, graphene; PoAP, poly(o-aminophenol); MWCNTs, multiwalled carbon nanotubes; BDD, boron-doped diamond; poly(p-ABSA), poly(p-aminobenzene sulfonic acid); AgNs, silver nanoparticles; NiO, nickel oxide; β-CD, β-cyclodextrin; PGE, pencil graphite electrode; poly(oPD-co-L-Dopa), copolymer deposition of o-phenylenediamine and levodopa; $AgVO_3$, silver vanadate; N-rGO, nitrogen-doped reduced graphene oxide; LSV, Linear Sweep Voltammetry; DPV, Differential Pulse Voltammetry; SWV, Square Wave Voltammetry

Table 2 Detection of LEV in soil and tap water samples

Sample	Added (nM)	MIP/$Ti_3C_2T_x$/ITO			HPLC–MS		
		Found (nM)	Recovery (%)	RSD (%)	Found (nM)	Recovery (%)	RSD (%)
Soil	0	0	–	–	0	–	–
	0.1	0.093	93.6	4.65	0.099	99.0	2.28
	1	0.949	94.9	3.69	0.976	97.6	1.40
	10	9.69	96.9	2.19	9.66	96.6	2.57
Tap water							
	0	0	–	–	0	–	–
	0.1	0.096	96.1	2.78	0.097	97.0	1.44
	1	1.04	104.5	3.11	0.983	98.3	2.30
	10	9.36	93.6	4.61	9.69	96.9	3.28

real samples, providing a promising approach for LEV monitoring in the environmental field.

Acknowledgements

Divya Hudda acknowledges the Delhi Technological University (DTU) for the financial support.

Author contributions

Divya Hudda: conceptualization, methodology, validation, and writing original draft. Devendra Kumar: Supervision, writing-review, and editing.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10853-024-10403-2.

References:

[1] Wang F, Zhu L, Zhang J (2014) Electrochemical sensor for levofloxacin based on molecularly imprinted polypyrrole–graphene–gold nanoparticles modified electrode. Sens Actuators, B Chem 192:642–647

[2] Tigari G, Manjunatha J, Souza ED, Raril C, Hareesha N, Charithra MM (2022) Electrochemical determination of levofloxacin drug at poly (clayton yellow)/carbon paste electrode. Monatshefte für Chemie-Chem Mon 153:311–319

[3] Tang L, Tong Y, Zheng R, Liu W, Gu Y, Li C, Chen R, Zhang Z (2014) Ag nanoparticles and electrospun CeO_2-Au composite nanofibers modified glassy carbon electrode for determination of levofloxacin. Sens Actuators, B Chem 203:95–101

[4] Han L, Zhao Y, Chang C, Li F (2018) A novel electrochemical sensor based on poly (p-aminobenzene sulfonic acid)-reduced graphene oxide composite film for the sensitive and selective detection of levofloxacin in human urine. J Electroanal Chem 817:141–148

[5] Jin Y, Xu G, Li X, Ma J, Yang L, Li Y, Zhang H, Zhang Z, Yao D, Li D (2021) Fast cathodic reduction electrodeposition of a binder-free cobalt-doped Ni-MOF film for directly sensing of levofloxacin. J Alloy Compd 851:156823

[6] Mohammad Hossein G, Alireza K, Hossein S, Mohammad Reza G, Mehdi R-N, Farhad A (2019) An electrochemical sensor based on poly (l-Cysteine)@AuNPs @ reduced graphene oxide nanocomposite for determination of levofloxacin. Microchem J 147:198–206

[7] González JAO, Mochón MC, de la Rosa FJB (2000) Spectrofluorimetric determination of levofloxacin in tablets, human urine and serum. Talanta 52:1149–1156

[8] Al-Momani IJ (2006) Flow injection spectrophotometric determination of the antibacterial levofloxacin in tablets and human urine. Anal Lett 39:741–750

[9] Liu YM, Cao JT, Tian W, Zheng YL (2008) Determination of levofloxacin and norfloxacin by capillary electrophoresis with electrochemiluminescence detection and applications in human urine. Electrophoresis 29:3207–3212

[10] Salem A, Mossa H, Barsoum BN (2005) Quantitative determinations of levofloxacin and rifampicin in pharmaceutical and urine samples using nuclear magnetic resonance

spectroscopy. Spectrochim Acta Part A Mol Biomol Spectrosc 62:466–472

[11] Shao X, Li Y, Liu Y, Song Z (2011) Rapid determination of levofloxacin in pharmaceuticals and biological fluids using a new chemiluminescence system. J Anal Chem 66:102–107

[12] Payán MR, López MÁB, Fernández-Torres R, González JAO, Mochón MC (2011) Hollow fiber-based liquid phase microextraction (HF-LPME) as a new approach for the HPLC determination of fluoroquinolones in biological and environmental matrices. J Pharm Biomed Anal 55:332–341

[13] Sun H, Wang H, Ge X (2012) Simultaneous determination of the combined drugs of ceftriaxone sodium, metronidazole, and levofloxacin in human urine by high-performance liquid chromatography. J Clin Lab Anal 26:486–492

[14] de Farias DM, de Faria LV, Lisboa TP, Matos MAC, Muñoz RAA, Matos RC (2020) Determination of levofloxacin in pharmaceutical formulations and urine at reduced graphene oxide and carbon nanotube-modified electrodes. J Solid State Electrochem 24:1165–1173

[15] Yan G, Zhang S, Han Z, Yi S, Wang Y, Zhu L, Zhang L (2024) Construction of a sensor array using TiO_2 NRs/CuO/rMIP with imprinted sites regulation as light-activated nanozyme to colorimetric detect fluoroquinolones. Sens Actuators, B Chem 419:136373

[16] Thakur D, Pandey CM, Kumar D (2022) Graphitic Carbon Nitride-Wrapped Metal-free PoPD-Based Biosensor for Xanthine Detection. ACS Omega 8:2328–2336

[17] Majdinasab M, Mishra RK, Tang X, Marty JL (2020) Detection of antibiotics in food: New achievements in the development of biosensors. TrAC, Trends Anal Chem 127:115883

[18] Li G, Wu J, Qi X, Wan X, Liu Y, Chen Y, Xu L (2022) Molecularly imprinted polypyrrole film-coated poly (3, 4-ethylenedioxythiophene): polystyrene sulfonate-functionalized black phosphorene for the selective and robust detection of norfloxacin. Materials Today Chemistry 26:101043

[19] Yan G, Han Z, Hou X, Yi S, Zhang Z, Zhou Y, Zhang L (2023) A highly sensitive TiO_2-based molecularly imprinted photoelectrochemical sensor with regulation of imprinted sites by Photo-deposition. J Colloid Interface Sci 650:1319–1326

[20] Han S, Sun R, Zhao L, Yan C, Chu H (2023) Molecularly imprinted electrochemical sensor based on synergistic interaction of honeycomb-like Ni-MOF decorated with AgNPs and N-GQDs for ultra-sensitive detection of olaquindox in animal-origin food. Food Chem 418:136001

[21] Hasseb AA, Shehab OR, El Nashar RM (2022) Application of molecularly imprinted polymers for electrochemical detection of some important biomedical markers and pathogens. Curr Opin Electrochem 31:100848

[22] Wu Y, Li G, Tian Y, Feng J, Xiao J, Liu J, Liu X, He Q (2021) Electropolymerization of molecularly imprinted polypyrrole film on multiwalled carbon nanotube surface for highly selective and stable determination of carcinogenic amaranth. J Electroanal Chem 895:115494

[23] Bai X, Zhang B, Liu M, Hu X, Fang G, Wang S (2020) Molecularly imprinted electrochemical sensor based on polypyrrole/dopamine@graphene incorporated with surface molecularly imprinted polymers thin film for recognition of olaquindox. Bioelectrochemistry 132:107398

[24] Wang H, Li H, Huang Y, Xiong M, Wang F, Li C (2019) A label-free electrochemical biosensor for highly sensitive detection of gliotoxin based on DNA nanostructure/MXene nanocomplexes. Biosens Bioelectron 142:111531

[25] Zhu X, Zhang Y, Liu M, Liu Y (2021) 2D titanium carbide MXenes as emerging optical biosensing platforms. Biosens Bioelectron 171:112730

[26] Zhao S, Nivetha R, Qiu Y, Guo X (2020) Two-dimensional hybrid nanomaterials derived from MXenes ($Ti_3C_2T_x$) as advanced energy storage and conversion applications. Chin Chem Lett 31(4):947–952

[27] Kalambate PK, Dhanjai A, Sinha Y, Li YS, Huang Y (2020) An electrochemical sensor for ifosfamide, acetaminophen, domperidone, and sumatriptan based on self-assembled MXene/MWCNT/chitosan nanocomposite thin film. Microchim Acta 187:402

[28] Iqbal A, Kwon J, Kim MK, Koo CM (2021) MXenes for electromagnetic interference shielding: experimental and theoretical perspectives. Mater Today Adv 9:100124

[29] Zhang Y, Zhang X, Cheng C, Yang Z (2020) Recent progress of MXenes as the support of catalysts for the CO oxidation and oxygen reduction reaction. Chin Chem Lett 31:931–936

[30] Nezhadali A, Rouki Z, Nezhadali M (2015) Electrochemical preparation of a molecularly imprinted polypyrrole modified pencil graphite electrode for the determination of phenothiazine in model and real biological samples. Talanta 144:456–465

[31] Duan D, Yang H, Ding Y, Li L, Ma G (2019) A three-dimensional conductive molecularly imprinted electrochemical sensor based on MOF derived porous carbon/carbon nanotubes composites and prussian blue nanocubes mediated amplification for chiral analysis of cysteine enantiomers. Electrochim Acta 302:137–144

[32] Qian T, Yu C, Zhou X, Ma P, Wu S, Xu L, Shen J (2014) Ultrasensitive dopamine sensor based on novel molecularly imprinted polypyrrole coated carbon nanotubes. Biosens Bioelectron 58:237–241

[33] Kumar D, Sharma RC (1998) Advances in conductive polymers. Eur Polymer J 34:1053–1060

[34] Peng C, Wei P, Chen X, Zhang Y, Zhu F, Cao Y, Wang H, Yu H, Peng F (2018) A hydrothermal etching route to synthesis of 2D MXene (Ti_3C_2, Nb_2C): enhanced exfoliation and improved adsorption performance. Ceram Int 44:18886–18893

[35] Liu Z, Guo R, Li F, Zheng M, Wang B, Li T, Luo Y, Meng L (2018) Reduced graphene oxide bridged, TiO_2 modified and Mn_3O_4 intercalated $Ti_3C_2T_x$ sandwich-like nanocomposite as a high performance anode for enhanced lithium storage applications. J Alloy Compd 762:643–652

[36] Chen J, Yuan X, Lyu F, Zhong Q, Hu H, Pan Q, Zhang Q (2019) Integrating MXene nanosheets with cobalt-tipped carbon nanotubes for an efficient oxygen reduction reaction. J Mater Chem A 7:1281–1286. https://doi.org/10.1039/C8TA10574J

[37] Laviron EJJ (1979) General expression of the linear potential sweep voltammogram in the case of diffusionless electrochemical systems. J Electroanal Chem Interfacial Electrochem 101:19–28

[38] Geng L, Huang J, Zhai H, Shen Z, Han J, Yu Y, Fang H, Li F, Sun X, Guo Y (2022) Molecularly imprinted electrochemical sensor based on multi-walled carbon nanotubes for specific recognition and determination of chloramphenicol in milk. Microchem J 182:107887

[39] Borowiec J, Yan K, Tin C-C, Zhang J (2015) Synthesis of PDDA functionalized reduced graphene oxide decorated with gold nanoparticles and its electrochemical response toward levofloxacin. J Electrochem Soc 162:H164–H169

[40] Wen W, Zhao D-M, Zhang X-H, Xiong H-Y, Wang S-F, Chen W, Zhao Y-D (2012) One-step fabrication of poly(o-aminophenol)/multi-walled carbon nanotubes composite film modified electrode and its application for levofloxacin determination in pharmaceuticals. Sens Actuators, B Chem 174:202–209

[41] Rkik M, Brahim MB, Samet Y (2017) Electrochemical determination of levofloxacin antibiotic in biological samples using boron doped diamond electrode. J Electroanal Chem 794:175–181

[42] Liu C, Xie D, Liu P, Xie S, Wang S, Cheng F, Zhang M, Wang L (2019) Voltammetric determination of levofloxacin using silver nanoparticles deposited on a thin nickel oxide porous film. Microchim Acta 186:1–10

[43] Ghanbari MH, Shahdost-fard F, Khoshroo A, Rahimi-Nasrabadi M, Ganjali MR, Wysokowski M, Rębiś T, Żółtowska-Aksamitowska S, Jesionowski T, Rahimi P, Joseph Y, Ehrlich H (2019) A nanocomposite consisting of reduced graphene oxide and electropolymerized β-cyclodextrin for voltammetric sensing of levofloxacin. Microchim Acta 186:438

[44] El Azab NF, Mahmoud AM, Trabik YA (2022) Point-of-care diagnostics for therapeutic monitoring of levofloxacin in human plasma utilizing electrochemical sensor mussel-inspired molecularly imprinted copolymer. J Electroanal Chem 918:116504

[45] Sharma TSK, Hwa K-Y (2021) Facile Synthesis of Ag/$AgVO_3$/N-rGO Hybrid Nanocomposites for electrochemical detection of levofloxacin for complex biological samples using screen-printed carbon paste electrodes. Inorg Chem 60:6585–6599

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

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

Check for updates

Computational Intelligence

WILEY

Computational Intelligence

ORIGINAL ARTICLE

Multimodal Integration of Mel Spectrograms and Text Transcripts for Enhanced Automatic Speech Recognition: Leveraging Extractive Transformer-Based Approaches and Late Fusion Strategies

Sunakshi Mehra | Virender Ranga | Ritu Agarwal

Department of Information Technology, Delhi Technological University, Delhi, India

Correspondence: Sunakshi Mehra (mehra.sunakshi623@gmail.com)

Received: 7 June 2024 | **Revised:** 12 October 2024 | **Accepted:** 14 November 2024

Keywords: feedforward neural network (FNN) | ImageNet | late fusion | pre-trained network | speech command recognition

ABSTRACT

This research endeavor aims to advance the field of Automatic Speech Recognition (ASR) by innovatively integrating multimodal data, specifically textual transcripts and Mel Spectrograms (2D images) obtained from raw audio. This study explores the less-explored potential of spectrograms and linguistic information in enhancing spoken word recognition accuracy. To elevate ASR performance, we propose two distinct transformer-based approaches: First, for the audio-centric approach, we leverage RegNet and ConvNeXt architectures, initially trained on a massive dataset of 14 million annotated images from ImageNet, to process Mel Spectrograms as image inputs. Second, we harness the Speech2Text transformer to decouple text transcript acquisition from raw audio. We pre-process Mel Spectrogram images, resizing them to 224×224 pixels to create two-dimensional audio representations. ImageNet, RegNet, and ConvNeXt individually categorize these images. The first channel generates the embeddings for visual modalities (RegNet and ConvNeXt) on 2D Mel Spectrograms. Additionally, we employ Sentence-BERT embeddings via Siamese BERT networks to transform Speech2Text transcripts into vectors. These image embeddings, along with Sentence-BERT embeddings from speech transcription, are subsequently fine-tuned within a deep dense model with five layers and batch normalization for spoken word classification. Our experiments focus on the Google Speech Command Dataset (GSCD) version 2, encompassing 35-word categories. To gauge the impact of spectrograms and linguistic features, we conducted an ablation analysis. Our novel late fusion strategy unites word embeddings and image embeddings, resulting in remarkable test accuracy rates of 95.87% for ConvNeXt, 99.95% for RegNet, and 85.93% for text transcripts across the 35-word categories, as processed by the deep dense layered model with Batch Normalization. We obtained a test accuracy of 99.96% for 35-word categories after using the late fusion of ConvNeXt + RegNet + SBERT, demonstrating superior results compared to other state-of-the-art methods.

1 | Introduction

Human–computer interaction (HCI) systems in various domains—such as media, contact centers, sports, audio monitoring, and more—rely heavily on audio data. However, current speech emotion recognition (SER) techniques face several challenges when implemented in efficient phone systems, including the need for intelligent feature selection and robust deep learning algorithms. To effectively address these ongoing SER issues, a comprehensive and balanced approach has been

© 2024 Wiley Periodicals LLC.

developed, focusing on recognizing both temporal and sequential emotional cues [1].

Recurrent Neural Networks (RNNs) have shown significant success in handling speech-related tasks such as voice recognition, speech enhancement, speech separation, and speech activity detection. However, training RNNs is challenging due to issues like vanishing and exploding gradients, which hinder the model's ability to capture long-term dependencies [2]. While exploding gradients can be mitigated through truncation techniques, vanishing gradients require a more comprehensive solution. Traditionally, the term "speech modulation spectrum" refers to the magnitude spectrum of the temporal envelope [3]. Distortions in temporal dynamics can have a substantial impact on speech recognition performance. Modern ASR systems typically generate speech features at a fixed sample rate of 100 Hz. To incorporate speech dynamics into ASR features, common methods include using delta and double-delta features in the frontend or directly feeding raw or smoothed speech temporal patterns (TRAPS) into the ASR system with a wide context [4]. Acoustic frontends for ASR often employ time-frequency masking techniques [5]. In single-channel scenarios, a time-frequency mask is applied to estimate clean speech, which is then passed to the backend ASR model, either in the complex spectral domain [6] or the feature domain [5]. Alternatively, frontend processing can be performed directly in the time domain [7]. It is important to note that frontends in ASR systems can lead to speech amplification due to a misalignment between the ASR's objective and its training criteria. This problem is further compounded by streaming limitations, which impose unidirectional and causal constraints on the model's architecture.

The key distribution of this paper is:

This paper presents two key contributions in the domain of audio-linguistic information fusion and classification. To rigorously evaluate and validate these contributions, experiments were performed using the GSCD version 2, a comprehensive corpus comprising 105,937 audio recordings in WAV format across 35-word categories. The study introduces two novel transformer-based methods, each designed to effectively harness the complementary strengths of audio and linguistic features for enhanced classification accuracy.

1. Mel Spectrogram Processing Using Image Transformers: The first method leverages RegNet and ConvNext architectures, originally trained on a vast dataset of annotated images, to process Mel spectrograms as image inputs. The audio data was transformed into two-dimensional Mel spectrograms, a common representation in speech processing, which were then converted into vectors using the compact versions of the RegNet and ConvNext transformer architectures.
2. Text Transcript Extraction Using Speech2Text Transformer: The second method employs the Speech2Text transformer to extract text transcripts directly from raw audio. These textual transcripts are then processed into embeddings using the SBERT model. The probabilistic scores generated from ConvNext and RegNet, along with the embeddings from SBERT processed on text transcriptions, were further passed to a proposed five-layered dense model. This model, enriched with batch normalization techniques, performs classification on the text and image embeddings generated by the three distinct models: Speech2Text, ConvNext, and RegNet.

By integrating linguistic information with audio-centric data, specifically combining word embeddings and image embeddings, the approach achieves impressive test accuracy rates. This work pioneers the integration of textual transcripts with Mel Spectrograms, exploring the potential of combining these different modalities to enhance ASR accuracy. The paper consists of five sections: Introduction, literature review on speech recognition, proposed architecture merging spectrogram and linguistic analysis, experimental plan and findings, and summary and conclusions with future research directions.

2 | Literature Survey

In recent years, there has been a growing interest in the application of neural networks for noise processing, particularly in the realm of voice enhancement. A prominent approach to noise filtering involves using a Deep Neural Network (DNN) with a multilayer architecture, primarily based on regression-based speech enhancement techniques [8]. It is important to note that in this endeavor, both non-parametric and parametric augmentation algorithms are commonly utilized. While non-parametric methods, though valuable, often lack the richness of statistical information, parametric methods may not fully exploit the potential of the underlying speech model [9].

In the pursuit of enhancing data quality, various augmentation techniques have emerged as valuable tools, including methods like speed perturbation [10] and tempo adjustment [11]. While these techniques offer only modest improvements, they have proven to be beneficial. It is essential to recognize that the development of a robust feature extractor is of paramount importance in this context. This feature extractor acts as the cornerstone in maintaining the recognition capabilities of an ASR system.

Acoustic features [12] that are both relevant and distinctive play a crucial role in differentiating between various speech classes. Additionally, they serve as an effective means of mitigating the adverse effects of extraneous noise and managing the complexities associated with speaker variability [13]. Ensuring robustness against noise has become a critical concern in the field of speech processing [14]. Historical studies have demonstrated the widespread use of Mel Frequency Cepstral Coefficients (MFCC) as a technique for parameterizing speech, supporting tasks such as speaker identification and speech recognition. These MFCC features are carefully designed to emulate the auditory characteristics of the human hearing system, proving their effectiveness across a wide range of speech-related applications.

Recent advancements in the field have introduced Gammatone Frequency Cepstral Coefficients (GFCC), which have shown superior performance compared to MFCC, particularly in noisy environments [15, 16]. However, it is important to note that in

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

scenarios with high Signal-to-Noise Ratios (SNR), MFCC often outperforms GFCC [17]. An intriguing alternative to both MFCC and GFCC is the use of Power Normalized Cepstral Coefficients (PNCC), which have demonstrated significantly improved accuracy in noisy conditions compared to MFCC [18]. The effectiveness of PNCC can be attributed to several key components, including power-law nonlinearity, asymmetric noise reduction, and temporal masking, all of which contribute to its superior performance.

In the context of speech recognition feature extraction using the PNCC algorithm, a bank of Gammatone filters [19] is employed to emulate the non-linearly increasing bandwidths akin to those found in the human auditory system. To further enhance feature extraction, Convolutional Neural Networks (CNNs) have proven successful in emulating structural localization and minimizing translational variance within the feature space [20]. For the Uzbek language and its dialects [21], a comprehensive approach is undertaken, which includes the development of an end-to-end Deep Neural Network-Hidden Markov Model (DNN-HMM) speech recognition model, as well as a hybrid Connectionist Temporal Classification (CTC)-Attention network. This innovative method leverages the CTC objective function during attention model training, resulting in reduced training time and improved speech recognition accuracy [21]. Addressing the challenge of variability in DNN-based ASR systems, feature transformation techniques such as feature Maximum Likelihood Linear Regression (f-MLLR) and Vocal Tract Length Normalization (VTLN) have been introduced [22, 23]. Speaker-level f-MLLR transforms [22], derived from (GMM-HMM)-based ASR systems [24], are commonly employed for this purpose.

3 | Proposed Methodology

In this section, we explore decision-level fusion [25], which integrates the outputs from two channels by averaging their respective weights. Previous studies [25, 26] have demonstrated that late fusion of features from various segments of the augmented input signal significantly enhances classification performance. Post-learning, or late fusion [25, 27, 28], is a widely adopted approach for merging multiple modalities into a unified representation [29, 30]. This technique aims to boost the overall performance of speech recognition systems by combining probabilistic scores through concatenation [31].

3.1 | Google Speech Command Dataset Version 2 (GSCD-2)

For the purpose of a speech recognition challenge, the GSCD version 2, generously provided by the TensorFlow and AIY teams, serves as the foundational dataset underpinning our research endeavors. The GSCD dataset presents a rich tapestry of challenges due to its diverse array of accents and speakers. Comprising audio files in WAV format, all sampled at a uniform rate of 16 kHz, this dataset effectively mirrors real-world complexities. Our experiments are thus uniquely positioned to address the intricacies of practical scenarios. We narrow our focus to 35 specific target categories, namely: “BACKWARD”, “BED”, “BIRD”, “CAT”, “DOG”, “DOWN”, “EIGHT”, “FIVE”, “FOLLOW”, “FORWARD”, “FOUR”, “GO”, “HAPPY”, “HOUSE”, “LEARN”, “LEFT”, “MARVIN”, “NINE”, “NO”, “OFF”, “ON”, “ONE”, “RIGHT”, “SEVEN”, “SHIELA”, “SIX”, “STOP”, “THREE”, “TREE”, “TWO”, “UP”, “VISUAL”, “WOW”, “YES”, and “ZERO”.

The allocation of data within each of these categories across the training, validation, and testing sets is visually represented in Figures 1, 2, and 3, providing a clear overview of the dataset’s partitioning for reference.

3.2 | Pre-Processing Audio Samples

The audio samples used in this experiment, provided in WAV format, undergo a dual-track pre-processing pipeline. The first track utilizes a Speech2Text transformer to transcribe the audio content into text. Simultaneously, the second track processes the audio with the Librosa library, transforming it into 2D Mel-spectrograms, saved as PNG images. The Mel-spectrogram generation is configured with the following parameters: a sample rate of 22,050 Hz, a non-equidistant fast Fourier transform (FFT) size of 2,058, and a hop length of 512. This robust pre-processing framework forms the backbone of our research, facilitating the extraction of both textual and spectral features from raw audio data, as illustrated in Figure 4.

3.2.1 | Leveraging Speech2Text Transformers for Accurate Transcript Generation

The “Speech2Text” pre-trained transformer [32] is employed to convert spoken audio into written text, referred to as text transcriptions. However, the resulting text transcripts are often concise. Speech transcript normalization involves refining this data by eliminating unnecessary elements such as stop words and punctuation marks. Additionally, it includes transforming numerical values into their verbal equivalents and converting all words to lowercase to enhance readability. Tokenization is carried out to break down the text into individual words, aiding comprehension. Text filtering is also applied to expedite processing and reduce document size, further enhancing efficiency. The “Speech2Text” model, introduced in 2020 as part of the Fairseq S2T framework, represents a transformative advancement in ASR and Speech Translation (ST). This model leverages a transformer-based sequence-to-sequence (seq2seq) architecture, featuring an encoder–decoder setup, for these tasks. A standout feature of Speech2Text is its efficient use of a convolutional down sampler. This down sampler significantly reduces the input speech data length by about 75% before it reaches the encoder. This reduction is pivotal for computational efficiency and improved performance. During training, Speech2Text employs the standard autoregressive cross-entropy loss, a common objective function for sequence-to-sequence models. This loss guides the model in generating precise transcripts and translations in an autoregressive manner—meaning it generates outputs step by step, considering prior predictions. The model expects input data in the form of a float tensor representing log-mel filter-bank features extracted from the speech

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

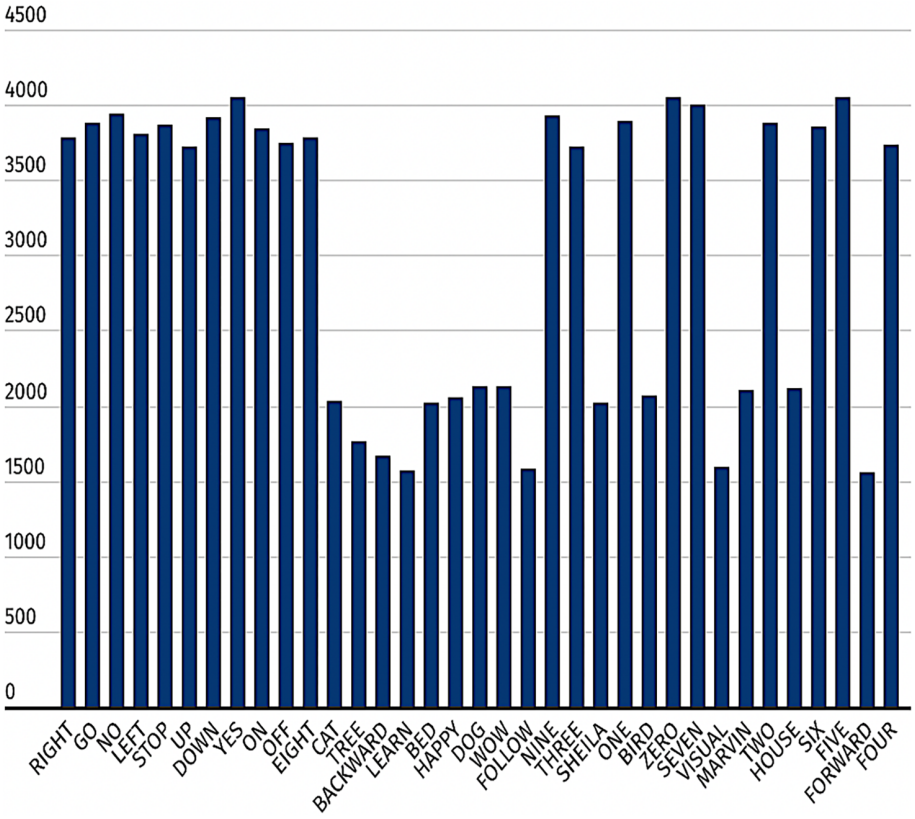


FIGURE 1 | Number of Training data/information in GSCD datasets.

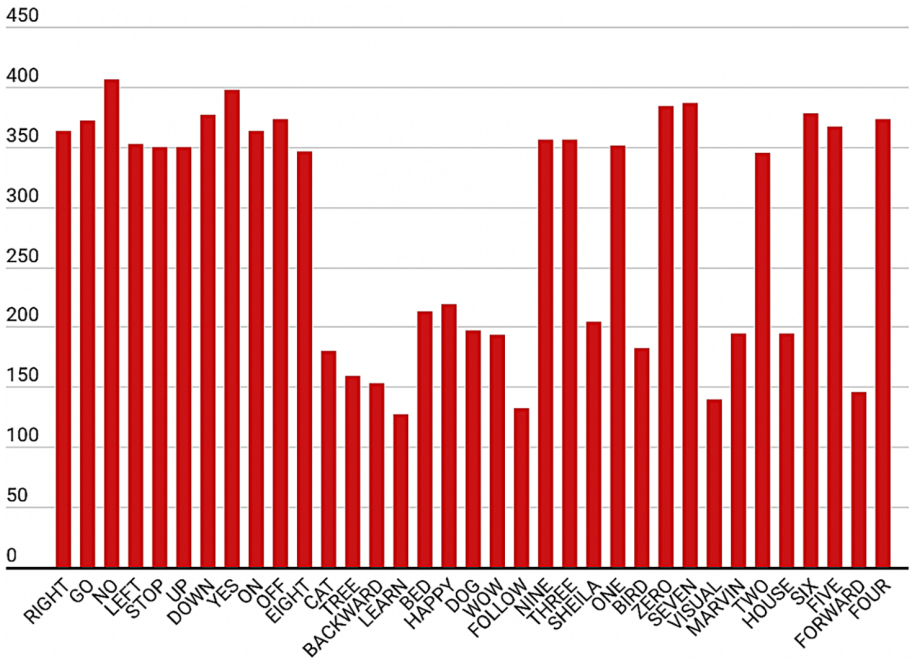


FIGURE 2 | Number of validation data/information in GSCD datasets.

signal. These features encapsulate vital acoustic information, enabling Speech2Text to produce accurate transcriptions and translations. In summary, the Speech2Text model, driven by a transformer-based seq2seq architecture, excels in ASR and ST. It efficiently reduces input length with a convolutional down sampler, utilizes autoregressive cross-entropy loss during training, and relies on log-mel filter-bank features for precise transcription and translation.

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

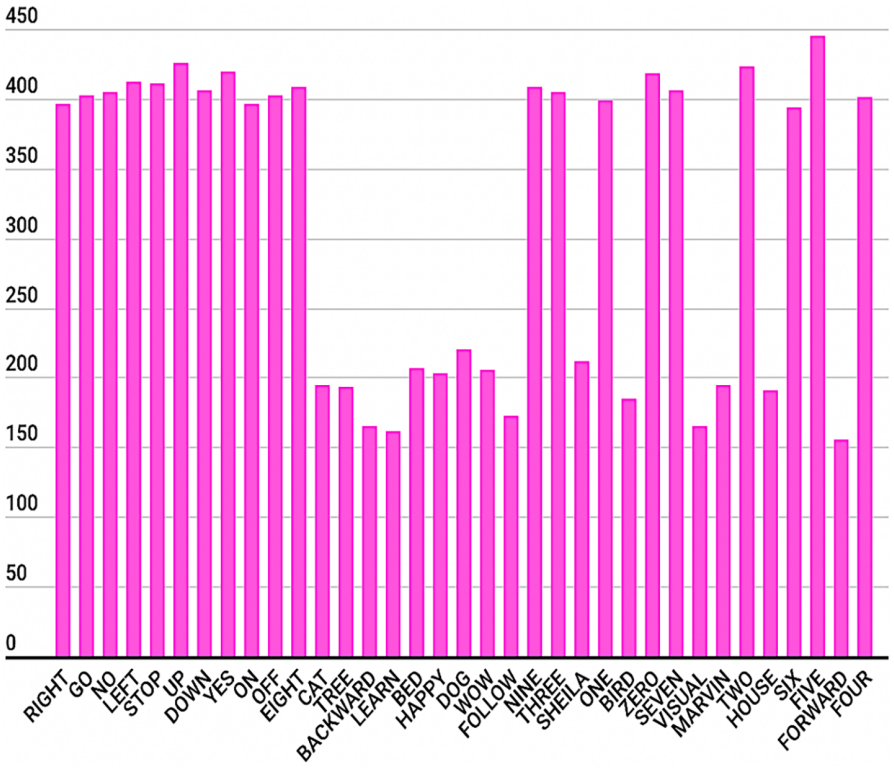


FIGURE 3 | Number of Testing data/information in GSCD datasets.

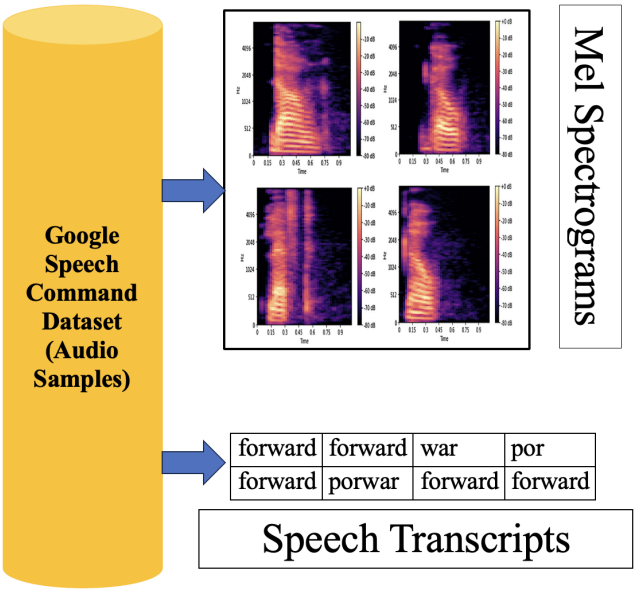


FIGURE 4 | Pre-processing of audio samples and their transformation into text and image formats.

Text transcriptions often contain a significant number of inaccuracies. Since we are dealing with isolated word recognition, we lack the context of preceding or succeeding words to aid in predicting the spoken word. It is the difficulty of transcribing text.

3.2.2 | Leveraging Image Transformers for Mel Spectrogram Processing

Subsequently, these audio signals are transformed into 2D Mel Spectrogram images using the Librosa library. The resulting images undergo several pre-processing steps, including translation, rotation, and resizing. In an iterative process, translation operations are applied to the images, which are then uniformly resized to 224×224 pixels to ensure consistency. The resized images are input into the ConvNeXt and RegNet models. The 2D Mel Spectrogram representations are utilized as input data for the ConvNeXt and RegNet models, as shown in Figures 5 and 6, respectively.

The ConvNeXt and RegNet model comprises sequential layers, ConvNeXt blocks, and patch merging, which are repeated. Specifically, in the middle of the model, there are several ConvNeXt blocks. ConvNeXt, a model pre-trained on large image datasets like ImageNet, serves as the benchmark for visual object recognition research. It encompasses a vast dataset of 1,281,167 training images, 50,000 validation images, and 100,000 test images. Renowned for its outstanding performance in vision tasks, ConvNeXt and RegNet has outperformed its predecessor, the Vision Transformer (ViT), in terms of both efficiency and accuracy. Consequently, ConvNeXt models have gained widespread adoption as the backbone of various vision-based model architectures. The Mel Spectrograms are then transformed into embeddings using the ConvNeXt model and RegNet model.

This model, pre-trained on ImageNet, excels in image classification and generates embeddings. ConvNeXt consists of several components, including Conv2d layers, LayerNorm, Linear layers, GELU activation functions, and Stochastic Depth blocks. These components are organized into a hierarchical structure. The network starts with feature extraction layers that consist of convolutional layers (Conv2d) followed by LayerNorm and activation functions. After feature extraction, the network utilizes multiple

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

CN Blocks. Each CN Block is composed of a sequence of operations, including Conv2d, Permute, LayerNorm, Linear, GELU, and Permute again. Stochastic Depth is applied to these blocks with varying dropout probabilities.

Between CN Blocks, there are downsampling layers that reduce the spatial dimensions of the feature maps. The network concludes with Adaptive Average Pooling, which reduces the spatial dimensions to 1×1, followed by a classifier. The classifier includes LayerNorm, flattening the output, and a Linear layer that maps the features to a 1000-dimensional vector. Finally, there is a head Linear layer that produces the network's final output. The RegNet architecture consists of several key components designed for efficient and effective feature extraction. At its core, it comprises a stem module that processes the input data, followed by multiple stages, each containing several residual bottleneck blocks.

The stem module initiates the network by applying a convolutional layer with 32 filters, followed by batch normalization and a ReLU activation function. This step is essential for initial feature extraction and dimensionality reduction. The main body of the network consists of multiple stages, and in each stage, there are several residual bottleneck blocks. These blocks incorporate bottleneck transformations that involve convolutional layers and batch normalization. The network architecture further benefits from Squeeze-and-Excitation (SE) blocks, which adaptively recalibrate channel-wise feature responses. This helps the

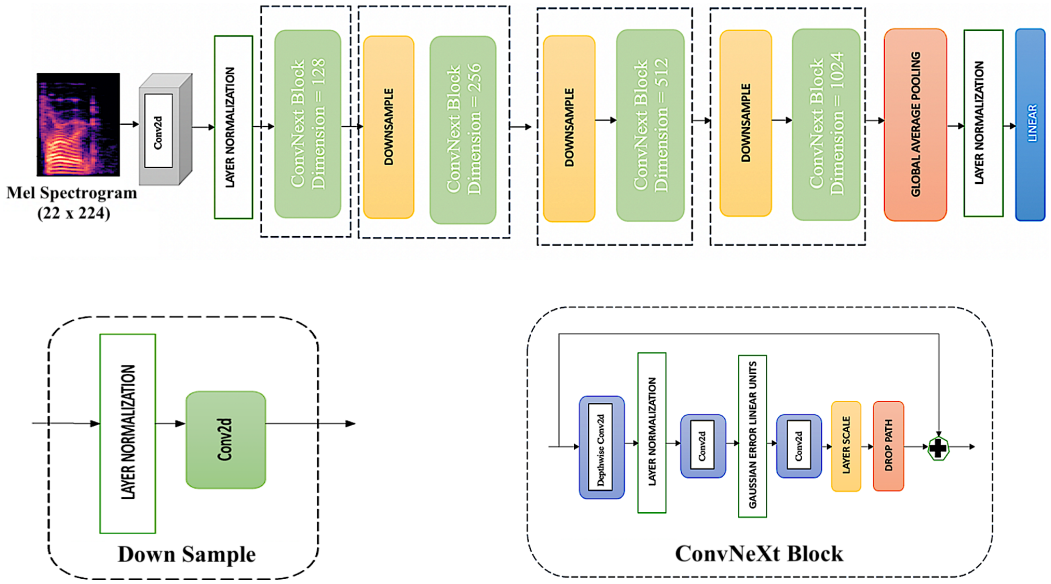


FIGURE 5 | The ConvNeXt network architecture is designed to extract features at multiple scales. Each stage includes a down-sample layer and a ConvNeXt block, along with layer normalization and Gaussian Error Linear Units (GELUs) to enhance the network's performance.

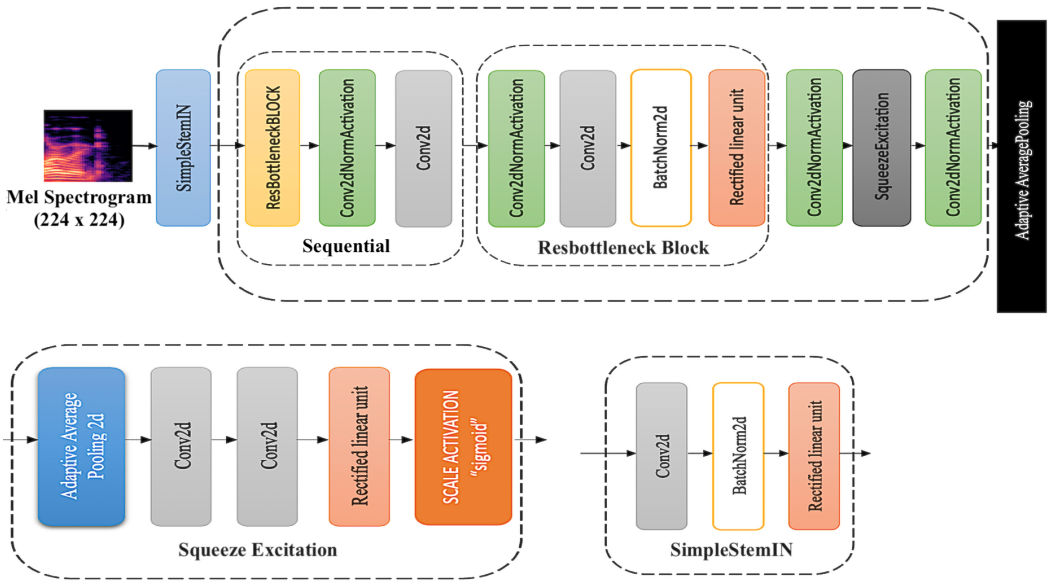


FIGURE 6 | The RegNet network architecture is designed for feature extraction at multiple scales. It achieves this through a combination of architectural components at each stage.

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

network focus on informative features. The architecture's depth and width are parameterized, making it highly customizable for various applications. The final stage involves global average pooling to reduce spatial dimensions, followed by fully connected layers for classification. In this specific configuration, it outputs 1000 classes for image classification. Furthermore, a separate head module is attached to the RegNet, which maps the extracted features to the desired output space. Overall, RegNet is designed to balance computational efficiency and performance, making it a versatile choice for various deep learning tasks, including image classification and feature extraction. Its modular structure allows for easy adaptation to different datasets and applications.

3.3 | Deep Neural Network Architecture With Batch Norm Layer

Each spoken word category is represented as a 768-dimensional vector with associated labels. Pre-trained scores derived from Mel Spectrograms using ConvNeXt, RegNet, and SentenceBERT embeddings are fed into a proposed neural network architecture, as illustrated in Figure 7.

The architecture begins with a flattening layer, followed by a dense layer with 1024 neurons and a "*tanh*" activation function. This is followed by a batch normalization layer, another dense layer with 512 neurons and a "*tanh*" activation function, and a subsequent batch normalization layer. The network continues with a dense layer of 256 neurons with "*tanh*" activation, followed by another batch normalization layer. The eighth layer consists of a dense layer with 64 neurons and "*tanh*" activation, followed by another batch normalization layer. The final layer comprises 35 neurons, from which the posterior probability scores are extracted.

Figure 8 illustrates how the three deep models in our fusion framework make predictions. We use a weighted average

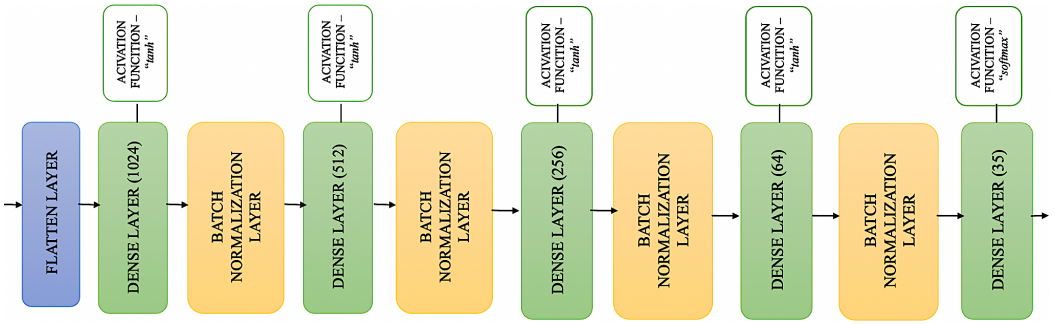


FIGURE 7 | The Proposed DNN architecture.

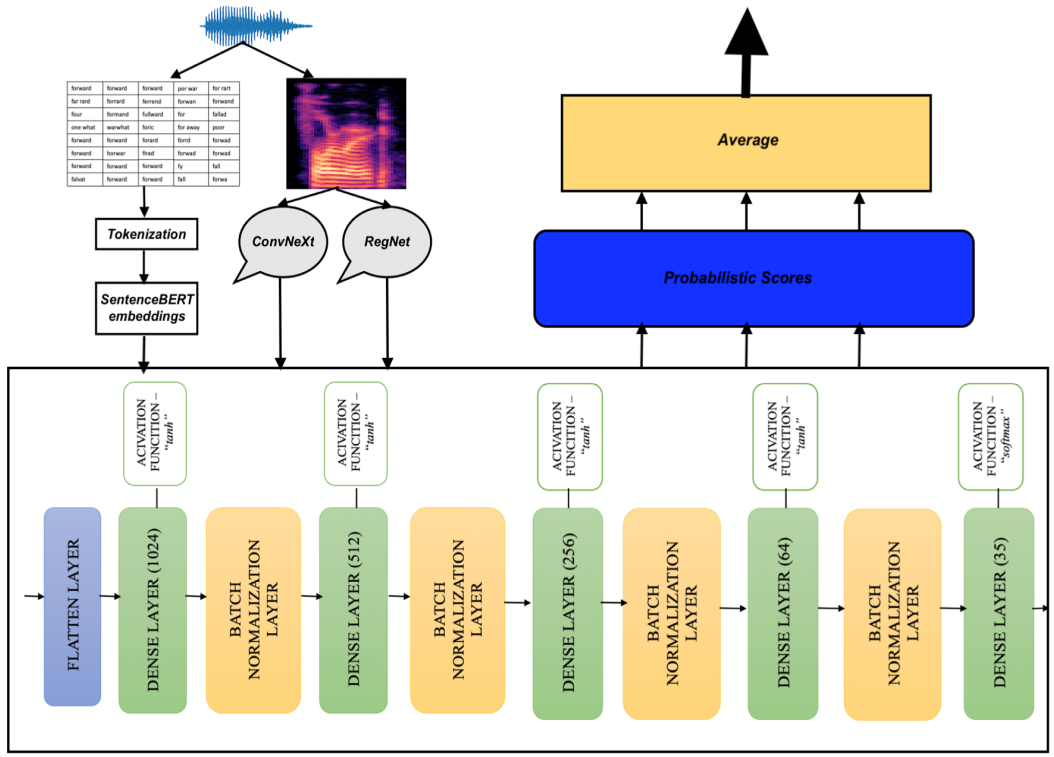


FIGURE 8 | The input to the proposed deep fusion framework consists of audio-text independent pairs.

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

function for decision fusion. In summary, Figure 8 visually represents the proposed architecture by integrating information from Figures 4–6. Next, let us delve into the details of our experimental setup.

4 | Experimental Results and Discussions

In our experimental setup, we harnessed the computational power of Google Colab Pro++ equipped with GPU acceleration, using Python as our programming language. In Table 1, the hyperparameters used in the experimentation are discussed in detail.

The hyperparameters chosen for our optimization process were guided by empirical testing and established practices in deep learning. The optimization process was steered by the Adam optimizer [33], leveraging a learning rate of 0.001. We embarked on an extensive training regimen spanning 100 epochs, with a modest batch size of 32. To introduce nonlinearity into the model, we applied the "*ReLU*" activation function. The embeddings processed through the previous channel are passed through a densely layered feed-forward network. The initial step involved flattening the data, followed by a dense layer boasting 1024 neurons and utilizing the "*tanh*" activation function. Subsequently, we introduced a batch normalization layer, succeeded by another dense layer with 512 neurons, employing the same "*tanh*" activation function. This pattern continued with additional batch normalization layers, followed by dense layers featuring 256 neurons and "*tanh*" activation, and yet another batch normalization layer. The eighth layer consisted of a dense layer with 64 neurons and "*tanh*" activation, followed by yet another batch normalization layer. Finally, the last layer was comprised of 35 neurons, from which we extracted posterior probability scores. In our optimization process, we utilized the Adam optimizer, known for its robust performance across various workloads. For training, we chose the sparse categorical cross-entropy loss function, which is well-suited for multi-class classification problems where class exclusivity is preserved. Figure 9 displays a t-SNE visualization of speech command recognition across 35-word categories using a ConvNeXt pre-trained transformer combined with a five-layer dense network featuring batch normalization. Figure 10 presents a similar visualization, but with a RegNet pre-trained transformer fused with the same five-layer dense network and batch normalization model. It is noteworthy that the t-SNE visualization for the Speech2Text model may lack the desired clarity, and the model's performance in distinguishing between the 35-word categories in text transcription with the dense batch normalization model appears to be suboptimal. These observations suggest potential areas for further investigation and model refinement to enhance the performance and visualization quality in speech command recognition and text transcription tasks across diverse word categories.

TABLE 1 | Hyperparameters for proposed five-layered dense architecture.

Framework	Values
Feature size of image	Mel spectrogram (224×224)
Feature size of text	SBERT (768×1)
Dense layer	Five layers (1024, 512, 256, 64, 35)
Dropout rate	0.3
Learning rate	0.001
Batch size	32
Epochs	100
Activation function	tanh

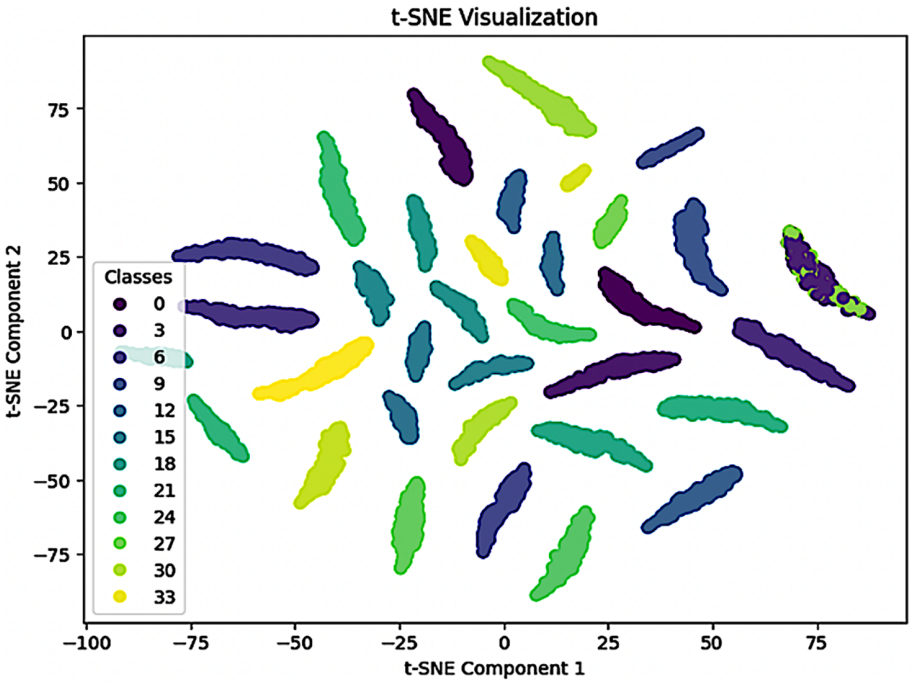


FIGURE 9 | Multiclass classification evaluation with t-SNE for Spoken Word Classification using ConvNeXt for 35-word categories.

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

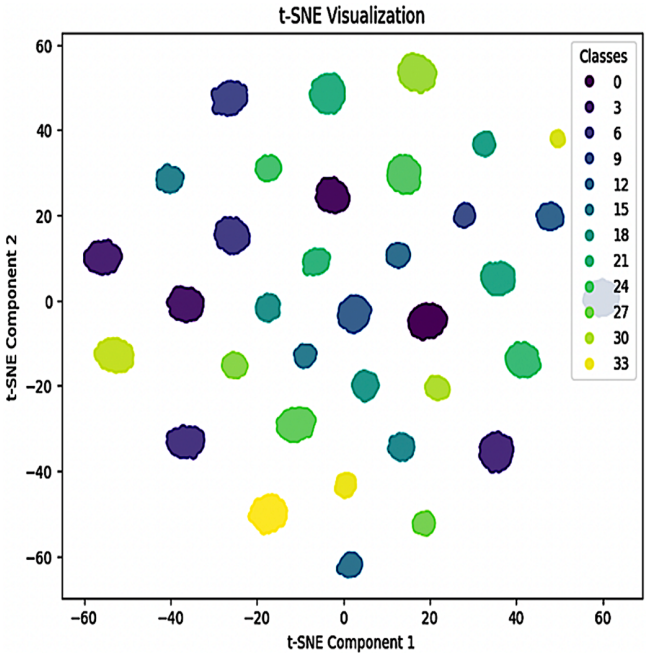


FIGURE 10 | Multiclass classification evaluation with t-SNE for Spoken Word Classification using RegNet for 35-word categories.

Our innovative late fusion strategy, which integrates word embeddings and image embeddings, has demonstrated exceptional performance in test accuracy. Specifically, our approach achieved test accuracies of 95.87% with the ConvNeXt model, 99.95% with the RegNet model, and 85.93% with text transcripts alone. When combining ConvNeXt, RegNet, and SentenceBERT (SBERT) embeddings, our method further enhanced accuracy, reaching an impressive 99.96% across the 35-word categories.

This outstanding performance surpasses that of other state-of-the-art methods currently available [12, 16, 25, 27, 34–41, 42–54]. In Table 2, we present a comprehensive comparison between our proposed model methodology and existing state-of-the-art approaches for recognizing 35-word categories. The results clearly demonstrate that our technique achieves a remarkable test accuracy of 99.96%, underscoring the effectiveness and superiority of our approach in a well-resourced data environment. This level of accuracy highlights the efficacy of our fusion strategy in leveraging the strengths of different embeddings to achieve superior performance in speech command recognition.

In the domain of speech recognition involving 35 distinct word-categories, several models have undergone rigorous testing, each demonstrating varying levels of accuracy. Noteworthy achievements include the TripletLoss res15 model proposed by Vygon and Mikhaylovskiy in 2021, which achieved 97.00% accuracy. Surpassing this, the Self-attention keyword spotting KWT 3 model introduced by Berg et al. in 2021 attained an accuracy of 97.69%. Additionally, the RNN neural attention model by de Andrade et al. in 2018 exhibited a commendable 93.90% accuracy. A standout performer among these models is M2D, presented by Niizumi et al. in 2023, boasting an impressive 98.50% accuracy. Other notable contenders include Eat S [38] at 98.15%,

TABLE 2 | Performance evaluation against the state-of-the-art for GSCD's 35-word categories.

Comparison results for 35-word categories	Accuracy (%)
TripletLoss-res15 [34]	97.00%
KWT-3 [35]	97.69%
RNN neural attention [36]	93.90%
M2D [37]	98.50%
Eat-S [38]	98.15%
AST [39]	98.11%
HTS-AT [40]	98.00%
ImportantAUG [41]	95.00%
KW-MLP [12]	97.56%
Deep-BiLSTM [27]	96.81%
Deep-BiLSTM-GRU [25]	98.90%
Proposed fusion approach (ConvNeXt + RegNet + SBERT)	99.96%

AST: Audio spectrogram transformer [39] at 98.11%, HTS AT: a hierarchical token semantic audio transformer [40] at 98.00%, ImportantAUG [41] at 95.00%, and KW MLP (Morshed and Ahsan, 2021) at 97.56% accuracy for recognizing these 35-word categories.

Our method outperforms state-of-the-art approaches on the GSCD dataset, as shown in Tables 2 and 3, achieving significant accuracy improvements across various word categories. With a test accuracy of 99.96% for 35-word categories and 99.99% for 10-word categories, this represents our highest performance to date, surpassing several leading models, including CNN, Mel Spectrogram with LSTM, EdgeCRNN, CNN-GFCCs, DenseNet-BiLSTM, LSTM for Urdu acoustic modeling, Deep Co-Training, and attention-based encoder–decoder models. Our transformer-based model demonstrates exceptional performance in speech command recognition, delivering superior accuracy in both the 10-word and 35-word categories.

These findings underscore the diverse range of model approaches and their varying levels of success in addressing the challenge of speech recognition across a broad vocabulary set [25, 27]. In the context of supervised learning and model evaluation, it is crucial to distinguish between train and test accuracy, as shown in Figures 11–13, since they serve different purposes in assessing a model's performance. Train accuracy measures how well a supervised learning model performs on the data it was trained on. It reflects the model's ability to fit the training data and quantifies how accurately the model predicts the target variable within its training dataset. In contrast, test accuracy evaluates the model's generalization capability. It assesses the model's performance on unseen data that was not used during training. Test accuracy indicates how well the model can make accurate predictions on new, previously unseen examples and is a crucial indicator of a model's real-world utility. A significant disparity between train and test accuracy can signal overfitting, where the model has memorized

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 | Performance evaluation against the state-of-the-art for the Google Speech Command Dataset's 10-word category.

Comparison results for 10-word categories	Accuracy (%)
MFCC + CNN [16]	93.28%
GFCC + CNN [42]	93.09%
MFCC + LSTM-RNN [43]	95.44%
MFCC + LSTM-RNN [44]	95.14%
MelSpec + LSTM [45]	95.07%
DenseNet + BiLSTM [46]	94.88%
RNN neural attention [36]	94.11%
EdgeCRNN [47]	98.20%
Semi supervised audio tagging [48]	95.58%
Attention based s2s model [49]	97.50%
TripletLoss-res15 [34]	98.38%
BC-ResNet-8 [50]	98.70%
KWT-3 [35]	98.49%
MatchboxNet-3 × 2 × 64 [51]	97.63%
ConvMixer [52]	98.21%
Embedding + head [53]	97.70%
Wav2KWS [54]	98.52%
Deep-BiLSTM [27]	98.53%
Deep-BiLSTM-GRU [25]	99.78%
Proposed fusion approach (ConvNeXt + RegNet + SBERT)	99.99%

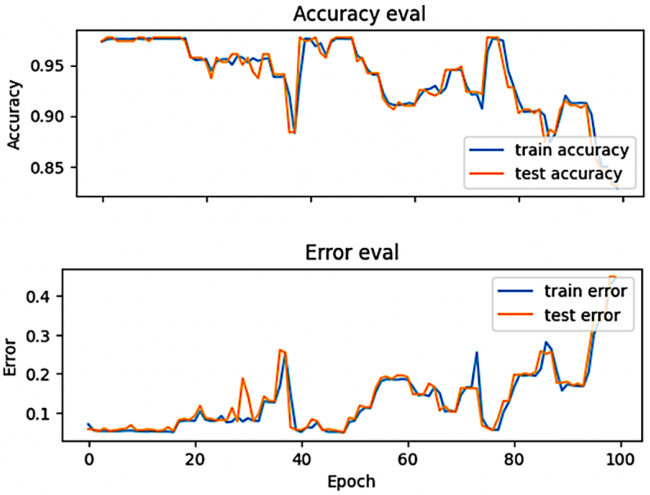


FIGURE 11 | Train vs test accuracy with train error vs test error is depicted above for results obtained from ConvNeXt with proposed neural network architecture.

the training data but struggles to generalize to new data. Monitoring the gap between these two accuracy metrics helps identify overfitting issues and guides model refinement. In practical applications, test accuracy is the more critical metric because it gauges a model's performance in making predictions on new, unseen data. High test accuracy is indicative of a robust and reliable model suitable for decision-making and real-world applications. Train accuracy alone can be misleading, as it may not reflect a model's actual performance when faced with new data. Therefore, both train and test accuracy are typically reported and compared to provide a comprehensive evaluation of a model's capabilities.

Train error measures a model's performance on the data it was trained on. It quantifies how well the model fits or predicts the target variable within the training dataset. Train error is typically lower than test error because the model has seen and learned from the training data. In contrast, test error assesses a model's performance on unseen data that it was not exposed to during training. It reflects the model's ability to generalize and make accurate predictions on new, previously unseen examples. Test error is a critical metric for evaluating a model's real-world performance and its utility for practical applications. Table 4 presents the Precision, Recall, and F1-scores for each word-category. In this context, we represent true positives, true negatives, false negatives, and false positives with the symbols tp, tn, fn, and fp, respectively. Precision, a vital metric, is calculated as the ratio of true positives to the total number of positives detected by the model, as described in Equation (1). It assesses the model's ability to accurately identify positive instances while minimizing false positives.

$$\text{Precision/positive predictive value (PPV)} = \frac{tp}{[tp + fp]} \quad (1)$$

Recall, sometimes called the true positive rate or sensitivity, holds a pivotal role as a metric. It is computed as the ratio of true positives to the overall actual positives, as depicted in Equation (2). This metric measures the model's aptitude in capturing all pertinent positive instances while striving to reduce false negatives.

$$\text{Sensitivity/recall} = \frac{tp}{[tp + fn]} \quad (2)$$

To gauge the model's performance comprehensively, we utilize the harmonic mean of precision and recall to compute the f1-Score, as expressed in Equation (3). This metric is of paramount importance as it offers a holistic evaluation of the model's accuracy by considering both precision and recall simultaneously.

$$\text{f1} - \text{Score} = \frac{2 \times precision \times recall}{[precision + recall]} \quad (3)$$

In our experimental setup, we observed remarkably high values for precision, recall, and f1-Score metrics for 35-word categories. These exceptional metrics collectively indicate a robust classification performance. As detailed in the specified mathematical formulas, the results of our statistical analysis are presented in Table 4. This table emphasizes that a significant proportion of the classes have been precisely classified, with only a few classes showing minor deviation.

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

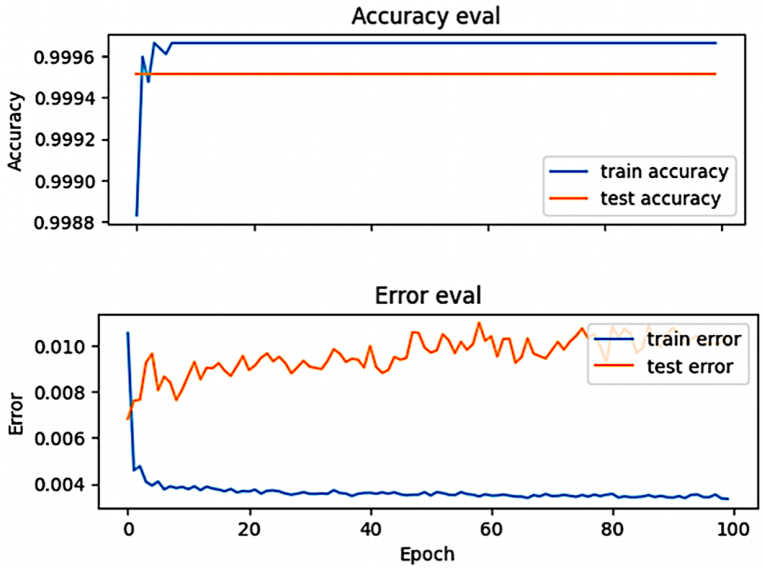


FIGURE 12 | Train vs test accuracy with train error vs test error is depicted above for results obtained from RegNet with proposed neural network architecture.

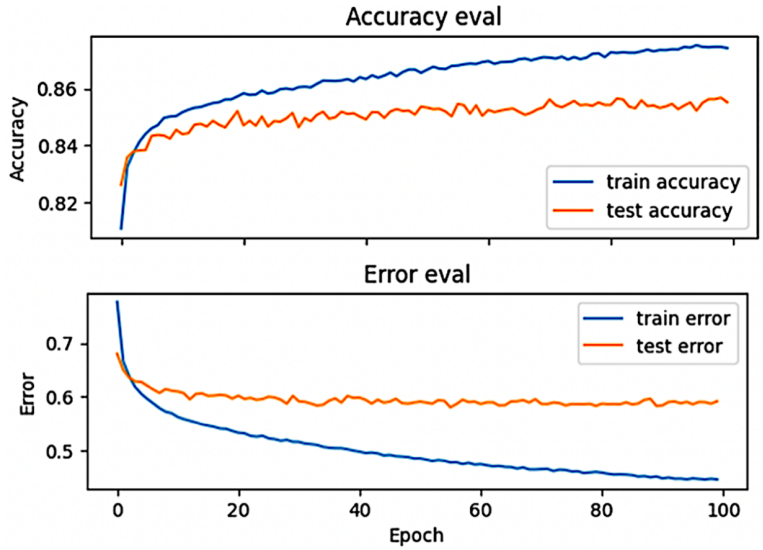


FIGURE 13 | Train vs test accuracy with train error vs test error is depicted above for results obtained from text transcript with proposed neural network architecture.

This underscores the effectiveness of our approach in achieving accurate classification results across the majority of categories, highlighting the success of our research.

Table 5 presents the classification report generated by our proposed deep dense model for all word categories. The report includes several evaluation metrics for each word class, specifically accuracy, negative predictive value (NPV), false positive rate (FPR), false negative rate (FNR), and false discovery rate (FDR).

NPV, represented mathematically as follows:

$$\text{Negative predictive value (NPV)} = \frac{tn}{[fn + tn]} \tag{4}$$

NPV signifies the percentage of cases with negative test results that are genuinely valid samples. In essence, it measures the proportion of subjects correctly identified as negative among all participants who received negative test results, including those incorrectly identified as positive. NPV achieves its highest value of 1 (100%) when a test is flawless, meaning it produces no false negative results. Conversely, it drops to its lowest value of 0 (0%) when the test fails to identify any true negative results. PPV, in an ideal test, attains its best score of 1 (100%), indicating that all positive test results are accurate. On the other end of the spectrum, its worst score is 0. The FNR represents the proportion of positive test results that incorrectly yield a negative outcome. It is also referred to as the conditional likelihood of obtaining a negative test result when the positives being tested for are present. The FDR anticipates the proportion of type I errors. In other words,

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 | Mathematical analysis per word category.

Categories	Recall	F1-score	Precision
Right	1.00	1.00	1.00
Go	1.00	1.00	1.00
No	1.00	1.00	1.00
Left	1.00	1.00	1.00
Stop	1.00	1.00	1.00
Up	0.99	1.00	1.00
Down	1.00	1.00	1.00
Yes	1.00	1.00	0.99
On	1.00	1.00	1.00
Off	1.00	1.00	1.00
Eight	1.00	1.00	1.00
Cat	1.00	1.00	1.00
Tree	1.00	1.00	1.00
Backward	0.99	1.00	1.00
Learn	1.00	1.00	0.99
Bed	1.00	1.00	1.00
Happy	1.00	1.00	1.00
Dog	1.00	1.00	1.00
Wow	1.00	1.00	1.00
Follow	1.00	1.00	1.00
Nine	1.00	1.00	1.00
Three	1.00	1.00	1.00
Sheila	1.00	1.00	1.00
One	1.00	1.00	1.00
Bird	1.00	1.00	1.00
Zero	1.00	1.00	1.00
Seven	1.00	1.00	1.00
Visual	1.00	1.00	1.00
Marvin	1.00	1.00	1.00
Two	1.00	1.00	1.00
House	1.00	1.00	1.00
Six	1.00	1.00	1.00
Five	1.00	1.00	1.00
Forward	1.00	1.00	1.00
Four	1.00	1.00	1.00

TABLE 5 | Classification report per word category.

35-word categories	TPR	TNR	PPV	NPV	FPR	FNR	FDR
Right	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Go	1.00	1.00	1.00	1.00	0.00	0.00	0.00
No	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Left	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Stop	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Up	0.99	1.00	1.00	0.99	0.00	0.02	0.00
Down	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Yes	1.00	0.99	0.99	1.00	0.94	0.00	0.02
On	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Off	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Eight	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Cat	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Tree	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Backward	0.99	1.00	1.00	0.99	0.00	0.06	0.00
Learn	1.00	0.99	0.99	1.00	0.92	0.00	0.06
Bed	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Happy	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Dog	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Wow	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Follow	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Nine	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Three	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Sheila	1.00	1.00	1.00	1.00	0.00	0.00	0.00
One	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Bird	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Zero	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Seven	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Visual	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Marvin	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Two	1.00	1.00	1.00	1.00	0.00	0.00	0.00
House	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Six	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Five	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Forward	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Four	1.00	1.00	1.00	1.00	0.00	0.00	0.00

it estimates the percentage of incorrect positive identifications among all positive test results. To summarize concisely, NPV is perfect when it is 1, PPV is ideal at 1, the FNR signifies the proportion of positives yielding negative results, and the FDR estimates the percentage of type I errors.

$$\text{False positive rate } (FPR) = \frac{fp}{[fp + tn]} \tag{5}$$

$$\text{False negative rate (FNR)} = \frac{fn}{[fn + tp]} \tag{6}$$

$$\text{False discovery rate (FDR)} = \frac{fp}{[fp + tp]} \tag{7}$$

Table 5 reveals that the majority of classes exhibit elevated specificity and NPV, indicating their ability to correctly identify true negatives. Furthermore, certain classes demonstrate robust sensitivity and PPV, signifying their proficiency in correctly detecting true positives. Across all classes, FPR values consistently remain at a minimal level, underscoring the model's ability to avoid false positives. Likewise, FNR and FDR values are generally on the lower end, demonstrating the model's effectiveness in minimizing false negatives and type I errors.

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Our success in achieving high accuracy was accomplished through the synergistic integration of ConvNeXt, RegNet, and texttranscript-based model Speech2Text. This achievement is particularly noteworthy considering the scale of our dataset, which totaled approximately 8 GB. Handling such a large dataset without GPU resources presented a significant challenge. The results underscore the impressive capability of ConvNeXt and RegNet models, which, despite having no prior exposure to Mel Spectrogram images, provided accurate estimations. In contrast, the text transcript-based approach showed relatively poor performance. It is important to note that consistently achieving high accuracy, even with pre-trained models, remains challenging. This study highlights the potential for improvement in the field of ASR, especially in the linguistic domain. Extracting precise text transcripts remains a complex and labour-intensive task. Our most notable results were achieved through the fusion of audio and text data, surpassing previous state-of-the-art methods. This success was enabled by leveraging the foundational knowledge encoded in pre-trained models. In summary, our research illustrates the potential for further advancements in ASR, emphasizing the value of multimodal approaches and the insights gained from pre-trained models.

TABLE 6 | Accuracy report per word category.

Categories	ACC
Right	1.00
Go	1.00
No	1.00
Left	1.00
Stop	1.00
Up	0.99
Down	1.00
Yes	1.00
On	1.00
Off	1.00
Eight	1.00
Cat	1.00
Tree	1.00
Backward	0.99
Learn	1.00
Bed	1.00
Happy	1.00
Dog	1.00
Wow	1.00
Follow	1.00
Nine	1.00
Three	1.00
Sheila	1.00
One	1.00
Bird	1.00
Zero	1.00
Seven	1.00
Visual	1.00
Marvin	1.00
Two	1.00
House	1.00
Six	1.00
Five	1.00
Forward	1.00
Four	1.00

Table 6 displays the results of our proposed fusion approach, which integrates audio and text information for each spoken word category. Overall, the accuracy is high, with only two categories—"Backward" and "Up"—showing relatively lower performance. Despite this, the results for these categories remain well-classified. For a deeper understanding of the fusion results, Figures 14 and 15 present confusion matrices for both 10-word and 35-word categories. These matrices illustrate the relationship between predicted values on the x-axis and actual values on the y-axis. In addition to accuracy, we have calculated several other classification metrics, including Macro Precision, Macro Recall, Macro F1-Score, and Matthews Correlation Coefficient (MCC). These metrics collectively highlight the robustness and effectiveness of our methodology for classification tasks. In summary, Table 6 provides a comprehensive overview of the fusion results across various spoken word categories, demonstrating high accuracy overall. The accompanying confusion matrices and additional metrics, as detailed in Figures 14 and 15, further underscore the reliability and quality of our classification methodology.

In Table 7, MCC is a metric that evaluates a classification model. It considers all four components of the confusion matrix to calculate a coefficient that ranges from −1 to 1. A higher MCC value indicates better model performance, with 1 representing perfect prediction, 0 indicating random prediction, and −1 suggesting total disagreement between predictions and actual values. It is a robust measure for imbalanced datasets.

$$\mathrm{MCC} = (\mathrm{tp} * \mathrm{tn} - \mathrm{fp} * \mathrm{fn}) / \sqrt{(\mathrm{tp} + \mathrm{fp})(\mathrm{tp} + \mathrm{fn})(\mathrm{tn} + \mathrm{fp})(\mathrm{tn} + \mathrm{fn})} \quad (8)$$

Due to diverse implementation complexities, working with multiple modalities—such as signal, audio, text, facial data, and motion—presents both exciting prospects and significant hurdles. One of the main obstacles is the computational energy required to train deep networks from scratch on large datasets. Our system employs autonomous learning processes tailored to each modality, providing a partial solution. It is important to note that the length of the input has a greater impact on computational complexity than the machine's processing speed.

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

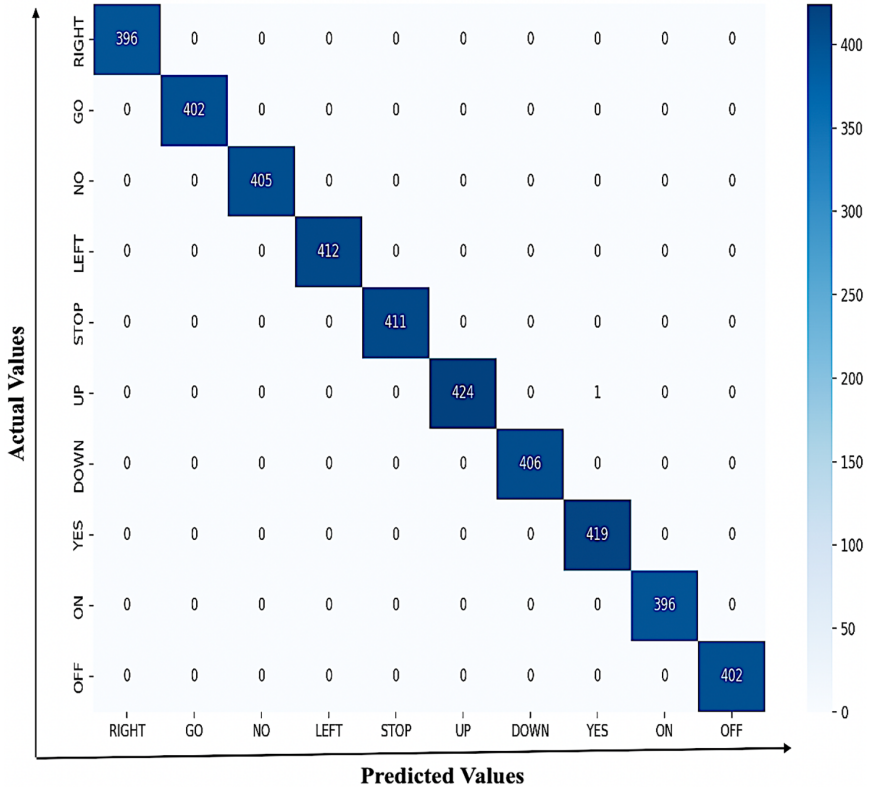


FIGURE 14 | Confusion matrix of proposed fusion approach for 10-word categories.

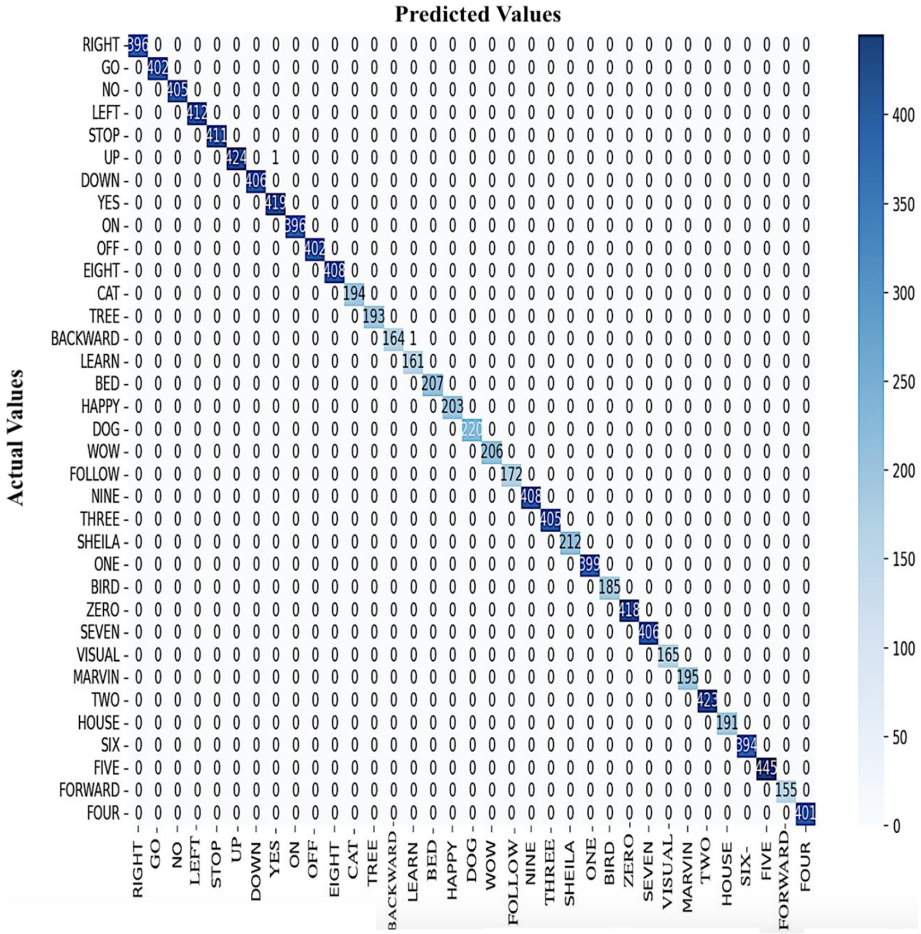


FIGURE 15 | Confusion matrix of proposed fusion approach for 35-word categories.

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 7 | Proposed fusion result metrics.

Method	ACC	Macro precision	Macro recall	Macro f1-score	Mathew's correlation coefficient
Proposed fusion result (ConvNeXt + RegNet + SBERT)	99.96	1.00	1.00	1.00	0.9998

5 | Conclusions

In conclusion, this research aims to advance the field of ASR [55, 56] by integrating multimodal data, specifically textual transcripts and Mel Spectrograms (2D images) derived from raw audio. By exploring the potential of combining spectrograms with linguistic information, this study seeks to enhance spoken word recognition accuracy. Two distinct transformer-based approaches were proposed: first, leveraging RegNet and ConvNeXt architectures—originally trained on a vast dataset of 14 million annotated images from ImageNet—to process Mel Spectrograms as image inputs; and second, utilizing the Speech2Text transformer to extract text transcripts from raw audio. Mel Spectrograms were pre-processed, resized to 224×224 pixels, and individually classified by ImageNet, RegNet, and ConvNeXt, generating embeddings for visual modalities. Sentence-BERT embeddings, created via Siamese BERT networks, transformed Speech2Text transcripts into vectors. These embeddings were then fine-tuned within a five-layer deep dense model, incorporating batch normalization, for spoken utterance classification. Experiments focused on the GSCD version 2, covering 35-word categories, and an ablation analysis was conducted to assess the impact of spectrograms and linguistic features. The novel late fusion strategy, which combines word embeddings and image embeddings, resulted in outstanding test accuracy rates: 95.87% for ConvNeXt, 99.95% for RegNet, and 85.93% for text transcripts across the 35-word categories, as processed by the deep dense layered model with batch normalization. The highest test accuracy, 99.96%, was achieved after applying the late fusion of ConvNeXt, RegNet, and SBERT, demonstrating superior results compared to other state-of-the-art methods. The low accuracy in text transcription is primarily due to the variability in speaker characteristics, such as speaking styles, tone, pitch, and intensity. Furthermore, differences in pronunciation and individual speaker idiosyncrasies exacerbate these inaccuracies, making consistent and accurate transcription more challenging. Despite employing the recommended approach, substantial accuracy improvements remain elusive for certain word categories. To achieve significant enhancements in recognition performance for these specific areas, further research and refinement are imperative. Addressing the challenges and achieving notable accuracy improvements in these word categories may necessitate additional in-depth study and experimentation.

In the future, artificial neural networks and transformers, initially developed for image analysis, hold significant potential for advancing speech recognition [57–62]. By leveraging CNNs and visual transformers to process spectrogram images, these models could enhance feature extraction and classification, leading to more accurate detection and interpretation of spoken language.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks From Overfitting," *Journal of Machine Learning Research* 15, no. 1 (2014): 1929–1958, https://doi.org/10.5555/2627435.2670313.

2. A. Iqbal and S. Aftab, "A Classification Framework for Software Defect Prediction Using Multi-Filter Feature Selection Technique and MLP," *International Journal of Modern Education & Computer Science* 12, no. 1 (2020): 18–25, https://doi.org/10.5815/ijmecs.2020.01.03.

3. H. Hermansky, "The Modulation Spectrum in the Automatic Recognition of Speech," *In 1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proceedings* (IEEE: 1997), 140–147, https://doi.org/10.1109/ASRU.1997.658998.

4. S. Sadhu and H. Hermansky, "Importance of Different Temporal Modulations of Speech: A Tale of Two Perspectives," *ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) (IEEE: 2023), 1–5, https://doi.org/10.48550/arXiv.2204.00065.

5. A. Narayanan and D. L. Wang, "Ideal Ratio Mask Estimation Using Deep Neural Networks for Robust Speech Recognition," *In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE: 2013), 7092–7096, https://doi.org/10.1109/ICASSP.2013.6639038.

6. Z.-Q. Wang, P. Wang, and D. L. Wang, "Complex Spectral Mapping for Single-and Multi-Channel Speech Enhancement and Robust ASR," *IEEE/ACM Transactions on Audio, Speech and Language Processing* 28 (2020): 1778–1787, https://doi.org/10.1109/TASLP.2020.2998279.

7. Y. Luo and N. Mesgarani, "Tasnet: Time-Domain Audio Separation Network for Real-Time, Single-Channel Speech Separation," *In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) (IEEE: 2018), 696–700, https://doi.org/10.1109/ICASSP.2018.8462116.

8. Y. Xu, D. Jun, L.-R. Dai, and C.-H. Lee, "A Regression Approach to Speech Enhancement Based on Deep Neural Networks," *IEEE/ACM Transactions on Audio, Speech and Language Processing* 23, no. 1 (2014): 7–19, https://doi.org/10.1109/TASLP.2014.2364452.

9. Z. Chen, "Noise Reduction of Bird Calls Based on a Combination of Spectral Subtraction, Wiener Filtering, and Kalman Filtering," *In 2022 IEEE 5th International Conference on Automation, Electronics and Electrical Engineering* (*AUTEEE*) (IEEE:2022), 512–517, https://doi.org/10.1109/AUTEEE56487.2022.9994282.

10. S. Liu, M. Geng, H. Shoukang, et al., "Recent Progress in the CUHK Dysarthric Speech Recognition System," *IEEE/ACM Transactions on Audio, Speech and Language Processing* 29 (2021): 2267–2281, https://doi.org/10.1109/TASLP.2021.3091805.

11. F. Xiong, J. Barker, and H. Christensen, "Phonetic Analysis of Dysarthric Speech Tempo and Applications to Robust Personalised

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Dysarthric Speech Recognition," *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) (IEEE: 2019), 5836–5840, https://doi.org/10.1109/ICASSP.2019.8683091.

12. M. M. Morshed and A. O. Ahsan, "Attention-Free Keyword Spotting," *arXiv preprint arXiv:2110.07749* (2021), https://doi.org/10.48550/arXiv.2110.07749.

13. S. Sadhu, R. Li, and H. Hermansky, "M-Vectors: Sub-Band Based Energy Modulation Features for Multi-Stream Automatic Speech Recognition," *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) (IEEE: 2019), 6545–6549, https://doi.org/10.1109/ICASSP.2019.8682710.

14. V. M. Sankari, U. Snekhalatha, M. Murugappan, M. E. H. Chowdhury, and Z. A. Chamkha, "Artificial Intelligence-Based Hearing Loss Detection Using Acoustic Threshold and Speech Perception Level," *Arabian Journal for Science and Engineering* 48 (2023): 1–17, https://doi.org/10.1007/s13369-023-07927-1.

15. P. Das and U. Bhattacharjee, "Robust Speaker Verification Using GFCC and Joint Factor Analysis," *Fifth International Conference on Computing, Communications and Networking Technologies* (*ICCCNT*) (IEEE: 2014), 1–4, https://doi.org/10.1109/ICCCNT.2014.6963092.

16. M. A. Haque, A. Verma, J. S. R. Alex, and N. Venkatesan, "Experimental Evaluation of CNN Architecture for Speech Recognition," in *First International Conference on Sustainable Technologies for Computational Intelligence: Proceedings of ICTSCI 2019* (Singapore: Springer, 2020), 507–514, https://doi.org/10.1007/978-981-15-0029-9_40.

17. X. Shi, H. Yang, and P. Zhou, "Robust Speaker Recognition Based on Improved GFCC," *In 2016 2nd IEEE International Conference on Computer and Communications* (*ICCC*) (IEEE: 2016), 1927–1931, https://doi.org/10.1109/CompComm.2016.7925037.

18. C. Kim and R. M. Stern, "Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition," *IEEE/ACM Transactions on Audio, Speech and Language Processing* 24, no. 7 (2016): 1315–1329, https://doi.org/10.1109/TASLP.2016.2545928.

19. A. Badi, K. Ko, and H. Ko, "Bird Sounds Classification by Combining PNCC and Robust Mel-Log Filter Bank Features," *Journal of the Acoustical Society of Korea* 38, no. 1 (2019): 39–46, https://doi.org/10.7776/ASK.2019.38.1.039.

20. V. Passricha and R. K. Aggarwal, "A Hybrid of Deep CNN and Bidirectional LSTM for Automatic Speech Recognition," *Journal of Intelligent Systems* 29, no. 1 (2019): 1261–1274, https://doi.org/10.1515/jisys-2018-0372.

21. A. Mukhamadiyev, I. Khujayarov, O. Djuraev, and J. Cho, "Automatic Speech Recognition Method Based on Deep Learning Approaches for Uzbek Language," *Sensors* 22, no. 10 (2022): 3683, https://doi.org/10.3390/s22103683.

22. F. Seide, G. Li, X. Chen, and Y. Dong, "Feature Engineering in Context-Dependent Deep Neural Networks for Conversational Speech Transcription," in *In 2011 IEEE Workshop on Automatic Speech Recognition & Understanding* (Waikoloa, HI: IEEE, 2011), 24–29, https://doi.org/10.1109/ASRU.2011.6163899.

23. L. F. Uebel and P. C. Woodland, "An Investigation Into Vocal Tract Length Normalisation," *Sixth European Conference on Speech Communication and Technology* (1999), https://doi.org/10.21437/Eurospeech.1999-553.

24. M. J. F. Gales, "Maximum Likelihood Linear Transformations for HMM-Based Speech Recognition," *Computer Speech & Language* 12, no. 2 (1998): 75–98, https://doi.org/10.1006/csla.1998.0043.

25. S. Mehra, V. Ranga, and R. Agarwal, "A Deep Learning Approach to Dysarthric Utterance Classification With BiLSTM-GRU, Speech Cue Filtering, and Log Mel Spectrograms," *Journal of Supercomputing* 80 (2024): 1–28, https://doi.org/10.1007/s11227-024-06015-x.

26. Q. Zhang, H. Zhang, K. Zhou, and L. Zhang, "Developing a Physiological Signal-Based, Mean Threshold and Decision-Level Fusion Algorithm (PMD) for Emotion Recognition," *Tsinghua Science and Technology* 28, no. 4 (2023): 673–685, https://doi.org/10.26599/TST.2022.9010038.

27. S. Mehra and S. Susan, "Deep Fusion Framework for Speech Command Recognition Using Acoustic and Linguistic Features," *Multimedia Tools and Applications* 82 (2023): 1–25, https://doi.org/10.1007/s11042-023-15118-1.

28. S. Mehra, V. Ranga, R. Agarwal, and S. Susan, "Speaker Independent Recognition of Low-Resourced Multilingual Arabic Spoken Words Through Hybrid Fusion," *Multimedia Tools and Applications* 83 (2024): 1–29, https://doi.org/10.1007/s11042-024-18804-w.

29. R. Das and T. D. Singh, "Image-Text Multimodal Sentiment Analysis Framework of Assamese News Articles Using Late Fusion," *ACM Transactions on Asian and Low-Resource Language Information Processing* 22 (2023): 1–30, https://doi.org/10.1145/3584861.

30. J. Zhu, C. Huang, and P. De Meo, "DFMKE: A Dual Fusion Multi-Modal Knowledge Graph Embedding Framework for Entity Alignment," *Information Fusion* 90 (2023): 111–119, https://doi.org/10.1016/j.inffus.2022.09.012.

31. S. Mehra, V. Ranga, and R. Agarwal, "Improving Speech Command Recognition Through Decision-Level Fusion of Deep Filtered Speech Cues," *Signal, Image and Video Processing* 18, no. 2 (2024): 1365–1373, https://doi.org/10.1007/s11760-023-02845-z.

32. C. Wang, Y. Tang, X. Ma, et al., "Fairseq S2T: Fast Speech-to-Text Modeling With Fairseq," *arXiv preprint arXiv:2010.05171* (2020), https://doi.org/10.48550/arXiv.2010.05171.

33. J. Duchi, E. Hazan, and Y. Singer, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," *Journal of Machine Learning Research* 12 (2011): 2121–2159, https://doi.org/10.5555/1953048.2021068.

34. R. Vygon and N. Mikhaylovskiy, "Learning Efficient Representations for Keyword Spotting With Triplet Loss," In Speech and Computer: 23rd International Conference, SPECOM 2021, St. Petersburg, Russia, September 27–30, 2021, Proceedings 23, pp. 773–785. Springer International Publishing, 2021, https://doi.org/10.1007/978-3-030-87802-3_69.

35. A. Berg, M. O'Connor, and M. T. Cruz, "Keyword Transformer: A Self-Attention Model for Keyword Spotting," *arXiv preprint arXiv:2104.00769* (2021), https://doi.org/10.21437/Interspeech.2021-1286.

36. D. C. De Andrade, S. Leo, M. L. D. S. Viana, and C. Bernkopf, "A Neural Attention Model for Speech Command Recognition," *arXiv preprint arXiv:1808.08929* (2018), https://doi.org/10.48550/arXiv.1808.08929.

37. D. Niizumi, D. Takeuchi, Y. Ohishi, N. Harada, and K. Kashino, "Masked Modeling Duo: Learning Representations by Encouraging Both Networks to Model the Input," *2023 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) (IEEE: 2023), 1–5, https://doi.org/10.1109/ICASSP49357.2023.10097236.

38. A. Gazneli, G. Zimerman, T. Ridnik, G. Sharir, and A. Noy, "End-to-End Audio Strikes Back: Boosting Augmentations Towards an Efficient Audio Classification Network," *arXiv preprint arXiv:2204.11479* (2022), https://doi.org/10.48550/arXiv.2204.11479.

39. Y. Gong, Y.-A. Chung, and J. Glass, "Ast: Audio Spectrogram Transformer," *arXiv preprint arXiv:2104.01778* (2021), https://doi.org/10.48550/arXiv.2104.01778.

40. K. Chen, D. Xingjian, B. Zhu, Z. Ma, T. Berg-Kirkpatrick, and S. Dubnov, "HTS-AT: A Hierarchical Token-Semantic Audio Transformer for Sound Classification and Detection," *ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) (IEEE: 2022), 649–650, https://doi.org/10.48550/arXiv.2202.00874.

41. V. A. Trinh, H. S. Kavaki, and M. I. Mandel, "Importantaug: A Data Augmentation Agent for Speech," *ICASSP 2022–2022 IEEE International*

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Conference on Acoustics, Speech and Signal Processing (*ICASSP*) (IEEE: 2022), 8592–8596, https://doi.org/10.1109/ICASSP43922.2022.9747003.

42. E. R. Abdelmaksoud, A. Hassen, N. Hassan, and M. Hesham, "Convolutional Neural Network for Arabic Speech Recognition," *Egyptian Journal of Language Engineering* 8, no. 1 (2021): 27–38, https://doi.org/10.21608/ejle.2020.47685.1015.

43. A. S. Wazir, M. Ba, and J. H. Chuah, "Spoken Arabic Digits Recognition Using Deep Learning," *In 2019 IEEE International Conference on Automatic Control and Intelligent Systems* (*I2CACIS*) (IEEE: 2019), 339–344, https://doi.org/10.1109/I2CACIS.2019.8825004.

44. T. Zia and U. Zahid, "Long Short-Term Memory Recurrent Neural Network Architectures for Urdu Acoustic Modeling," *International Journal of Speech Technology* 22 (2019): 21–30, https://doi.org/10.1007/s10772-018-09573-7.

45. I. Lezhenin, N. Bogach, and E. Pyshkin, "Urban Sound Classification Using Long Short-Term Memory Neural Network," *In 2019 Federated Conference on Computer Science and Information Systems* (*FedCSIS*) (2019), 57–60, https://doi.org/10.15439/2019F185.

46. M. Zeng and N. Xiao, "Effective Combination of DenseNet and BiLSTM for Keyword Spotting," *IEEE Access* 7 (2019): 10767–10775, https://doi.org/10.1109/ACCESS.2019.2891838.

47. Y. Wei, Z. Gong, S. Yang, K. Ye, and Y. Wen, "EdgeCRNN: An Edge-Computing Oriented Model of Acoustic Feature Enhancement for Keyword Spotting," *Journal of Ambient Intelligence and Humanized Computing* 13 (2022): 1–11, https://doi.org/10.1007/s12652-021-03022-1.

48. L. Cances and T. Pellegrini, "Comparison of Deep co-Training and Mean-Teacher Approaches for Semi-Supervised Audio Tagging," *ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) (IEEE: 2021), 316–365.

49. B. Higy and P. Bell, "Few-Shot Learning With Attention-Based Sequence-to-Sequence Models," *arXiv preprint arXiv:1811.03519* (2018), https://doi.org/10.48550/arXiv.1811.03519.

50. B. Kim, S. Chang, J. Lee, and D. Sung, "Broadcasted Residual Learning for Efficient Keyword Spotting," *arXiv preprint arXiv:2106.04140* (2021).

51. S. Majumdar and B. Ginsburg, "Matchboxnet: 1d Time-Channel Separable Convolutional Neural Network Architecture for Speech Commands Recognition," *arXiv preprint arXiv:2004.08531* (2020), https://doi.org/10.21437/Interspeech.2020-1058.

52. D. Ng, Y. Chen, B. Tian, Q. Fu, and E. S. Chng, "Convmixer: Feature Interactive Convolution With Curriculum Learning for Small Footprint and Noisy Far-Field Keyword Spotting," *ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) (IEEE: 2022), 3603–3607.

53. J. Lin, K. Kilgour, D. Roblek, and M. Sharifi, "Training Keyword Spotters With Limited and Synthesized Speech Data," *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) (IEEE: 2020), 7474–7478, https://doi.org/10.48550/arXiv.2002.01322.

54. D. Seo, O. Heung-Seon, and Y. Jung, "Wav2kws: Transfer Learning From Speech Representations for Keyword Spotting," *IEEE Access* 9 (2021): 80682–80691, https://doi.org/10.1109/ACCESS.2021.3078715.

55. D. Singh and R. Chandel, "FPGA-Based Hardware-Accelerated Design of Linear Prediction Analysis for Real-Time Speech Signal," *Arabian Journal for Science and Engineering* 48 (2023): 1–15, https://doi.org/10.1007/s13369-023-07926-2.

56. A. S. Dhanjal and W. Singh, "A Comprehensive Survey on Automatic Speech Recognition Using Neural Networks," *Multimedia Tools and Applications* 83, no. 8 (2024): 23367–23412, https://doi.org/10.1007/s11042-023-16438-y.

57. J. Zhang, C. Li, Y. Yin, J. Zhang, and M. Grzegorzek, "Applications of Artificial Neural Networks in Microorganism Image Analysis: A Comprehensive Review From Conventional Multilayer Perceptron to Popular Convolutional Neural Network and Potential Visual Transformer," *Artificial Intelligence Review* 56, no. 2 (2023): 1013–1070, https://doi.org/10.1007/s10462-022-10192-7.

58. X. Li, M. Chen Li, M. Rahaman, et al., "A Comprehensive Review of Computer-Aided Whole-Slide Image Analysis: From Datasets to Feature Extraction, Segmentation, Classification and Detection Approaches," *Artificial Intelligence Review* 55, no. 6 (2022): 4809–4878, https://doi.org/10.1007/s10462-021-10121-0.

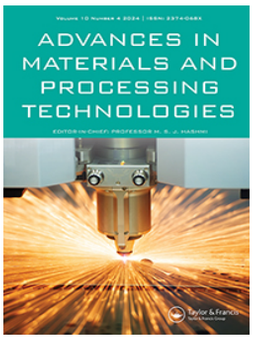
59. W. Liu, C. Li, X. Ning, et al., "CVM-Cervix: A Hybrid Cervical Pap-Smear Image Classification Framework Using CNN, Visual Transformer and Multilayer Perceptron," *Pattern Recognition* 130 (2022): 108829, https://doi.org/10.1016/j.patcog.2022.108829.

60. Z. Fan, W. Xiangchen, C. Li, et al., "CAM-VT: A Weakly Supervised Cervical Cancer Nest Image Identification Approach Using Conjugated Attention Mechanism and Visual Transformer," *Computers in Biology and Medicine* 162 (2023): 107070, https://doi.org/10.1016/j.compbiomed.2023.107070.

61. A. Chen, C. Li, M. Shuojia Zou, et al., "SVIA Dataset: A New Dataset of Microscopic Videos and Images for Computer-Aided Sperm Analysis," *Biocybernetics and Biomedical Engineering* 42, no. 1 (2022): 204–214, https://doi.org/10.1016/j.bbe.2021.12.010.

62. Q. Nie, C. Li, J. Yang, et al., "OII-DS: A Benchmark Oral Implant Image Dataset for Object Detection and Image Classification Evaluation," *Computers in Biology and Medicine* 167 (2023): 107620, https://doi.org/10.1016/j.compbiomed.2023.107620.

14678640, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coin.70012 by Delhi Technological University, Wiley Online Library on [25/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



Advances in Materials and Processing Technologies

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tmpt20

Numerical optimization of cold gas dynamic spray process parameters through response surface methodology

Chirag Singhal & Qasim Murtaza

To cite this article: Chirag Singhal & Qasim Murtaza (11 Dec 2024): Numerical optimization of cold gas dynamic spray process parameters through response surface methodology, Advances in Materials and Processing Technologies, DOI: 10.1080/2374068X.2024.2439710

To link to this article: https://doi.org/10.1080/2374068X.2024.2439710

Published online: 11 Dec 2024.

Submit your article to this journal

Article views: 14

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tmpt20

ADVANCES IN MATERIALS AND PROCESSING TECHNOLOGIES
https://doi.org/10.1080/2374068X.2024.2439710

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Check for updates

Numerical optimization of cold gas dynamic spray process parameters through response surface methodology

Chirag Singhal[a] and Qasim Murtaza[b]

[a]Department of Mechanical Engineering, ZHCET, AMU, Aligarh, India; [b]Department of Mechanical and Production & Industrial Engineering, Delhi Technological University, Delhi, India

ABSTRACT

Copper-Zinc and its alloys are extensively used for the corrosion protection of the metal substrate surfaces like steel. Cold gas dynamic spraying (CGDS) is a material coating technique in which metal particles directly adhere to the substrate surface at a relatively low temperature. Numerical simulation (CFD) is a better alternative to understanding CGDS as compared to experimental data due to less expenditure and less energy consumption in CFD. Optimisation of CGDS can be done either by varying the process parameters or by improving the design of the de-Laval nozzle. In this study, the Response Surface Methodology (Central Composite Design) was used to optimise cold spray deposition efficiency by using a 2-D axisymmetric model by varying the process parameters such as carrier gas temperature, stand-off distance, and particle size. Various models were used to predict the critical velocity and erosion velocity. All the process parameters under consideration were found significant in analysing the model by ANOVA. Copper-Zinc Alloy (C93200) particle size is varied from 2.96 µm to 87.04 µm and carrier gas temperature is varied from 287.78K to 708.22K. With the increase in carrier gas temperature and particle size diameter, critical velocity decreases resulting in improved deposition efficiency.

ARTICLE HISTORY
Received 6 April 2024
Accepted 4 December 2024

KEYWORDS
Cold spray; critical velocity; erosion velocity; deposition efficiency; optimisation; coating

1. Introduction

The coatings of copper-zinc and its alloys are very widely used for the corrosion resistance of the steel structure and degradation resistance in a moist environment. Cold dynamic gas spraying is a pretty novel emerging spray technique in which the feedstock particles of materials are accelerated to a high speed ranging from 300–1200 m/s with the help of a supersonic de-Laval nozzle. The sprayed particles get deposited with the substrate due to plastic deformation [1–3]. This coating is done at a relatively lower temperature as compared to thermal spraying so that the powder particles remain below their melting temperature. In cold spraying, the deposition of the material particle depends upon the kinetic energy of the accelerated particle and the plasticity of the substrate rather than their thermal energy. Coatings obtained by cold spray are oxygen-free metallic coatings because of the relatively lower temperature of particles. It also

CONTACT Qasim Murtaza csamumech@gmail.com; qasimmurtaza@dce.ac.in Department of Mechanical and Production & Industrial Engineering, Delhi Technological University, Delhi 110042, India

© 2024 Informa UK Limited, trading as Taylor & Francis Group

ensures that the powder feedstock retains its original properties. The cold spray coating technique reduces the problem of melting material particles. Due to the advantage of low operating temperature, the oxygen-sensitive materials can also be coated easily using the cold spray. Therefore, CGDS has gained so much attention due to its advantages over other thermal spraying techniques [4–8].

In cold spray, the deposition of the particles depends upon the nozzle design, throat area and process parameters. Critical velocity and temperature of the particles are the deciding factors for the deposition of the particles which depends on the material properties of the substrate [7,8]. Critical velocity is a threshold value that suggests the deposition of the material particles. The velocity of the impacting particle is also an important parameter for the evolution of the coating [9,10] and is a function of carrier gas species, carrier gas temperature, the pressure of injection of the particles and stand-off distance of the substrate from the nozzle exit [11–14].

The simulation of the cold gas dynamic spray has been going on extensively by researchers to predict the critical velocity of particles [11–22]. Li et al. changed the stand-off distance, decreasing the temperature of the main gas and substrate diameter [4]. The maximum gas temperature was found at the throat section of the de-Laval nozzle. It was concluded that the temperature gradient remains almost the same on decreasing the inlet temperature and increasing the stand-off distance although the temperature of the substrate surface went on decreasing steadily [4]. Suo et al. varied the nozzle expansion ratio, the diameter of the nozzle inlet, the dimensions of the nozzle injector, and also the converging and diverging lengths of the de-Laval nozzle [1]. It was found that a change in the height of the injector affected the distribution of particles and as the nozzle height increased the distribution of particles broadens, flattens in the lateral direction and also the particle mean velocity was affected. An increase in the throat height led to a decrease in the particle mean velocity along the lateral direction [1]. Wang et al. modified the ratio of nozzle exit radius (r) to substrate radius (R) and substrate thickness [9]. It was seen that when the r/R ratio was less than 1, high gas velocities did not impact fully on the substrate and part of the gases passed near the substrate edges. They further reported that the bow shock thickness was relatively small in front of the substrate surface which directly led to the copper particle acceleration and the copper particle impact. As the r/R increased, the thickness of the bow shock went on increasing and at a limiting value, the bow shock became an independent function of the r/R ratio. The temperature of the substrate showed a decreasing trend with an increase in the substrate thickness [9].

Numerical simulation is reported to have various advantages over experimental processing as it is easier and more economical. Presently reported work confirms that the computational fluid dynamics (CFD) results are comparable to the experimental observations such as particle acceleration and flow of gaseous species in the cold spray [23–25]. In the numerical simulation, process parameters can be changed easily according to our needs without investing anything. Improvements in the computational power are being done so the modelling of cold spray can be simulated in very less time.

Therefore, the present work aims to provide a numerical solution of Cold Spray deposition by varying its process parameters such as carrier gas temperature, stand-off distance and particle size. To optimise the deposition efficiency of Cold Spray particles, Response Surface Methodology (Central Composite Design) model is used. This statistically optimised process parameters by response surface methodology (central

composite design) reduce the cost of experimentation. Through this work, we are in a position to suggest process parameters of Cold Spray system that can be used to get maximum deposition efficiency of coating particles in the experimental environment using the simulated observations. The results are demonstrated by a case study of Cold Spray coating carried out on the circular substrate of steel using the Copper-Zinc Alloy (C93200) spherical solid particles in 2-D axisymmetric model.

2. Numerical methodology

2.1. Geometry of the domain

The schematic plan and dimensions (Table 1) of the CGDS nozzles are shown in Figure 1. Boundary conditions applied to the computational domain are shown in Figure 2 and the operating conditions are tabulated in Table 2. The nozzle had circular cross-section with an inlet diameter and outlet diameter of 18.2 mm and 6 mm respectively. The flat substrate plate of steel (Table 3) was circular in shape with a diameter of 100 mm and a thickness of 6 mm. Computational domain was taken as the 2D axisymmetric to save computational time. The computational domain was meshed into 132,334 quadrilateral cells. The grid at near wall regions was refined to capture the viscosity effects in affected region. The heat transfer between the gas and the nozzle wall was not taken into consideration. The substrate surface was taken as the coupled interface between the gas and solid phase to consider the fluid-solid coupling.

Table 1. Dimensions of MOC nozzle.

Configuration	Dimensions (mm)
Inlet	18.2
Barrel length	10
Outlet diameter	6
Injector length	25
Injector diameter	2
Throat diameter	2.7
Convergent length	54
Divergent length	170
Substrate diameter	100
Substrate thickness	6

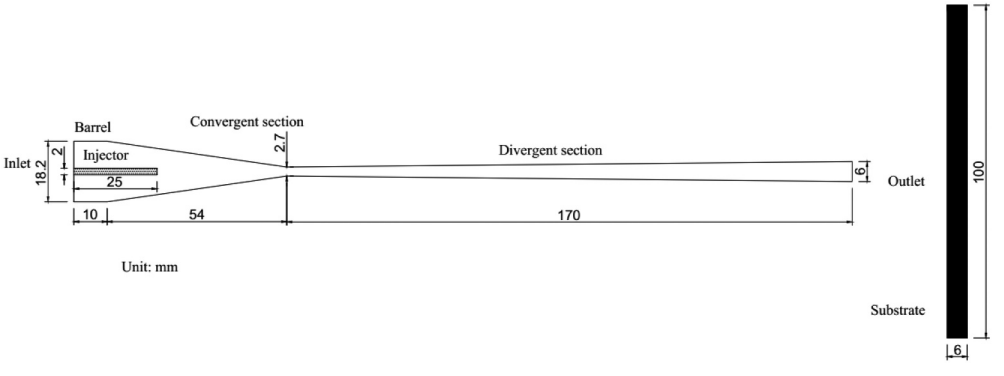


Figure 1. Schematic of the nozzle and the substrate model used in simulation.

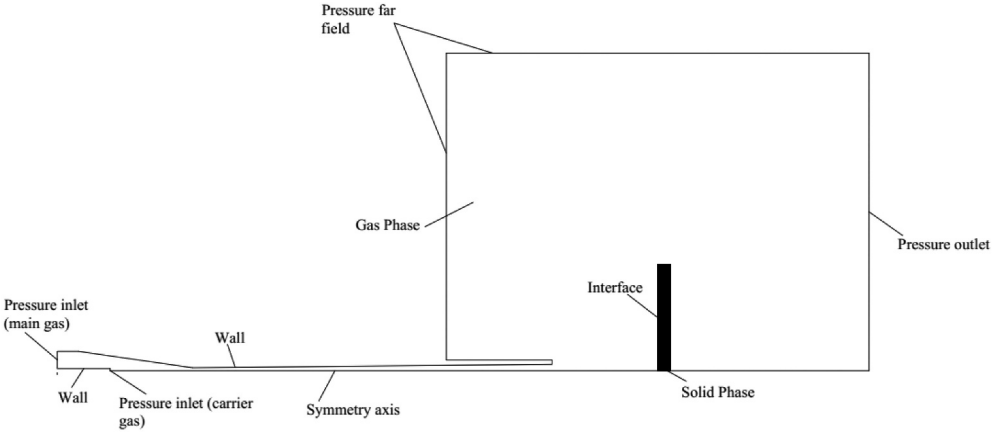


Figure 2. Schematic of the boundary conditions applied to the model used in simulation.

Table 2. Operating parameters in simulation.

	Dimensions
Gas Inlet pressure	3.0 MPa
Injection pressure	3.1 MPa
Gas inlet temperature	923 K
Ambient pressure	0.101 MPa

Table 3. Properties of the powder particle and substrate used in simulation.

Property	Copper-Zinc Alloy (C93200)	Substrate Material Steel
Density (kg/m^3)	8910	8030
Specific heat density (J kg^{-1} K^{-1})	377.1	502.48
Melting Point (K)	1127K	
Thermal conductivity (W/m K)		16.27

2.2. Gas phase and solid phase

Nitrogen (Table 4) was chosen as a carrier gas and main gas because it is an inert gas, it is freely available in the environment that can be processed very easily and is inexpensive. The carrier gas is governed by the ideal gas law which takes care of the compressibility effects. The two-dimensional steady compressible flow is governed by the equations which are written in tensor equation:

Continuity equation:

Table 4. Properties of the gas used in simulation.

Property	Nitrogen (N_2)
Thermal conductivity	0.0242 W/m-K
Specific heat capacity	1040.67 J/kg-K
Viscosity	1.663e-05 kg/m-s
Molecular Weight	28 kg kmol^{-1}

$$\frac{\partial(\rho u_i)}{\partial x} = 0 \tag{1}$$

Momentum equation:

$$\frac{\partial(\rho u_i u_j)}{\partial x} = -\frac{\partial P}{\partial x} + \frac{\partial(\tau_{ij})}{\partial x} \tag{2}$$

Energy equation:

$$\frac{\partial(\rho e u_i)}{\partial x_i} = -\frac{\partial P u_i}{\partial x_i} + \frac{\partial(u_j \tau_{ij} - q_i)}{\partial x_i} \tag{3}$$

where ρ,u,e,P,T,τ,q are the gas density, velocity, internal energy, pressure, temperature, viscous shear stress and heat flux respectively. Pressure-based solver was used with realisable k-ε turbulence model to capture the turbulence in the flow. The realisable k-ε model predicts more accurately the spreading rate of the jets and dissipation rate ε and is predicted for the transport of the mean-square vorticity fluctuations as compared to the standard k-ε turbulence model. The non-equilibrium wall function was used for the near-wall flow treatment due to its potential to deal with the large pressure gradient in the impinging jet.

In the simulation of CGDS, Copper-Zinc alloy (C93200) particles were used as the powder and the particles were considered spherical in nature (Table 3). The mass flow rate of discrete phase was taken as 0.0005 kg/sec and the particles were injected with a velocity of 30 m/sec. To compute particle motion, discrete phase modelling (DPM) was used which is available in FLUENT. The interaction between the particles was neglected. The high Mach number law is used to compute the particle drag force.

$$\frac{du_p}{dt} = F_d(u - u_p) + \frac{g_x(\rho_p - \rho)}{\rho_p} + F_x \tag{4}$$

where u is the gas velocity, $u_p$is the particle velocity, ρ_p is the particle density, g_x is the x component of inertia coefficient, F_x is an additional acceleration (force/unit particle mass) term, $[g_x(u - u_p)]/\rho_p$is inertia force per unit particle mass, $F_d(u - u_p)$ is the drag force per unit particle mass and F_d can be expressed as:

$$F_d = \frac{3\rho}{4\rho_p d_p} C_d |u - u_p| \tag{5}$$

where $d_p$is the particle diameter, C_d is the drag coefficient. The Stochastic-Tracking type model was also implemented in Fluent to capture the effects of turbulence. In this approach, the Discrete Random Walk (DRW) was also used to predict the fluctuations in the total particle velocity and it's the particle trajectory.

3. Prediction critical velocities, erosion velocity and deposition efficiency

Critical velocity and the erosion velocity are the deciding factor for the adherence of the powder particles. To predict critical velocity (Eq 6) Assadi [17] has given a model

$$CV1 = V_{cr} = V_{cr}^{ref} \frac{0.42(d_p/d_p^{ref})^{0.5}(1 - T_p/T_m)^{\frac{1}{2}} + 1.19(1 - 0.73T_p/T_m)^{\frac{1}{2}}}{0.65 + (d_p/d_p^{ref})^{0.5}} \quad (6)$$

In equation 6, V_{cr} is the critical velocity for copper, $T_p$is a particle temperature of copper, T_m is a copper particle melting temperature, d_p is diameter of copper particle, $d_p^{ref}$is a reference particle diameter (10 μm for copper) and $V_{cr}^{ref}$is 650 m/s of 10 μm copper particle as a reference critical velocity. In this model, the particle injecting temperature, particles dimensions and melting point of the particle are taken care of. However, Assadi (eq 6) model has a disadvantage that a predefined critical velocity is only for copper particles deposition and its effectiveness is confined to the materials whose properties are comparable to copper. So, for any other material an experimental data is needed.

For incorporating all these parameters, Moridi [26] has given a hybrid approach to predict the critical (Eq 7) and the erosion velocities (Eq 8) for the material powder particles. In this model, both the Assadi and the Moridi models are incorporated into a single equation.

$$CV2 = V_{cr}^{Cu} = (707 + 90\ln(WP)) \frac{0.42(d_p/d_p^{ref})^{0.5}(1 - T_p/T_m)^{\frac{1}{2}} + 1.19(1 - 0.73T_p/T_m)^{\frac{1}{2}}}{0.65 + (d_p/d_p^{ref})^{0.5}} \quad (7)$$

$$EV = V_{erosion}^{Cu} = (1364 + 218\ln(WP)) \frac{0.42(d_p/d_p^{ref})^{0.5}(1 - T_p/T_m)^{\frac{1}{2}} + 1.19(1 - 0.73T_p/T_m)^{\frac{1}{2}}}{0.65 + (d_p/d_p^{ref})^{0.5}} \quad (8)$$

where WP is weight porosity, $d_p^{ref}$is a reference particle diameter (25 μm for copper). In this model, the shortcoming of Assadi model, i.e. a critical reference velocity, is removed by using the weight porosity of the copper particles. Weight porosity is the porosity caused in the material particles when the particles are prepared through gas atomisation. The weight porosity is a material property and its data is easily available. In the current analysis, weight porosity was taken as 0.25. Secondly, hybrid model (eq 7, 8) can also be easily used to calculate the critical velocity and the erosion velocity for very low powder particle size.

Deposition efficiency is the efficiency which gives a quantitative analysis about the deposition of the cold spray metal particles on the substrate.

$$DE = \int_0^\infty tf(d_p)d(d_p) \quad (9)$$

where $d(d_p)$ is the differential of $d_p$and $t$is a size-dependent function, which is calculated in terms of the characteristics of the particle velocities as follows:

t = 1if v_{cr}<v_{pi}<$v_{erosion}$ (9(b))

0if v_{pi}<v_{cr} or v_{pi}>$v_{erosion}$

where $v_{erosion}$ is the upper value of the velocity for the particle impact, further than there would be erosion of the substrate surface and no deposition, v_{pi} is the particle impact velocity before 1 mm from substrate surface in the axial direction.

4. Predictive Model for responses

In the current study, to predict the responses a statistical technique, response surface methodology, (RSM) was used. RSM predictive model was used to predict the deposition efficiency of CGDS by varying the process parameters. RSM is an economical, practical and relatively easy to use technique for prediction of experiments. To correlate and simulate the cold spray process parameters and the deposition responses of coating, a second-order quadratic model was selected to forecast the responses based on the simulated values in the analysis. The responses are the function of particle size diameter(A), stand-off distance(B), carrier gas temperature(C), and its function can be expressed as follows:

$$Responses = f(A, B, C) \tag{10}$$

The model used here inculcates the effects of the main cold spray process factors and its interaction. The prediction model used here is polynomial in nature which can be written as follows:

$$R = j_0 + j_1(A) + j_2(B) + j_3(C) + j_{11}(A^2) + j_{22}(B^2) + j_{33}(C^2) + j_{13}(AC) + j_{12}(AB) + j_{23}(BC) \tag{11}$$

where $j_0$is the average of responses, and j_1,j_{22},. . .,j_{12} are the coefficients which depend upon their respective main and interaction effects of the cold spray process parameters. To estimate the regression coefficients of cold spray process parameters, using central composite rotatable design approach was used. The regression coefficients were calculated using the Design Expert V 10.0 statistical software and the final regression model to estimate the deposition responses is given below:

Deposition Efficiency (%)

$$\begin{aligned} &= 255.45847 - 1.47311(A) - 0.94231(B) - 0.57075(C) + \\ &0.013466(A^2) + 0.00993915(B^2) + 0.000560147(C^2) + 0.0002732(AC) + \\ &0.00385667(AB) + 0.000212667(BC) \end{aligned} \tag{12}$$

To validate the adequacy of the regression model developed, empirical relationship analysis of variance (ANOVA) was used [27]. In ANOVA technique, the model F value and the related probability values were analysed to check the significance of the empirical regression relationships.

5. Results and discussion

5.1. Flow analysis

The temperature profile of steel and aluminium for stand-off distance 50 mm along the axial direction is shown in Figure 3. On applying the same process parameters and numerical methodology, taken by Yin et al., well-validated results were obtained with a very less variation on comparing the reported result [24]. In Figure 4, carrier gas temperature contour is shown for the defined CGDS nozzle along the carrier gas flow direction at different carrier gas temperature and operating process parameters are used as per the run 2,6,13 conditions tabulated in Table 5. Carrier gas temperature

significantly affects the characteristics of gas flow field inside the nozzle as well as outside the nozzle as shown in Figure 4. On decreasing the carrier gas temperature, interaction between the carrier gas and the main gas increases. The interaction of both gases results in the momentum and heat exchange at the lower temperature causing the reduction in the temperature of the driving gas inside the nozzle. Significant effect of the carrier gas interaction is seen in the convergent section of the nozzle. The decrease in the temperature of driving gas results in the decrease of temperature on the substrate centre at the stagnation point as observed from Figure 4(a–c). The temperature drop results in the decrease of deposition efficiency and bonding between the alloy particles inside the coating [28–32]. Temperature variation is clearly seen at the centre line region due to the interaction of the main gas and carrier gas as the injector is located on the central line axis. At lower carrier gas temperature, the driving gas is almost heterogeneous but the heterogeneity of the carrier gas and the main gas changes to homogeneity with an increase in temperature of the carrier gas. In Figure 5, temperature profile along the axial direction for substrate at different stand-off distance is shown and operating process parameters are used as per the run 3, 5, 6, 16 conditions tabulated in Table 5. The temperature profile (Figure 5) confirms the interaction of carrier gas and the main gas in terms of temperature. The temperature rises in the convergent section but it falls rapidly near the throat and then reduces smoothly upto the substrate because near the substrate, whole kinetic energy gets converted into thermal energy and results in the increase of the temperature due to the creation of bow shock. So evident from Figure 5 that the higher carrier gas results in the higher driving gas temperature.

The velocity profiles are shown at the different carrier gas temperature at constant driving gas inlet temperature in Figure 6 and operating process parameters are used as

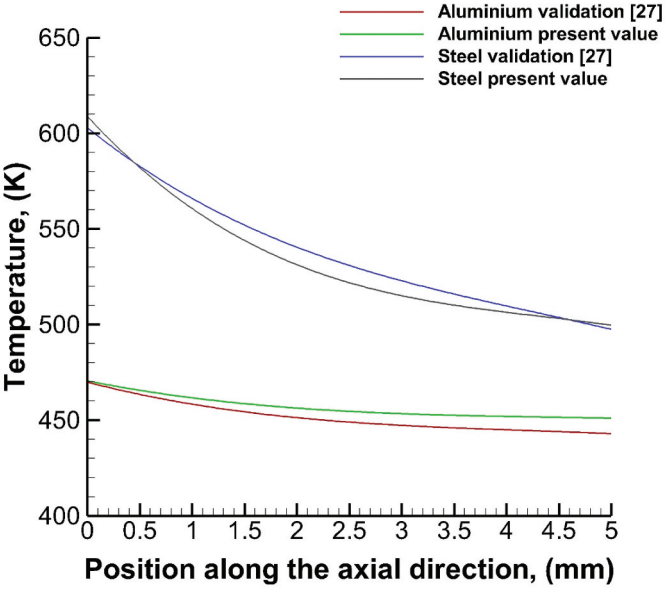


Figure 3. Validated results of temperature profile along the axial direction for stand-off distance 50 mm [24].

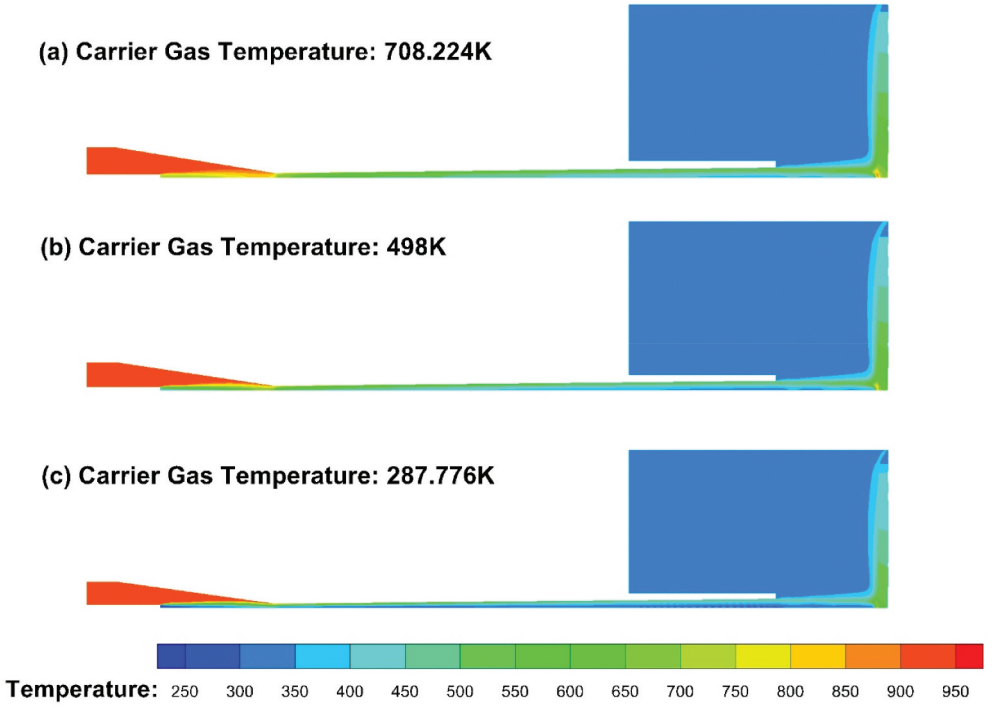


Figure 4. Temperature contour at different carrier gas temperatures for run 2, 6, 13.

Table 5. Operating process parameters used in simulation.

Run	Factor 1 A:Particle size diameter (μm)	Factor 2 B: Stand-off distance (mm)	Factor 3 C: Carrier gas temperature (K)	Critical Velocity Assadi, CV1 (m/s)	Critical Velocity Assadi +Moridi, CV2 (m/s)	Erosion Velocity Assadi +Moridi, EV (m/s)	Deposition Efficiency, DE1(%) (CV1+EV)	Response1 Deposition Efficiency, DE2(%) (CV2+EV)
1	45	60.23	498	381.84	406.5	741.19	66.67	68.16
2	87.04	35	498	341.96	362.2	665.22	88.89	87.21
3	45	9.77	498	389.14	414.23	740.21	75.00	73.21
4	45	35	498	387.13	407.2	750.19	66.67	68.23
5	45	35	498	391.44	412.39	744.19	66.67	65.12
6	45	35	708.22	330.12	356.35	645.66	88.89	89.45
7	70	20	373	381.65	406.27	732.42	90.00	82.05
8	20	50	373	482.61	510.82	927.41	90.00	85.12
9	70	50	623	325.02	351.90	638.2	95.23	93.12
10	70	50	373	386.45	400.47	731.22	93.14	89.12
11	20	20	623	416.05	444.68	812.25	87.5	85.16
12	70	20	623	324.42	346.10	639.5	88.89	86.79
13	45	35	498	383.04	405.23	741.29	66.67	68.16
14	20	20	373	482.12	502.12	912.46	90.00	86.17
15	20	50	623	412.2	438.08	809.95	90.00	88.04
16	45	35	287.78	430.75	455.94	837.27	90.00	88.78
17	45	35	498	382.14	410.89	737.19	66.67	69.48
18	45	35	498	383.44	404.09	753.24	66.67	63.76
19	45	35	498	388.54	413.89	751.64	66.67	63.23
20	2.96	35	498	625.15	637.82	1154.1	90.00	89.12

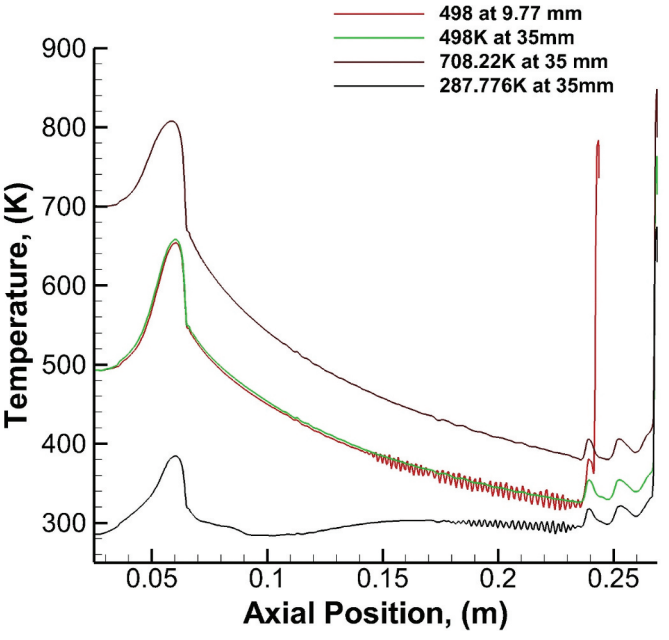


Figure 5. Temperature profile at different carrier gas temperature in axial direction at the different stand-off distance for run 3, 5, 6, 16.

per the run 2, 6, 13 conditions tabulated in Table 5. The velocity of the driving gas decreases with the decrease in the carrier gas temperature due to the interaction of main gas and the carrier gas. The conversion of pressure energy to the velocity energy takes place near the throat. As an outcome, driving gas velocity increases very rapidly in the throat region and then increases smoothly in the divergent section upto the substrate. Near the substrate, there is creation of the bow shock. Bow shock is the deciding factor, which defines the acceleration and the impact velocity of the alloy particle (C93200).

Figure 7 gives the quantitative data of the velocities at different temperature of the carrier gas in the axial direction over the central line and operating process parameters are used as per the run 3, 5, 6, 16 conditions tabulated in the Table 5. The fluctuating component velocity prediction done by the DRW is clearly shown in the runs at different conditions [32–41]. This results in more accurate turbulence incorporation of the gas particles with the powder particles.

Both Assadi (eq 6) and (Moridi + Assadi) (eq 7) approach were used to predict the critical velocity in the current analysis. For predicting erosion velocity, (Moridi + Assadi) (eq 8) was used. The deposition efficiency was calculated by using both the critical velocity and the same erosion velocity. Prediction of deposition efficiency can be done more accurately and precisely by using these models.

5.2. Statistical analysis

In the current study, Rosin-Rammler particle distribution was seen to get more realistic results. The particles were considered with distribution in the range of $\pm 0.1 d_p$ from mean

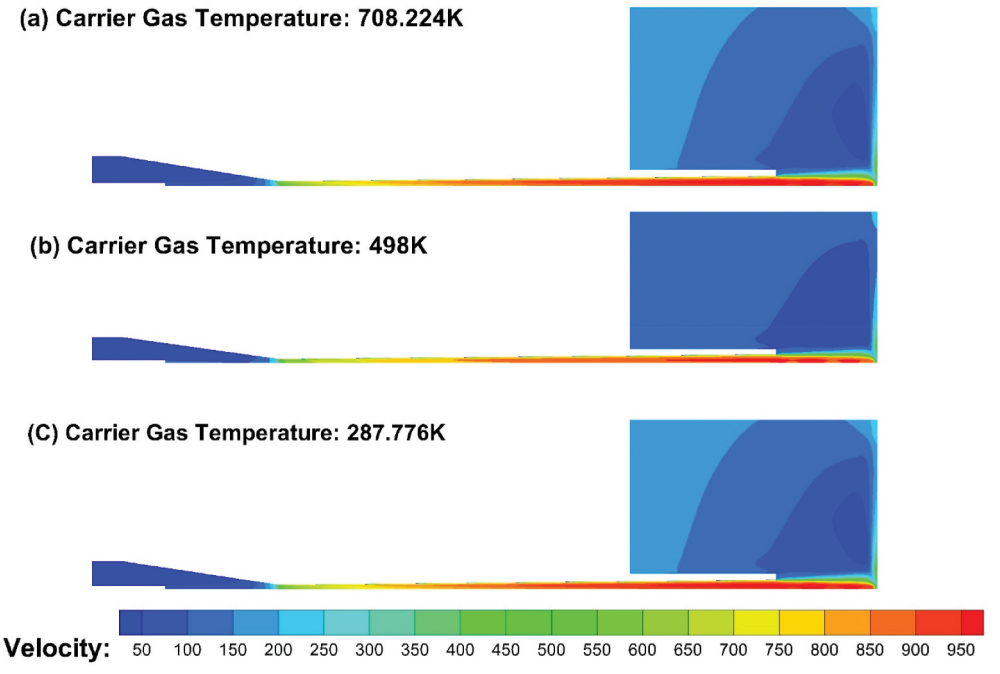


Figure 6. Velocity contour at different carrier gas temperatures contour at different carrier gas temperatures run 2, 6, 13.

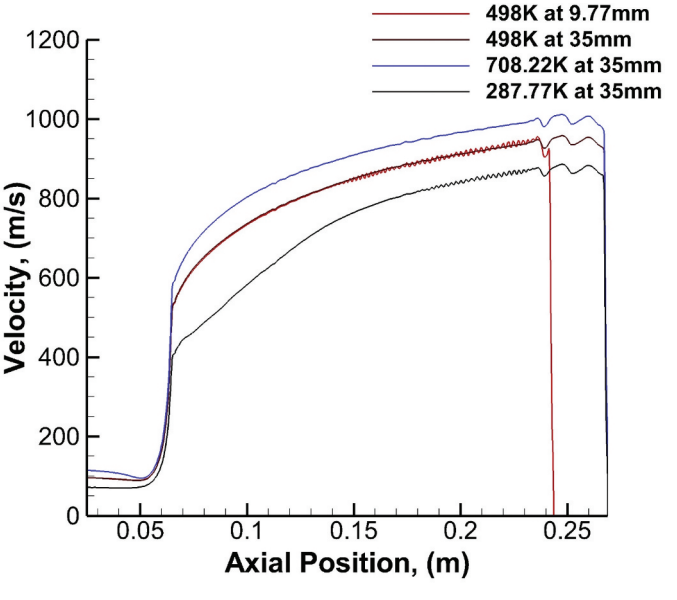


Figure 7. Velocity profile at different carrier gas temperature in axial direction at the different stand-off distance run 3, 5, 6, 16.

particle size (d_p) to capture feature of particle distribution and moving close towards the experimental environment. In the present work optimization of the process parameters CGDS which are carrier gas temperature, stand-off distance and particle size diameter is done based on reviewing the literature data which suggests that these parameters are significant. In this optimization, Central Composite Design of Response surface methodology was chosen to make the runs in the statistical order [42]. RSM central composite design technique, with reduced number of runs from 27 to 20 runs, was designed by this method with the help of Design Expert Software V 10. The respective responses were calculated and designed runs are tabulated in Table 5. Statistical analysis of main and the interaction effects was quantified using an ANOVA test. The lower limits for the particle size diameter (A), stand-off distance (B) and carrier gas temperature (C) are 20 µm, 20 mm and 373 K, and higher limits are 70 µm, 50 mm and 623 K, respectively.

The ANOVA analysis of deposition efficiency (Table 6) shows that the model is significant as the p-value obtained is < 0.0001 and the lack of fit is also non-significant. ANOVA shows that the A^2, B^2, C^2 are significant. The interaction of stand-off distance and the particle size diameter (Figure 8) with deposition efficiency (%) is shown in the surface plot. At the constant carrier gas temperature (498K) the deposition efficiency is minimum at the particle size 45 µm and stand-off distance 35 mm because in this range powder particles are unable to achieve the critical velocity. This may have occurred due to the deviation of the powder particles from their path which are unable to achieve the critical velocity. The deposition efficiency is almost maximum at all the four corners of the plot because when the powder particle size and stand-off distance are least and critical velocity is more, the powder particles achieve velocity more than from it. Secondly, both the parameters are significant. As a result we get more deposition efficiency [28,42]. But, in the case when the powder particle size and stand-off distance are maximum the critical velocity is minimum (eq 6-7) the resulting deposition efficiency is maximum.

Table 6. ANOVA for the deposition efficienc.

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F	
Model	1978.07	9	219.79	18.14	<0.0001	Significant
A-Particle size diameter	0.84	1	0.84	0.069	0.7982	
B-Standoff distance	3.32	1	3.32	0.27	0.6119	
C-Carrier gas temperature	10.16	1	10.16	0.84	0.3815	
AB	16.73	1	16.73	1.38	0.2672	
AC	5.83	1	5.83	0.48	0.5037	
BC	1.27	1	1.27	0.10	0.7526	
A^2	1020.84	1	1020.84	84.25	<0.0001	
B^2	72.07	1	72.07	5.95	0.0349	
C^2	1103.94	1	1103.94	91.11	<0.0001	
Residual	121.17	10	12.12			
Lack of Fit	86.61	5	17.32	2.51	0.1681	Not significant
Pure Error	34.56	5	6.91			
Corrected Total	2099.25	19				
Std. Dev.	3.48					
Mean	79.47					
C.V. %	4.38					
R^2	0.9423					
Adjusted R-Squared	0.8903					
Predicted R-Squared	0.6617					
Adequate Precision	10.646					

df: degrees of freedom; CV: coefficient of variation; F: Fisher ratio; p: probability.

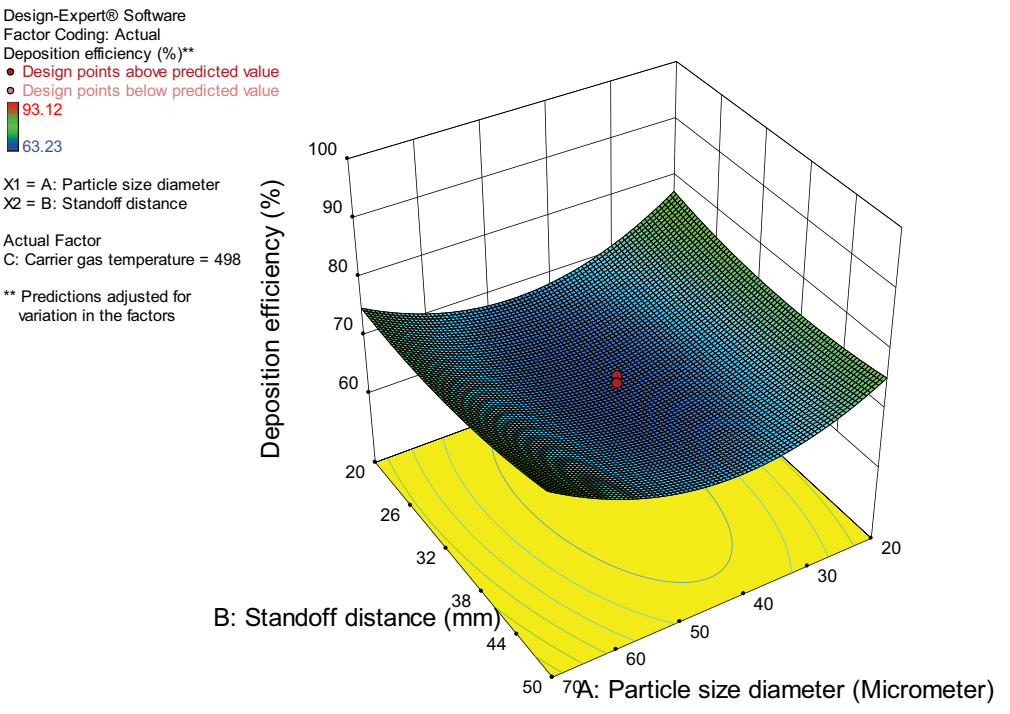


Figure 8. The surface plots of deposition efficiency (%) with interaction of stand-off distance and particle size diameter.

Figure 9 shows the surface plot of interaction of stand-off distance and carrier gas temperature at constant particle size with deposition efficiency (%). The deposition efficiency is maximum at extreme of both interactions because at the constant particle size diameter the critical velocity (eq. 6-7) is minimum for the particle.

Figure 10 shows the surface plot of interaction of carrier gas temperature and particle size diameter with deposition efficiency (%). In Figure 10 the deposition efficiency is maximum when both the carrier gas temperature and the particle size diameter are maximum because at this condition the critical velocity is minimum (eq 6-7). In the case, when particle size diameter is 45 µm and the carrier gas velocity is 498K the deposition efficiency is minimum because in this case some of the powder particles are deflected from their respective path of momentum gain as a result some of powder particles packets unable of to achieve the critical velocity. At all the corners, deposition efficiency is almost maximum. When the powder particle size is minimum and the carrier gas temperature increases, the deposition efficiency increases because the critical velocity is decreasing except in the middle of the temperature range [28,42].

5.3. Statistical optimization

In this study, optimisation of cold spray process parameters was carried out so that the actual experimentation could be done by using these parameters. This

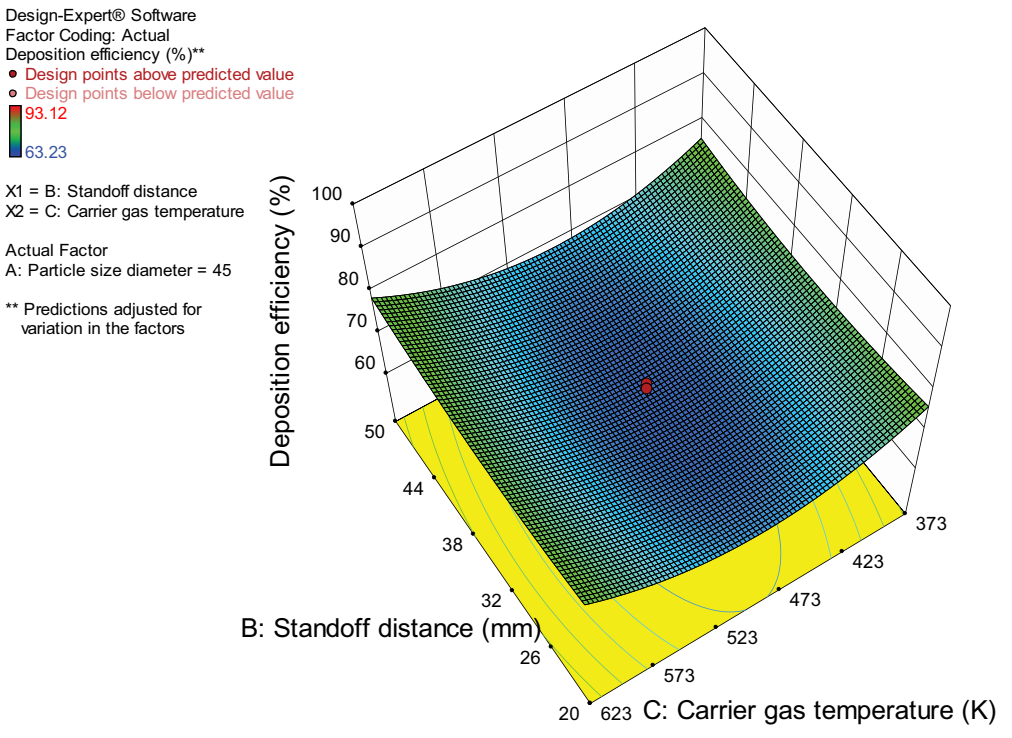


Figure 9. The surface plots of deposition efficiency (%) with interaction of stand-off distance and carrier gas temperature.

would reduce the experimentation cost and time consumption. The optimisation of this model was done by searching the combination of various process parameters with the help of software and taking care of the optimisation criteria whereas numerical optimisation was done with the aim of combining the goals into an overall desirability function. In this optimisation propagation of error (POE) was also considered to reduce the main factor. All process parameters were considered in range and the deposition efficiency was maximised in range of the responses obtained and POE was minimised. For numerical optimisation Eq-12 was analysed using the response i.e. deposition efficiency (%) and POE in deposition efficiency was used and the optimised results are tabulated in the Table 7.

6. Conclusion

In the present study, 2-D axisymmetric model of Cold Spray Deposition Efficiency was successfully optimised by Response Surface Methodology (Central Composite Design). All the process parameters, carrier gas temperature, stand-off distance and particle size, were significant. Whereas temperature of carrier gas is significantly affects the temperature and velocity profile of cold spray region. Turbulence effects are also captured by the help of DRW model resulting in more accurate particle velocity and trajectory prediction. The maximum

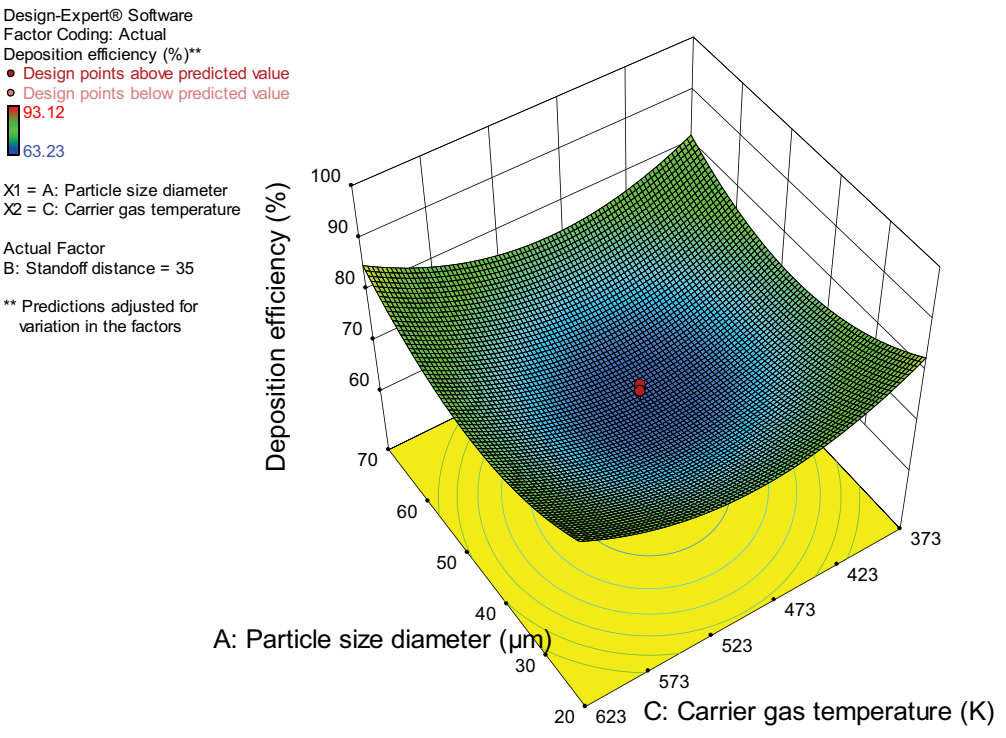


Figure 10. Showing the surface plots of deposition efficiency (%) with interaction of carrier gas temperature and particle size diameter.

Table 7. Optimized process parameters and response.

Process parameters and response	Numerical Optimization
A: Particle size diameter (µm)	23.57
B: Stand-off distance (mm)	50
C: Carrier gas temperature (K)	623
R: Deposition efficiency (%)	82.97
POE in deposition efficiency (%)	3.48
Desirability	0.731

deposition efficiency in the case of Copper-Zinc Alloy (C92300) is achieved at carrier gas temperature 623K, stand-off distance 50 mm and particle size 70 µm.

Acknowledgements

Authors would like to acknowledge Mohd. Danish Ansari first author's colleague, Mehul Varshney, PG student, Shahid Husain, Mohd. Imran Ansari research scholars and Computational Fluid Dynamics Lab (CFD Lab) of Mechanical Engineering Department as without their support and facilities this work is not possible.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] Suo XK, Liu TK, Li WY, et al. Numerical study on the effect of nozzle dimension on particle distribution in cold spraying, surf. Coat Technol. 2013;220:107–111. doi: 10.1016/j.surfcoat.2012.09.029

[2] Karimi M, Fartaj A, Rankin G, et al. Numerical simulation of the cold gas dynamic spray process. J Therm Spray Technol. 2006;15(4):518–523. doi: 10.1361/105996306X146866

[3] Papyrin A. Cold spray technology. Adv Mater Process. 2001;159:49–51.

[4] Li WY, Yin S, Guo X, et al. An investigation on temperature distribution within the substrate and nozzle wall in cold spraying by numerical and experimental methods. J Therm Spray Technol. 2012;21(1):41–48. doi: 10.1007/s11666-011-9685-2

[5] Moridi A, Hassani-Gangaraj SM, Guagliano M, et al. Cold spray coating: review of material systems and future perspectives, surf. Eng. 2014;36(6):369–395. doi: 10.1179/1743294414Y.0000000270

[6] Rokni MR, Nutt SR, Widener CA, et al. Review of relationship between particle deformation, coating microstructure, and properties in high-pressure cold spray. J Therm Spray Technol. 2017;26(6):1308–1355. doi: 10.1007/s11666-017-0575-0

[7] Li WY, Yin S, Wang XF. Numerical investigations of the effect of oblique impact on particle deformation in cold spraying by the SPH method, Appl. Surf Sci. 2010;256(12):3725–3734. doi: 10.1016/j.apsusc.2010.01.014

[8] Meyer M, Yin S, Lupoi R. Particle In-flight velocity and dispersion measurements at increasing particle feed rates in cold spray. J Therm Spray Tech. 2017;26(1–2):60–70. doi: 10.1007/s11666-016-0496-3

[9] Yin S, Wang X, Li W, et al. Numerical study on the effect of substrate size on the supersonic Jet flow and temperature distribution within the substrate in cold spraying. J Therm Spray Technol. 2012;21(3–4):628–635. doi: 10.1007/s11666-011-9711-4

[10] Li CJ, Li WY. Deposition characteristics of titanium coating in cold spraying. Surf And Coat Technol. 2003;167(2–3):278–283. doi: 10.1016/S0257-8972(02)00919-2

[11] Grujicic M, Zhao CL, Tong C, et al. Analysis of the impact velocity of powder particles in the cold-gas dynamic-spray process. Mater Sci Eng A. 2004;368(1–2):222. doi: 10.1016/j.msea.2003.10.312

[12] Irissou E, Ilinca F, Wong W et al. Investigation on the Effect of Helium-to-Nitrogen Ratio As Propellant Gas Mixture on the Processing of Titanium Coating Using Cold Gas Dynamic Spray. In: Proceedings of the International Thermal Spray Conference; Germany; 2011 Sep; p. 66.

[13] Jodoin B, Raletz F, Vardelle M. Cold spray modeling and validation using an optical diagnostic method. Surf & Coat Technol. 2006;200(14–15):4424–4432. doi: 10.1016/j.surfcoat.2005.02.209

[14] Li W-Y, Zhang C, Guo XP, et al. Effect of standoff distance on coating deposition characteristics in cold spraying. Mater And Des. 2008;29(2):297–304. doi: 10.1016/j.matdes.2007.02.005

[15] Champagne VK, Helfritch DJ, Dinavahi SPG, et al. Theoretical and experimental particle velocity in cold spray. J Therm Spray Technol. 2011;20(3):425–431. doi: 10.1007/s11666-010-9530-z

[16] Lupoi R. Current design and performance of cold spray nozzles: experimental and numerical observations on deposition efficiency and particle velocity. Surf Eng. 2014;30(5):316–322. doi: 10.1179/1743294413Y.0000000214

[17] Assadi H, Schmidt T, Richter H, et al. On parameter selection in cold spraying. J Therm Spray Technol. 2011;20(6):1161–1176. doi: 10.1007/s11666-011-9662-9

[18] Cinca N, List A, Gärtner F, et al. Influence of spraying parameters on cold gas spraying of iron aluminide intermetallics. Surf & Coat Technol. 2015;268:99–107. doi: 10.1016/j.surfcoat.2014.07.070

[19] Li WY, Liao H, Wang HT, et al. Optimal design of a convergent-barrel cold spray nozzle by numerical method, Appl. Surf Sci. 2006;253(2):708–713. doi: 10.1016/j.apsusc.2005.12.157

[20] Li CJLWY. Optimization of spray conditions in cold spraying based on the numerical analysis of particle velocity. Trans Nonferr Met Soc China. 2004;14:43–48.
[21] Yin S, Wang XF, Li WY, et al. Numerical study on the effect of substrate angle on particle impact velocity and normal velocity component in cold gas dynamic spraying based on CFD. J Therm Spray Technol. 2010;19(6):1155–1162. doi: 10.1007/s11666-010-9510-3
[22] Tarun Goyal RSW, Sidhu TS, Sidhu TS. Effect of parameters on coating density for cold spray process. Mater Manuf Processes. 2012;27(2):193–200. doi: 10.1080/10426914.2011.566906
[23] Sova A, Okunkova A, Grigoriev S, et al. Velocity of the particles accelerated by a cold spray micronozzle: experimental measurements and numerical simulation. J Therm Spray Technol. 2013;22(1):75–80. doi: 10.1007/s11666-012-9846-y
[24] Yin S, Wang XF, Li WY, et al. Examination on substrate preheating process in cold gas dynamic spraying. J Therm Spray Technol. 2011;20(4):852–859. doi: 10.1007/s11666-011-9623-3
[25] Sova A, Maestracci R, Jeandin M, et al. Kinetics of composite coating formation process in cold spray: modelling and experimental validation. Surf & Coat Technol. 2017;318:309–314. doi: 10.1016/j.surfcoat.2016.06.084
[26] Moridi A, Hassani-Gangaraj SM, Guagliano M. A hybrid approach to determine critical and erosion velocities in the cold spray process, Appl. Surf Sci. 2013;273:617–624. doi: 10.1016/j.apsusc.2013.02.089
[27] Ramachandran CS, Balasubramanian V, Ananthapadmanabhan PV. Multiobjective optimization of atmospheric plasma spray process parameters to deposit yttria-stabilized zirconia coatings using response surface methodology. J Therm Spray Technol. 2011;20(3):590–607. doi: 10.1007/s11666-010-9604-y
[28] Yin S, Suo X, Liao H, et al. Significant influence of carrier gas temperature during the cold spray process, surf. Eng. 2014;30(6):443–450. doi: 10.1179/1743294414Y.0000000276
[29] Singhal C, Murtaza Q, Parvej, et al. Simulation of critical velocity of cold spray process with different turbulence models. Mater Today Proc. 2018;5(91):17371–17379. doi: 10.1016/j.matpr.2018.04.150
[30] Faheem A, Hasan F, Murtaza Q. Material perspective and deformation pattern of micro-sized metallic particle using cold gas dynamic spray. Adv Comput Methods In Manuf. 2019:393–401.
[31] Khan M, Zunaid M, Murtaza Q. Simulation of cold spray coating for powder pre-heat and impact velocity. Mater Today Proc. 2021;46(20):10837–10844. doi: 10.1016/j.matpr.2021.01.780
[32] Faheem A, Tyagi A, Pandey SM, et al. A sustainable ecofriendly additive manufacturing approach of repairing and coating on the substrate: cold spray. Australian J Mech Eng. 2022;21(5):1–18. doi: 10.1080/14484846.2021.2023379
[33] Faheem A, Tyagi A, Hasan F, et al. Residual stress investigation in the metallic coating approach of micro-sized particles on the substrate: cold spray additive manufacturing. Adv Mater And Process Technol. 2022;8(4):1–18. doi: 10.1080/2374068X.2022.2079250
[34] Seraj RA, Zadeh AA, Assadi H, et al. Effect of substrate on the properties of cold sprayed coating of WC-10Ni. Adv Mater Process Technol. 2020;8(1):1–14. doi: 10.1080/2374068X.2020.1794228
[35] Singh S, Raman RKS, Berndt CC, et al. Influence of cold spray parameters on bonding mechanisms. A Rev Met. 2021;11(12):1–26. doi: 10.3390/met11122016
[36] Kamaraj M, Radhakrishnan VM. Cold spray coating diagram: bonding properties and construction methodology. J Therm Spray Technol. 2019;28(4):756–768. doi: 10.1007/s11666-019-00853-5
[37] Fardan A, Berndt CC, Ahmed R. Numerical modelling of particle impact and residual stresses in cold sprayed coatings: a review. Surf & Coat Technol. 2021;409:1–38. doi: 10.1016/j.surfcoat.2021.126835

[38] Singhal C, Murtaza Q. Numerical study using nano-particles by varying the expansion ratio and height of cold spray nozzle to predict deposition efficiency, ITSC 2018. Paper no.: itsc2018p0286. 2018. p. 286–293.
[39] Gabor T, Akin S, Ju MBG. Numerical studies on cold spray gas dynamics and powder flow in circular and rectangular nozzles. J Manuf Processes. 2024;114:232–246. doi: 10.1016/j.jmapro.2024.02.005
[40] Tsai JT, Akin S, Bahr DF, et al. A predictive modeling approach for cold spray metallization on polymers. Surf And Coat Technol. 2024;483:130711. doi: 10.1016/j.surfcoat.2024.130711
[41] Zhang ZL, Gitardi D, Scarafile M, et al. Experimental and numerical investigations of pore formation mechanisms in cold spray. J Manuf Processes. 2024;127:488–510. doi: 10.1016/j.jmapro.2024.08.006
[42] Murtaza Q, Stokes J, Ardhaoui M. Experimental analysis of spray dryer used in hydroxyapatite thermal spray powder. J Therm Spray Technol. 2012;21(5):963–974. doi: 10.1007/s11666-012-9791-9

arXiv:2412.17403v1 [math.CV] 23 Dec 2024

ON COEFFICIENT PROBLEMS FOR S_ρ^*

S. SIVAPRASAD KUMAR*, ARYA TRIPATHI, AND SNEHAL PANNU

Abstract. Logarithmic and inverse logarithmic coefficients play a crucial role in the theory of univalent functions. In this study, we focus on the class of starlike functions $\mathcal{S}_\rho^*$, defined as

$$\mathcal{S}_\rho^* = \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec \rho(z),\ z \in \mathbb{D} \right\},$$

where $\rho(z) := 1 + \sinh^{-1}(z)$, which maps the unit disk $\mathbb{D}$ onto a petal-shaped domain. This investigation aims to establish bounds for the second Hankel and Toeplitz determinants, with their entries determined by the logarithmic coefficients of f and its inverse f^{-1}, for functions $f \in \mathcal{S}_\rho^*$.

1 Introduction

Let $\mathcal{A}$ denote the class of all normalized analytic functions f defined on the open unit disk $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$, such that $f(0) = 0$ and $f'(0) = 1$. Each function $f \in \mathcal{A}$ admits a Taylor series expansion of the form:

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n. \tag{1.1}$$

Let $\mathcal{S} \subset \mathcal{A}$, where $\mathcal{S}$ represents the collection of analytic univalent functions. In particular, the class $\mathcal{S}^*$, comprising starlike functions, maps the unit disk $\mathbb{D}$ onto a starlike domain, represented as:

$$\mathcal{S}^* = \left\{ f \in \mathcal{A} : \operatorname{Re} \frac{zf'(z)}{f(z)} > 0, \quad z \in \mathbb{D} \right\}.$$

Let f and g be two analytic functions, we say that f is subordinate to g, denoted by $f \prec g$, if there exists a Schwarz function w with $w(0) = 0$ and $|w(z)| < 1$ such that $f(z) = g(w(z))$. In 1992, Ma and Minda [8] introduced a broader subclass of $\mathcal{S}^*$, denoted by $\mathcal{S}^*(\varphi)$, and defined as:

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

where φ is an analytic univalent function satisfying the conditions $\operatorname{Re} \varphi(z) > 0$, with $\varphi(\mathbb{D})$ being starlike with respect to $\varphi(0) = 1$, $\varphi'(0) > 0$, and the domain $\varphi(\mathbb{D})$ being symmetric about the real axis. Various subclasses of $\mathcal{S}^*$ arise when the choice of φ is varied in the class $\mathcal{S}^*(\varphi)$.

Arora and Kumar [1] introduced a class of starlike functions associated with the petal-shaped domain $\rho(\mathbb{D})$, by choosing $\varphi(z) := \rho(z)$, where $\rho(z) = 1 + \sinh^{-1}(z)$, defined by:

*Corresponding author

2020 *Mathematics Subject Classification.* 30C45 · 30C50 .

Key words and phrases. Univalent functions · Starlike functions · Petal-shaped domain · Hankel determinant · Toeplitz determinant · Logarithmic coefficient · Inverse Logarithmic coefficient.

$$\mathcal{S}^*_\rho = \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec 1 + \sinh^{-1}(z), \quad z \in \mathbb{D} \right\}.$$

Clearly, the function $\rho(z)$ is defined over the branch cuts along the segments $(-\infty, -i) \cup (i, \infty)$ on the imaginary axis. Consequently, the function is analytic in the unit disk $\mathbb{D}$, geometrically, $\rho(\mathbb{D}) = \Omega_\rho := \{\Omega \in \mathbb{C} : |\sinh(\Omega - 1)| < 1\}$. Certain coefficient problems of the analytic functions associated with the class $\mathcal{S}^*_\rho$ were discussed in [6]. The function $f_0(z)$ serves as an extremal function for the family $\mathcal{S}^*_\rho$ which helps in obtaining the sharp results, given by

$$f_0(z) = z \exp\left(\int_0^z \frac{\sinh^{-1}(t)}{t}\, dt \right) = z + z^2 + \frac{1}{2}z^3 + \frac{1}{9}z^4 - \frac{1}{72}z^5 - \frac{1}{225}z^6 + \dots . \tag{1.2}$$

The logarithmic coefficients γ_n associated with $f \in \mathcal{S}$ are defined as follows:

$$F_f(z) = \log\left(\frac{f(z)}{z} \right) = 2\sum_{n=1}^{\infty} \gamma_n(f) z^n; \quad z \in \mathbb{D} \setminus \{0\}, \quad \log 1 = 0.$$

We denote $\gamma_n(f)$ by γ_n. For functions in class $\mathcal{S}$, the sharp bounds for the logarithmic coefficients when $n = 1$ and $n = 2$ are:

$$|\gamma_1| \le 1, \quad |\gamma_2| \le \frac{1}{2} + \frac{1}{e^2}.$$

A significant motivation behind studying logarithmic coefficients lies in estimating the sharp bounds for the class $\mathcal{S}$. Currently, these sharp bounds have only been established for γ_1 and γ_2. The problem of determining the sharp bounds for γ_n when $n \ge 3$ is still open. The Koebe 1/4-theorem, enables us to define the inverse function $F_f \in \mathcal{A}$ in a neighborhood of the origin with the Taylor series expansion:

$$F(w) := f^{-1}(w) = w + \sum_{n=2}^{\infty} A_n w^n, \quad |w| < 1. \tag{1.3}$$

The logarithmic inverse coefficients Γ_n, for $n \in \mathbb{N}$, associated with the function F, are defined by the relation

$$F_{f^{-1}}(w) := \log\left(\frac{F(w)}{w} \right) = 2\sum_{n=1}^{\infty} \Gamma_n w^n, \quad |w| < \frac{1}{4}.$$

Ponnusamy *et al.* [13] established sharp bounds for the initial logarithmic coefficients of the inverse function for certain subclasses of $\mathcal{S}^*$.

Hankel matrices (and determinants) have emerged as fundamental elements in different areas of mathematics, finding a wide array of applications [15]. For a function $f \in \mathcal{A}$, the q^{th} Hankel determinant $H_{q,n}(F_f)$, where $q, n \in \mathbb{N}$, with entries as logarithmic coefficients, is given by:

$$H_{q,n}(F_f/2) = \begin{vmatrix} \gamma_n & \gamma_{n+1} & \cdots & \gamma_{n+q-1} \\ \gamma_{n+1} & \gamma_{n+2} & \cdots & \gamma_{n+q} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n+q-1} & \gamma_{n+q} & \cdots & \gamma_{n+2(q-1)} \end{vmatrix}. \tag{1.4}$$

Kowalczyk *et al.* [5] investigated the Hankel determinants formed using logarithmic coefficients. Specifically, the expression $H_{2,1}(F_f/2)$ shares a notable resemblance to $H_{2,1}(f) = a_2a_4 - a_3^2$, where $f \in \mathcal{A}$. Hankel determinants are beneficial, for example, in determining whether certain

coefficient functionals related to functions are bounded in $\mathbb{D}$ and if they achieve sharp bounds, see [14]. Furthermore, Mundalia and Kumar [9] explored the problem of logarithmic coefficients for certain subclasses of close-to-convex functions. The logarithmic coefficients for a function $f \in \mathcal{S}$ are represented as follows:

$$\gamma_1 = \frac{1}{2}a_2, \quad \gamma_2 = \frac{1}{2}\left(a_3 - \frac{1}{2}a_2^2\right), \quad \gamma_3 = \frac{1}{2}\left(a_4 - a_2a_3 + \frac{1}{3}a_2^3\right). \tag{1.5}$$

In 2016, Ye and Lim [15] established that any $n \times n$ matrix over $\mathbb{C}$ can be represented, in general, as a product of certain Toeplitz or Hankel matrices. Toeplitz matrices, along with their determinants, play a significant role in both applied and theoretical mathematics. Their applications extend to various fields such as analysis, quantum physics, image processing, integral equations, and signal processing. A defining feature of Toeplitz matrices is that their elements remain constant along each diagonal. Giri and Kumar [4] explored bounds on Toeplitz determinants for different subclasses of normalized univalent functions in higher-dimensional spaces. Inspired by these contributions, we define the Toeplitz determinant of the logarithmic coefficients for $f \in \mathcal{S}$ as follows:

$$T_{q,n}(F_f/2) = \begin{vmatrix} \gamma_n & \gamma_{n+1} & \cdots & \gamma_{n+q-1} \\ \gamma_{n+1} & \gamma_n & \cdots & \gamma_{n+q-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n+q-1} & \gamma_{n+q-2} & \cdots & \gamma_n \end{vmatrix}. \tag{1.6}$$

Motivated further by the above concepts, we begin the examination of the Hankel determinant $H_{q,n}(F_{f^{-1}}/2)$ and the Toeplitz determinant $T_{q,n}(F_{f^{-1}}/2)$, where the elements are the logarithmic coefficients of the inverse functions [12] of $f^{-1} \in \mathcal{S}$. The Hankel determinant is given by

$$H_{q,n}(F_{f^{-1}}/2) = \begin{vmatrix} \Gamma_n & \Gamma_{n+1} & \cdots & \Gamma_{n+q-1} \\ \Gamma_{n+1} & \Gamma_{n+2} & \cdots & \Gamma_{n+q} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{n+q-1} & \Gamma_{n+q} & \cdots & \Gamma_{n+2(q-1)} \end{vmatrix}, \tag{1.7}$$

and the Toeplitz determinant $T_{q,n}(F_{f^{-1}}/2)$ is given by

$$T_{q,n}(F_{f^{-1}}/2) = \begin{vmatrix} \Gamma_n & \Gamma_{n+1} & \cdots & \Gamma_{n+q-1} \\ \Gamma_{n+1} & \Gamma_n & \cdots & \Gamma_{n+q-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{n+q-1} & \Gamma_{n+q-2} & \cdots & \Gamma_n \end{vmatrix}. \tag{1.8}$$

This work focuses on deriving bounds for the second Hankel and Toeplitz determinants, where the entries are the logarithmic coefficients associated with the functions f and f^{-1}. Here, f belongs to the class $\mathcal{S}_\rho^*$.

2 Preliminary results

The Carathéodory class $\mathcal{P}$ and its coefficient bounds are crucial in determining Hankel and Toeplitz determinant bounds. Let $\mathcal{P}$ denote the class of functions consisting of p, such that

$$p(z) = 1 + \sum_{n=1}^{\infty} p_n z^n, \tag{2.1}$$

which are analytic in the open unit disk $\mathbb{D}$ and satisfy $\text{Re}\, p(z) > 0$ for any $z \in \mathbb{D}$. Here $p(z)$ is called the Carathéodory function, or functions with positive real part [2, 11]. It is known that $p_n \leq 2$, $n \geq 1$, for a function $p \in \mathcal{P}$. In this section, we provide crucial lemmas that will be utilized to establish the main results of this paper.

Lemma 2.1. *[7, 10, 11]: If $p \in \mathcal{P}$ is the form (2.1), then*

$$p_1 = 2\zeta_1, \tag{2.2}$$

$$p_2 = 2\zeta_1^2 + 2(1-\zeta_1^2)\zeta_2, \tag{2.3}$$

$$p_3 = 2\zeta_1^3 + 4(1-\zeta_1^2)\zeta_1\zeta_2 - 2(1-\zeta_1^2)\zeta_1\zeta_2^2 + 2(1-\zeta_1^2)(1-|\zeta_2|^2)\zeta_3,$$

where $\zeta_1 \in [0,1]$ and $\zeta_2, \zeta_3 \in \overline{\mathbb{D}} := \{z \in \mathbb{C} : |z| \leq 1\}$.

Lemma 2.2. *[3]: If A, B, $C \in \mathbb{R}$, let us consider*

$$Y(A,B,C) := \max\{|A + Bz + Cz^2| + 1 - |z|^2, \quad z \in \overline{\mathbb{D}}\}$$

1. If $AC \geq 0$, then

$$Y(A,B,C) = \begin{cases} |A| + |B| + |C|, & |B| \geq 2(1-|C|), \\ 1 + |A| + \frac{B^2}{4(1-|C|)}, & |B| < 2(1-|C|). \end{cases}$$

2. If $AC < 0$, then

$$Y(A,B,C) = \begin{cases} 1 - |A| + \frac{B^2}{4(1-|C|)}, & -4AC(C^{-2}-1) \leq B^2 \wedge |B| < 2(1-|C|), \\ 1 + |A| + \frac{B^2}{4(1+|C|)}, & B^2 < \min\{4(1+|C|)^2, -4AC(C^{-2}-1)\}, \\ R(A,B,C), & \textit{Otherwise}, \end{cases}$$

where

$$R(A,B,C) = \begin{cases} |A| + |B| - |C|, & |C|(|B| + 4|A|) \leq |AB|, \\ -|A| + |B| + |C|, & |AB| \leq |C|(|B| - 4|A|), \\ (|C| + |A|)\sqrt{1 - \frac{B^2}{4AC}}, & \textit{Otherwise}. \end{cases}$$

3 Second Hankel Determinant

By applying the definition of the Hankel determinant, we obtain the expression for second Hankel determinant given by,

$$H_{2,1}(f) = a_1 a_3 - a_2^2.$$

Since $a_1 := 1$, it follows that $H_{2,1}(f) = a_3 - a_2^2$. By varying the coefficients, we can compute the second Hankel determinant for different values of the coefficients. For instance, using equation (1.4), we can derive the result for logarithmic coefficients, given by

$$H_{2,1}(F_f/2) = \begin{vmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & \gamma_3 \end{vmatrix} = \gamma_1\gamma_3 - \gamma_2^2 = \frac{1}{4}\left(a_2a_4 - a_3^2 + \frac{1}{12}a_2^4\right). \tag{3.1}$$

It is important to note that $H_{2,1}(F_f/2)$ remains invariant under rotation. Since, for the function $f_\theta(z) := e^{-i\theta}f(e^{i\theta}z)$, where $f \in \mathcal{S}$ and $\theta \in \mathbb{R}$, we have the following relation:

$$H_{2,1}\left(F_{f_\theta}/2\right) = \frac{e^{4i\theta}}{4}\left(a_2a_4 - a_3^2 + \frac{1}{12}a_2^4\right) = e^{4i\theta}H_{2,1}\left(F_f/2\right). \tag{3.2}$$

Using (1.7), we can derive the logarithmic coefficients for the inverse functions, given by

$$H_{2,1}(F_{f^{-1}}/2) = \begin{vmatrix} \Gamma_1 & \Gamma_2 \\ \Gamma_2 & \Gamma_3 \end{vmatrix} = \Gamma_1\Gamma_3 - \Gamma_2^2 = \frac{1}{48}\left(13a_2^4 - 12a_2^2a_3 - 12a_3^2 + 12a_2a_4\right). \tag{3.3}$$

Similarly, it can be observed that $H_{2,1}(F_{f^{-1}}/2)$ is also invariant under rotation.

3.1 Sharp Bound Estimation of $|H_{2,1}(F_f/2)|$ for $\mathcal{S}_\rho^*$

In this section, we will primarily focus on estimating the sharp bound for the Hankel determinant associated with the logarithmic coefficients for starlike functions ($|H_{2,1}(F_f/2)|$) defined in the petal-shaped domain $\rho(\mathbb{D})$.

Theorem 3.1. *Let $f \in \mathcal{S}_\rho^*$, then*

$$\frac{1}{72} \leq |H_{2,1}(F_f/2)| \leq \frac{1}{16}. \tag{3.4}$$

The above inequality is sharp.

Proof. Since $f \in \mathcal{S}_\rho^*$, we can find a Swcharz function $w(z)$ such that

$$\frac{zf'(z)}{f(z)} = 1 + \sinh^{-1}(w(z)). \tag{3.5}$$

Assume that $w(z) = (p(z)-1)/(p(z)+1)$ where, $p \in \mathcal{P}$ is given by (2.1). Using the expansion of f from (1.1) and p from (2.1), we get

$$a_2 = \frac{1}{2}p_1,\ a_3 = \frac{1}{4}p_2,\ a_4 = \frac{1}{144}\left(-p_1^3 - 6p_1p_2 + 24p_3\right). \tag{3.6}$$

Using (1.5), (3.1), (3.6), we obtain the following expression

$$\begin{aligned} \mathcal{L} &:= \gamma_1\gamma_3 - \gamma_2^2 \\ &= \left(\frac{1}{2}a_2\right)\left(\frac{1}{2}\left(a_4 - a_2a_3 + \frac{1}{3}a_2^3\right)\right) - \left(\frac{1}{2}\left(a_3 - \frac{1}{2}a_2^2\right)\right)^2 \\ &= \frac{1}{4}\left(a_2a_4 - a_3^2 + \frac{1}{12}a_2^4\right) \qquad (3.7) \\ &= \frac{1}{2304}\left(p_1^4 - 12p_1^2p_2 - 36p_2^2 + 48p_1p_3\right) \qquad (3.8) \end{aligned}$$

By the Lemma (2.1) and (3.8), we get an equation in terms of the $\zeta_i^{'}$s, where $\zeta_i \in \overline{\mathbb{D}}$ for $i = 1, 2, 3$:

$$\mathcal{L} = \frac{1}{144}\left(-9\zeta_2^2 + 6\zeta_1^2\zeta_2^2 + \zeta_1^4(-2+3\zeta_2^2) + 12\zeta_1\zeta_3 - 12\zeta_1^3\zeta_3 + 12\zeta_1(-1+\zeta_1^2)\zeta_3|\zeta_2|^2\right) \tag{3.9}$$

Since $\zeta_1 \in [0,1]$ from Lemma (2.1), therefore the above expression in (3.9) leads to,

$$|\mathcal{L}| = \begin{cases} \frac{|\zeta_2|^2}{16} \leq \frac{1}{16}, & \zeta_1 = 0, \\ \\ \frac{1}{72}, & \zeta_1 = 1. \end{cases}$$

By utilising thr triangle inequality in (3.9), for $\zeta_1 \in (0,1)$ and $|\zeta_3| \leq 1$, we obtain the following inequality:

$$\begin{aligned} |\mathcal{L}| &\leq \frac{1}{144}|-9\zeta_2^2 + 6\zeta_1^2\zeta_2^2 + \zeta_1^4(-2+3\zeta_2^2) + 12\zeta_1\zeta_3 - 12\zeta_1^3\zeta_3| + 12\zeta_1(-1+\zeta_1^2)\zeta_3|\zeta_2|^2 \\ &= \frac{1}{12}\zeta_1(1-\zeta_1^2)\ \Psi(A,B,C), \end{aligned} \tag{3.10}$$

where

$$\Psi(A,B,C) := |A + B\zeta_2 + C\zeta_2^2| + 1 - |\zeta_2^2|,$$

with

$$A = \frac{-\zeta_1^3}{6(1-\zeta_1^2)}, \quad B = 0, \quad C = -\left(\frac{\zeta_1^2+3}{4\zeta_1}\right). \tag{3.11}$$

We now examine the following cases in the context of Lemma 2.2, based on the various cases involving A, B, and C, which are given in (3.11).

I. Suppose $\zeta_1 \in X = (0,1)$. Clearly, we can observe that $AC \geq 0$, also

$$|B| - 2(1-|C|) = -2 + \frac{3}{2\zeta_1} + \frac{\zeta_1}{2} > 0, \ \ \zeta_1 \in X.$$

Therefore, by using Lemma 2.2, we get,

$$\Psi(A,B,C) = |A| + |B| + |C|.$$

Now, by utilising above equation, (3.10) can be further reduced in the following manner:

$$\begin{aligned} |\mathcal{L}| &\leq \frac{1}{12}\zeta_1(1-\zeta_1^2)\left(|A| + |B| + |C|\right) \\ &= \frac{1}{12}\zeta_1(1-\zeta_1^2)\left(\left|\frac{-\zeta_1^3}{6(1-\zeta_1^2)}\right| + \left|\frac{-\zeta_1^2-3}{4\zeta_1}\right|\right) \\ &= \frac{1}{144}(9 - 6\zeta_1^2 - \zeta_1^4) \\ &\leq \frac{1}{16}. \end{aligned}$$

Thus, from the above result we can see that inequality (3.4) holds. To demonstrate the sharpness of (3.4), we consider the function $f_1 \in \mathcal{S}_\rho^*$ for the upper bound, given by:

$$f_1(z) = z\exp\left(\int_0^z \frac{\sinh^{-1}(t^2)}{t}\,dt\right) = z + \frac{z^3}{2} + \frac{z^5}{8} - \frac{z^7}{144} + \dots\ .$$

By comparing the coefficients, we can easily find the values for a_2, a_3, and a_4, which upon further substitution into (3.7) yields the equality:

$$|H_{2,1}(F_f/2)| = \frac{1}{16}.$$

For the lower bound case, we consider the extremal function $f_0(z)$, defined in (1.2), which completes the proof. ∎

3.2 Sharp bound of $|H_{2,1}(F_{f^{-1}}/2)|$ for $\mathcal{S}_\rho^*$

In this section, we will focus on estimating the sharp bound of $|H_{2,1}(F_{f^{-1}}/2)|$ within $\rho(\mathbb{D})$.

Theorem 3.2. *Let $f \in \mathcal{S}_\rho^*$, then*

$$\frac{1}{16} \leq |H_{2,1}(F_{f^{-1}}/2)| \leq \frac{1}{9}. \tag{3.12}$$

The above inequality is sharp.

Proof. Let us suppose that $f \in \mathcal{S}_\rho^*$ satisfy (3.5). Ponnusamy et al. [3] examined the logarithmic coefficients of the inverses of univalent functions, defined as

$$\Gamma_1 = \frac{-1}{2}a_2,\ \Gamma_2 = \frac{-1}{2}a_3 + \frac{3}{4}a_2^2,\ \Gamma_3 = \frac{-1}{2}a_4 + 2a_2a_3 - \frac{5}{3}a_2^3. \tag{3.13}$$

By utilising (3.3), (3.13), and Lemma 2.1, we express $\Gamma_1\Gamma_3 - \Gamma_2^2$ in terms of $\zeta_i^{'}s$ by substituting $p_i^{'}s$ with $\zeta_i^{'}s$ $(i = 1, 2, 3)$:

$$\begin{aligned}
\mathcal{L} &:= \Gamma_1\Gamma_3 - \Gamma_2^2 \\
&= \frac{1}{48}\left(13a_2^4 - 12a_2^2a_3 - 12a_3^2 + 12a_2a_4\right) \\
&= \frac{1}{2304}\left(37p_1^4 - 48p_1^2p_2 - 36p_2^2 + 48p_1p_3\right) \\
&= \frac{1}{144}(6\zeta_1^2(-3+\zeta_2)\zeta_2 - 9\zeta_2^2 + \zeta_1^4(16 + 18\zeta_2 + 3\zeta_2^2) + 12\zeta_1\zeta_3 - 12\zeta_1^3\zeta_3 \\
&\quad +12\zeta_1(-1+\zeta_1^2)\zeta_3|\zeta_2|^2).
\end{aligned} \tag{3.14}$$

As $|\zeta_3| \leq 1$, the above inequality provides

$$|\mathcal{L}| = \begin{cases} \frac{|\zeta_2|^2}{16} \leq \frac{1}{16}, & \zeta_1 = 0, \\ \frac{1}{9}, & \zeta_1 = 1. \end{cases}$$

Let $\zeta_1 \in (0, 1)$. By applying the triangle inequalitiy in (3.14) and the fact that $|\zeta_3| \leq 1$, we obtain

$$|\mathcal{L}| = \frac{1}{12}\zeta_1(1-\zeta_1^2)\ \Psi(A, B, C), \tag{3.15}$$

where

$$\Psi(A, B, C) := |A + B\zeta_2 + C\zeta_2^2| + 1 - |\zeta_2^2|,$$

with

$$A = \frac{4\zeta_1^3}{3(1-\zeta_1^2)},\ B = -\frac{3}{2}\zeta_1,\ C = -\left(\frac{3+\zeta_1^2}{4\zeta_1}\right). \tag{3.16}$$

In light of Lemma 2.2 and the expressions of A, B and C obtained in (3.16), we study the following cases. Let us assume that $\zeta_1 \in X$ where $X = (0, 1)$. Clearly, we can observe that $AC < 0$ for $\zeta_1 \in X$. This case demands the following sub-cases:

(a) Simply for each $\zeta_1 \in X$

$$T_1(\zeta_1) := |B| - 2(1 - |C|) = 2\zeta_1 - 2 + \frac{3}{2\zeta_1} > 0,$$

implies

$$|B| > 2(1 - |C|).$$

$$T_2(\zeta_1) := -4AC\left(\frac{1}{C^2} - 1\right) - B^2 = \frac{-\zeta_1^2(225 + 11\zeta_1^2)}{12(3 + \zeta_1^2)} \leq 0,$$

implies

$$-4AC\left(\frac{1}{C^2} - 1\right) \leq B^2.$$

We can infer from the above computations that $T_1(X^*) \cap T_2(X^*)$ is empty. Consequently, for every $\zeta_1 \in X$, this case does not occur, as stated by Lemma 2.2.

(b) In view of $\zeta_1 \in X$, the relation $4(1 + |C|)^2$ and $-4AC\left(\frac{1}{C^2} - 1\right)\big)$ become

$$T_3(\zeta_1) := 4(1 + |C|)^2 = \frac{(3 + 4\zeta_1 + \zeta_1^2)^2}{4\zeta_1^2} > 0,$$

$$T_4(\zeta_1) := -4AC\left(\frac{1}{C^2} - 1\right) = \frac{4\zeta_1^2(-9 + \zeta_1^2)}{3(3 + \zeta_1^2)} < 0.$$

Therefore, $\min\{T_3(\zeta_1), T_4(\zeta_1)\} = T_4(\zeta_1)$.

From case (a), we know that:

$$-4AC\left(\frac{1}{C^2} - 1\right) \leq B^2.$$

Thus this case fails for $\zeta_1 \in X$, as stated by Lemma 2.2.

(c) For $\zeta_1 \in X$ we assume that $T_5(\zeta_1) := |AB| - |C|(|B| + 4|A|)$. Therefore,

$$T_5(\zeta_1) = \frac{27 + 78\zeta_1^2 - 25\zeta_1^4}{24(-1 + \zeta_1^2)} < 0$$

implies

$$|AB| < |C|(|B| + 4|A|).$$

Thus the case fails to exist for $\zeta_1 \in X$, as stated by Lemma 2.2.

(d) For $\zeta_1 \in X$ we assume that

$$T_6(\zeta_1) := |AB| - |C|(|B| - 4|A|) = \frac{27 - 114\zeta_1^2 - 89\zeta_1^4}{24(-1 + \zeta_1^2)} \leq 0.$$

We observe that the above equation holds true for $0 \leq \zeta_1 \leq \zeta' = \left(\sqrt{\frac{-57+6\sqrt{157}}{89}}\right) \approx 0.451959 \in X$. Therefore, by using Lemma 2.2, we get,

$$\Psi(A,B,C) = R(A,B,C) = -|A| + |B| + |C|.$$

Using (3.15) to proceed towards our desired result,

$$\begin{aligned} |\mathcal{L}| &\leq \frac{1}{12}\zeta_1(1-\zeta_1^2)(-|A|+|B|+|C|) \\ &= \frac{1}{144}\left(9+12\zeta^2-37\zeta_1^4\right) = \phi(\zeta_1), \end{aligned} \tag{3.17}$$

where

$$\phi(t) := \frac{1}{144}\left(9+12t^2-37t^4\right)$$

Since $\phi'(t) = 0$ for $t \in (0,1)$ holds true for the only possible value of $t_0 = \sqrt{6/37} < \zeta_1'$, we deduce that ϕ is increasing in $[0,t_0]$ and decreasing in $[t_0,\zeta_1']$. Therefore, for $0 < t < \zeta_1'$, we obtain

$$\phi(\zeta_1) \leq \phi(t_0) \approx 0.0692568.$$

Hence, from (3.17) we see that

$$|\mathcal{L}| \leq \phi(\zeta_1) \leq \phi(t_0) \approx 0.0692568 < 1/9.$$

(e) By applying Lemma 2.2 for $\zeta_1' < \zeta_1 < 1$, we get

$$\begin{aligned} \mathcal{L} &\leq \frac{1}{12}\zeta_1(1-\zeta_1^2)(|C|+|A|)\sqrt{1-\frac{B^2}{4AC}} \\ &= \frac{1}{576}\sqrt{\frac{75-11\zeta_1^2}{3+\zeta_1^2}}\left(9-6\zeta_1^2+13\zeta_1^4\right) = \phi_2(\zeta_1) \end{aligned}$$

where

$$\phi_2(t) := \frac{1}{576}\sqrt{\frac{75-11t^2}{3+t_1^2}}\left(9-6t^2+13t^4\right)$$

It is observed that ϕ_2 is an increasing function such that $\phi_2'(t) \geq 0$ for $\zeta_1' < t < 1$, therefore after a simple computation, we get

$$|\mathcal{L}| \leq \phi_2(\zeta_2) \leq \phi_2(\zeta_1') \approx 0.451959 < 1/9.$$

Based on the results obtained, we can see that the inequality (3.12) holds true.

To demonstrate the sharpness of the bound, we can utilize the functions $f_0(z)$ and $f_1(z)$ for the upper and lower bounds respectively, similar to those employed in Theorem 3.1. This concludes the proof. ∎

4 Second Toeplitz Determinant

By applying the definition of the Toeplitz determinant, we obtain the following expression for the second Toeplitz determinant:

$$T_{2,1}(f) = a_1^2 - a_2^2.$$

By varying the coefficients, the Toeplitz determinant can be computed for different sets of coefficients. For instance, by using equation (1.6), we can determine the determinant for logarithmic coefficients:

$$T_{2,1}(F_f/2) = \begin{vmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & \gamma_1 \end{vmatrix} = \gamma_1^2 - \gamma_2^2 = \frac{1}{16}\left(4a_2^2 - a_2^4 - 4a_3^2 + 4a_2^2a_3\right). \tag{4.1}$$

Similarly, by applying equation (1.8), we can compute the Toeplitz determinant for logarithmic coefficients of inverse functions:

$$T_{2,1}(F_{f^{-1}}/2) = \begin{vmatrix} \Gamma_1 & \Gamma_2 \\ \Gamma_2 & \Gamma_1 \end{vmatrix} = \Gamma_1^2 - \Gamma_2^2 = \frac{1}{16}\left(-9a_2^4 + 4a_2^2 - 4a_3^2 + 12a_2^2a_3\right). \tag{4.2}$$

Let $T_{2,1}$ be a functional and F_{f_θ}, $F_{f_{\theta^{-1}}}$ be functions. Then, we can observe the rotational invariant characteristic of the functional $T_{2,1}$ with respect to the functions F_{f_θ} and $F_{f_{\theta^{-1}}}$ will be similar to $H_{2,1}(F_f/2)$ and $H_{2,1}(F_{f^{-1}}/2)$, as found in section 3.

4.1 Sharp bound of $|T_{2,1}(F_f/2)|$ for $\mathcal{S}^*_\rho$

This section explores the sharp bound estimation of the Toeplitz determinant linked to the logarithmic coefficients of starlike functions ($|T_{2,1}(F_f/2)|$) defined on $\rho(\mathbb{D})$.

Theorem 4.1. *Let $f \in \mathcal{S}^*_\rho$, then*

$$\frac{1}{16} \leq |T_{2,1}(F_f/2)| \leq \frac{1}{2}.$$

The above inequality is sharp.

Proof. Let us suppose that $f \in \mathcal{S}^*_\rho$ be of the form (1.1). Then there exist a Schwarz function, $w(z) = (p(z)-1)/(p(z)+1)$ such that it satisfies (3.5). From (4.1) and (3.6) we obtain:

$$\begin{aligned} T_{2,1}(F_f/2) &= \gamma_1^2 - \gamma_2^2 \\ &= \frac{1}{16}\left(-a_2^4 + 4a_2^2 - 4a_3^2 + 4a_2^2a_3\right) \qquad (4.3) \\ &= \frac{1}{256}\left(-p_1^4 - 4p_2^2 + 4p_1^2(4+p_2)\right). \qquad (4.4) \end{aligned}$$

Using (2.2) and (2.3) from Lemma 2.1, we derive the relation:

$$2p_2 = p_1^2 + (4 - p_1^2)\zeta_2,$$

where $|\zeta_2| \leq 1$. By substituting the value of p_2 in (4.4) and assuming $\zeta_2 := \zeta$, we derive the following equation:

$$T_{2,1}(F_f/2) = \frac{1}{256}\left(-p_1^4\zeta^2 - 16\zeta^2 + 16p_1^2 + 8p_1^2\zeta^2\right).$$

In view of triangle inequality, the above equation can be expressed as:

$$|T_{2,1}(F_f/2)| \leq \frac{1}{256}\left(|p_1^4||\zeta^2| + 16|\zeta^2| + 16|p_1^2| + 8|p_1^2||\zeta^2|\right).$$

Given the rotational invariance of $\mathcal{P}$, we restrict our analysis to non-negative values of p_1. Since $|p_1| \leq 2$, we have $0 \leq p_1 \leq 2$. Letting $p := |p_1|$, this implies $p \in [0, 2]$. Thus,

$$|T_{2,1}(F_f/2)| \leq \frac{1}{256}\left(p^4 + 24p^2 + 16\right).$$

Hence for $p \in [0, 2]$,

$$\frac{1}{16} \leq |T_{2,1}(F_f/2)| \leq \frac{1}{2}.$$

To complete the proof, it remains to verify the sharpness of the bound. Thus, in order to establish the upper bound, consider the analytic function $f_2 \in \mathcal{S}_\rho^*$, defined by:

$$f_2(z) = z\exp\left(\int_0^z \frac{\sinh^{-1}(\sqrt{8}it^2)}{t}\,dt\right) = z + \sqrt{2}iz^3 - z^5 + \frac{\sqrt{2}iz^7}{9} + \dots\ .$$

For the lower bound, we examine the analytic function $f_3 \in \mathcal{S}_\rho^*$, defined by:

$$f_3(z) = z\exp\left(\int_0^z \frac{\sinh^{-1}(it^2)}{t}\,dt\right) = z + \frac{iz^3}{2} - \frac{z^5}{8} + \frac{iz^7}{144} + \dots\ .$$

By comparing the coefficients, we can easily determine the values of a_2 and a_3. Substituting these into (4.3) yields the desired equality, completing the proof. ∎

4.2 Sharp bound of $|T_{2,1}(F_{f^{-1}}/2)|$ for $\mathcal{S}_\rho^*$

This section will center on estimating the sharp bound around the petal shaped domain for $|T_{2,1}(F_{f^{-1}}/2)|$, where $f \in \mathcal{S}_\rho^*$.

Theorem 4.2. *Let $f \in \mathcal{S}_\rho^*$, then*

$$\frac{1}{16} \leq |T_{2,1}(F_{f^{-1}}/2)| \leq \frac{5}{4}.$$

The above inequality is sharp.

Proof. Let $f \in \mathcal{S}_\rho^*$ be of the form (1.1) satisfying the equation (3.5).In view of (4.2) and (3.13), a simple computation shows that

$$\begin{aligned} T_{2,1}(F_{f^{-1}}/2) &= \Gamma_1^2 - \Gamma_2^2 \\ &= \frac{1}{16}\left(-9a_2^4 + 4a_2^2 - 4a_3^2 + 12a_2^2a_3\right) \\ &= \frac{1}{256}\left(-9p_1^4 - 4p_2^2 + 4p_1^2(4 + 3p_2)\right). \end{aligned}$$

Using Lemma 2.1 in above equation, we obtain

$$T_{2,1}(F_{f^{-1}}/2) = \frac{1}{256}\left(16p_1^2 - 4p_1^4 + 16p_1^2\zeta - 4p_1^4\zeta - 16\zeta^2 + 8p_1^2\zeta^2 - p_1^4\zeta^2\right).$$

In view of triangle inequality, we see above equation can be expressed as:

$$|T_{2,1}(F_{f^{-1}}/2)| \leq \frac{1}{256}\left(16|p_1^2| + 4|p_1^4| + 16|p_1^2||\zeta| + 4|p_1^4||\zeta| + 16|\zeta^2| + 8|p_1^2||\zeta^2| + |p_1^4||\zeta^2|\right).$$

Assume $p := |p_1|$ where $p_1 \geq 0$ by analogous reasoning in Theorem 4.1 and for the case $|\zeta| \leq 1$, we obtain the following result by taking the help of Lemma 2.1.

$$|T_{2,1}(F_{f^{-1}}/2)| \leq \frac{1}{256}\left(9p^4 + 40p^2 + 16\right).$$

Since $p \in [0,2]$. Therefore,

$$\frac{1}{16} \leq |T_{2,1}(F_{f^{-1}}/2)| \leq \frac{5}{4}. \tag{4.5}$$

To demonstrate the sharpness of the upper bound, consider the analytic function $f_4 \in \mathcal{S}^*_\rho$, defined by:

$$f_4(z) = z \exp\left(\int_0^z \frac{\sinh^{-1}(\sqrt{20}it^2)}{t}\, dt\right) = z + \sqrt{5}iz^3 - \frac{5z^5}{2} + \frac{5}{18}\sqrt{5}iz^7 + \dots .$$

For this function, a straightforward calculation yields

$$|T_{2,1}(F_{f^{-1}}/2)| = \frac{5}{4}.$$

For the sharpness of the lower bound, consider the function f_2, defined in Theorem 4.1. Thus, we achieve equality in (4.5) which concludes our proof. ∎

Conflict of interest: The authors declare no conflict of interest.
Authors' Contribution: Each author contributed equally to the research and preparation of the manuscript.

References

[1] Arora, K., Kumar, S. S.: Starlike functions associated with a petal shaped domain, Bulletin of the Korean Mathematical Society. Korean Mathematical Society, **59**(4), (2022), 993–1010.

[2] Carathéodory, C.: Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen, Mathematische Annalen, **64**(1), (1907), 95–115.

[3] Choi, J. H., Kim, Y. C., Sugawa, T.: A general approach to the Fekete-Szegö problem, Journal of the Mathematical Society of Japan, **59**(3), (2007), 707–727.

[4] Giri, S., Kumar, S. S.: Toeplitz determinants in one and higher dimensions. Acta Mathematica Scientia, **44**, (2024), 1931–1944.

[5] Kowalczyk, B., Lecko, A.: The Fekete-Szegö inequality for close-to-convex functions with respect to a certain starlike function dependent on a real parameter, Journal of Inequalities and Applications, (2014), 1–16.

[6] Kumar, S. S., Verma, N.: Coefficient problems for starlike functions associated with a petal shaped domain, arXiv:2208.14644, (2022).

[7] Libera, R. J., Złotkiewicz, E. J., Early coefficients of the inverse of a regular convex function, Proceeding of the American Mathematical Society, **85**(2), (1982), 225–230.

[8] Ma, W. C., Minda, D.: A unified treatment of some special classes of univalent functions, Proceedings of the Conference on Complex Analysis, International Press Inc., (1992).

[9] Mundalia, M., Kumar, S. S.: Coefficient problems for certain close-to-convex functions, Bulletin of the Iranian Mathematical Society, **49**(1), p.5, (2023).

[10] Pommerenke, C.: On the Hankel determinants of univalent functions, Mathematika, **14**(1), (1967), 108–112.

[11] Pommerenke, C.: Univalent functions, Vandenhoeck & Ruprecht, Göttingen, (1975).

[12] Ponnusamy, S., Sharma, N. L., Wirths, K. J.: Logarithmic coefficients of the inverse of univalent functions, Results in Mathematics, **73**, (2018), 1–20.

[13] Ponnusamy, S., Sharma, N. L., Wirths, K. J.: Logarithmic coefficients problems in families related to starlike and convex functions, Journal of the Australian Mathematical Society, **109**(2), (2020), 230–249.

[14] Verma, N., Kumar, S. S.: A conjecture on $H_3(1)$ for certain starlike functions, Mathematica Slovaca, **73**(5), (2023), 1197–1206.

[15] Ye, K., Lim, L. H.: Every matrix is a product of Toeplitz matrices, Foundations of Computational Mathematics, **16**(3), (2016), 577–598.

Department of Applied Mathematics, Delhi Technological University, Bawana Road, Delhi-110042, INDIA

Email address: spkumar@dce.ac.in

Department of Applied Mathematics, Delhi Technological University, Bawana Road, Delhi-110042, INDIA

Email address: aryatripathiiii@gmail.com

Department of Applied Mathematics, Delhi Technological University, Bawana Road, Delhi-110042, INDIA

Email address: s02nehal@gmail.com

Physica Scripta

INCLUSIVE PUBLISHING TRUSTED SCIENCE

PURPOSE-LED PUBLISHING™

PAPER

On novel Hexa-compound combination synchronization over nineteen *n*-dimensional category-B chaotic systems and electronic circuit schematic

To cite this article: Dinesh Khattar *et al* 2025 *Phys. Scr.* **100** 015280

View the article online for updates and enhancements.

You may also like

- Dynamics of Transiently Chaotic Neural Network and Its Application to Optimization
Yang Li-Jiang, Chen Tian-Lun and Huang Wu-Qun
- A New 4-D Chaotic System with Self-Excited Two-Wing Attractor, its Dynamical Analysis and Circuit Realization
A Sambas, S Vaidyanathan, S Zhang et al.
- Hybrid synchronisation method based on inverse generalised and inverse projected high dimensional discrete chaotic systems
Ya'nan Xing, Jian Zeng, Wenjie Dong et al.

This content was downloaded from IP address 14.139.251.98 on 30/12/2024 at 07:12

Physica Scripta

PAPER

On novel Hexa-compound combination synchronization over nineteen *n*-dimensional category-B chaotic systems and electronic circuit schematic

RECEIVED
30 September 2024

REVISED
13 November 2024

ACCEPTED FOR PUBLICATION
12 December 2024

PUBLISHED
26 December 2024

Dinesh Khattar[1], **Naokant Deo**[2] **and Mukul Sirohi**[2]

[1] Department of Mathematics, Kirori Mal College, University of Delhi, Delhi-110 007, India
[2] Department of Applied Mathematics, Delhi Technological University, Delhi-110 042, India

E-mail: mukul7sirohi@gmail.com

Keywords: chaos, synchronization, category-B, circuit schematic, bifurcation, hexa-compound combination, Lyapunov exponents

Abstract

Synchronization of chaotic models involving multiple drives and responses has numerous practical applications in cryptography and information processing. Existing research on synchronizing multiple chaotic systems is currently limited to twelve components. This study introduces a novel higher-order synchronization method, hexa compound combination, that synchronizes an assembly of nineteen *n*-dimensional chaotic models. Well-known synchronization methods, such as double compound, triple compound, and quad compound, serve as particular instances of this new strategy. Thus, our research significantly advances the understanding of multi-leveled chaos synchronization. In addition, we also present a non-uniformly conservative system classified into a rare category B, analyze its dynamic properties, and utilize it for achieving the proposed synchronization. Numerical results are provided through graphical representations to illustrate the efficacy of the new synchronization approach by comparing it with other techniques. Furthermore, we emulate the corresponding virtual schematic circuit of the newly designed system to evaluate its real-world applicability and utility.

1. Introduction

In recent decades, chaos theory has drawn significant attention from experts in virtually all possible domains. In nonlinear dynamics, a deterministic dynamical system is considered chaotic if it exhibits extreme sensitivity to any change in its initial circumstances with several other dynamical properties [1]. An important fact to mention here is that regardless of how small these changes are, any ambiguity in the initial state of these models can create significant gaps in our attempt to predict their future dynamics; that is, we can only predict these systems if we know their initial state with infinite precision, which is manually unattainable, resulting them to have a dual character of both determinism and unpredictability [2]. Chaotic systems display remarkable symmetries when graphed despite their unpredictability and intricate structures. Researchers often explore the patterns within this geometry, such as unique, hidden attractors, self-excited, multi-scroll, multi-stability, etc [3–8]. Chaos theory offers a new lens to view our universe, revealing that seemingly ordinary models can generate intriguing yet often strange behaviors. Today, this theory is considered fundamental, impacting numerous facets of our world and bridging various fields of study—underscoring the pervasive nature of chaos.

Chaos synchronization, which involves coordinating the states of two or more chaotic systems in a coupling manner, remarkably contributed to the practicality and applicability of chaotic systems. The research on chaos synchronization is quite interesting because chaotic models should oppose any form of coordination due to their intrinsic attribute of sensitive dependence on initial conditions. However, this phenomenon is formally documented by Pecora and Carroll, who explored potential applications of chaos in the late 1980s [9]. They embedded a message within a chaotic drive signal for secure transmission to a response system. Over the last few decades, numerous such strategies for synchronizing chaotic models have been extensively researched [10–25].

© 2024 IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Most traditional synchronization schemes are limited to a single drive-response configuration. Inquiring whether such investigations can be expanded to accommodate multiple drive-response models becomes compelling and natural. Thus, synchronization concepts incorporating three or more chaotic systems have emerged as an extensive area of research. In 2000, a dual synchronization technique was proposed [26]. The combination synchronization approach [27] for three chaotic systems was explored in study published in 2011. This investigation pioneered the demonstration of the state variables of a response system being effectively synchronized with the cumulative state variables of two driving systems. In 2013, a compound synchronization strategy for four chaotic systems was introduced [28]. Unlike the combination scheme, compound synchronization involves both the summation and multiplication of the state variables from the drive, enabling the output signal to achieve higher complexity and randomness. Then, the scheme of compound combination was proposed to overcome the disadvantage of a single response system in compound synchronization. In 2018, double compound combination synchronization (DCCS) was researched [29], a composite sum of two compound combinations with four responses and combining three compound sums; the authors examined the concept of triple compound synchronization (TCS) for eight chaotic systems [30]. Then, further extensions of these methods, including quad compound combination [31] over the ten chaotic models and the notion of penta compound combination [32] over twelve chaotic models, are also well-researched topics. Today, it is known that synchronization techniques are vital in enhancing the potential information transfer security in various fields, including computing, neural networking, image processing, communication fields, etc [33–40]. Examining multi-leveled synchronization schemes, researchers found enhanced complexity, randomness, and signal ranges within the drive dynamics compared to traditional single configurations, thereby opening many new avenues for research exploration and applications.

Further, the trace of the Jacobian matrix primarily classifies dynamical systems as either dissipative or conservative. In rare occurrences, this trace depends on state variables, resulting in the system showing a non-uniformly conservative behavior [41–45]. As a result, it becomes challenging to determine whether the model is conservative or dissipative owing to this variable dependence. Hence, estimating the LEs is essential to assert the system's nature further. Many studies overlook this variable by performing linear approximations, while others determine these exponents without considering the trace of the Jacobi matrix. In 2019, Jafari *et al* [46] devised a generalized categorization for conservative dynamical models, separating them into A, B, C, and D categories. The same study highlighted that groups A and C are well-established and widely known [47–49]. However, classes B and D are rare and need to be researched more than their counterparts [50–52]. Cang *et al* [53] recently presented a 3D non-uniformly dissipative model with two equilibrium points falling under category B, expanding our understanding of conservative systems. Inspired by this, our research introduces a novel, non-uniformly conservative category B system. We investigated the novel system using attractor visuals, volume contraction, and bifurcation analysis and evaluated its conservative behavior. We also thoroughly reviewed its Lyapunov exponents and general stability, particularly near equilibrium points. To showcase the practicality of this system, we also offer a detailed schematic design of its electronic representation. In addition, adhering to the combinatorial sum definition of DCCS and TCS, we study a novel higher-order synchronization, namely hexa compound combination (HCCS), over nineteen chaotic models. The literature on synchronizing multiple drive-response is currently limited to twelve chaotic models. We use nineteen n-dimensional category B chaotic systems to implement the hexa compound combination methodology comprising seven drives and twelve responses. Various nonlinear controllers are built utilizing Lyapunov's stability concept to carry out the proposed technique. Using MATLAB®, numerical operations are performed, and systematic plots are shown visually to validate the analytical results.

Following is the list of contributions and novelty of this research:

- A new 3D non-uniformly conservative model that falls under the rare category B is introduced.
- The system is analyzed utilizing a theoretical description of its attractors visuals, assessing the volume contraction and conservatism, Lyapunov spectra, and bifurcation analysis.
- A detailed electronic circuit schematic design of the model is proposed, and its emulated results are illustrated.
- The system reveals three equilibrium points in the examined chaotic scenario and possesses a line equilibrium of nature $(x, 0, 0)$, where $x \in \mathbb{R}$ for its conservative case.
- A novel synchronization approach, namely, hexa compound combination, is also examined and implemented to coordinate a total of nineteen n-dimensional chaotic models—significantly advancing the existing literature on higher-order synchronizations that primarily synchronizes only twelve models.

- We design controllers based on Lyapunov's stability principles, enhancing and proving the robustness of the novel synchronization method and ensuring reliable performance across other chaotic, hyper-chaotic, and higher-dimensional systems.
- Utilizing MATLAB®, we performed extensive numerical simulations and generated systematic plots, visually validating our analytical findings and demonstrating the effectiveness of our proposed techniques.

This research investigation proceeds as follows: A new 3D Category B chaotic structure with a non-conservative flow is investigated, followed by a thorough nonlinear analysis in section 2. Section 3 outlines a novel comprehensive problem formulation for achieving hexa compound combinations over nineteen n-dimensional chaotic systems. Section 4 examines the proposed 3D system to this upgraded scheme and discusses several special cases, including the penta compound combination. Numerical simulations and results from the previously outlined sections are reported and addressed in section 5. In section 6, a comparison between the hexa compound combination and the penta compound combination method is conducted. The electronic circuit blueprint for the 3D Category B chaotic system, along with captured outputs, is presented in section 7. Finally, section 8 delivers the conclusion of this investigation.

2. System description and analysis

In 2019, Jafari *et al* [46] created a general classification for conservative dynamical models. It had four groups (A, B, C, and D) based on the Jacobian matrix trace and the associated Lyapunov exponents. They also emphasized that categories A and C are well established and widely known. However, models belonging to classes B and D are not as prevalent as their counterparts. This inspired us to develop a novel, non-uniformly conservative system that falls into the above-mentioned category B. The following array of equations describe the novel deterministic model:

$$\begin{cases} \frac{dx}{dt} = a(y - x) + z, \\ \frac{dy}{dt} = by - x(z + dy), \\ \frac{dz}{dt} = xy - cz - y. \end{cases} \tag{1}$$

We name this 3D (continuous) autonomous dynamical model as '$\mathscr{M}-$system' comprising a, b, c, and d. Attractors of $\mathscr{M}-$system using $(a, b, c, d) = (36, 25, 3, 0.5)$ are presented in figure 1. An attractor can be perceived as a collection of points onto which the system seems to gravitate, regardless of its original state, when viewed in phase space [54]. To understand it mathematically, let x_0 be the initial value of $\mathscr{M}-$system and $x(t)$ be the state at which x_0 arrives after time (t) then a non-empty subset $\mathcal{A}$ of phase-space is characterized as attractor:

• If $x(t') \in \mathcal{A}$ for some t', then $x(t) \in \mathcal{A}\ \forall\ t > 0$. This attribute is many times regarded as the invariance property of the attractor.

• If $\exists$ a basin of attraction around $\mathcal{A}$ such that all the trajectories originating in ultimately settles down to $\mathcal{A}$.

• Such an $\mathcal{A}$ should be minimal, i.e. the above two features cannot hold for any other proper subset of $\mathcal{A}$.

Each trajectory of the $\mathscr{M}-$system in its chaotic regime exhibits a long-term aperiodic conduct. It implies that as $t \to \infty$, there are trajectories that never converge to bits, fixed points, quasiperiodic, or periodic orbits, notably for specified and examined parameters.

2.1. Lyapunov exponents

To understand Lyapunov exponents (LEs) [55], imagine the evolution of a minuscule sphere representing the dynamics of a chaotic system in $\mathbb{R}^n-$space. Now, choose the same system with slight alterations in its starting conditions. The sphere will gradually transform into an infinitesimally stretched ellipsoid during its evolutionary process. Consider σ_k, in which $k = 1, 2, \ldots, n$, the measure of the k-th primary axis of this ellipsoid. The expression $\sigma_k(t) \approx \sigma_k(0)e^{L_k t}$, wherein L_k symbolizes the LEs of the dynamical system. When t is large enough, the diameter of the ellipsoid is mostly decided by the most positive L_k, also known as the largest LE. The LEs of the $\mathscr{M}-$system are computed on MATLAB® using the well-grounded Wolf algorithm [56], and their values are $LE_1 \approx 1.334$, $LE_2 \approx 0$, and $LE_3 \approx -12.663$. For $(a, b, c, d) = (36, 25, 3, 0.5)$, the category B system has one negative, one positive, and one LE converging zero. The $\mathscr{M}-$system has chaotic dynamics with $LE_1 + LE_2 + LE_3 < 0$ and largest exponent, $LE_1 = 1.334 > 0$ for $t \in [0, 1000]$.

2.2. Volume contraction and non-uniformly conservatism

Assume that the $\mathscr{M}-$system is represented with $\dot{X}(t) = F(X)$. Choose any closed but arbitrary surface $C(t)$ containing volume $Vol(t)$ in state-space without any other specific constraint. Let the points on C as starting

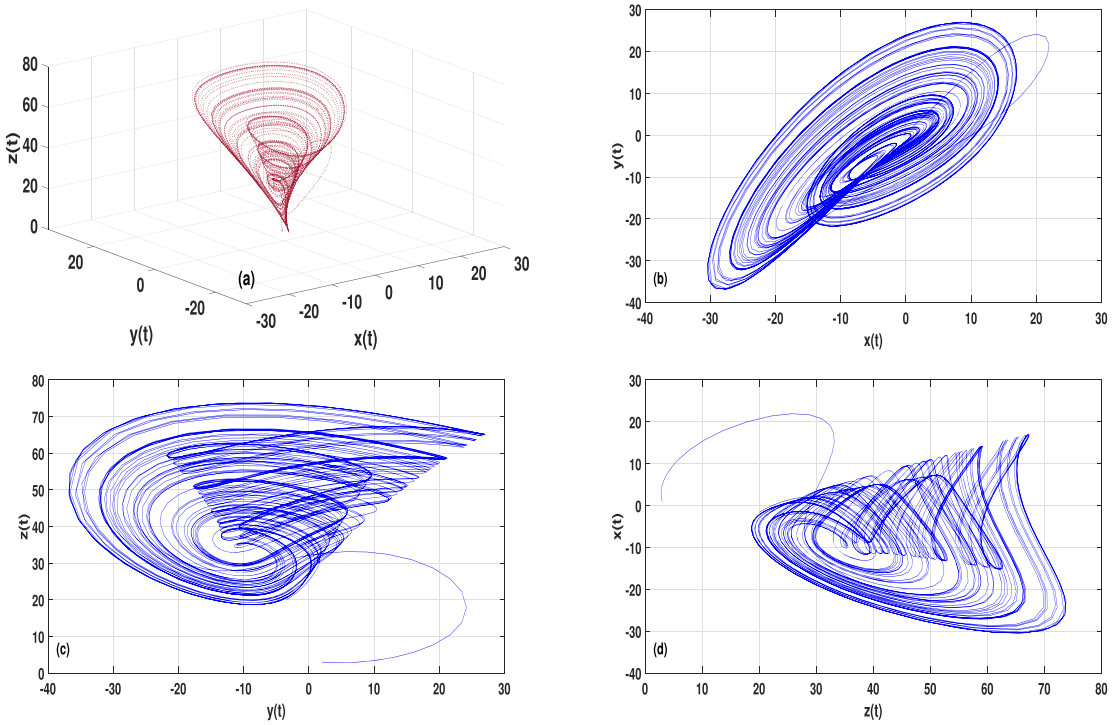


Figure 1. Phase-portraits of the $\mathscr{M}-$system in: (**a**) space $x-y-z$, (**b**) plane $x-y$, (**c**) plane $y-z$, and (**d**) plane $z-x$.

conditions in the $\mathscr{M}-$system trajectories, and allow them to evolve for a tiny duration (dt). Subsequently, the surface C transforms and becomes the new surface, say, $C(t+dt)$. Let $\hat{n}$ represent the outward normal vector emerging from surface C. Since F depicts the instantaneous velocity of the points, $(F\cdot\hat{n})$ should denote the outside normal element for this velocity. Thus, over a small period dt, a patch with an area $d(Ar)$ will cover a volume of $(F\cdot\hat{n})\,(dt)\,d(Ar)$. Thus,

$$Vol(t+dt) = \bar{V}ol(t) + Vol(t). \tag{2}$$

where $\bar{V}ol(t)$ is the volume generated by infinitesimal sections of a surface, summed across all sections, i.e.

$$Vol(t+dt) = \left(\int_C (F\cdot\hat{n})\,(dt)\,d(Ar)\right) + Vol(t). \tag{3}$$

which further transforms to

$$\frac{Vol(t+dt)-Vol(t)}{dt} = \frac{d(Vol)}{dt} = \int_C (F\cdot\hat{n}) \quad d(Ar). \tag{4}$$

Utilizing the divergence theorem [57], we get

$$\frac{d(Vol)}{dt} = \int_{Vol} (\nabla\cdot F)\,d(Vol). \tag{5}$$

As $(\nabla\cdot F) = b-a-dx-c$. So, when parameter $d=0$ then

$$\Rightarrow\ Vol(t) = Vol(0)e^{-(a+c-b)t}. \tag{6}$$

The volume of $\mathscr{M}-$system shrinks exponentially fast when $(a+c-b)>0$, i.e. if we begin with an enormous mass that represents starting conditions, it subsequently compresses to a limiting mass of zero volume. In another sense, all trajectories originating from the initial mass gradually converge to a confining set consisting of limit cycles, fixed points, or strange attractors. The possible solutions for the $\mathscr{M}-$system in this case are limited due to this pronounced volume restriction. As $(\nabla\cdot F) = tr(J)$, the $\mathscr{M}-$system majorly shows a conservative nature [58] whenever $a+c=b$ and $d=0$. However, interestingly, when $d\neq 0$, the system also reveals conservatism while keeping all parameters zero and $d=0.00255\bar{}$. The LEs in this special case with initial conditions $(0,-2,0)$ for 6635 iterations are $(0.097, 0, -0.097)$ proving $(\nabla\cdot F)=0$. This is in fact true for $L_\delta = \{(0,\gamma,0): \gamma\in(-2-\delta,-2+\delta)\}$ where $\delta>0$ is arbitrarily small, that is, $\lim_{\gamma\to-2}\sum_{i=1}^{3} LE_i(0,\ \gamma,\ 0)\approx 0$ for L_δ. The conservative nature of the $\mathscr{M}-$system can be visualized through figure 2.

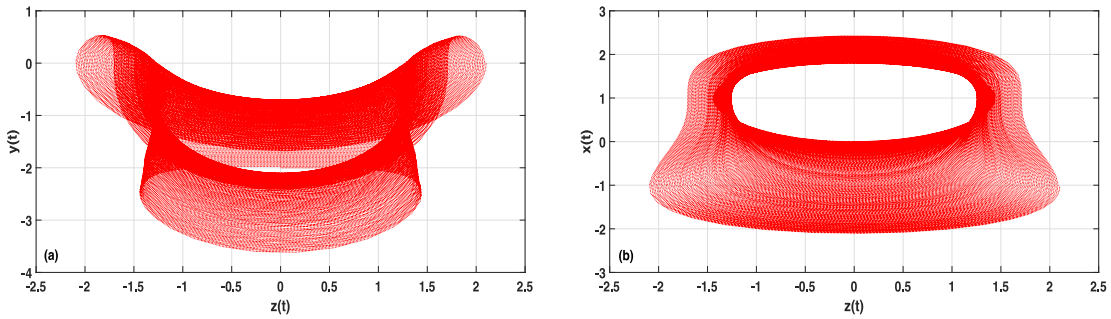


Figure 2. The conservative nature of the $\mathscr{M}-$system with parameter $d = 0.002\bar{5}$ in: (**a**) $z-y$ plane and (**b**) $z-x$ plane.

Table 1. The local stability of equilibrium points in the $\mathscr{M}-$system.

Equilibrium	Eigenvalues of matrix (J)	Nature
E_1	$(-36, -3, 25)$	Unstable
E_2	$(-20.93, 1.36+15.77\,\iota, 1.36–15.77\,\iota)$	Unstable
E_3	$(-20.46, 5.45+15.55\,\iota, 5.45–15.55\,\iota)$	Unstable

2.3. Local stability analysis

This part addresses the regional stability of the equilibrium points for the integer-order chaotic, $\mathscr{M}-$system. Chaotic systems commonly exhibit unstable equilibrium points. The equilibrium spots of the $\mathscr{M}-$system can be determined by computing the following three equations:

$$\begin{cases} z + a(y - x) = 0, \\ by - x(dy + z) = 0, \\ xy - cz - y = 0. \end{cases} \tag{7}$$

For $(a, b, c, d) = (36, 25, 3, 0.5)$, there are three equilibrium's for the $\mathscr{M}-$system in $\mathbb{R}^3$, given by

$$\begin{cases} E_1 = (0,\ 0,\ 0), \\ E_2 \approx (8.413,\ 7.873,\ 19.457), \\ E_3 \approx (-8.913,\ -9.814,\ 32.434). \end{cases} \tag{8}$$

The linearized (J) matrix for the $\mathscr{M}-$system is

$$J = \begin{bmatrix} -a & a & 1 \\ -z - dy & b - dx & -x \\ y & x - 1 & -c \end{bmatrix}. \tag{9}$$

The trace of (J) depends on variable x with non-uniformly conservative behavior, in addition to having three locally unstable equilibrium points depicted in table 1. However, for conservative case the $\mathscr{M}-$system also possesses line equilibrium of type $(x, 0, 0)$, where $x \in \mathbb{R}$. All the mentioned factors indicates that the $\mathscr{M}-$system belongs to the category B family.

2.4. Bifurcation analysis

In nonlinear analysis, bifurcation diagrams [59] help to understand the dependence of a dynamical system's orbital nature over its parameters. The bifurcation analysis of the $\mathscr{M}-$system is scrutinized through figure 3 by charting the regional maximums of variable x and simultaneously altering the value of parameter a over $[30, 33.2]$. These bifurcation charts can be verified using the LEs spectrum provided in figure 4. When parameter a is varied, $\mathscr{M}-$system switches between different regimes: periodicity, quasi-periodicity, and chaos.

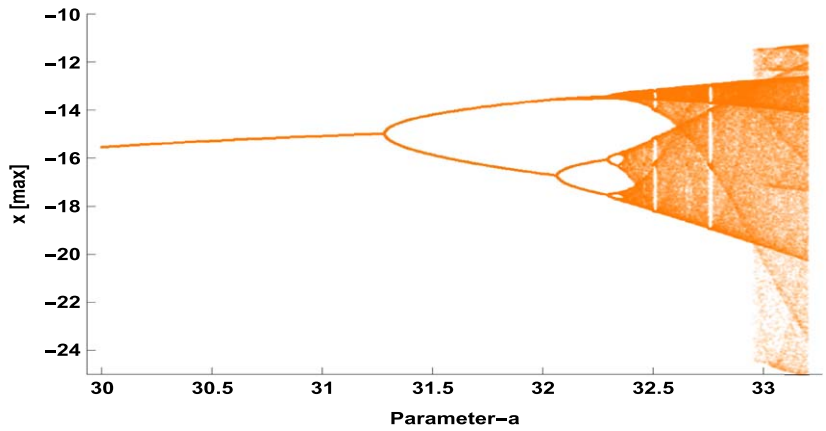


Figure 3. Bifurcation in the $\mathscr{M}-$system for $30\leqslant a\leqslant 33.2$.

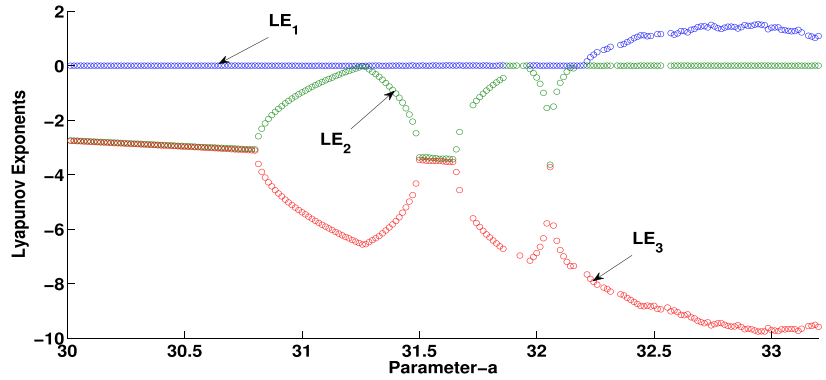


Figure 4. The Lyapunov spectrum of the $\mathscr{M}-$system on [30, 33.2].

3. Problem formulation

A novel synchronization approach of hexa compound combination (HCCS) between seven drives and twelve responses of n-dimensional chaotic structures is discussed in this section. To begin, consider four drive models given by

$$\begin{cases} \dot{X}_1(t) = F_1(X_1(t)), \\ \dot{X}_2(t) = F_2(X_2(t)), \\ \dot{X}_3(t) = F_3(X_3(t)), \\ \dot{X}_4(t) = F_4(X_4(t)). \end{cases} \tag{10}$$

and, another triplet of drive systems given by

$$\begin{cases} \dot{X}_5(t) = F_5(X_5(t)), \\ \dot{X}_6(t) = F_6(X_6(t)), \\ \dot{X}_7(t) = F_7(X_7(t)). \end{cases} \tag{11}$$

where $F_\varrho(X_\varrho) = \text{diag}(F_{\varrho 1}(X_\varrho), F_{\varrho 2}(X_\varrho), \ldots, F_{\varrho n}(X_\varrho))$ represents the diagonal matrices comprising of smooth nonlinear vector functions and $X_\varrho = \text{diag}(X_{\varrho 1}, X_{\varrho 2}, \ldots, X_{\varrho n})$ represents the states of drives in (10) and (11). Now, suppose the corresponding twelve response models as

$$\begin{cases} \dot{Y}_1(t) = G_1(Y_1(t)) + \Theta_1, \\ \dot{Y}_2(t) = G_2(Y_2(t)) + \Theta_2, \\ \vdots \\ \dot{Y}_{12}(t) = G_{12}(Y_{12}(t)) + \Theta_{12}. \end{cases} \tag{12}$$

where each $Y_\xi = \text{diag}\,(Y_{\xi 1}, Y_{\xi 2}, \ldots, Y_{\xi n})$ represents the state vectors of response models in (12), $G_\xi(Y_\xi) = \text{diag}\,(G_{\xi 1}(Y_\xi), G_{\xi 2}(Y_\xi), \ldots, G_{\xi n}(Y_\xi))$ represents the diagonal matrix of smooth nonlinear vector functions, and $\Theta_\xi = \text{diag}(\Theta_{\xi 1}, \Theta_{\xi 2}, \ldots, \Theta_{\xi n})$ are twelve controllers that need to be designed in order to achieve HCCS.

Definition 3.1. The **Hexa Compound Combination** synchronization is said to attain in-between nineteen chaotic models, if there exists nineteen non-zero matrices of the form $M_{\bar{\varrho}}$, M_ϱ, and N_ξ in $\mathbb{R}^{n\times n}$ with

$$\lim_{t\to\infty} \| \underbrace{\left(\sum_{\varrho=1}^{4} M_\varrho X_\varrho\right)\left(\sum_{\bar{\varrho}=5}^{7} M_{\bar{\varrho}} X_{\bar{\varrho}}\right) - \left(\sum_{\xi=1}^{12} N_\xi Y_\xi\right)}_{E(t)} \| = 0. \tag{13}$$

where $\|.\|$ is the vectornorm, $M_{\bar{\varrho}}$ represent the scale drive matrices, M_ϱ represents the base drive matrices, and N_ξ are the responses matrices.

Remark 3.2. The synchronization error mentioned within the definition (3.1) may also be phrased as $E(t) = \sum_{i=1}^{6} \tilde{E}_i(t)$, and can be interpret as the sums of six individual errors in the synchronization of compound combinations. Thus, we can translate the novel approach (13) as a hexa compound combination (HCCS). In particular, $\tilde{E}_1 = [M_5X_5(M_1X_1 + M_2X_2) - N_1Y_1 - N_2Y_2]$, $\tilde{E}_2 = [M_5X_5(M_3X_3 + M_4X_4) - N_3Y_3 - N_4Y_4]$, $\tilde{E}_3 = [M_6X_6(M_1X_1 + M_2X_2) - N_5Y_5 - N_6Y_6]$, $\tilde{E}_4 = [M_6X_6(M_3X_3 + M_4X_4) - N_7Y_7 - N_8Y_8]$, $\tilde{E}_5 = [M_7X_7(M_1X_1 + M_2X_2) - N_9Y_9 - N_{10}Y_{10}$, and $\tilde{E}_6 = [M_7X_7(M_3X_3 + M_4X_4) - N_{11}Y_{11} - N_{12}Y_{12}]$.

Remark 3.3. If any one component of $M_{\bar{\varrho}}X_{\bar{\varrho}}(M_1X_1 + M_2X_2)$ or $M_{\bar{\varrho}}X_{\bar{\varrho}}(M_3X_3 + M_4X_4)$ is neglected for $(\bar{\varrho} = 5, 6, 7)$ in addition to any two of $N_\xi = 0, \xi = 1, 2, \ldots, 12$ then HCCS reduces into synchronization scheme of penta compound combinations (PCCS).

Remark 3.4. If any one of the $M_{\bar{\varrho}} = 0$, $\bar{\varrho} = 5, 6, 7$ in addition to any four of $N_\xi = 0, \xi = 1, 2, \ldots, 12$, then HCCS converts to the problem of quad compound combination synchronization (QCCS).

Remark 3.5. If we take any two $M_\varrho = 0, \varrho = 1, 2, \ldots, 4$ in addition to any six of $N_\xi = 0, \xi = 1, 2, \ldots, 12$, then HCCS converts to the problem of triple compound combination synchronization (TCCS).

Remark 3.6. If we take any two $M_{\bar{\varrho}} = 0$, $\bar{\varrho} = 5, 6, 7$ in addition to any eight of $N_\xi = 0, \xi = 1, 2, \ldots, 12$, then HCCS reduces to the synchronization statement of double compound combinations (DCCS).

Remark 3.7. If any six of $N_\xi = 0, \xi = 1, 2, \ldots, 12$, then HCCS converts to hexa compound synchronization (HCS).

Remark 3.8. If five component from $M_{\bar{\varrho}}X_{\bar{\varrho}}(M_1X_1 + M_2X_2)$ and $M_{\bar{\varrho}}X_{\bar{\varrho}}(M_3X_3 + M_4X_4)$ are zero for $(\bar{\varrho} = 5, 6, 7)$ in addition to any eleven of $N_\xi = 0, \xi = 1, 2, \ldots, 12$ then HCCS reduces into compound combination synchronization (CCS).

3.1. Results

We aim to construct the controllers to achieve hexa compound combination synchronization on nineteen n-dimensional models and discuss some important results in this section. Choose the components of drives scaling matrices $M_\varrho = \text{diag}(m_{\varrho 1}, m_{\varrho 2}, \ldots, m_{\varrho n})$; for $\varrho = 1, 2, \ldots, 4$, $M_{\bar{\varrho}} = \text{diag}(m_{\bar{\varrho}1}, m_{\bar{\varrho}2}, \ldots, m_{\bar{\varrho}n})$; for $\bar{\varrho} = 5, 6, 7$, and $N_\xi = \text{diag}(n_{\xi 1}, n_{\xi 2}, \ldots, n_{\xi n})$; for $\xi = 1, 2, \ldots, 12$. The drive to response errors in HCCS configuration are as:

$$\begin{cases} E_1 = \left(\sum_{\varrho=1}^{4} m_{\varrho 1}X_{\varrho 1}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho}1}X_{\bar{\varrho}1}\right) - \sum_{\xi=1}^{12} n_{\xi 1}Y_{\xi 1}, \\ E_2 = \left(\sum_{\varrho=1}^{4} m_{\varrho 2}X_{\varrho 2}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho}2}X_{\bar{\varrho}2}\right) - \sum_{\xi=1}^{12} n_{\xi 2}Y_{\xi 2}, \\ \vdots \\ E_n = \left(\sum_{\varrho=1}^{4} m_{\varrho n}X_{\varrho n}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho}n}X_{\bar{\varrho}n}\right) - \sum_{\xi=1}^{12} n_{\xi n}Y_{\xi n}. \end{cases} \tag{14}$$

Using equations in (10)–(12), the HCCS errors designated by (14) converts to

$$\begin{cases} \dot{E}_1 = \left(\sum_{\varrho=1}^{4} m_{\varrho 1} F_{\varrho 1}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 1} X_{\bar{\varrho} 1}\right) + \left(\sum_{\varrho=1}^{4} m_{\varrho 1} X_{\varrho 1}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 1} F_{\bar{\varrho} 1}\right) \\ \qquad -\sum_{\xi=1}^{12} n_{\xi 1} G_{\xi 1} - \underbrace{\sum_{\xi=1}^{12} n_{\xi 1} \Theta_{\xi 1}}_{\Theta_1(t)}, \\ \dot{E}_2 = \left(\sum_{\varrho=1}^{4} m_{\varrho 2} F_{\varrho 2}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 2} X_{\bar{\varrho} 2}\right) + \left(\sum_{\varrho=1}^{4} m_{\varrho 2} X_{\varrho 2}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 2} F_{\bar{\varrho} 2}\right) \\ \qquad -\sum_{\xi=1}^{12} n_{\xi 2} G_{\xi 2} - \underbrace{\sum_{\xi=1}^{12} n_{\xi 2} \Theta_{\xi 2}}_{\Theta_2(t)} \\ \vdots \\ \dot{E}_n = \left(\sum_{\varrho=1}^{4} m_{\varrho n} F_{\varrho n}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} n} X_{\bar{\varrho} n}\right) + \left(\sum_{\varrho=1}^{4} m_{\varrho n} X_{\varrho n}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} n} F_{\bar{\varrho} n}\right) \\ \qquad -\sum_{\xi=1}^{12} n_{\xi n} G_{\xi n} - \underbrace{\sum_{\xi=1}^{12} n_{\xi n} \Theta_{\xi n}}_{\Theta_n(t)}. \end{cases} \tag{15}$$

Theorem 3.9. *If the controllers $\Theta_p(t)'s$ are formulated as*

$$\begin{cases} \Theta_1(t) = \left(\sum_{\varrho=1}^{4} m_{\varrho 1} F_{\varrho 1}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 1} X_{\bar{\varrho} 1}\right) + \left(\sum_{\varrho=1}^{4} m_{\varrho 1} X_{\varrho 1}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 1} F_{\bar{\varrho} 1}\right) \\ \qquad -\sum_{\xi=1}^{12} n_{\xi 1} G_{\xi 1} + K_1 E_1, \\ \Theta_2(t) = \left(\sum_{\varrho=1}^{4} m_{\varrho 2} F_{\varrho 2}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 2} X_{\bar{\varrho} 2}\right) + \left(\sum_{\varrho=1}^{4} m_{\varrho 2} X_{\varrho 2}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 2} F_{\bar{\varrho} 2}\right) \\ \qquad -\sum_{\xi=1}^{12} n_{\xi 2} G_{\xi 2} + K_2 E_2, \\ \vdots \\ \Theta_n(t) = \left(\sum_{\varrho=1}^{4} m_{\varrho n} F_{\varrho n}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} n} X_{\bar{\varrho} n}\right) + \left(\sum_{\varrho=1}^{4} m_{\varrho n} X_{\varrho n}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} n} F_{\bar{\varrho} n}\right) \\ \qquad -\sum_{\xi=1}^{12} n_{\xi n} G_{\xi n} + K_n E_n. \end{cases} \tag{16}$$

where $\mathrm{K} = diag(\mathrm{K}_1, \mathrm{K}_2, \ldots, \mathrm{K}_n)$ *is a strictly positive definite matrix in* $\mathbb{R}$. *Then, the desired alignment of HCCS is attained theoretically between nineteen n-dimensional arbitrary chaotic systems.*

Proof. To adhere the necessary conditions of Barbalat's lemma [60], we opt the Lyapunov's function as:

$$V(E_1, E_2, \ldots, E_n) = \frac{1}{2}\sum_{j=1}^{n} E_j^2. \tag{17}$$

Since $V(E) = V(E_1, E_2, \ldots, E_n) > 0$, the opted quadratic function is positive definite. Thus

$$\dot{V} = \left(\sum_{j=1}^{n} E_j \cdot \dot{E}_j\right). \tag{18}$$

Utilizing (15), we get

$$\begin{cases} \dot{V} = \sum_{j=1}^{n} E_j \cdot [(m_{1j} F_{1j} + m_{2j} F_{2j} + m_{3j} F_{3j} + m_{4j} F_{4j})(m_{5j} X_{5j} + m_{6j} X_{6j} + m_{7j} X_{7j}) \\ \quad + (m_{1j} X_{1j} + m_{2j} X_{2j} + m_{3j} X_{3j} + m_{4j} X_{4j})(m_{5j} F_{5j} + m_{6j} F_{6j} + m_{7j} F_{7j}) \\ \quad - (n_{1j} G_{1j} + n_{2j} G_{2j} + n_{3j} G_{3j} + n_{4j} G_{4j} + n_{5j} G_{5j} + n_{6j} G_{6j} + n_{7j} G_{7j} \\ \quad + n_{8j} G_{8j} + n_{9j} G_{9j} + n_{[10]j} G_{[10]j} + n_{[11]j} G_{[11]j} + n_{[12]j} G_{[12]j}) - \Theta_j]. \end{cases} \tag{19}$$

Substituting from (16) into (19), we obtain

$$\dot{V} = -\left(\sum_{j=1}^{n} K_j E_j^2\right) < 0. \tag{20}$$

It essentially proves the negative definiteness of the function $\dot{V}$. As a result, it establishes the fact that for each n; $\lim_{t\to\infty} \|E_n\| = 0$ and nineteen n-dimensional arrays accomplishes the hexa compound combination technique.

□

4. Illustrative example

In this part, the technique outlined in section 3 will be implemented over 3D chaotic systems that was initially described in section 2. A total of nineteen category B chaotic models with different initial values will be synchronized with the proposed method of hexa compound combinations. With the aid of this case study, we shall validate the theoretical findings and later check their viability in MATLAB®. To initiate, consider the seven 3D drive category B systems as

$$\begin{cases} \dot{X}_{11} = a(X_{12} - X_{11}) + X_{13}, \\ \dot{X}_{12} = bX_{12} - X_{11}(X_{13} + dX_{12}), \\ \dot{X}_{13} = X_{11}X_{12} - cX_{13} - X_{12}. \\ \dot{X}_{21} = a(X_{22} - X_{21}) + X_{23}, \\ \dot{X}_{22} = bX_{22} - X_{21}(X_{23} + dX_{22}), \\ \dot{X}_{23} = X_{21}X_{22} - cX_{23} - X_{22}. \\ \vdots \\ \dot{X}_{71} = a(X_{72} - X_{71}) + X_{73}, \\ \dot{X}_{72} = bX_{72} - X_{71}(X_{73} + dX_{72}), \\ \dot{X}_{73} = X_{71}X_{72} - cX_{73} - X_{72}. \end{cases} \tag{21}$$

Now suppose the corresponding twelve 3D category B response models be

$$\begin{cases} \dot{Y}_{11} = a(Y_{12} - Y_{11}) + Y_{13} + \Theta_{11}, \\ \dot{Y}_{12} = bY_{12} - Y_{11}(Y_{13} + dY_{12}) + \Theta_{12}, \\ \dot{Y}_{13} = Y_{11}Y_{12} - cY_{13} - Y_{12} + \Theta_{13}. \\ \dot{Y}_{21} = a(Y_{22} - Y_{21}) + Y_{23} + \Theta_{21}, \\ \dot{Y}_{22} = bY_{22} - Y_{21}(Y_{23} + dY_{22}) + \Theta_{22}, \\ \dot{Y}_{23} = Y_{21}Y_{22} - cY_{23} - Y_{22} + \Theta_{23}. \\ \vdots \\ \dot{Y}_{[12]1} = a(Y_{[12]2} - Y_{[12]1}) + Y_{[12]3} + \Theta_{[12]1}, \\ \dot{Y}_{[12]2} = bY_{[12]2} - Y_{[12]1}(Y_{[12]3} + dY_{[12]2}) + \Theta_{[12]2}, \\ \dot{Y}_{[12]3} = Y_{[12]1}Y_{[12]2} - cY_{[12]3} - Y_{[12]2} + \Theta_{[12]3}. \end{cases} \tag{22}$$

Define the error by

$$\begin{cases} e_1 = \sum_{\xi=1}^{12} n_{\xi 1}Y_{\xi 1} - \left(\sum_{\varrho=1}^{4} m_{\varrho 1}X_{\varrho 1}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 1}X_{\bar{\varrho} 1}\right), \\ e_2 = \sum_{\xi=1}^{12} n_{\xi 2}Y_{\xi 2} - \left(\sum_{\varrho=1}^{4} m_{\varrho 2}X_{\varrho 2}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 2}X_{\bar{\varrho} 2}\right), \\ e_3 = \sum_{\xi=1}^{12} n_{\xi 3}Y_{\xi 3} - \left(\sum_{\varrho=1}^{4} m_{\varrho 3}X_{\varrho 3}\right)\left(\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 3}X_{\bar{\varrho} 3}\right). \end{cases} \tag{23}$$

with components of scaling matrices in drives as $M_{\varrho} = \text{diag}(m_{\varrho 1}, m_{\varrho 2}, m_{\varrho 3})$; $\varrho = 1, 2, \ldots, 4$, $M_{\bar{\varrho}} = \text{diag}(m_{\bar{\varrho} 1}, m_{\bar{\varrho} 2}, m_{\bar{\varrho} 3})$; for $\bar{\varrho} = 5, 6, 7$, and for responses as $N_{\xi} = \text{diag}(n_{\xi 1}, n_{\xi 2}, n_{\xi 3})$; $\xi = 1, 2, \ldots, 12$.

Theorem 4.1. *The twelve response systems in (22) are aligned in novel synchronization of hexa compound combinations with the seven drive systems in (21), if we choose controllers Θ_1, Θ_2, and Θ_3 as*

$$\begin{cases} \Theta_1 = \left[\sum_{\varrho=1}^{4} m_{\varrho 1}(a(X_{\varrho 2} - X_{\varrho 1}) + X_{\varrho 3})\right]\left[\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 1}X_{\bar{\varrho} 1}\right] + \left[\sum_{\varrho=1}^{4} m_{\varrho 1}X_{\varrho 1}\right] \\ \quad \left[\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 1}(a(X_{\bar{\varrho} 2} - X_{\bar{\varrho} 1}) + X_{\bar{\varrho} 3})\right] - \sum_{\xi=1}^{12} n_{\xi 1}(a(Y_{\xi 2} - Y_{\xi 1}) + Y_{\xi 3}) - k_1 e_1, \\ \Theta_2 = \left[\sum_{\varrho=1}^{4} m_{\varrho 2}(bX_{\varrho 2} - X_{\varrho 1}(X_{\varrho 3} + dX_{\varrho 2}))\right]\left[\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 2}X_{\bar{\varrho} 2}\right] + \left[\sum_{\varrho=1}^{4} m_{\varrho 2}X_{\varrho 2}\right] \\ \quad \left[\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 2}(bX_{\bar{\varrho} 2} - X_{\bar{\varrho} 1}(X_{\bar{\varrho} 3} + dX_{\bar{\varrho} 2}))\right] - \sum_{\xi=1}^{12} n_{\xi 2}(bX_{\xi 2} - X_{\xi 1}(X_{\xi 3} + dX_{\xi 2})) \\ \quad -k_2 e_2, \\ \Theta_3 = \left[\sum_{\varrho=1}^{4} m_{\varrho 3}(X_{\varrho 1}X_{\varrho 2} - cX_{\varrho 3} - X_{\varrho 2})\right]\left[\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 3}X_{\bar{\varrho} 3}\right] + \left[\sum_{\varrho=1}^{4} m_{\varrho 3}X_{\varrho 3}\right] \\ \quad \left[\sum_{\bar{\varrho}=5}^{7} m_{\bar{\varrho} 3}(X_{\bar{\varrho} 1}X_{\bar{\varrho} 2} - cX_{\bar{\varrho} 3} - X_{\bar{\varrho} 2})\right] - \sum_{\xi=1}^{12} n_{\xi 3}(Y_{\xi 1}Y_{\xi 2} - cY_{\xi 3} - Y_{\xi 2}) - k_3 e_3. \end{cases} \tag{24}$$

where k_1, k_2, and k_3 are positive real numbers. Then, HCCS is attained between seven drives and tweleve responses.

Proof. We again define a function to ensure that it follows the prerequisites of Barbalat's lemma. Assume the following function:

Table 2. Initial values of the drive-response systems for all the numerical simulations.

States	Initial conditions	States	Initial conditions
(X_{11}, X_{12}, X_{13})	(6, 2, 8)	(X_{21}, X_{22}, X_{23})	(4, 3, 1)
(X_{31}, X_{32}, X_{33})	(5, 3, 1)	(X_{41}, X_{42}, X_{43})	(4, 1, 7)
(X_{51}, X_{52}, X_{53})	(7, 2, 3)	(X_{61}, X_{62}, X_{63})	(1, 4, 7)
(X_{71}, X_{72}, X_{73})	(3, 4, 1)	(Y_{11}, Y_{12}, Y_{13})	(5, 9, 1)
(Y_{21}, Y_{22}, Y_{23})	(6, 2, 1)	(Y_{31}, Y_{32}, Y_{33})	(2, 4, 3)
(Y_{41}, Y_{42}, Y_{43})	(2, 1, 5)	(Y_{51}, Y_{52}, Y_{53})	(2, 3, 1)
(Y_{61}, Y_{62}, Y_{63})	(5, 2, 1)	(Y_{71}, Y_{72}, Y_{73})	(3, 2, 5)
(Y_{81}, Y_{82}, Y_{83})	(1, 2, 7)	(Y_{91}, Y_{92}, Y_{93})	(6, 1, 9)

$$V(e_1, e_2, e_3) = \frac{1}{2}\left(\sum_{j=1}^{3} e_j^2\right). \tag{25}$$

The quadratic function defined above is positive definite, with $V(e) = V(e_1, e_2, e_3) > 0$ and

$$\dot{V} = \left(\sum_{j=1}^{3} e_j \cdot \dot{e}_j\right). \tag{26}$$

Using equations in (21), (22), and (23), we obtain

$$\begin{aligned}\dot{V} = \sum_{j=1}^{3} e_j \cdot [&(n_{1j}G_{1j} + n_{2j}G_{2j} + n_{3j}G_{3j} + n_{4j}G_{4j} + n_{5j}G_{5j} + n_{6j}G_{6j} \\ &+ n_{7j}G_{7j} + n_{8j}G_{8j} + n_{9j}G_{9j} + n_{10j}G_{10j} + n_{11j}G_{11j} + n_{12j}G_{12j} + \Theta_j) \\ &- (m_{1j}F_{1j} + m_{2j}F_{2j} + m_{3j}F_{3j} + m_{4j}F_{4j}) \cdot (m_{5j}X_{5j} + m_{6j}X_{6j} + m_{7j}X_{7j}) \\ &- (m_{1j}X_{1j} + m_{2j}X_{2j} + m_{3j}X_{3j} + m_{4j}X_{4j}) \cdot (m_{5j}F_{5j} + m_{6j}F_{6j} + m_{7j}F_{7j})].\end{aligned} \tag{27}$$

Substituting (24) into (27), we get

$$\dot{V} = -\left(\sum_{j=1}^{3} k_j e_j^2\right) < 0. \tag{28}$$

which shows that function $\dot{V}$ is negative definite. Thus, it proves that $\lim_{t\to\infty} \|e(t)\| = 0$. Therefore, the desired synchronization of hexa compound combination is attained between nineteen 3D category B chaotic models. □

5. Numerical simulations

Numerical validation for all acquired theoretical hexa compound combination results obtained in section 4 is essential to assess the efficiency, viability, and authenticity of all these derived theories and analyses. The initial conditions used for extracting the simulation results of each category B chaotic system are orderly arranged in table 2; in addition to the values of the last three responses as: $(Y_{[10]1}, Y_{[10]2}, Y_{[10]3}) = (3, 1, 5)$, $(Y_{[11]1}, Y_{[11]2}, Y_{[11]3}) = (7, 2, 5)$, and $(Y_{[12]1}, Y_{[12]2}, Y_{[12]3}) = (5, 1, 7)$.

6. Comparative analysis

When it comes to multi-leveled synchronization schemes on compound combinations, there are double compound combinations (DCCS), triple compound combinations (TCCS), quad compound combinations (QCCS), and finally, the most advanced penta compound combinations (PCCS). The concept of compound combinations was initially proposed to overcome the disadvantages of a single response system in compound synchronization [28], and this applies to all the aforementioned schemes. In general, DCCS [29] utilizes four drives and four responses. In [30], the authors achieved TCS over eight chaotic systems, followed by quad compound combination [31] over ten models. The most recent penta compound combination [32] was achieved over twelve models.

In our research, we extend the concept of PCCS to a higher-order synchronization, namely hexa compound combinations (HCCS), involving a total of nineteen chaotic models. We utilize the remark 3.3 and neglect the

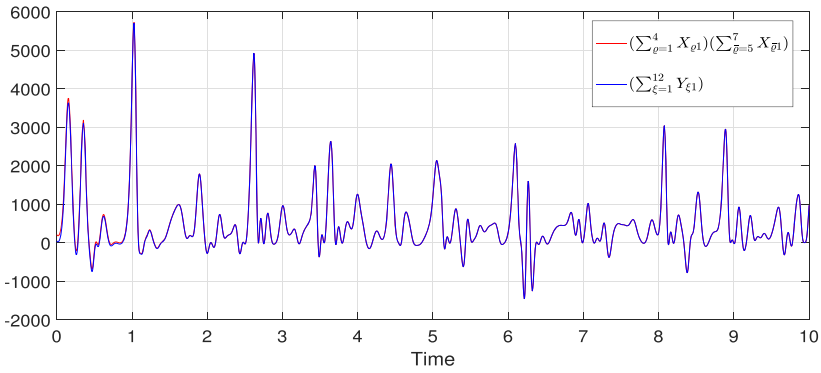


Figure 5. First states of the hexa compound combination, i.e. $(\sum_{\varrho=1}^{4} X_{\varrho 1})(\sum_{\bar{\varrho}=5}^{7} X_{\bar{\varrho} 1})$ with $(\sum_{\xi=1}^{12} Y_{\xi 1})$ after the control implementation.

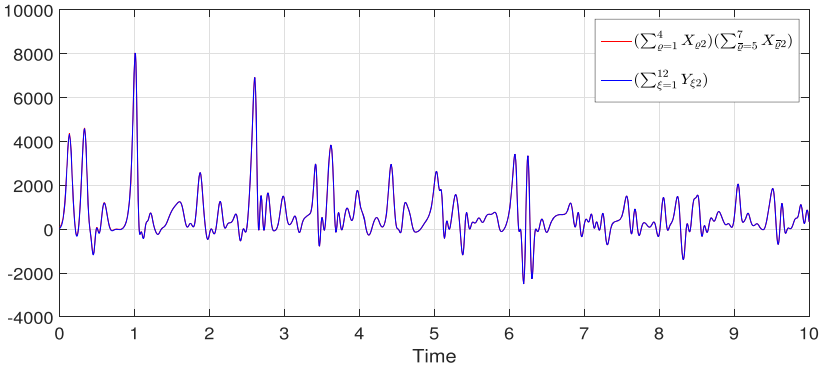


Figure 6. Second states of the hexa compound combination, i.e. $(\sum_{\varrho=1}^{4} X_{\varrho 2})(\sum_{\bar{\varrho}=5}^{7} X_{\bar{\varrho} 2})$ with $(\sum_{\xi=1}^{12} Y_{\xi 2})$ after the control implementation.

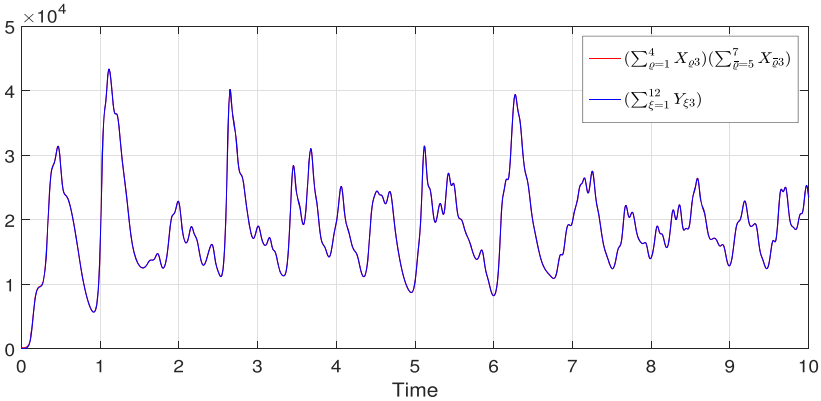


Figure 7. Third states of the hexa compound combination, i.e. $(\sum_{\varrho=1}^{4} X_{\varrho 3})(\sum_{\bar{\varrho}=5}^{7} X_{\bar{\varrho} 3})$ with $(\sum_{\xi=1}^{12} Y_{\xi 3})$ after the control implementation.

factor $M_{\bar{\varrho}} X_{\bar{\varrho}} (M_3 X_3 + M_4 X_4)$ for $\bar{\varrho} = 7$, in addition to $N_\xi = 0$ for $\xi = 11$ and 12, to achieve the special case of penta compound combination synchronization (PCCS). We compare PCCS with our proposed hexa compound combination synchronization (HCCS) scheme using numerical results. The states of the drive-response configuration for HCCS after control addition are provided in figures 5, 6, and 7, while the states for PCCS are

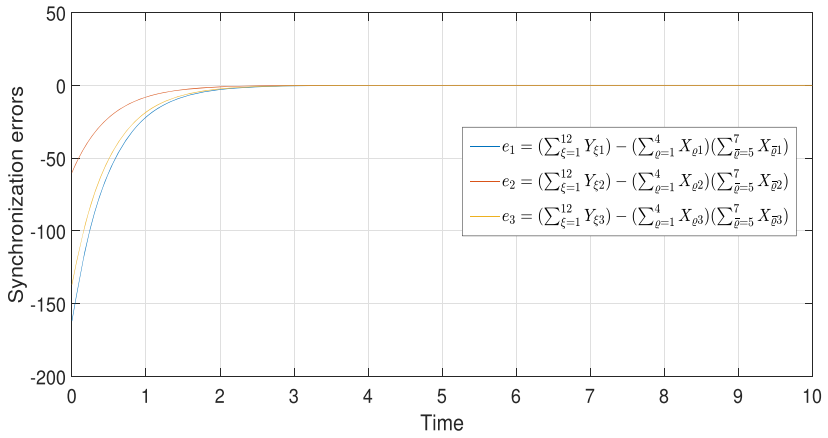


Figure 8. Errors variation for the hexa compound combination scheme.

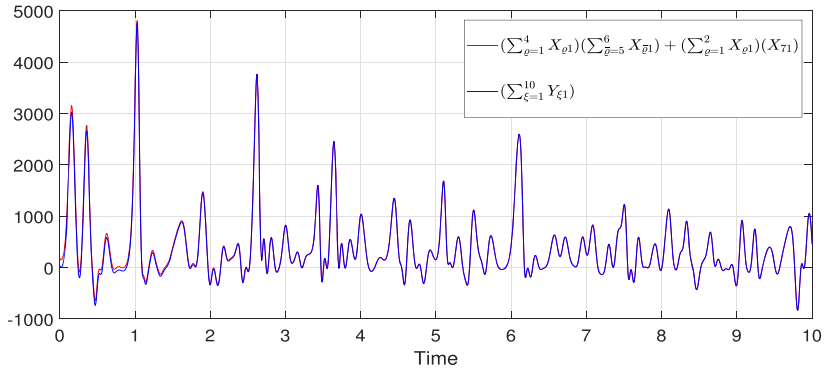


Figure 9. First states of the penta compound combination, i.e. $(\sum_{\varrho=1}^{4} X_{\varrho 1})(\sum_{\bar{\varrho}=5}^{6} X_{\bar{\varrho} 1}) + (\sum_{\varrho=1}^{2} X_{\varrho 1})(X_{71})$ with $(\sum_{\xi=1}^{10} Y_{\xi 1})$ after the control implementation.

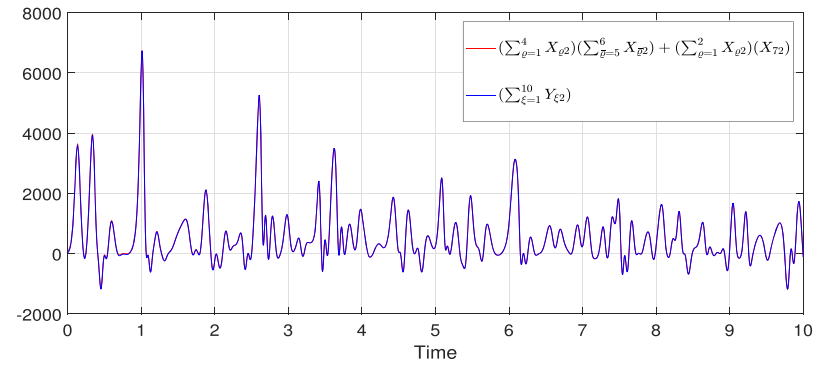


Figure 10. Second states of the penta compound combination, i.e. $(\sum_{\varrho=1}^{4} X_{\varrho 2})(\sum_{\bar{\varrho}=5}^{6} X_{\bar{\varrho} 2}) + (\sum_{\varrho=1}^{2} X_{\varrho 2})(X_{72})$ with $(\sum_{\xi=1}^{10} Y_{\xi 2})$ after the control implementation.

shown in figures 9, 10, and 11. The synchronization errors for HCCS and PCCS are presented in figures 8 and 12, respectively. It is observed that HCCS exhibits more complicated and varied state dynamics than PCCS. Additionally, twelve response systems are directly involved on the receiver's part in HCCS. As a result, the proposed HCCS approach enhances the security of broadcasted information, making it an improved version of previously published research on compound combinations.

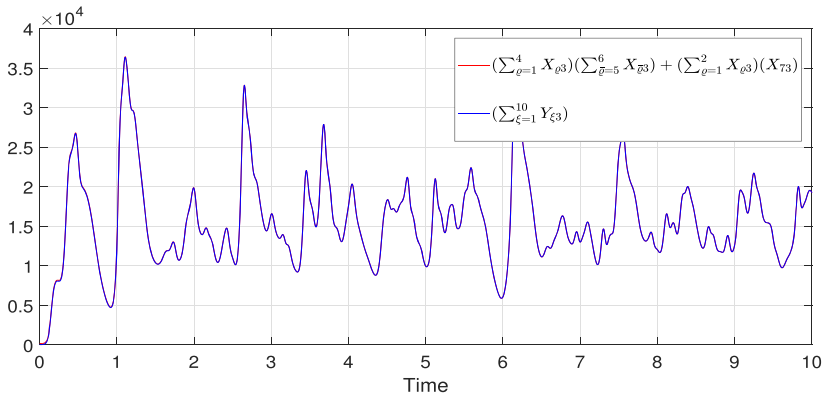


Figure 11. Third states of the penta compound combination, i.e. $(\sum_{\varrho=1}^{4} X_{\varrho 3})(\sum_{\bar{\varrho}=5}^{6} X_{\bar{\varrho} 3}) + (\sum_{\varrho=1}^{2} X_{\varrho 3})(X_{73})$ with $(\sum_{\xi=1}^{10} Y_{\xi 3})$ after the control implementation.

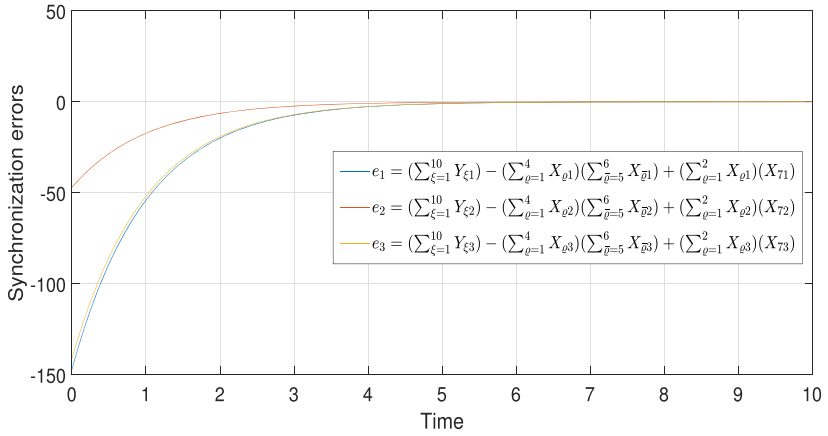


Figure 12. Errors variation for the penta compound combination scheme.

7. Electronic circuit design

In this section, we design an electronic schematic [61–63] representation for the $\mathscr{M}-$system on NI Multisim^{TM}. These designs are important as they demonstrate the practical applicability of the studied chaotic system [64]. We require fundamental analog elements such as amplifiers, multipliers, resistors (R), and capacitors (C) to replicate the voltage dynamics. Each amplifier's voltage-controlled current source (VCCS) supply is firmed at $\pm$ 15 volts. Employing Ohm's and Kirchhoff's law with scaling factors $\varpi = 1$ and $\varpi_0 = 10$, the $\mathscr{M}-$system transforms to the following circuit representation:

$$\begin{cases} \dot{X} = \frac{\varpi}{R_1 C_1} Y - \frac{\varpi}{R_2 C_1} X + \frac{\varpi}{R_3 C_1} Z, \\ \dot{Y} = \frac{\varpi}{R_6 C_2} Y - \frac{\varpi}{\varpi_0 R_4 C_2} XZ - \frac{\varpi}{\varpi_0 R_5 C_2} XY, \\ \dot{Z} = \frac{\varpi}{\varpi_0 R_7 C_3} XY - \frac{\varpi}{R_8 C_3} Z - \frac{\varpi}{R_9 C_3} Y. \end{cases} \tag{29}$$

The voltages that flow through each unity microfarad capacitor C_1, C_2, and C_3 are symbolized with X, Y, and Z, respectively. On comparing (29) with the $\mathscr{M}-$system, we computed the values of each R_i for $i = 1, 2, \ldots, 9$, described in the schematic figure while supplementary resistors R_s for $s = 10, 11, \ldots, 15$ are fixed at 1000 $k\Omega$.

When the circuit in figure 13 is connected with an external oscilloscope, the captured output are placed in figure 14. The designed circuit outcomes aligns with the numerical results that we obtained on MATLAB®.

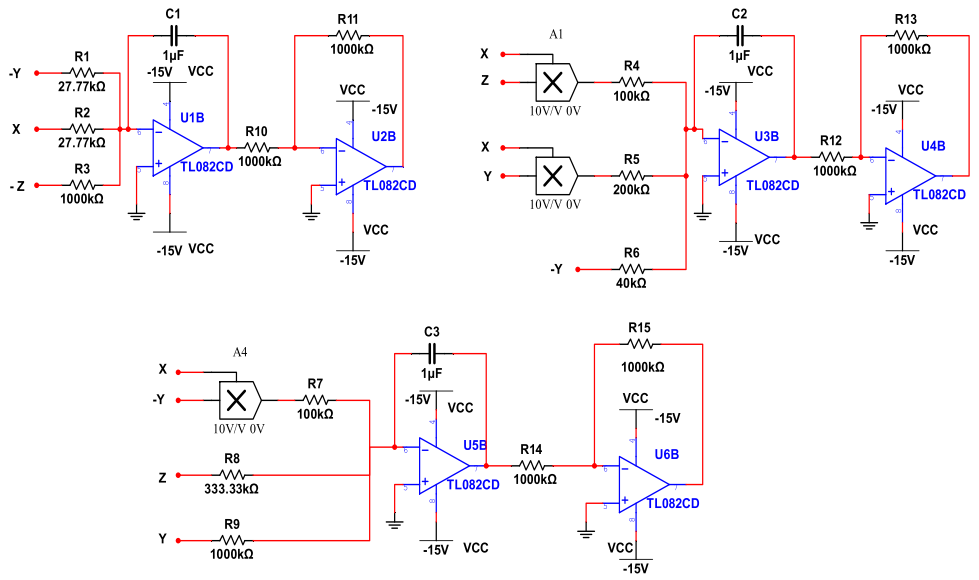


Figure 13. Electronic circuit representation of the $\mathscr{M}-$system.

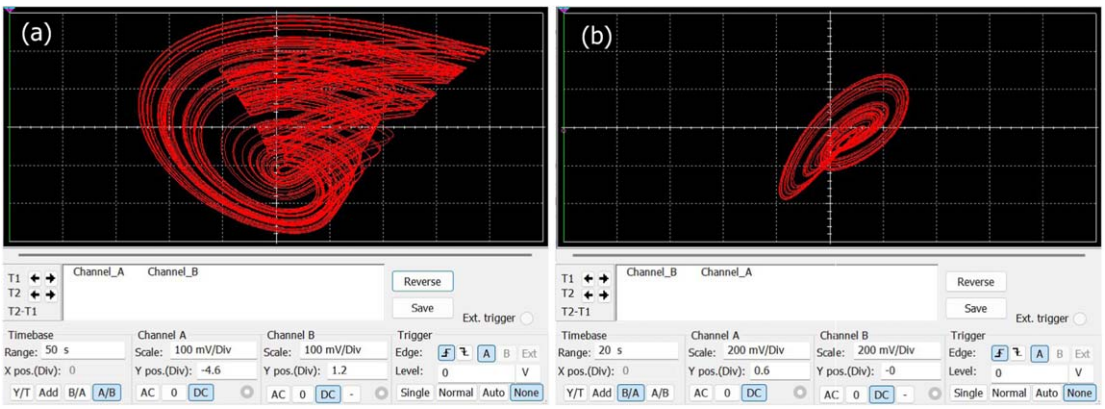


Figure 14. Output captured on oscilloscope for circuit schematic of the $\mathscr{M}-$system in: (a) $Y-Z$ plane and (b) $X-Y$ plane.

8. Conclusions

We introduce a new 3D chaotic system that falls under the rare category B, featuring three equilibrium points. The trace of its Jacobi matrix depends on the state variable, demonstrating the novel system's non-uniform conservative behavior. We also establish the underlying aperiodic attribute of each system trajectory through a theoretical discussion on its attractors and subsequently visualizing them using phase portrait analysis. We derive the system's dissipation-induced volume contraction and confirm its conservative character using Lyapunov exponents. The new category B system also possesses line equilibrium, i.e. infinite equilibrium for its conservative case. Further, it is proved that periodic, chaotic, and quasi-periodic orbits of the system are conditioned to the parameter a, and this feature is examined in-depth through bifurcation charts. These acquired bifurcation transition graphs are verified by switching $30 \leqslant a \leqslant 33.2$ on the Lyapunov plane. We investigate a generalized novel synchronization, hexa compound combination over n-dimensional category B chaotic systems, consisting of seven drives and twelve responses. We demonstrate that the proposed synchronization method is significantly more advanced than the penta compound combination strategy. We used the Lyapunov theory to define the synchronization errors and their corresponding nonlinear controllers. The research concludes that various synchronization methods on compound combinations, such as double compound, triple compound, quad compound, and penta compound, are either simplified versions or are just particular cases of this new scheme. We prove and compare this strategy with the penta version by performing

various numerical simulations on MATLAB®. We found a good agreement between all the theoretical and computational findings. The suggested scheme is promising because a combined array of nineteen drive-response frameworks releases a much higher signal complexity than many reported schemes. The scheme presents several avenues for optimization and future exploration, including using simpler models with fewer nonlinear terms to enhance cost-effectiveness and examining alternative control strategies. Such investigations will provide deeper insights into the synchronization process. Future work could extend this multi-level synchronization to hyper-chaotic or integrate it with multi-switches, although this may nonlinearly increase the computational challenges. Furthermore, the proposed methodology can effectively encompass time-delay and fractional-order systems, which merit further investigation. Time delay is significant in various applications, making it essential to explore the conditions under which the results of this study can be applied to diverse contexts, including neural networks.

Acknowledgments

The authors express their sincere gratitude to the editors and anonymous reviewers for their valuable suggestions and insightful comments, which significantly improved the quality of this manuscript. The corresponding author also wishes to thank Delhi Technological University, Delhi, for their support.

Conflict of interest

No potential conflict of interests to declare.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

ORCID iDs

Mukul Sirohi https://orcid.org/0000-0001-5229-0298

References

[1] Field M and Golubitsky M 2009 *Symmetry in Chaos: A Search for Pattern in Mathematics, Art, and Nature* (Society for Industrial and Applied Mathematics, Philadelphia, USA)
[2] Strogatz S H 2018 *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (CRC Press)
[3] Bouali S 2012 A novel strange attractor with a stretched loop *Nonlinear Dynamics* **70** 2375–81
[4] Yang Y, Huang L, Xiang J, Bao H and Li H 2021 Generating multi-wing hidden attractors with only stable node-foci via non-autonomous approach *Physica Scripta* **96** 125220
[5] Yang Y, Huang L, Kuznetsov N V and Lai Q 2023 Design and implementation of grid-wing hidden chaotic attractors with only stable equilibria *IEEE Transactions on Circuits and Systems I: Regular Papers* (https://doi.org/10.1109/TCSI.2023.3312489)
[6] Selmi K, Bouallegue K and Soufi Y 2024 A new class of chaotic attractors using different activation functions in neuron with multi dendrites *Cognitive Neurodynamics* **24** 1–20
[7] He P, Liu H, Li G, Xu X and Gu Y 2023 A general method for generating multi-scroll and multi-wing chaotic systems and its implementation of attractor reproduction *Physica Scripta* **98** 085232
[8] Al-Azzawi SF and Al-Hayali MA 2022 Coexisting of self-excited and hidden attractors in a new 4D hyperchaotic Sprott-S system with a single equilibrium point *Archives of Control Sciences* **31** 37–56
[9] Pecora L and Carroll T 1990 Synchronization in chaotic systems *Physics Review Letters* **64** 821–4
[10] Lopes L M, Fernandes S and Grcio C 2015 Complete synchronization and delayed synchronization in couplings *Nonlinear Dynamics* **79** 1615–24
[11] Singh P P, Singh J P and Roy B K 2014 Synchronization and anti-synchronization of Lu and BhalekarGejji chaotic systems using nonlinear active control *Chaos, Solitons & Fractals* **69** 31–9
[12] Yu F, Kong X, Yao W, Zhang J, Cai S, Lin H and Jin J 2024 Dynamics analysis, synchronization and FPGA implementation of multiscroll Hopfield neural networks with non-polynomial memristor *Chaos, Solitons & Fractals* **179** 114440
[13] Bekiros S, Yao Q, Mou J, Alkhateeb A F and Jahanshahi H 2023 Adaptive fixed-time robust control for function projective synchronization of hyperchaotic economic systems with external perturbations *Chaos, Solitons & Fractals* **172** 113609
[14] Zhou P, Ma J and Xu Y 2023 Phase synchronization between neurons under nonlinear coupling via hybrid synapse *Chaos, Solitons & Fractals* **169** 113238
[15] Yu F, Qian S, Chen X, Huang Y, Liu L, Shi C, Cai S, Song Y and Wang C 2020 A new 4D four-wing memristive hyperchaotic system: dynamical analysis, electronic circuit design, shape synchronization and secure communication *International Journal of Bifurcation and Chaos* **30** 2050147
[16] Zaqueros-Martinez J, Rodriguez-Gomez G, Tlelo-Cuautle E and Orihuela-Espina F 2023 Fuzzy synchronization of chaotic systems with hidden attractors *Entropy* **25** 495

[17] Chakraborty A and Veeresha P 2024 Investigating the dynamics, synchronization and control of chaos within a transformed fractional SamardzijaGreller framework *Chaos, Solitons & Fractals* **182** 114810
[18] Yu F, Li L, He B, Liu L, Qian S, Zhang Z, Shen H, Cai S and Li Y 2021 Pseudorandom number generator based on a 5D hyperchaotic four-wing memristive system and its FPGA implementation *The European Physical Journal Special Topics* **230** 1763–72
[19] Sirohi M 2023 Qualitative analysis of a novel 5D chaotic system based on Boualis system and its application in private communication via adaptive control *Boletn de la Sociedad Matemtica Mexicana* **29** 26
[20] Saeed N A, Saleh H A, El-Ganaini W A, Kamel M and Mohamed M S 2023 On a New ThreeDimensional Chaotic System with Adaptive Control and Chaos Synchronization *Shock and Vibration* **2023** 1969500
[21] Khattar D, Agrawal N and Sirohi M 2024 Qualitative analysis of a new 6D hyper-chaotic system via bifurcation, the Poincar notion, and its circuit implementation *Indian Journal of Physics* **98** 259–73
[22] Khan A, Khattar D and Prajapati N 2017 Multiswitching compound antisynchronization of four chaotic systems *Pramana* **89** 1–0
[23] Hailong Z, Ding Z and Wang L 2024 Predefined-time multi-switch combination-combination synchronization of fractional-order chaotic systems with time delays *Physica Scripta* **99** 105223
[24] Shukla V K, Joshi M C, Mishra P K and Xu C 2024 Adaptive fixed-time difference synchronization for different classes of chaotic dynamical systems *Physica Scripta* **99** 095264
[25] Khattar D, Agrawal N and Sirohi M 2023 Dynamical analysis of a 5D novel system based on Lorenz system and its hybrid function projective synchronisation using adaptive control *Pramana* **97** 76
[26] Liu Y and Davids P 2000 Dual synchronization of chaos *Physics Review* E **61** R2176
[27] Runzi L, Yinglan W and Shucheng D 2011 Combination synchronization of three classic chaotic systems using active backstepping design *Chaos: An Interdisciplinary Journal of Nonlinear Science* **21**
[28] Sun J, Shen Y, Yin Q and Xu C 2013 Compound synchronization of four memristor chaotic oscillator systems and secure communication *Chaos: An Interdisciplinary Journal of Nonlinear Science* **23**
[29] Mahmoud G M, Abed-Elhameed T M and Farghaly A A 2018 Double compound combination synchronization among eight n-dimensional chaotic systems *Chinese Physics* B **27** 080502
[30] Yadav V K, Prasad G, Srivastava M and Das S 2022 Triple compound synchronization among eight chaotic systems with external disturbances via nonlinear approach *Differential Equations and Dynamical Systems* **30** 549–572
[31] Jahanzaib L S, Trikha P and Baleanu D 2021 Analysis and application using quad compound combination anti-synchronization on novel fractional-order chaotic system *Arabian Journal for Science and Engineering* **46** 1729–42
[32] Mahmoud E E, Trikha P, Jahanzaib L S and Matoog R T 2021 Chaos control and Penta-compound combination anti-synchronization on a novel fractional chaotic system with analysis and application *Results in Physics* **24** 104130
[33] Zhao X, Liu J, Chen G, Chai L and Wang D 2022 Dynamics and synchronization of complex-valued ring networks *International Journal of Bifurcation and Chaos* **32** 2230011
[34] Jin J, Chen W, Ouyang A, Yu F and Liu H 2023 A time-varying fuzzy parameter zeroing neural network for the synchronization of chaotic systems *IEEE Transactions on Emerging Topics in Computational Intelligence* **8** 364–376
[35] Aliabadi F, Majidi M H and Khorashadizadeh S 2022 Chaos synchronization using adaptive quantum neural networks and its application in secure communication and cryptography *Neural Computing and Applications* **34** 6521–33
[36] Xu G, Zhao S and Cheng Y 2021 Chaotic synchronization based on improved global nonlinear integral sliding mode control *Computers & Electrical Engineering* **96** 107497
[37] Yu N and Zhu W 2021 Event-triggered impulsive chaotic synchronization of fractional-order differential systems *Applied Mathematics and Computation* **388** 125554
[38] Liu J, Zhang J and Wang Y 2022 Secure communication via chaotic synchronization based on reservoir computing *IEEE Transactions on Neural Networks and Learning Systems* **35** 285–99
[39] Wang L, Mao X, Wang A, Wang Y, Gao Z, Li S and Yan L 2020 Scheme of coherent optical chaos communication *Optics Letters* **45** 4762–5
[40] Sarkar A 2021 Secure exchange of information using artificial intelligence and chaotic system guided neural synchronization *Multimedia Tools and Applications* **80** 18211–41
[41] Singh J P and Roy B K 2018 Five new 4-D autonomous conservative chaotic systems with various type of non-hyperbolic and lines of equilibria *Chaos, Solitons & Fractals* **114** 81–91
[42] Bayani A, Rajagopal K, Khalaf A J, Jafari S, Leutcho G D and Kengne J 2019 Dynamical analysis of a new multistable chaotic system with hidden attractor: Antimonotonicity, coexisting multiple attractors, and offset boosting *Physics Letters A* **2383** 1450–6
[43] Rech P C 2022 Self-excited and hidden attractors in a multistable jerk system *Chaos, Solitons & Fractals* **164** 112614
[44] Gokyildirim A, Kocamaz U E, Uyaroglu Y and Calgan H 2023 A novel five-term 3D chaotic system with cubic nonlinearity and its microcontroller-based secure communication implementation *AEU-International Journal of Electronics and Communications* **160** 154497
[45] Ramadoss J, Kengne R, Tokoue Ngatcha D, Kamdoum Tamba V, Rajagopal K and Motchongom Tingue M 2022 A ThreeDimensional Autonomous System with a Parabolic Equilibrium: Dynamical Analysis, Adaptive Synchronization via Relay Coupling, and Applications to Steganography and Chaos Encryption *Complexity* **2022** 8362836
[46] Jafari S, Sprott J C and Dehghan S 2019 Categories of conservative flows *International Journal of Bifurcation and Chaos* **29** 1950021
[47] Cang S, Wu A, Wang Z and Chen Z 2017 On a 3-D generalized Hamiltonian model with conservative and dissipative chaotic flows *Chaos, Solitons & Fractals* **99** 45–51
[48] Wang Q, Yan S, Wang E, Ren Y and Sun X 2023 A simple Hamiltonian conservative chaotic system with extreme multistability and offset-boosting *Nonlinear Dynamics* **111** 7819–30
[49] Sprott J C 2010 *Elegant Chaos: Algebraically Simple Chaotic Flows* (World Scientific)
[50] Sprott J C 2014 A dynamical system with a strange attractor and invariant tori *Physics Letters* A **378** 1361–3
[51] Sprott J C 2015 Symmetric time-reversible flows with a strange attractor *International Journal of Bifurcation and Chaos* **25** 1550078
[52] Abdul-Kareem K N and Al-Azzawi S F 2024 Electronic circuit and image encryption for a new 3D nonuniformly conservative system *Indian Journal of Physics* **1** 1–6
[53] Cang S, Wang L, Zhang Y, Wang Z and Chen Z 2022 Bifurcation and chaos in a smooth 3D dynamical system extended from Nos-Hoover oscillator *Chaos, Solitons & Fractals* **158** 112016
[54] Ott E 1981 Strange attractors and chaotic motions of dynamical systems *Reviews of Modern Physics* **53** 655
[55] Bryant P, Brown R and Abarbanel H D 1990 Lyapunov exponents from observed time series *Physical Review Letters* **65** 1523
[56] Wolf A, Swift J B, Swinney H L and Vastano J A 1985 Determining Lyapunov exponents from a time series *Physica D: Nonlinear Phenomena* **16** 285–317

[57] Stolze C H 1978 A history of the divergence theorem *Historia Mathematica* **5** 437–42
[58] Heidel J and Zhang F 2007 Nonchaotic and chaotic behavior in three-dimensional quadratic systems: Five-one conservative cases *International Journal of Bifurcation and Chaos* **17** 2049–72
[59] Bleher S, Grebogi C and Ott E 1990 Bifurcation to chaotic scattering *Physica D: Nonlinear Phenomena* **46** 87–121
[60] Khalil H K 2002 *Nonlinear Systems* (Prentice Hall)
[61] Wang N, Xu D, Jin Z, Zhang G, Chen M and Xu Q 2025 State-extension modulation yields infinite attractors *Nonlinear Dynamics* **113** 1761–1773
[62] Mobayen S, Mostafaee J, Alattas K A, Ke M T, Hsueh Y H and Zhilenkov A 2024 A new hyperchaotic system: circuit realization, nonlinear analysis and synchronization control *Physica Scripta* **99** 105204
[63] Wang N, Xu D, Kuznetsov N V, Bao H, Chen M and Xu Q 2023 Experimental observation of hidden Chuas attractor *Chaos, Solitons & Fractals* **170** 113427
[64] Cuomo K M and Oppenheim A V 1993 Circuit implementation of synchronized chaos with applications to communications *Physical Review Letters* **71** 65

Check for updates

Received: 27 August 2024 | Revised: 5 November 2024 | Accepted: 12 November 2024

DOI: 10.1002/cjce.25566

RESEARCH ARTICLE

THE CANADIAN JOURNAL OF CHEMICAL ENGINEERING

Optimizing passive micromixers: Enhancing mixing efficiency through computational fluid dynamics and metaheuristic algorithms

Md Gulam Mustafa | **Mohammad Zunaid** | **Samsher Gautam**

Department of Mechanical Engineering, Delhi Technological University, Delhi, India

Correspondence
Md Gulam Mustafa, Department of Mechanical Engineering, Delhi Technological University, Delhi, India.
Email: gm09mes31@gmail.com

Abstract

This paper examines the optimization of micromixers, highlighting their essential role in improving the efficiency of chemical reactions and fluid mixing at a micro-scale, and using metaheuristic algorithms for a novel passive micromixer featuring bends and twists with offsets. The twists and bends cause constant changes in the direction of liquid flow, leading to chaotic advection that enhances species mixing while minimizing pressure loss. Although increasing channel length generally improves mixing performance, the proposed design's performance was higher than the reference channel for Reynolds numbers (Re) greater than 100. Key findings include achieving an excellent mixing performance of 87.76% at a Reynolds number of 400, with a bend angle of 60° and a twist factor of 4. The Harris hawk optimization (HHO) algorithm proved the most effective for optimizing microfluidic channel designs. This approach offers significant advantages for cost-effectively optimizing microfluidic channel designs.

KEYWORDS

CFD, chaotic advection, Haris hawk optimization, machine learning, passive micromixer, response surface method

1 | INTRODUCTION

Micro-mixers have received a lot of research interest in recent years, particularly in the fields of pharmaceuticals,[1] reactive polymerization,[2] nanoparticle precipitation,[3] and chemical processes.[4] The applications of micromixer technology include chemical synthesis,[5] polymerization,[2] DNA analysis,[6] biological screening enzyme tests,[7] protein folding,[8] and detection/analysis of chemical or biochemical content.[9] Microscale approaches are employed in chemical synthesis to construct microreactors. The core concept behind these mixers revolves around the interaction of two reactant streams within a limited area, resulting in spontaneous mixing without requiring any stirring apparatus. Nevertheless, achieving the desired mixing requires a relatively intricate channel design since molecular diffusion plays a primary role in small-scale mixing.[10] Here, mixing is a multiscale phenomenon since it assimilates a variety of phenomena, encompassing flow advection at a distance of 10 m, liquid diffusivity at a distance of 1 nm, and chemical reaction at a distance of 0.1 nm.[11]

Various numerical[12–16] and experimental[17–19] studies on the blending of fluids in the microchannel of T-junctions have been investigated. Substantial reviews on micro-channel mixing techniques have also been done.[20–23] Mixing at T-junctions has been investigated concerning aspects like the impact of the mixer's aspect ratio, flow regimes, fluid speed, and strategies to improve mixing efficiency. Ashar Sultan et al.[24]

 © 2024 Canadian Society for Chemical Engineering.

observed four distinct flow patterns in the T-junction micromixer:

- Segregated flow regime: At low Reynolds numbers, two fluid streams flow in parallel from the source to the exit without significant mixing, maintaining a stable flow regime.
- Vortex flow regime: Each jet contains vortices, and their rotation axis is parallel to the axis of the micromixer channel.
- Engulfment flow regime: Under continuous flow conditions, the fluid patterns deviate from parallel to the channel axis and rotate across it, transporting the fluid from one end to another.
- Chaotic flow regime: It can be recognized by the evolution of a vortex street that travels through a micromixer, allowing the two jets to mix quickly. Vortices developed immediately after the jet impingement diminish immediately and have a diameter roughly half that of the mixing channel width in the chaotic flow regime, which is a dynamic flow regime. The fluid from the two intake jets is encompassed by the vortex pathway that starts from the jet impingement point and extends toward the outlet.

The flow mode in a T-junction microchannel[25] can have essential implications for microfluidic applications, such as in chemical and biological assays or microreactors. It can affect the mixing efficiency, residence time, and reaction kinetics, among other factors. Therefore, it is essential to warily design and optimize the microchannel geometry and flow conditions to achieve the desired flow regime. A micromixer's ability to mix well is a crucial design consideration since it is best suited for quick, complicated chemical reactions, where the mixing quality significantly impacts the yields. Passive[26,27] and active[28] mixing augmentation can be used to improve mixing. For example, we may utilize the flow's inherent energy by, for instance, engineering the channel's geometry to extend and adhere to the flow and induce advection in a micromixer cross-channel.

Additionally, many academics have proposed various kinds of geometries, such as twist,[25] split–recombination,[29] serpentine,[30] and zigzag.[31] Mixing augmentation may be accomplished by thoroughly engineering the micromixer channel. Hence, the reactants flow via channels with unique forms, produce vortices, and disrupt the flow stability, resulting in a chaotic flow regime.

On the other hand, with an active micromixer, external pressures and artefacts are applied to improve mixing. Pulsation of the flow mechanical rotation,[32] vibration,[33] flow stimulation,[34] electrical,[35] and microwaves[36] are all methods for introducing this external force. However, most suggested methods complicate design, production procedures, and costs, making them unworkable for various applications. Leveraging fractal patterns, as explored by,[35] increases fluid contact and improves mixing precision, which is ideal for applications like liposome manufacturing. The study also employs a cost-effective, three-objective optimization approach, incorporating heating and electrophoresis, which reduces equipment needs and operational expenses.[13] Additionally, the inclusion of electrophoretic sorting makes the device highly adaptable across different fluid types, from synthetic to biological, providing a versatile solution in biomedical applications.

Ortega-Casanova[4] employed the optimization of the geometry of a T-type micromixer with response surface method (RSM). In their investigation, the authors identified degrees of freedom and employed computer simulations to adjust geometry parameters until they achieved an optimal configuration iteratively. de Maionchi et al.[37] enhanced this concept using machine learning techniques. They employed a neural network algorithm to teach the model using the assessed parameters, and a genetic algorithm was employed to discover the optimal configuration.

Additionally, Nikpour and Mohebbi[38] combined computational fluid dynamics, RSM, and multi-objective optimization, known for their capability to achieve high-precision optimization of equipment and processes, which can also be employed to attain optimal process conditions. Their efforts resulted in satisfactory outcomes for the design of passive Y-type micromixers with obstructions, aligning with previous related studies.

Yoshimura et al.[39] employed CFD simulation and a multi-objective genetic algorithm[40–43] based on Kriging to optimize a grooved micromixer. Their investigation explored the impact of convex and concave grooves on mixing and identified characteristics that enhanced mixing using the genetic algorithm. Chen and Lv[44] conducted a concise overview of prior studies that used various optimization methods for distinct micro-mixer design concepts. Their research incorporated Cantor fractal theory to create barriers for a passive micromixer and identify the configuration leading to optimal mixing efficiency and minimal pressure variation.

Improving the design of micromixers can enhance the efficiency of these devices and encourage more research on microdevice production techniques. Due to the swift advancement of manufacturing techniques, constructing prototypes of optimized micromixers in the following years is becoming increasingly viable and cost-effective. By taking advantage of the opportunity to build inexpensive prototypes, testing and confirming the effectiveness of the optimized designs is necessary.

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The paper layout is as follows. First, the novel offset model structure was described, which includes equations for the conservation of species, momentum, and mass during mixing. Further, this analytical model was solved using Ansys Fluent, based on finite volumes and the sensitiveness of numerical to finite elements. Then, the 'Results' section discusses the mathematical study of the influence of Reynolds, bends and twists on η, and optimization with the constructal theory (numerical-ML approach). Constructal theory efficiently replaces chaos, complexity, and unpredictability to explain nature's development and variety. It accomplishes so using a physical rule of nature that is predictable and quantifiable. Finally, the design's benefits and shortcomings are emphasized, and conclusions are taken in light of the data shown.

2 | METHODOLOGY

2.1 | Geometric structure

In this section, the parameter of our proposed noble offset hybrid T-micromixer channel is discussed in Figure 1 with mesh size of 2,758,398. An offset T-model micromixer with twists and bends (OMTB) was studied in this work, which has an axial channel distance of 3000 μm, width of 800 μm, depth 100 μm, and offset 100 μm.

2.2 | Governing equation and numerical simulation

This study is based on the works of Dundi et al. and Wong et al.,[45,46] the OMTB micro-channel designed with a depth and width of 100 μm having offset at mid of the inlet of 50 μm. Figure 2 shows different combinations of simulation models. With the combination of computational with machine learning tools (Haris hawk optimization [HHO] and RSM), we were able to find an optimal model to get the max mixing index and min pressure drops. This unique approach of the numerical-ML method is much faster than the conventional optimization method using CFD.

Various numerical analysis of the passive micromixer was conducted by varying bend angles (between 30° and 70°), twists (1–4), and Reynolds number (Re) (0.05–450), as shown in Figure 2. All simulations relate to the betterment of mixing performance (η) and optimization of the mixing model for max η and min ΔP conducted using HHO and RSM combing CFD analysis. The primary goal of this study is to determine how the offset micromixer's bend and twist affect various Re, η, and ΔP. The investigation aimed to comprehend the impact of both design and operation on the overall performance of mixing processes, aiming to optimize the micromixer. Moreover, the findings indicated that as the Re increases, the cost of mixing energy escalates with higher values of twist and bend. This study compares RSM[47] and HHO[48] methods and creates neural networks utilizing ANN to optimize the micromixer model. Ansys Fluent was employed for all simulations utilizing finite volume analysis.

This study employed Ansys Fluent 2022 to analyze the η and ΔP of offset-micromixers. The bundled code utilizes the finite volume approach with a coupling solver to solve steady continuity, convection–diffusion equations, and steady incompressible Navier–Stokes.

$$\nabla . V = 0 \tag{1}$$

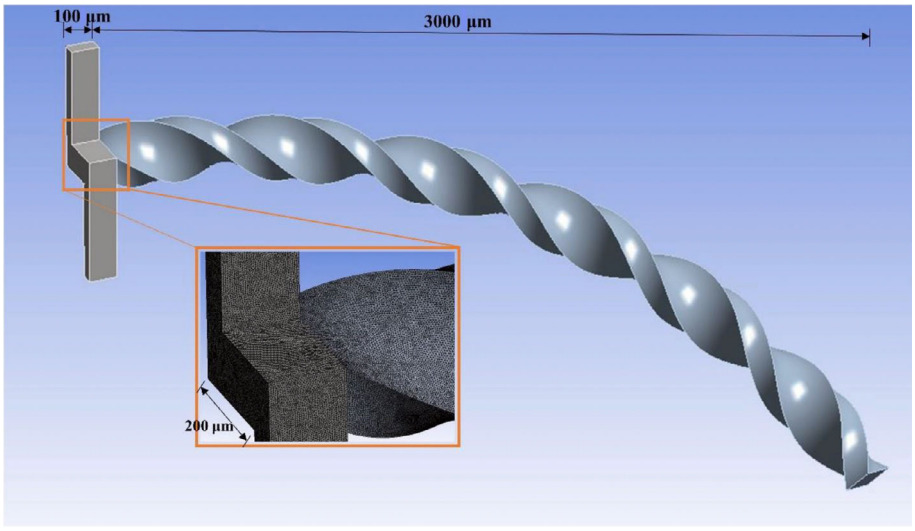


FIGURE 1 Proposed offset micromixer model.

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

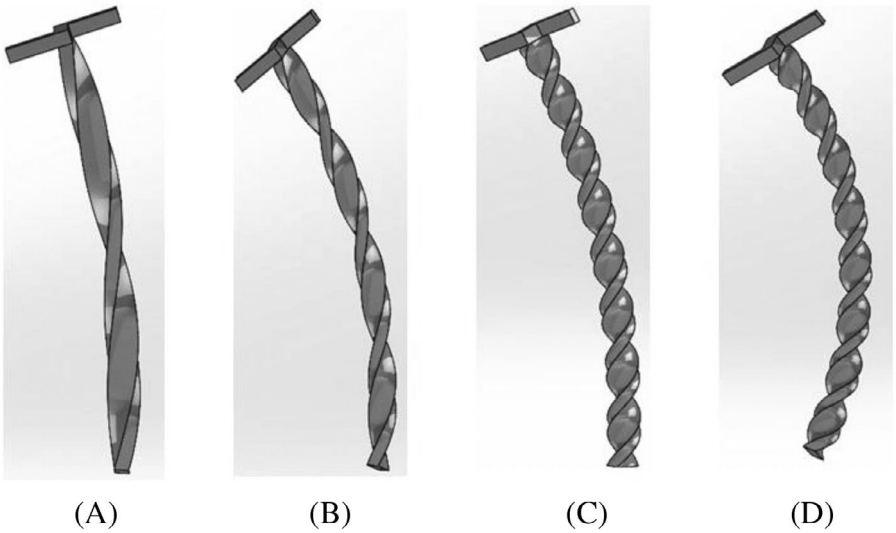


FIGURE 2 Proposed model with different combinations of bend and twist: (A) 1 twist and 30° bend; (B) 2 twists and 50° bend; (C) 3 twists and 30° bend; (D) 4 twists and 70° bend.

$$\rho(V.\nabla V) = -\nabla P + \mu\nabla^2 V \tag{2}$$

$$(\nabla.V)C = D\nabla^2 C \tag{3}$$

In the simulation context, V denotes the velocity vector, ρ stands for density, P represents pressure, μ indicates dynamic viscosity, D refers to the diffusion coefficient (2×10^{-9} m^2/s), and C represents the mole fraction of the species. The governing equations of flow (Equations (1)–(3)) solved for uniform laminar flow for Newtonian fluid. The upper admittance lacked solute with a $C = 0$ mol/m^3 concentration, whereas the bottom entry included a $C = 1$ mol/m^3 concentration.

The input boundary conditions were gauze pressure at the output using pressure–velocity coupled methods and a pressure-based solver. The values of viscosity, coefficient of diffusion, and density are $\mu =$ 1e-03 Pa·s, $\gamma =$ 2e-09 m^2/s, and ρ = 998.2 kg/m^3. The mixing process of the two fluids occurs as they move along the micromixer, and the following equation can be used to determine its value, known as the η[49]:

$$\eta = 1 - \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{C_i - \overline{C}}{\overline{C}}\right)^2} \tag{4}$$

The scope of η varies from 0 to 1. When the mixing is intense, η approaches 1, and when there is an insignificant blending of two fluids, it approaches 0. Values of η equal to 1 and 0 correspond to complete mixing and complete lack of mixing (unmixable), respectively.

The ratio of $^{\Delta P}/_{\eta}$ is used to compare the performance of micromixers effectively. In this mathematical equation, obtaining a high η and low ΔP and a low $^{\Delta P}/_{\eta}$ are essential elements for a micromixer, which leads to a minuscule amount of pressure drop; $^{\Delta P}/_{\eta}$ refers to mixing cost, which is thus used in the efficiency calculation of micromixers and is termed 'performance index (PI)'.

To meet the Reynolds number requirement for each simulation, the inlet velocity was assigned a specific, constant value for both fluids entering the system. Boundary conditions at inlet 1 and 2 were defined as velocity inlets, while the outlet boundary condition was set to atmospheric pressure. All channel walls were modelled with no-slip boundary conditions. The SIMPLEC algorithm was applied for pressure–velocity coupling, and spatial discretization was performed accordingly.

The meta-heuristic algorithm optimized the microchannel to show the fluid dynamics precisely. Finding a function's maximum or minimum point in this stage involves using the model for $^1/_{\eta}$, ΔP and $^{\Delta P}/_{\eta}$.

2.3 | Model validation

Using the Batchelor scale, we can predict the minimal scale of the concentration field and estimate the finest concentration structures using the equation. $\lambda_B = \frac{\eta}{\sqrt{Sc}}$, where λ_B denotes concentration, η is Kolmogorov scale length, and Sc is Schmidt number. This study indicated that the microchannel width of 800 μm may be used to approximate η in viscous flow. Based on these parameters, the value of λ_B is 23.67 μm. The grid of 3,341,380 cells has a maximum edge length of 18.43 μm, accurately predicting concentration dispersion.

The findings from an earlier experiment conducted by Dundi et al.[45] using the identical mixing unit model and controlling variables serve as the validation. Figure 1 shows a micromixer model. The primary mixing is done in the micromixer. Two fluids with varying

19390019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

concentrations of 0 mol/m^3 and 1 mol/m^3 value were fed separately into the inlets (1 and 2) and the microchannel. The main objective is to increase mixing effectiveness by altering the design domain's flow channel's form. At the outflow, the mixture of fluid is obtained. Figure 3 shows the validation of OMTB with experimental work by Dundi et al.[45] graphically and through mass fraction contour comparison; from this, it can be concluded that the results are in agreement.

3 | META-HEURISTIC ALGORITHM

Several concurrent, frequently conflicting aspects influence engineering, mathematical, and economic optimization problems. Therefore, choosing optimization parameters that bring the outcomes as close to the ideal scenario as possible is crucial. In this study, the performance of the micromixer is optimized for many objectives at once using the maximum. η and the least ΔP. The objective function and constraints for this multi-objective optimization problem are as follows:

$$\begin{cases} \text{Max}: \eta = fn_1(\text{Re}, \text{Bend}.\text{Twist}) \\ \text{Min}: \Delta P = fn_2(\text{Re}.\text{Bend}, \text{Twist}) \\ \text{Subject}: \\ 0.05 \le \text{Re} \le 450 \\ 30^0 \le \text{Bend} \le 70^0 \\ 1 \le \text{Twist} \le 4 \end{cases} \quad (5)$$

Optimization of the micromixer is derived from meta-heuristic algorithms that were selected and coded in MATLAB software: RSM and HHO, whose steps are shown in Figure 4 as a flowchart.

3.1 | Response surface methodology (RSM)

RSM involves a mathematical model, typically a polynomial equation, that describes the relationship between the input variables and the response variable(s) of interest (Equation (6)). The model is fitted to experimental data from a series of designed experiments. These experiments are typically carried out using a factorial design, where different combinations of input variables are tested to observe their impact on the response variable.

$$Y = b_o + \sum_{i=1}^{3} b_i x_i + \sum_{i=1}^{3} b_i x_i^2 + \sum_{i=1j=1}^{3} b_i x_i x_j \quad (6)$$

RSM aims to determine the optimal input variables for maximizing or minimizing the output variable(s) within a specified range. ANOVA was used to assess the relevance of each input variable's role in influencing output responses and the model's dependability. From the 3D response surface plot, the interaction influence of input parameters on output was observed.

RSM with a desirability function is used to find the optimal solution for a multi-objective response. The solution with the highest combined desirability factor is considered optimal in the desirability approach. The individual desirability (dn) for an output parameter aimed at maximization is calculated using Equation (7),

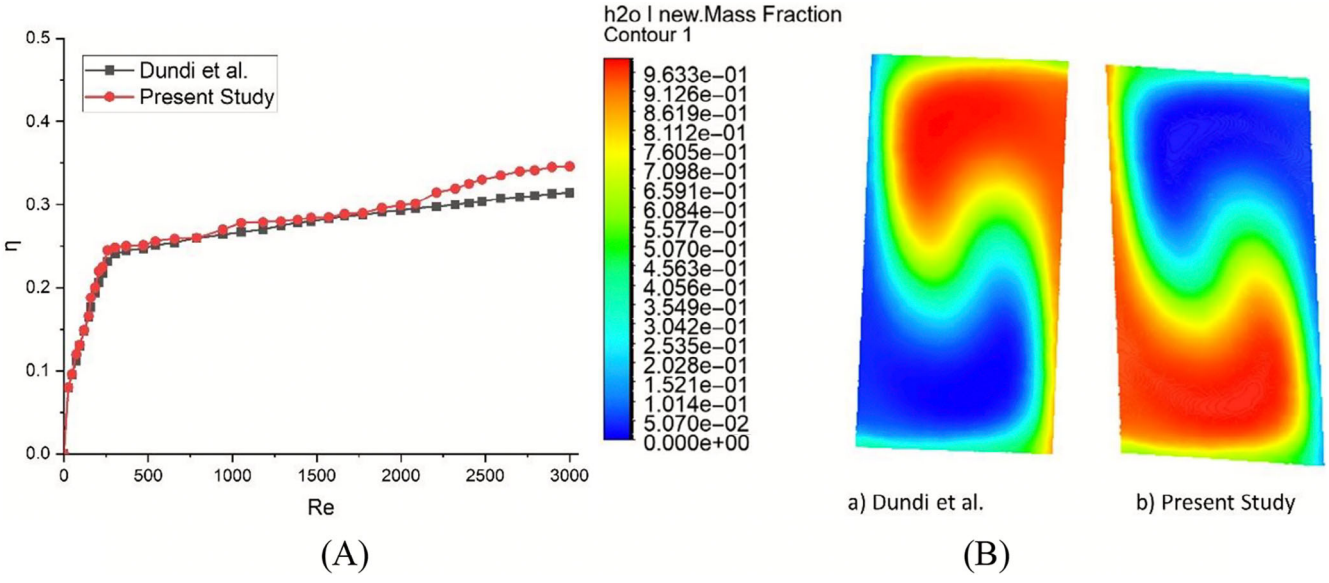


FIGURE 3 Proposed model Validation graphically and with mass fraction at the outlet from Dundi et al.[45]

19390019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

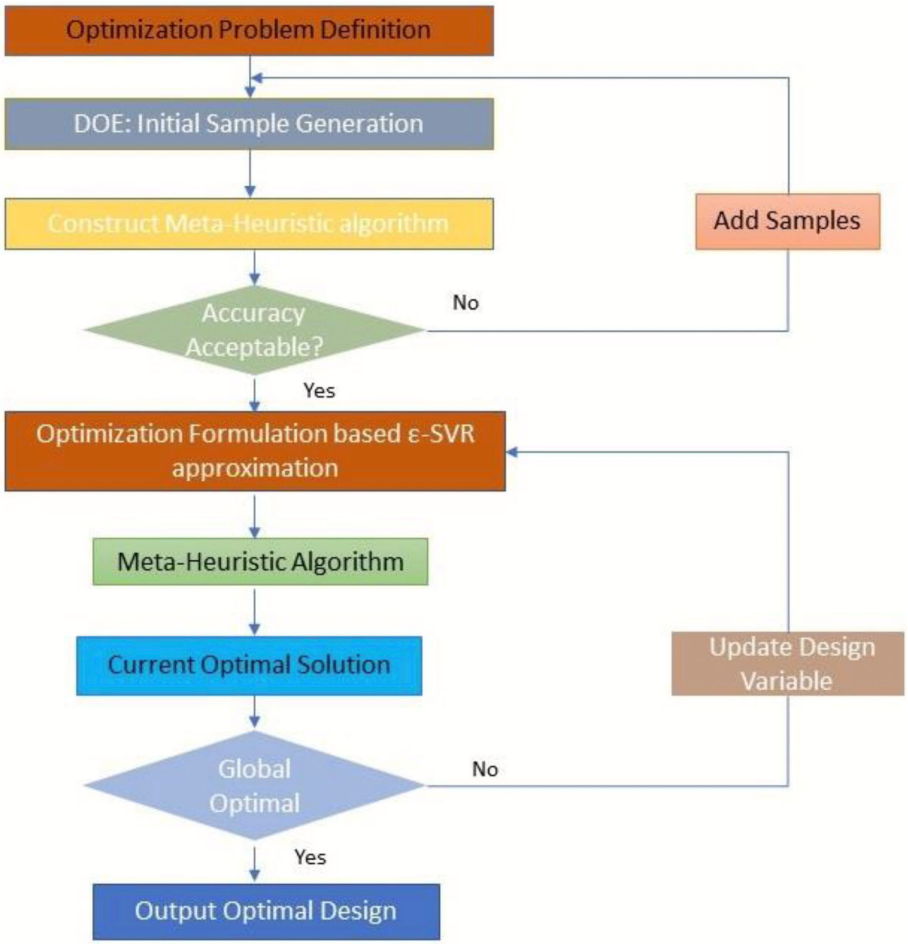


FIGURE 4 Optimization flow chart.

while for minimization, Equation (8) is used. The combined desirability is determined as described in Equation (9).

$$d_n = \begin{bmatrix} 0 \; n < L_n \\ \dfrac{n - L_n}{G_n - L_n} \; L_n < n < G_n \\ 1 \; n > G_n \end{bmatrix} \tag{7}$$

$$d_n = \begin{bmatrix} 0 \; n > H_n \\ \dfrac{n - H_n}{G_n - H_n} \; G_n < n < H_n \\ 1 \; n < G_n \end{bmatrix} \tag{8}$$

$$\text{Desirability} = [\pi(d_n w_n)]^{\frac{1}{w}}, \tag{9}$$

here, n, G_n, L_n, H_n, and r_n denote the predicted value, goal value, lower suitable value, higher suitable value, and desirability function weight for the nth output, respectively. w_n represents the importance of the nth output, and $W = \sum w_n$. The desirability factor ranges from 0 to 1, where 1 indicates the most desired ideal option and 0 represents the least desired.

The simulation results are statistically analyzed using ANOVA, a method based on the principle of general variance, employed to assess differences among all the findings in a sample. Using ANOVA is quite helpful and illuminating when comparing data across many categories. The p-value, F-value, R^2, Adj R^2, and Pred R^2 values in the ANOVA test, which represent the variance of the data from the mean, are indications of the validity and quality of the results.

3.2 | Artificial neural network (ANN)

An analytical tool called ANN is used to validate the correlation between input and output variables and verify the data's predictive regression. The multilayer perceptron (MLP) neural network architecture and the backpropagation technique are used in this ANN study. MATLAB R2020a is used for coding and execution. MLP planning, as shown in Figure 5, is in the format of A-H1-H2-Z, where A and Z indicate the input and output layers, and H1 and H2 are hidden layers. For each output, a network is built to train the DOE findings. Each

19390019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

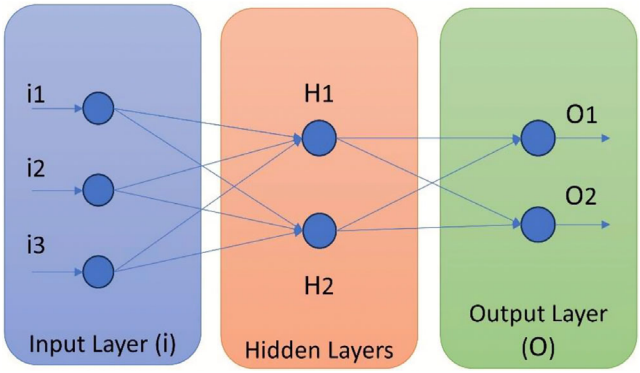


FIGURE 5 Artificial neural network (ANN) schematic.

hidden layer neuron in the network ranges in value from 1 to 10, and many neuron combinations are used during training. Then, the neuronal pairings with the lowest RMSE are chosen. The best neuron combination that was discovered is then utilized to build a new network for each output. Mean square error (MSE) was used as the performance function, while Levenberg Marquardt (trainlm) was used as the training function. The hidden layer uses the Tan-sigmoid, and the output layer is handled by Log-sigmoid functions of transfer, respectively. From datasets, 70% was used to train the network, 15% was used to test, and the rest was used to validate.

3.3 | Harris's hawk optimization (HHO)

HHO is a metaheuristic optimization algorithm inspired by the hunting behaviour of Harris's hawks, a species of raptors. The algorithm mimics the cooperative hunting strategy these birds employ to search for and capture prey.

In HHO, the optimization problem is represented as a population of hawks. Every hawk symbolizes a prospective solution to the problem, and its location within the search space corresponds to the values of the decision variables. The hawks collaborate and share information to improve the overall search process. Given that HHO is a population-based, gradient-free optimization method, any optimization issue may be addressed with it with the correct formulation. All of the HHO stages are depicted in Figure 6, and they are discussed in the following sections.[48]

3.3.1 | Exploration phase

Here, hawks explore and trail prey using their instinct. The Harris hawks in HHO huddle at isolated spots and use one of two methods to watch for prey to appear.

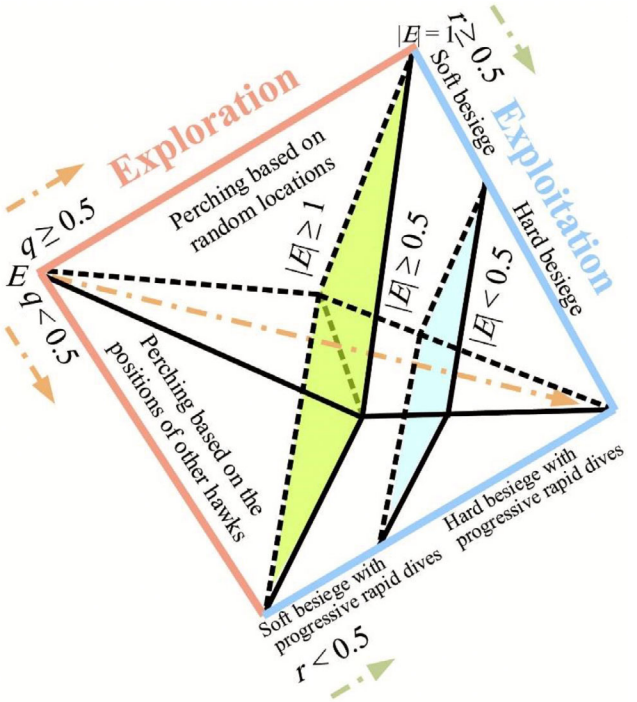


FIGURE 6 Various stages of Harris hawk optimization (HHO).[48]

$$X(t+1)=\begin{cases} X_{\text{random}}(t)-r_1|X_{\text{random}}(t)-2r_2X(t)| & q\le 0.5 \\ X_{\text{prey}}(t)-X_m(t)-r_3(Lb+r_4(\text{Ub}-\text{Lb})) & q<0.5 \end{cases} \quad (10)$$

$X(t+1)$ denote the Hawks position vector for iteration. X_{prey} denotes prey location, $X(t)$ symbolizes hawks' current location, and r_1, r_2, r_3, r_4, and q are arbitrary numbers between [0,1], which changes with iteration. Variables of upper and lower bonds are symbolized by Ub and Lb. $X_{\text{random}}(t)$ is the arbitrarily chosen hawk's location and X_m is the average current location of the hawks' population.

$$X_m(t)=\frac{1}{N}\sum_{i=1}^{N}X_j(t) \quad (11)$$

$X_i(t)$ is every hawk location in iteration, and N is the total hawk.

3.3.2 | The transformation from exploration to exploitation

The HHO technique exploits prey's behaviour to escape. Hence, the energy of prey diminishes. The energy of prey is represented as follows to illustrate this step:

19390019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$E = 2E_o\left(1 - \frac{t}{T}\right) \tag{12}$$

E in Equation (12) denotes the energy of prey fleeing, and E_o is the initial energy state.

3.3.3 | Exploitation phase

When hawk attacks prey, they try to escape, and the probability of fleeing is $|r| < 0.5$, getting caught $|r| \geq 0.5$.

3.3.4 | Soft besiege

Here $r \geq 0.5$ and $|E| \geq 0.5$, prey jumps arbitrarily here and there to flee. This action can be explained in the following equation:

$$X(t+1) = \Delta X(t) - E\left[JE_{\text{prey}}(t) - X(t)\right] \tag{13}$$

$$\Delta X(t) = X_{\text{prey}}(t) - E[\Delta X(t)] \tag{14}$$

Equations (13) and (14) depict the exploits of this model. ΔX(t) indicates the distinction between the prey location and the current location of iteration t.

3.3.5 | Hard besiege

Here $r \geq 0.5$ and $|E| < 0.5$, prey wear out and its present location is depicted in equation:

$$X(t+1) = X_{\text{prey}}(t) - E[\Delta X(t)] \tag{15}$$

3.3.6 | Soft besiege with progressive dive

In this stage, $|E| \geq 0.5$ and $r < 0.5$, prey has sufficient energy to flee. Hawks' movement is described by the following equation:

$$Y = X_{\text{prey}}(t) - E\left|X_{\text{prey}}(t) - X(t)\right| \tag{16}$$

3.3.7 | Hard besiege with progressive dives

Here $|E| < 0.5$ and $r < 0.5$, not able to prey, so hawks ambush and kill the prey. For this challenging besiege, the conditions are as follows:

$$X(t+1) = \begin{cases} Y \text{ if } F(Y) < F(t) \\ Z \text{ if } F(Z) < F(t) \end{cases} \tag{17}$$

$$Y = X_{\text{prey}}(t) - E\left|X_{\text{prey}}(t) - X_m(t)\right| \tag{18}$$

$$Z = Y + S \times \text{Lf} \tag{19}$$

4 | RESULTS AND DISCUSSION

In this study, numerical simulations of OMTB micromixers (Figures 7 and 8) that were built for Re in the range of 0.05–450 at various geometric parameters were carried out.

4.1 | Effect of Re on η

The most important parameter in the microchannel is the velocity with which the fluids were fed, which is considered as the Reynolds number, η at different Reynolds values, which is displayed through the micromixer in Figure 7.

Due to the low velocity at Re = 0.05, there is a significant residence time. Only molecular diffusion (N_A) can mix fluid particles. When the molecule diffusion channel is shortened, mixing is greatly accelerated. For Re = 5, the molecular diffusion process still produces mixing, but because the residence period is shorter than it is for Re = 0.05, molecular diffusion slowed down.

Rapid mixing results from the convection process and inertial force defeating molecular diffusion from Re 5. Re can be increased as much as possible without affecting the mixing rate. Molecular diffusion is reduced at Re = 1 and mixing is diminished. As a result, the mixing region between the two fluids is smaller, which accounts for the difference between mixing at Re = 0.05 and Re = 5.

While achieving a favourable η at high Reynolds numbers is favourable, the significant ΔP associated with these values is undesirable. Surprisingly, the mixing performance for Reynolds numbers of 0.05 and 100 is nearly identical.

4.2 | Effect of bend and twists on mixing performance

This research aims to determine how a twisted mixing channel with a bend affects an offset T-junction micromixer's performance. Micromixer models that have offsets with varied turns in twist along mixing channels (1–4) and bend angles (30°–70°).

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

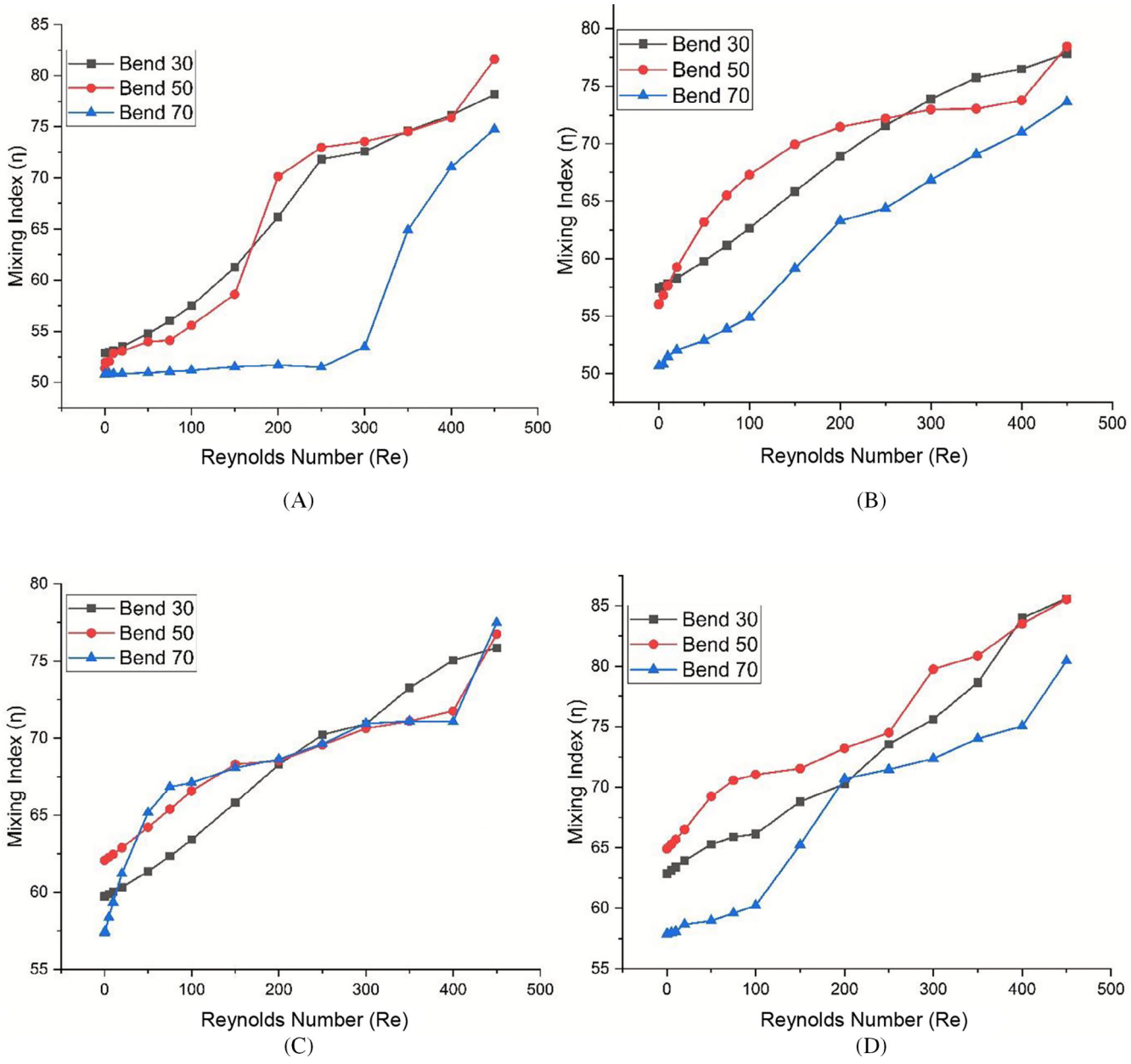


FIGURE 7 η versus Re for Twists 2, 3, and 4.

Figure 9 illustrates the velocity streamlines along the micromixer. For higher $L_m/D_h = 20$, the simple T-channel exhibits negligible secondary flow, but the channel having the bend and twist exhibits secondary flow that is detectable and increases with twist and bend. Bends can cause centrifugal forces that push fluids toward the outer edge of the curve, creating secondary flow patterns that enhance mixing across different fluid layers. Conversely, twists can induce rotational or helical flows, which help stretch and fold fluid interfaces, increasing the interfacial area and improving mixing efficiency. These effects are more pronounced in microfluidic environments, where laminar flow dominates, and even slight geometric changes can lead to significant mixing enhancements due to the increased influence of viscous forces over inertial ones.

The fluid flow of particles along the micromixer for the offset geometries is depicted in Figure 10. The tracking line, however, begins to bend around the micromixer after twisting is included. The impulsiveness intensifies as the bend and twist of mixing channels rise. Figure 11 shows the mass fraction at the outlet of OMTB, such as for (a, b, and c) for bend 30°, (d, e, and f) for 50°, and (g, h, and i) for 70°, for twists 2, 3, and 4, respectively.

Convective and diffusion mass transfer are the primary mixing mechanisms in the passive micromixer.

19390019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

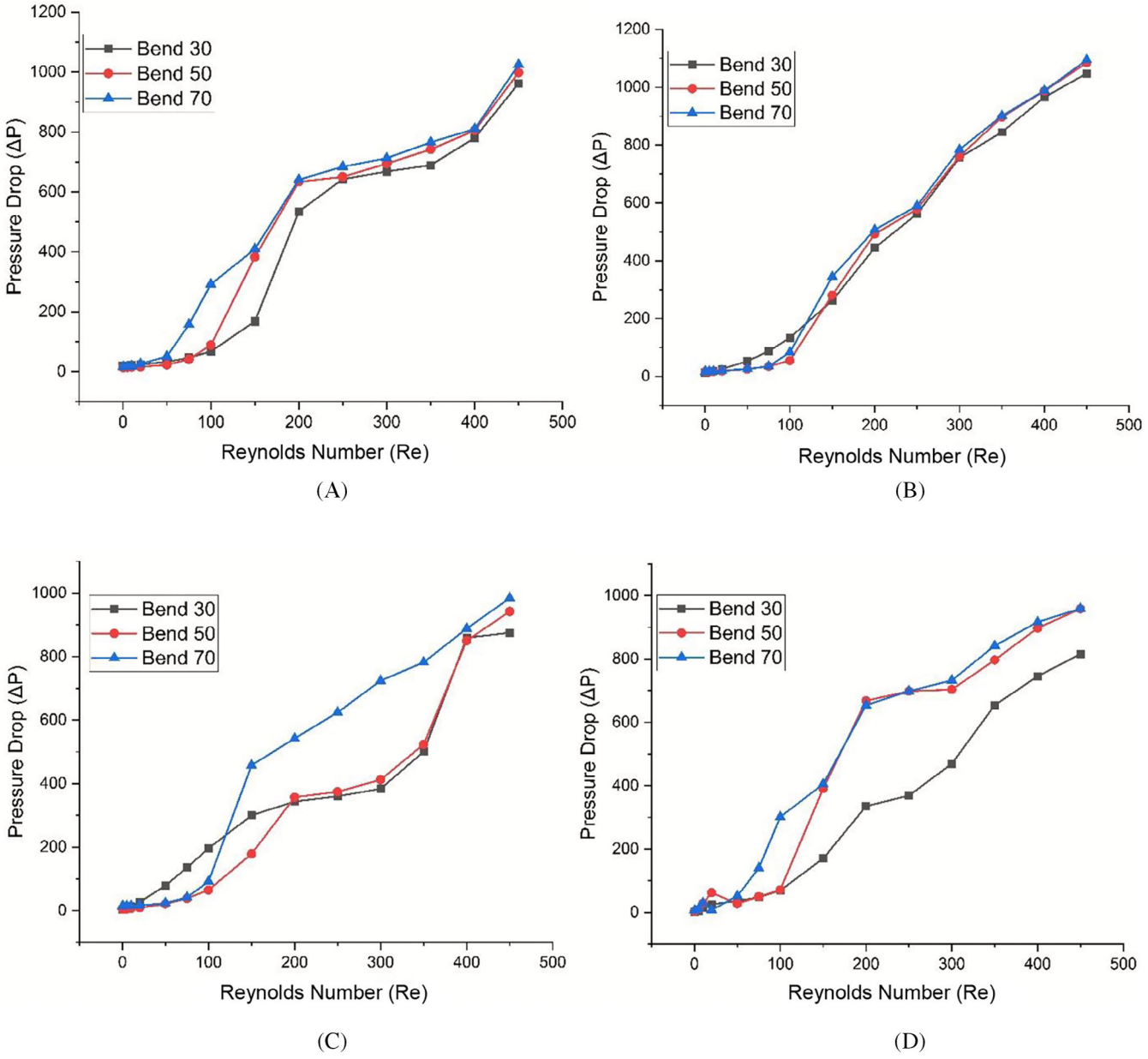


FIGURE 8 ΔP versus Re for Twists 2, 3, and 4.

Diffusion would be the primary factor controlling mixing performance without the secondary flow the twisted mixing channel created in the micromixer. Convective mass transfer becomes essential for the micromixers with bend and twist because of the increased secondary flow, which eventually leads to more excellent mixing performance.

As demonstrated, micromixers with twisted mixing channels perform better than their counterparts with no twist, especially between Reynolds numbers of 0.05 and 100. The mixing index shows a pronounced improvement in the range of Re from 0.05 to 450. Additionally, the twisting and bending-induced secondary flow is prominent here, mimicking the improved blending of micromixers.

Increased twist in a flow leads to more significant disruptions, resulting in improved mixing rates due to the addition of advective effects to laminar flow. ΔP in the micromixers increases after the introduction of whirling, which also improves mixing. In Figures 7 and 8, the ΔP and associated η were investigated for the microchannel with twists and bend channels at the specified Re. The rate for the pressure decrease is minimal despite the significant improvement in the η for Re at its mid-range.

Consequently, especially for the micromixer with an increased number of twists, a notable enhancement in performance is observed within the intermediate range of Re (betwixt 0.05 and 450). This performance improvement

19390019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

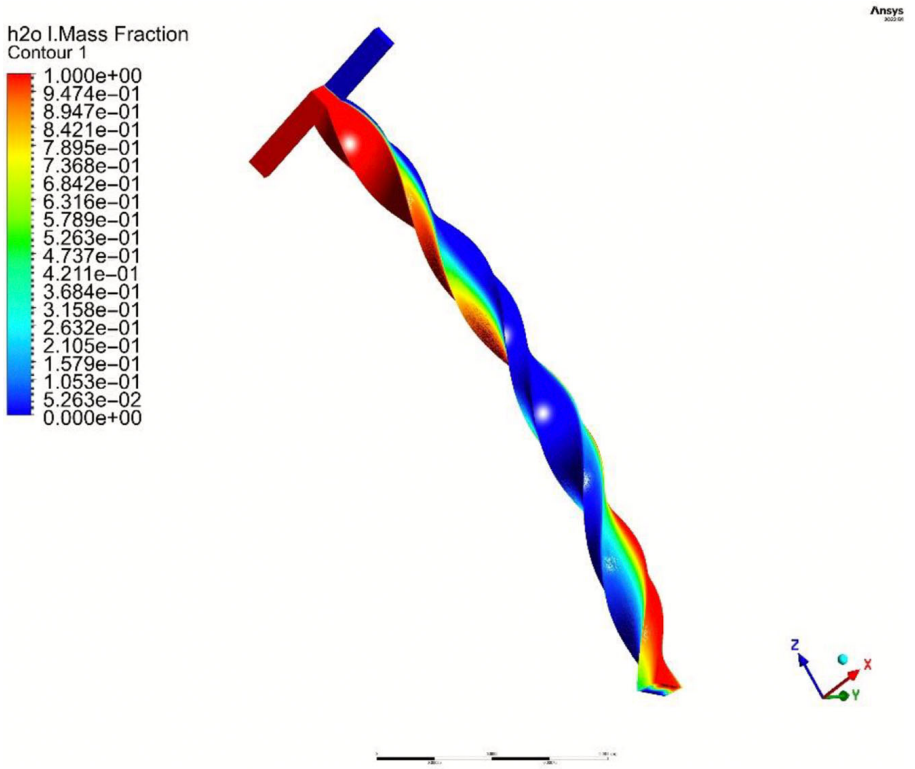


FIGURE 9 Velocity streamline across offset T-model micromixer with twists and bends (OMTB).

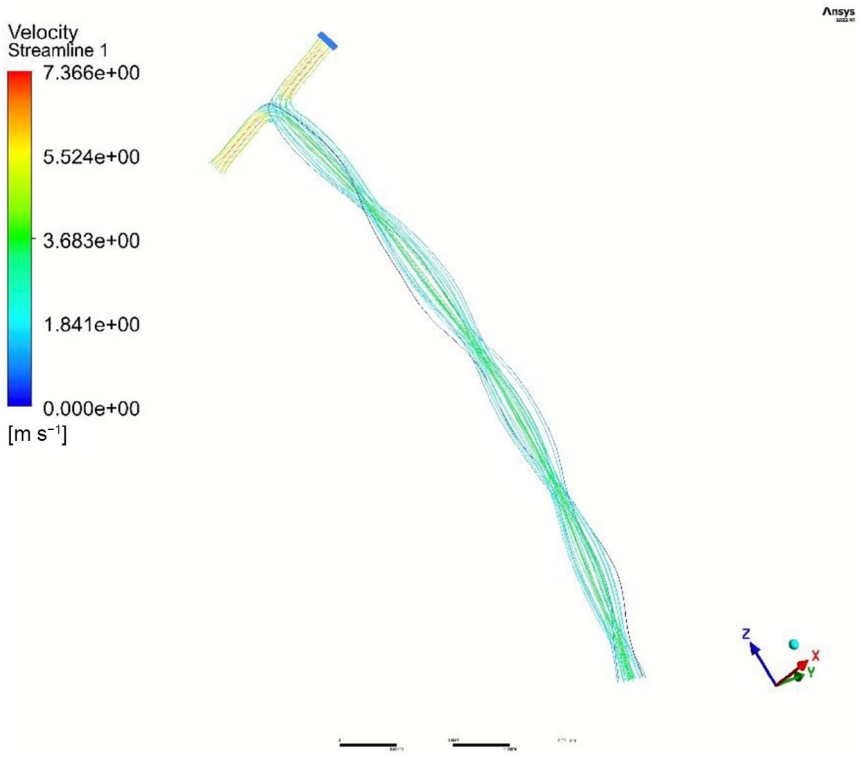


FIGURE 10 Mass fraction contour of offset T-model micromixer with twists and bends (OMTB).

index correlates with the mixing magnification achieved per unit increase in ΔP penalty. As demonstrated, introducing bending and twisting to micromixer channels can sometimes improve the performance index by more than three times, suggesting the potential of OMTB microchannel applications in the industry.

The results indicate no recirculation downstream due to the low inertia in the flow, with a Reynolds number

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25586 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Re) of 0.5. The curve occurs due to twists and bends in the streamlines, which causes a change in velocity direction. This phenomenon is crucial for effectively mixing substances, as it facilitates the movement of one material toward another.

The analysis of ΔP is of utmost importance as it signifies the pumping energy needed to sustain the flow within the micromixer. A low ΔP is essential for ensuring an efficient micromixer. Figure 8 compares ΔP and presents the results for each micromixer at various Re. It is observed that in all the micromixers, the ΔP increases with Re (Table 1).

The concentration profiles at the micromixer's intake suggest that the mixing between the two fluids is adequate close to the channel's inlet, according to the data. Upon reaching the first twist, a minor disruption in the flow ensues, resulting in an augmentation of its viscosity. The intensity of this process increases as the mixture passes through the 'mixing unit' (a configuration of barriers), resulting in increasingly pronounced disruptions in the flow. At the second twist, the mixing is more substantial than the first, indicating the impact of the impediments on the species mixing.

Observing the outlet area shows that the central axis interface is green, which signifies a complete mixing process (100%). In contrast to the original region, there are no longer distinct areas consisting solely of one species (red or dark blue). Instead, there is now a gradient of mixed species across the cross-section of the channel.

The disparity between the η and ΔP values acquired at the ideal locations from the response surfaces through simulation is minimal, particularly for η. In contrast, ΔP exhibits more pronounced disparities, even when employing more accurate measurements, as depicted in Figure 8, although it remains satisfactory. The schematic demonstrates that the channel length of each micromixer design varies according to the increased twists and bends. Thus, the localized pressure dips are more evenly spread.

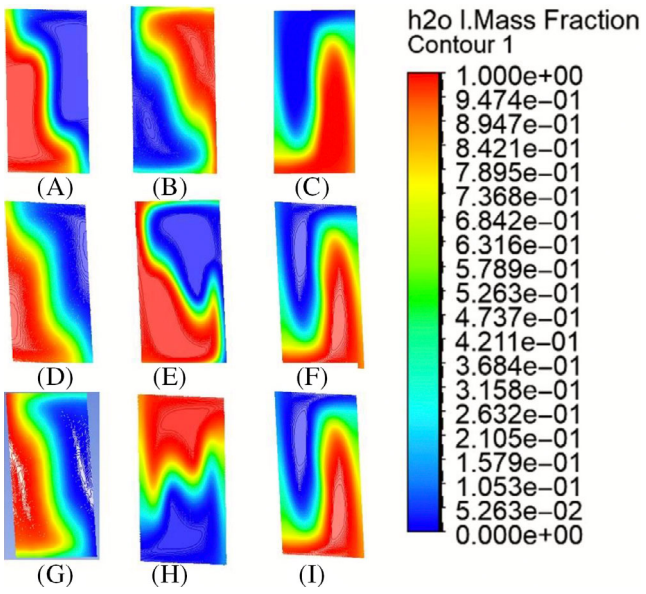


FIGURE 11 Mass fraction contour of offset T-model micromixer with twists and bends (OMTB) at the outlet.

4.3 | Optimization

The goal is to concurrently determine the twist and bend that will provide the best possible circumstances for optimal mixing performance (η) and minimal pressure drop (ΔP), which cannot be attained using the previously outlined technique. The Python Platypus package implemented a multi-objective optimization technique based on HHO. Along with the previously employed limitation condition, the three-goal functions (maximizing η, minimizing [ΔP and $^{\Delta P}/_{\eta}$]) were defined and considered. With output limitation from the obj, the number of obj and constraints are considered to calculate the various input variables. Finally, by removing the strange ones, a workable answer was developed.

4.3.1 | RSM modelling

RSM subsequently helps to equate the impact of variables and their relationships on the outcomes. The upper and lower limits of the input parameters are presented to the RSM, along with the selection parameters shown in Table 2. The recursive behaviour between the variables

TABLE 2 Lowest value of optimization value.

Obf	Min	η	ΔP	Twists	Bend	Re
η	50.64	50.64	-	2	70°	0.05
ΔP (Pa)	13.28	-	13.28 Pa	3	30°	0.05
$^{\Delta P}/_{\eta}$	26.22	52.86	13.28 Pa	3	30°	0.05

	Twist	Bend	Re	η	ΔP (Pa)	$\Delta P/\eta$
No offset[50]	4	50°	450	85.512%	998.06	1167.15
With offset	4	60°	400	87.76%	974.20	1109.86

TABLE 1 Comparison of model with and without offset micromixer model.

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

for bend and twists and how they impacted the η and ΔP are shown in Figures 12 and 13. For all bend and twist values, the η increases with increasing Re, as seen in Figure 14. By analyzing the charts in Figure 14, it can be deduced that a decrease in Re in all values of bend and twist leads to a fall in ΔP and that a reduction of bends and twists also prevents the pressure drop from growing, which is the optimum situation for micromixers (Table 3).

The ANOVA Table 4 model for the output parameter is statistically significant, with p-values less than 0.05 for both the model and the current inputs. All output responses had high R^2 and adjusted R^2 values, indicating that the model provides data that is very similar to the experimental data. The adj. R^2 and predicted (Pred.) R^2 values were less than 2% for all replies, showing high prediction reliability. The lack of fit is insignificant as the p value is more prominent than 0.05.

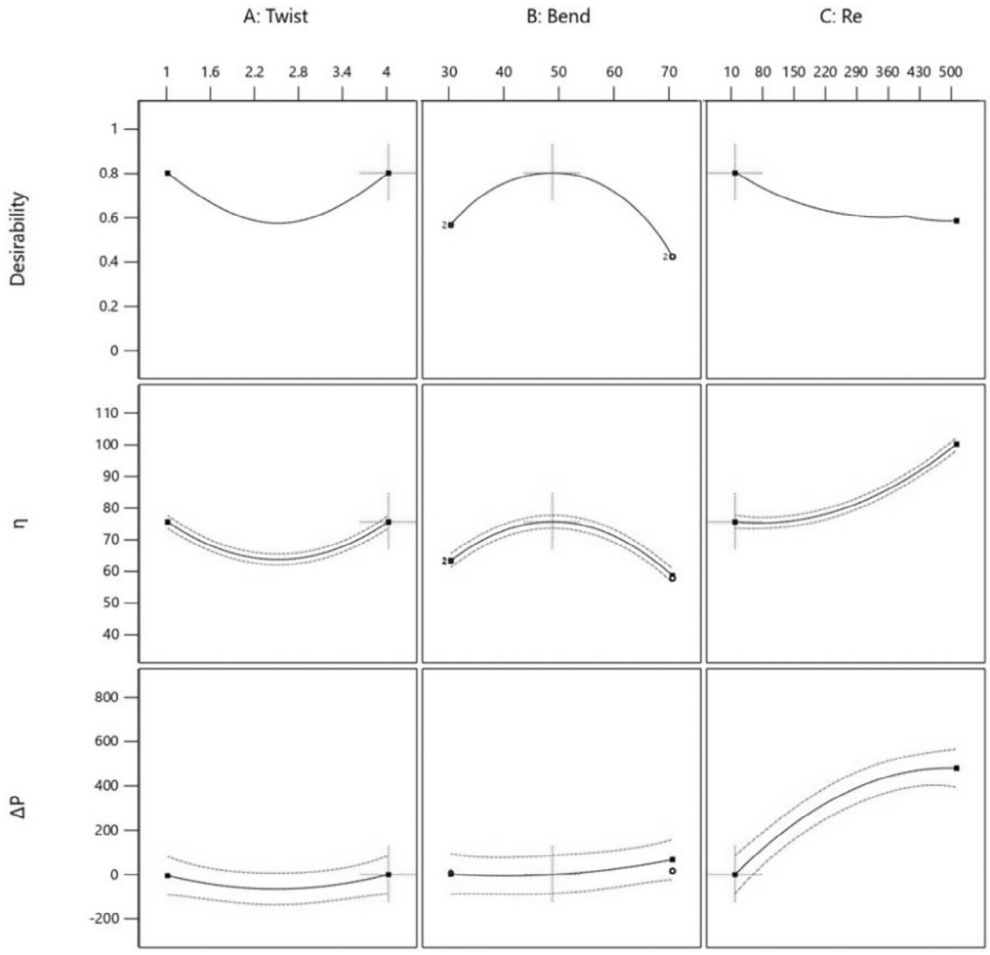


FIGURE 12 Surface response effect for η: (A) Re and bend; (B) bend and twist; (C) Re and twist.

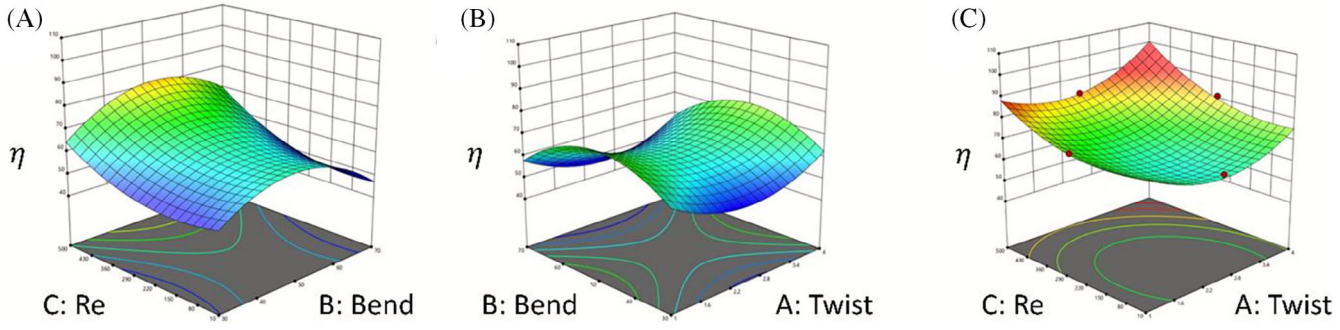


FIGURE 13 Surface response for ΔP: (A) Re and bend; (B) bend and twist; (C) Re and twist.

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

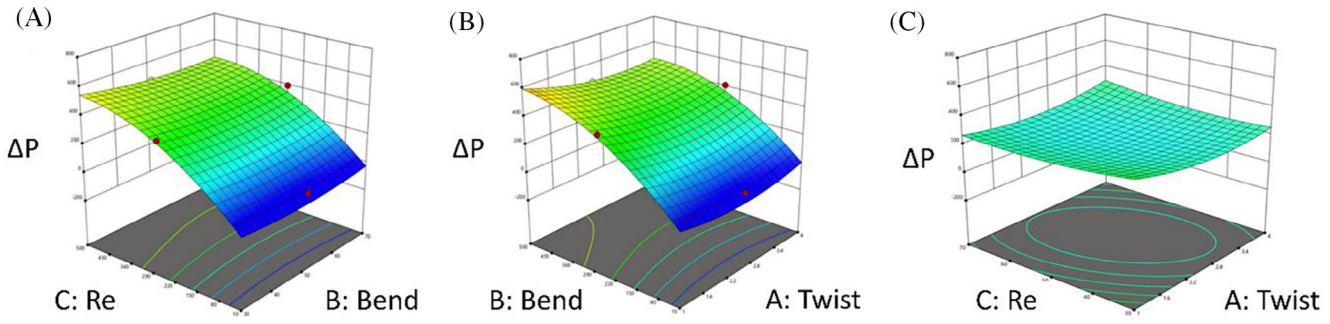


FIGURE 14 All factor desirability of response surface method (RSM).

TABLE 3 Design of experiment (DOE) with responses.

Twists	Bends	Re	η	ΔP
4	70°	10	57.8	16.78301
2.5	70°	255	52.2336	696.8079
2.5	50°	255	55.9549	362.1438
1	30°	450	70.92	1020.14
1	30°	10	58.3499	12.7164
4	70°	450	89.02	528.6035
2.5	50°	255	55.9549	362.1438
1	50°	255	82.2306	447.0021
2.5	50°	10	69.5533	12.71834
4	30°	10	63.12	12.73183
4	50°	255	87.3669	680.9073
1	70°	10	58.36	12.81532
2.5	30°	255	71.2162	402.6061
2.5	50°	450	82.6814	477.5227
1	70°	450	76.3117	532.9106
4	30°	450	80.3	479.4351

Re-alterations have been significantly more effective than other variables in minimizing the pressure drop. According to the data gathered from the simulations, the variables' interactions are the same for each of the three distinct baffles; the only difference is in the range in which the surfaces are placed.

4.4 | ANN

The dataset was formed using the simulation results shown in Figures 7 and 8 for ANN training, creating a prediction model. The twist and bend angles were included as input variables, and η and ΔP were included as output variables to organize and prepare the data collected from the simulations for use in the neural network. The data was split into three sets for the training process, with 70% being utilized for training, 15% for testing, and 15% for validation. This distribution makes the input and output variables handled similarly during the training process with an equal impact on the model. Figure 15 shows the linear fitting with R^2 values 0.95529 and 0.97209 for η and ΔP, respectively. Linear fit graphs show the link between observed and expected values. In our case, observed values are taken from simulations, while predicted data are gathered from the HHO response surface. The proximity of our observation sets to this line indicates good accuracy in responding to areas.

4.5 | HHO modelling

The HHO optimization process involves five phases: initial population, fitness function, selection, crossover, and mutation. The HHO is responsible for evaluating and choosing the solution based on the objective function derived from the ANN.

With a trained dataset (using ANN), multi-objective function optimization was enabled using HHO. Mainly, three obj were formulated for optimization, that is, (i) $1/\eta$, (ii) ΔP, (iii) $1/\eta + \Delta P$, and (iv) $\Delta P/\eta$. The performance index (PI), denoted as $\Delta P/\eta$, is a measure that specifically pertains to the cost associated with mixing, also known as the mixing cost (mc). With the availability of data and resources, the system or application requirements were considered while choosing this objective function. The created neural network model has produced accurate and trustworthy results using this objective function. A penalty value is imposed on the function's output if the link between the inputs and outputs is not met. The penalty must be much higher than those generally obtained at the exit to ensure this condition is far from the function's minimum.

The proper goal function is first provided, and then the evolutionary algorithm is executed. The sample size and the number of iterations for this method must be

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

precisely specified since they substantially impact its overall performance.

There was shown to be a link between max η and ΔP. Similarly, the connection between ΔP min and the smallest value η was found. In this study, in addition to the optimal point determined by the minimum of $^{\Delta P}/_{\eta}$, we also take into consideration the points with a minimum of $^{\Delta P}/_{\eta}$ and a maximum of η. We obtained the twist and bend values that match the minimum of $^{\Delta P}/_{\eta}$. As was previously documented, ΔP declines while η stays high in this location.

Figure 16 juxtaposes the findings of the present research with other published works. The dotted lines indicate the pressure differential, while the solid lines represent the mixing index. It is essential to mention that Cortes-Quiroz et al. and Solehati et al.[19,51] optimized their mixer only for a given Re. The graph illustrates that the present study exhibits a greater η at the Reynolds number than prior published works.

The simulations performed better in fitting ΔP than η, suggesting that the mixing performance is more responsive to the formulation. Ultimately, the models achieved satisfactory metrics to determine the extent of the task. The table in Table 4 displays the most favourable values of η and their corresponding ΔP (Tables 2 and 5).

The optimization findings revealed that the optimized case's errors for the prediction of η, ΔP, and $^{\Delta P}/_{\eta}$ were 0.378%, 1.4%, and 1.9%, respectively. These values are accurate to the actual value. The restrictions of the

TABLE 4 ANOVA F-value and p-value for η and ΔP.

	η		ΔP	
Source	***F*-value**	***p*-value**	***F*-value**	***p*-value**
Model	458.48	< 0.0001	64.91	< 0.0001
A-Twist	245.70	< 0.0001	6.31	< 0.0001
B-Bend	20.59	< 0.0001	1.80	< 0.0001
C-Re	1240.73	< 0.0001	537.55	< 0.0001
AB	14.08	0.1386	2.59	0.1386
AC	7.48	0.0211	7.46	0.0211
BC	164.58	0.3472	0.9728	0.3472
Std. Dev.	0.7160		36.69	
Mean	69.12		316.15	
C.V %	1.04		11.60	
R^2	0.9976		0.9832	
Adjusted R^2	0.9954		0.9680	
Adequate precision	78.9605		28.4307	

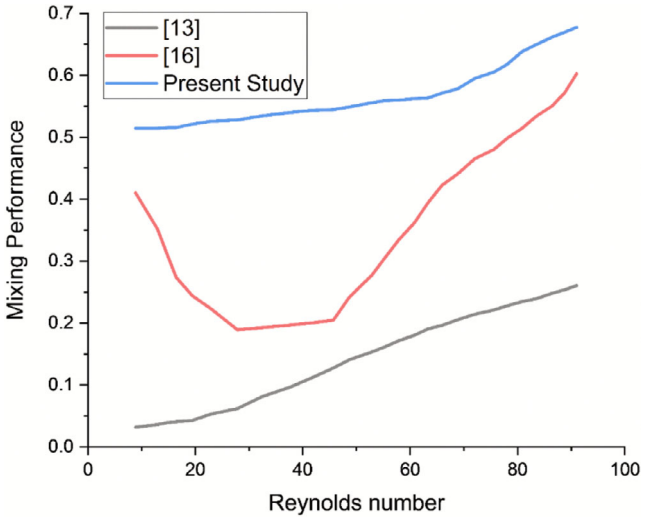


FIGURE 16 The juxtaposition of the findings from the present study with previous publications.

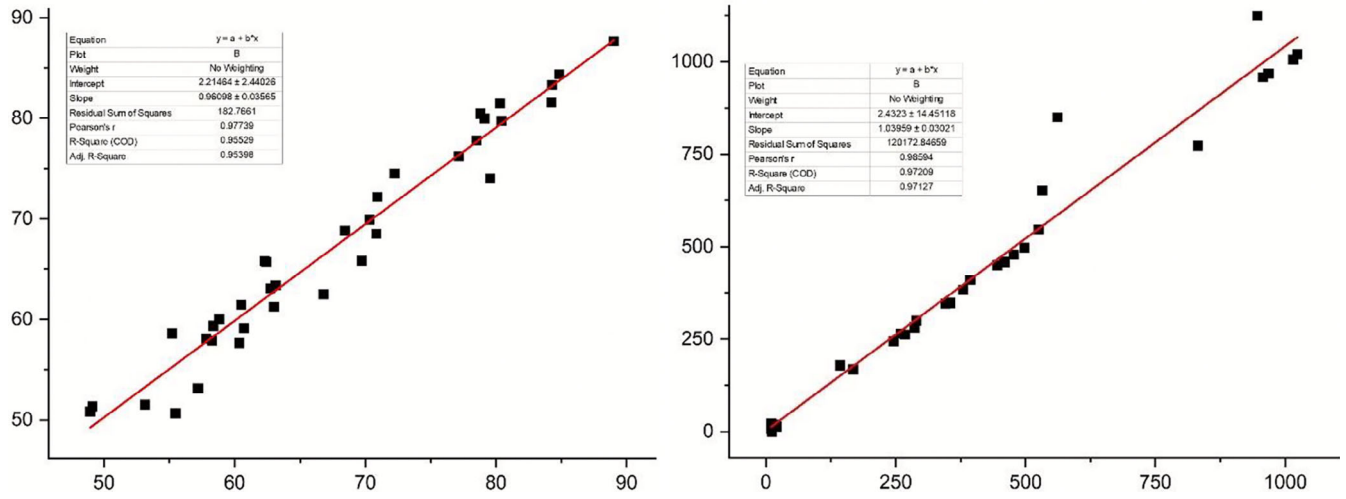


FIGURE 15 Linear fit plot for η and ΔP.

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Obf	Max	η (%)	ΔP	Twists	Bend	Re
η(%)	87.76	87.76	-	4	60°	400
ΔP (Pa)	1094.55	-	1094.55	3	70°	450
$\Delta P/\eta$	1280.02	85.51	1094.55	3	70°	450

TABLE 5 The highest value of optimization results.

	Twists	Bends	Re	η	ΔP (Pa)	$\Delta P/\eta$
Simulation	4	50°	450	85.512%	998.06	1167.15
Optimal	4	60°	400	87.76%	974.20	1109.86

TABLE 6 Comparison of simulation and optimized results for offset T-model micromixer with twists and bends (OMTB).

TABLE 7 Comparison of response surface method (RSM) and Harris Hawk optimization (HHO) optimality for offset T-model micromixer with twists and bends (OMTB).

	Twist	Bend	Re	η	ΔP (Pa)	$\Delta P/\eta$
RSM	4	48°	450	79.31%	998.65	1259.17
HHO	4	60°	400	87.76%	974.20	1109.86

multi-variable optimization approach utilized may be to blame for the mistake in the maximum and lowest situations. Tables 6 and 7 compare the best results from simulation and optimization for the variables.

5 | CONCLUSION

This study utilized the principles of constructal theory to undertake an evolutionary design of passive micromixers. The constructal design method, in conjunction with the RSM, HHO, and CFD, serves as a framework for improving the performance of a given configuration by gradually introducing additional degrees of freedom and progressing toward more optimal designs. In this investigation, the best values of parameters impacting the performance of the micromixer were explored with CFD simulations and metaheuristic algorithms to obtain the maximum η and minimal ΔP. A neural network was trained and tested using the gathered data, and the results showed more than 99% accuracy for predictions η and 98.5% for ΔP predictions. A global optimization strategy based on the twist and bend parameter was used to obtain the ideal model.

The optimized OMTB model's simulation-based verification revealed slight inaccuracies of 0.378% and 1.4% in the calculated values of max η (87.76%) and min ΔP (13.28 Pa). It shows a micromixer with precise η and ΔP values from this optimization procedure.

The findings of this investigation led to the following conclusions.

- The CFD simulations indicate that mixing is enhanced at both low and very high Reynolds numbers (Re). However, pressure drops increase at high Re, countering the optimization goal. This occurs due to elevated transverse velocities within the channel bends and twists, which enhance mixing. A higher transverse velocity factor intensifies chaotic convection, increasing mixing efficiency (η) while aiming to minimize pressure drop (ΔP).
- The response surfaces also demonstrated that, after crossing the threshold point, the mixing improved with increasing Re in all values of bend and twists. With an overall desirability of 0.898 for a micromixer with bends and twists with $\eta = 79.31\%$ and $\Delta P = 988.65$ Pa, at input variable of $\mathrm{Re} = 450$, $\mathrm{bend} = 48°$, and $\mathrm{twists} = 4$ was obtained.
- HHO outperformed RSM, optimizing for optimal parameters. The preferred design, with parameters, are $\mathrm{Re} = 400$, $\mathrm{bend} = 60^{\circ}$, and $\mathrm{twist} = 4$, was determined for optimality, having $\eta = 87.76\%$ and $\Delta P = 974.20$ Pa.
- A better ratio of the $\Delta P/\eta$ achieved using RSM rather than through optimization with HHO. HHO and RSM came up with results of 1109.86 and 1259.17, respectively. The best input factor values from the RSM and HHO approaches were used in the CFD simulation.

Future research can optimize micromixers by exploring alternative passive designs or applying pillars or obstacles to improve performance across different fluidic systems. Additionally, multi-objective optimization algorithms that consider factors like pressure drop, cost, and mixing efficiency could lead to more well-rounded designs. Expanding the study to include non-Newtonian and complex fluids would allow for broader applications, particularly in industries like biomedicine.

NOMENCLATURE

C concentration
w channel width

19390019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

L	channel length
d	channel depth
η	mixing performance
Q	flow rate
Obf	objective function
Re	Reynolds number
Sc	Schmidt number
u	fluid velocity
ΔP	pressure drops
μ	fluid viscosity
γ	fluid diffusion coefficient
ρ	fluid density

AUTHOR CONTRIBUTIONS

Md Gulam Mustafa: Conceptualization; methodology; validation; writing – original draft; software. **Mohammad Zunaid:** Visualization; resources; software. **Samsher Gautam:** Supervision.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available upon request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Md Gulam Mustafa https://orcid.org/0000-0003-1947-5703

REFERENCES

[1] S. J. Haswell, R. J. Middleton, B. O'Sullivan, V. Skelton, P. Watts, P. Styring, *Chem. Commun.* **2001**, *5*, 391.
[2] A. Nagaki, K. Kawamura, S. Suga, T. Ando, M. Sawamoto, J. I. Yoshida, *J. Am. Chem. Soc.* **2004**, *126*, 14702.
[3] R. Keçili, S. Büyüktiryaki, C. M. Hussain, *Handbook of Nanomaterials in Analytical Chemistry: Modern Trends in Analysis*, Elsevier, Amsterdam, The Netherlands **2019**, p. 185.
[4] J. Ortega-Casanova, *Chem. Eng. Process.: Process Intesif.* **2017**, *117*, 18.
[5] G. Kunti, A. Bhattacharya, S. Chakraborty, *Phys. Fluids* **2018**, *30*, 062004.
[6] V. E. Papadopoulos, I. N. Kefala, G. Kaprou, G. Kokkoris, D. Moschou, G. Papadakis, E. Gizeli, A. Tserepi, *Microelectron. Eng.* **2014**, *124*, 42.
[7] J. Charmet, P. Arosio, T. P. J. Knowles, *J. Mol. Biol.* **2018**, *430*, 565.
[8] O. Bilsel, C. Kayatekin, L. A. Wallace, C. R. Matthews, *Rev. Sci. Instrum.* **2005**, *76*, 014302.
[9] K. Petkovic, G. Metcalfe, H. Chen, Y. Gao, M. Best, D. Lester, Y. Zhu, *Lab Chip* **2017**, *17*, 169.
[10] L. Wang, D. Liu, X. Wang, X. Han, *Chem. Eng. Sci.* **2012**, *81*, 157.
[11] C. Duan, W. Wang, Q. Xie, *Biomicrofluidics* **2013**, *7*, 26501.
[12] S. Zhang, X. Chen, Z. Wu, Y. Zheng, *Int. J. Heat Mass Transfer* **2019**, *133*, 1065.
[13] H. Lv, X. Chen, X. Li, Y. Ma, D. Zhang, *Int. Commun. Heat Mass Transfer* **2022**, *131*, 105867.
[14] X. Chen, J. Shen, *Int. J. Heat Mass Transfer* **2017**, *106*, 593.
[15] J. G. Romero, J. A. Moreno, Á. A. M. Aguilar, *Automatica* **2021**, *130*, 109710.
[16] Z. Y. Gao, G. Guo, *Science China Information Sciences* **2021**, *64*, 179204.
[17] S. R. Bazaz, A. A. Mehrizi, S. M. Javid, *presented at Can. Soc. Mech. Eng. (CSME) Int. Conf.*, Ottawa, Canada **2016**.
[18] Y. Yang, X. Wei, W. Yao, J. Lan, *Sens. Actuators, A* **2023**, *364*, 114819.
[19] C. A. Cortes-Quiroz, A. Azarbadegan, M. Zangeneh, *Sens. Actuators, B* **2014**, *202*, 1209.
[20] C. Y. Lee, C. L. Chang, Y. N. Wang, L. M. Fu, *Int. J. Mol. Sci.* **2011**, *12*, 3263.
[21] M. Bayareh, M. N. Ashani, A. Usefian, *Chem. Eng. Process.* **2020**, *147*, 107771.
[22] W. Han, X. Chen, *J. Chem. Technol. Biotechnol.* **2020**, *95*, 1622.
[23] H. Sadique, Q. Murtaza, *Int. J. Heat Mass Transfer* **2022**, *194*, 123063.
[24] M. Ashar Sultan, C. P. Fonte, M. M. Dias, J. C. B. Lopes, R. J. Santos, *Chem. Eng. Sci.* **2012**, *73*, 388.
[25] J. C. Kurnia, A. Ahmadihosseini, A. P. Sasmito, *Chem. Eng. Process.* **2022**, *181*, 109171.
[26] G. Kunti, A. Bhattacharya, S. Chakraborty, *Electrophoresis* **2017**, *38*, 1310.
[27] S. H. Ding, W. H. Chen, K. Q. Mei, *IEEE Trans Ind Electron* **2020**, *67*, 1222.
[28] E. H. T. Cunegatto, F. S. F. Zinani, C. Biserni, L. A. Rocha, *Int. J. Heat Mass Transfer* **2023**, *215*, 124519.
[29] W. Raza, S. Hossain, K. Y. Kim, *Sens. Actuators, B* **2018**, *258*, 381.
[30] X. Chen, T. Li, H. Zeng, Z. Hu, B. Fu, *Int. J. Heat Mass Transfer* **2016**, *98*, 131.
[31] X. Chen, T. Li, *Chem. Eng. J.* **2017**, *313*, 1406.
[32] D. Owen, M. Ballard, A. Alexeev, P. J. Hesketh, *Sens. Actuators, A* **2016**, *251*, 84.
[33] H. Van Phan, M. B. Coşkun, M. Şeşen, G. Pandraud, A. Neild, T. Alan, *Lab Chip* **2015**, *15*, 4206.
[34] M. Hidaka, M. Kojima, S. Sakai, *Lab Chip* **2024**, *24*, 4571.
[35] X. Chen, X. Chen, *Anal. Methods* **2024**, *16*, 2368.
[36] C. Lv, Y. Liu, G. Li, T. Yang, H. Zhao, X. Chen, *Case Studies in Thermal Engineering* **2024**, *60*, 104741.
[37] D. O. de Maionchi, L. Ainstein, F. P. dos Santos, M. B. de Souza Júnior, *Int. J. Heat Mass Transfer* **2022**, *194*, 123110.
[38] M. Nikpour, A. Mohebbi, *Chem. Eng. Process.* **2022**, *170*, 108713.
[39] M. Yoshimura, K. Shimoyama, T. Misaka, S. Obayashi, *Microfluid. Nanofluid.* **2019**, *23*, 1.
[40] S. Ahmad, R. M. Singari, R. S. Mishra, *Mater. Manuf. Processes* **2021**, *36*, 843.
[41] O. Elhaki, K. Shojaei, *Ocean Eng.* **2018**, *167*, 239.
[42] H. M. Lu, T. Uemura, D. Wang, *Mobile Netw Appl* **2020**, *25*, 1008.
[43] H. Zhang, Y. Xu, R. Luo, Y. Mao, *China Communications* **2023**, *20*, 70.
[44] X. Chen, H. Lv, *Sci. Rep.* **2022**, *12*, 4162.
[45] T. M. Dundi, V. R. K. Raju, V. P. Chandramohan, *Australian Journal of Mechanical Engineering* **2021**, *19*, 363.
[46] S. H. Wong, M. C. L. Ward, C. W. Wharton, *Sens. Actuators, B* **2004**, *100*, 359.

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[47] J. Rahmannezhad, S. A. Mirbozorgi, *Int. J. Heat Mass Transfer* **2019**, *140*, 483.
[48] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, H. Chen, *Future Generation Computer Systems* **2019**, *97*, 849.
[49] H. Lv, X. Chen, X. Wang, X. Zeng, Y. Ma, *Int. J. Heat Mass Transfer* **2022**, *183*, 122159.
[50] M. G. Mustafa, M. Zunaid, S. Gautam, *Chem. Eng. Process.* **2023**, *190*, 109436.
[51] N. Solehati, J. Bae, A. P. Sasmito, *Comput. Fluids* **2014**, *96*, 10.

How to cite this article: M. G. Mustafa, M. Zunaid, S. Gautam, *Can. J. Chem. Eng.* **2024**, 1. https://doi.org/10.1002/cjce.25566

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25566 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2024.3510812

Optimizing Secure Data Transmission in 6G-Enabled IoMT Using Blockchain Integration

Aruna Malik, Vikas Tyagi, Samayveer Singh, *Senior Member, IEEE*, Rajeev Kumar, Huaming Wu, *Senior Member, IEEE,* and Sukhpal Singh Gill

***Abstract*—The advent of sixth-generation (6G) technology is poised to revolutionize connectivity, particularly by enhancing the integration of Internet of Medical Things (IoMT) devices. This advancement offers ultra-fast data transmission, low latency, and high mobility but also brings the challenge of ensuring secure and energy-efficient communication. To solve these challenges, this paper introduces a novel hybrid greylag goose-based optimized clustering (HGGOC) algorithm. It merges the efficiency of greylag goose optimization with the precision of the golden sine strategy. The Lévy flight mechanism guides the algorithm to optimize cluster head selection in 6G-enabled IoMT networks. The integration of blockchain technology further strengthens data security and transparency. Simulation results show that HGGOC surpasses existing methods, delivering up to 64% improvement in network stability, a 47% increase in node lifetime, and a 59% boost in energy efficiency and data throughput. These findings position HGGOC as a promising solution for sustainable communication in 6G-enabled IoMT environments.**

***Index Terms*—Internet of Medical Things, Blockchain, Data Dissemination, Energy Efficiency, Sustainable Communication**

I. Introduction

In recent years, advancements in sensing technologies have underscored the growing need for reliable deployment and integration of Internet of Medical Things (IoMT) devices. As these devices become more prevalent, modern communication technologies, particularly wireless networks, have become the backbone for connecting IoMT systems [1]. However, despite the adoption of fifth-generation (5G) and beyond-5G technologies to support these networks, they face significant challenges such as limited data rates and insufficient capacity. These limitations make it difficult to handle advanced applications like holographic communications. These limitations have spurred growing interest in the development of sixth-generation (6G) technology, which promises to address these shortcomings and offer enhanced capabilities [2].

6G technology is characterized by its potential for ultra-high data rates and minimal end-to-end delays, which are essential for a wide range of applications, including forest fire detection, IoMT, and intelligent transportation systems [3]. These diverse applications form a triangular coverage model, illustrating the vast utility of 6G technology in smart city scenarios [4]. However, a significant challenge with 6G and IoMT devices lies in their high energy consumption and limited battery life, which can impede their ability to maintain intelligent and seamless connectivity [5]. This situation necessitates the development of energy-efficient, or sustainable communication strategies for managing the connectivity of a vast array of 6G-enabled IoMT devices. Nevertheless, many key challenges exist: In large-scale IoMT networks, managing energy consumption is crucial due to the limited battery power of numerous devices, necessitating energy-efficient communication protocols like the one proposed in our work. Additionally, ensuring the security and integrity of medical data is a significant challenge, particularly due to the decentralized nature of IoMT devices, which we address through blockchain integration [6]. As these networks grow, scalability becomes increasingly complex, requiring efficient management of cluster head selection, network stability, and reliable data transmission. Moreover, in healthcare applications, maintaining low latency while ensuring energy efficiency and security presents a critical technical challenge [7]. In addressing these challenges, our work focuses on optimizing the cluster head (CH) selection process for cluster-based routing in IoMT networks. The existing research, such as [8], [9]–[11] have explored various methods like genetic algorithms for CH selection. These studies highlight the importance of a metaheuristic approach in optimizing CH selection. Additionally, strategies like the use of multiple data-collecting sinks [12]–[14], can reduce the load on relay nodes.

To advance this field, the proposed work introduces a novel approach that combines the exploratory aspects of the greylag goose-based optimized clustering (GGOC). It also incorporates the exploitation capabilities of the golden sine strategy guided by the Lévy flight mechanism. The motivation behind this unique combination of proposed hybrid greylag goose-based optimized clustering (HGGOC) is deeply rooted in addressing the ever-growing challenges of data dissemination in 6G-enabled IoMT networks. In these networks, the pursuit of excellence is intricately intertwined with the imperatives of energy efficiency and cost-effectiveness. The proposed HGGOC, inspired by the greylag geese's remarkable navigational capabilities, epitomizes optimization exploration. We aim to explore uncharted territories within the vast data space by harnessing these geese's inherent curiosity and adaptability, seeking optimal solutions that traditional methods might overlook. This exploration element is vital in ensuring that our approach is not confined to local optima but embraces the full spectrum of possibilities, thereby enhancing the robustness and adaptability of our optimization strategy. Complementing the GGOC, the golden sine strategy guided by the Lévy flight mechanism brings a sophisticated level of exploitation

A. Malik, V. Tyagi, and S. Singh are with the Department of Computer Science & Engineering, National Institute of Technology Jalandhar, Punjab 144008 India. E-mail: {malika, vikast.cs.19, samays}@nitj.ac.in

R. Kumar is with the Department of Computer Science & Engineering, Delhi Technological University, New Delhi, 110042, India. E-mail: rajeevkumar@dtu.ac.in

H. Wu is with the Center for Applied Mathematics, Tianjin University, Tianjin, China. Email: whming@tju.edu.cn

S. S. Gill is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, United Kingdom. Email: s.s.gill@qmul.ac.uk

(Corresponding author: Huaming Wu)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:27:21 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2024.3510812

TABLE I
SUMMARY OF STATE-OF-THE-ART CLUSTER-BASED ROUTING APPROACHES FOR ENERGY-EFFICIENT WSNS

Research	Protocol	Methodology	Advantages	Research Gaps
Sahoo *et al.* [15]	PSO-based energy efficient clustering & sink mobility (PSO-ECSM)	Particle swarm optimization & mobile sink approach	Stability & prolonged network lifetime	Reducing end-to-end delay using a hybrid routing method
Alghamdi *et al.* [16]	Energy efficient protocol for optimized CH selection	Dragonfly & Firefly algorithm	Improved statistics for node longevity, network convergence & energy consumption	Achieving a balance between computational cost in the algorithms & calculations for further exploration
Prakash *et al.* [17]	Energy-optimization route & cluster head selection	M-PSO & GA	Used probability to select CHs & GA for optimal path	Re-clustering challenges need more investigation
Osamy *et al.* [18]	Load-balancing aware clustering Scheme (LBACS)	Graph theory concepts	Designed clusters with load balancing	Maintain load balancing, data flow & connectivity under changing network conditions
Loganathan *et al.* [19]	Load-balanced clustering for energy-aware routing (LBCE-EAR)	Clusters are formed as concentric circles	Cluster diameters are measured	Load balancing is achieved to maintain energy efficiency
Prasad *et al.* [20]	BAT algorithm	Barabasi–Albert (BA) model & Topsis (T) technique	Reduced intracluster & intercluster distances by retaining best performance attachments	Only varying the cluster size considered to achieve load balancing
Verma *et al.* [8]	Intelligent Clustering approach for ITS (IC-ITS)	Hybrid optimization of GA and BAT	Ensures energy efficiency & load balancing	Only CH rotation strategy considered for load balancing
Singh *et al.* [9]	Optimized data aggregation algorithm	GA	Improves data aggregation & reduces energy consumption	All direct information requires multiple paths & fixed topologies
Nandan *et al.* [12]	Optimized GA-based data collection method	GA	Improves intracluster efficiency & reduces hop count	Movable sink strategies

to our hybrid approach. Inspired by the harmonic nature of the golden ratio, the golden sine strategy imbues our optimization with precision and efficiency. Coupled with the Lévy flight mechanism, which introduces stochasticity and long-range jumps, our strategy balances thorough exploration and targeted exploitation. This dynamic interplay ensures that our optimization adapts to the evolving nature of data dissemination challenges, optimizing resource utilization and network performance. The 6G-enabled IoMT networks prioritize security and privacy alongside optimization strategies. The proposed HGGOC approach integrates blockchain technology to enhance data dissemination in these networks. Blockchain's decentralized and tamper-resistant nature ensures secure and transparent data management, reducing vulnerabilities and enabling accountable network operations. This work focuses on advancing green communication in 6G-enabled massive IoMT networks. The main contributions of this paper are outlined as follows:

- We propose a cluster-based routing method for data dissemination that incorporates a novel hybrid metaheuristic technique, the Greylag Goose-based Optimized Clustering (HGGOC) algorithm. This technique combines elements of the Greylag Goose Optimization (GGO) algorithm with the golden sine strategy, guided by the Lévy flight mechanism, to enhance the selection of cluster heads (CHs) within the network. The main contributions of this work are the reduction of energy consumption and the extension of the lifespan of 6G-enabled IoMT devices.
- We develop a fitness function for the proposed HGGOC that considers multiple factors for CH selection. These factors include residual energy, node and sink proximity, remaining energy levels of 6G-enabled IoMT devices, and cluster energy. This multifaceted approach ensures a more comprehensive and efficient CH selection process.
- Integrating blockchain technology with 6G-enabled IoMT systems significantly enhances security and privacy. This integration capitalizes on blockchain's robust features to provide secure and transparent data management across widespread IoMT networks for data dissemination.
- We compared its simulation results with those of recent cluster-based routing methods, including IC-ITS [8], OptiGA-ASR-DDC [9], OptimizedGA-MS-In&Out-DDC-ASR [12], and OptiGeA-MoMS-In&Out-DDC [14]. Our findings demonstrate that the proposed HGGOC method outperforms these existing approaches regarding efficiency and effectiveness in 6G-enabled IoMT networks.

The structure of this paper is as follows: Section II reviews related studies. Section III describes the network model assumptions. Section IV details the proposed hybrid clustering approach and its fitness evaluation. Section V demonstrates how blockchain technology is integrated into the proposed work. Section VI presents the results and compares them with state-of-the-art techniques. Finally, Section VII concludes this paper and suggests directions for future research.

II. RELATED WORK

A key feature of 6G is its utilization of terahertz frequency bands, which promise to offer significantly higher bandwidth up to three times greater than that of 5G. Recent research has extensively explored the core technologies and components associated with 6G. A recurring theme in these studies concerns energy consumption, which significantly limits the capabilities of 6G-enabled IoMT devices. This challenge highlights the urgent need for optimized solutions to enhance energy efficiency.

Pokhrel *et al.* [1] introduced a cluster-based intelligent routing protocol that employs multiple gateways to alleviate network hotspot problems. Verma *et al.* [8] developed intelligent and secure clustering in wireless sensor network (WSN)-based intelligent transportation systems. They utilized a hybrid optimization method called GABAT that integrates the strengths of genetic algorithm (GA) and BAT algorithm (BA), to reduce the distance data must travel. However, GABAT introduces higher overhead and complexity when managing large-scale sensor networks. Singh *et al.* [9] introduced a dynamic data transmission method for communicable diseases in IoMT environments, leveraging a GA. Their method optimizes energy usage in IoMT devices by employing GA-based evolutionary processing that adapts to the dynamic range of sensors. It also includes a defined deployment area for IoMT devices to address energy depletion issues, incorporates multiple sinks, and utilizes a direct data collection strategy. While this approach enhances system performance and reduces

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:27:21 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2024.3510812

extensive data packet movement, GA still suffers from slower convergence.

Nandan *et al.* [12] presented a green data collection and transmission method for IoT-based WSNs utilizing an optimized GA. This approach focuses on reducing intra-cluster distances, efficiently using node energy, and minimizing hop counts. It also included direct data transmission with movable sinks to lessen communication distances and hotspot issues, and a dynamic sensing range to decrease CH sensing range overlap and save transmission energy. Singh *et al.* [14] introduced a secure and energy-efficient system focusing on communicable diseases. Enhanced by a genetic algorithm, this system utilizes a WSN with OptiGeA protocol, designed specifically for IoT-enabled WSNs in healthcare. OptiGeA optimizes CH selection by considering node energy, density, distance, and capacity. However, the cost of the implication and complexity is high. Sahoo *et al.* explored the optimization of sink mobility with their particle swarm optimization-based energy-efficient clustering based on sink mobility (PSO-ECSM), but this led to increased overhead due to the moving sink [15]. In another approach, Ali *et al.* integrated novel rank-based clustering (NRC) with their ARSH-FATI-based cluster-head (CH) selection method, aiming to reduce energy consumption [21]. It involves significant computational complexity. Behera *et al.* [22] proposed an improved residual energy selection (I-RES) approach, focused solely on residual energy for CH selection, leaving room for further improvement. It focuses solely on residual energy for cluster-head selection, which may lead to suboptimal network performance as other important factors like node density, distance, and communication overhead are not considered in the selection process. Tariq *et al.* discussed an energy-efficient routing protocol for improving the link quality [23]. This approach introduces computational overhead and increased complexity in dynamic, large-scale IoT environments. Table I shows the summary of state-of-the-art cluster-based routing approaches for energy-efficient WSNs.

III. System Model

In this study, 6G-enabled IoMT devices are deployed in the monitoring scenarios by varying initial energy levels among 6G-enabled IoMT devices. The network is composed of N 6G-enabled IoMT devices with three levels of energy heterogeneity. The network includes a mix of advanced, intermediate, and normal 6G-enabled IoMT devices with each type characterized by distinct initial energy levels. The advanced 6G-enabled IoMT devices have higher initial energy than intermediate 6G-enabled IoMT devices, and both have more energy than normal nodes, following the energy inequality $E_{\text{nrm}} < E_{\text{int}} < E_{\text{adv}}$. The total number of each type of 6G-enabled IoMT devices is advanced (n_{adv}), intermediate (n_{int}), and normal (n_{nrm}) reflects a descending order, with normal 6G-enabled IoMT devices being the most numerous, satisfying the inequality $n_{\text{nrm}} > n_{\text{int}} > n_{\text{adv}}$. The initial energy for each 6G-enabled IoMT device type is denoted as E_0, and the total energy in the network is represented by E_t. The advanced 6G-enabled IoMT devices have their energy defined as $E_0 \times (1+\omega) \times n_{\text{adv}}$, and intermediate 6G-enabled IoMT devices as $E_0 \times (1+\Psi) \times n_{\text{int}}$, with normal 6G-enabled IoMT devices at $E_0 \times n_{\text{nrm}}$. Thus, the total energy for the network is the sum of the energies of advanced, intermediate, and normal 6G-enabled IoMT devices, resulting in $E_T = E_{\text{adv}} + E_{\text{int}} + E_{\text{nrm}}$. The total energy is $E_T = E_0 \times N \times (1 + \Psi \times k_o + k \times \omega)$, combining the energy contributions from all levels. The energy ratios Ψ and ω correspond to the energy fractions of advanced and intermediate 6G-enabled IoMT devices, respectively. The proportions of advanced and intermediate 6G-enabled IoMT devices within the network are indicated by k and k_o. There are several assumptions about the network setup of the proposed work which are outlined as follows:

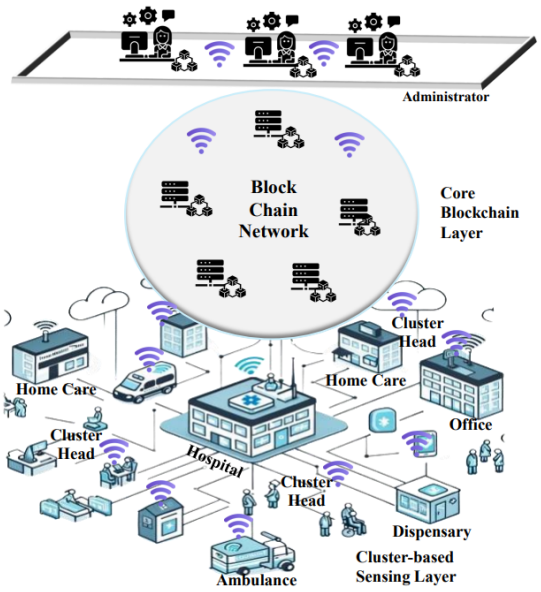


Fig. 1. Secure framework for a 6G-Enabled IoMT network.

- The proposed work assumes a static network configuration, despite the capability of 6G's to support high mobility in IoMT devices. Both the data-collecting sink and the 6G-enabled IoMT devices are considered stationary throughout the network operation.
- It is assumed that all 6G-enabled IoMT devices in the network have uniform configurations and operational characteristics, though nodes possess heterogeneous energy capabilities. This assumption simplifies the analysis and implementation of the proposed method.
- The data-collecting sink is assumed to operate without energy limitations, in contrast to the IoMT devices, which are subject to energy constraints.
- The IoMT devices in the network are assumed to be wireless sensor nodes equipped with 6G communication capabilities, enhancing their connectivity and data transmission efficiency.

In this work, we considered the same radio energy consumption model for transmitting and receiving data between the 6G-enabled IoMT devices and sinks as discussed in the existing works [8], [9], [12], [14]. Fig. 1 illustrates a secure data transmission framework in a 6G-enabled Internet of Medical Things (IoMT) network using blockchain technology. The framework consists of three parts: the 6G-enabled IoMT network, the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:27:21 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2024.3510812

blockchain network, and the administrator domain. The 6G-enabled IoMT network includes devices in homes, hospitals, and ambulances, leveraging 6G technology for fast, low-latency communication. The blockchain network ensures secure and transparent data management through smart contracts, making the data unalterable and traceable. The administrator domain oversees the network, managing its operation and addressing any issues. Data flows from IoMT devices to a sink, which collects and sends the data to the blockchain for secure storage and processing. The administrator ensures the system operates smoothly, monitors the network, and ensures adherence to policies.

IV. Proposed Hybrid Greylag Goose-based Optimized Clustering (HGGOC)

The optimized clustering method, inspired by the greylag goose, initiates by generating a diverse set of individuals through a stochastic process. Each individual proposes a potential solution that is subject to evaluation for potential inclusion in the pool of solutions to address the problem. The dynamic group behavior of the native GGO algorithm [24] assigns each member to either an exploration or an exploitation group. Throughout each iteration, the quantity of potential solutions in each group is dynamically adjusted based on the solution identified as the most optimal. The goose explorer embarks on a journey to investigate novel locations proximate to its current position. The responsibility for refining the existing solutions lies with the exploitation team. After each cycle, the GGO evaluates participants to identify those who have achieved the highest fitness level and duly acknowledges their accomplishments. The following steps elaborate on the exploration and exploitation phases in intricate detail.

A. Exploration Phase: Migration

This process involves the goose explorer actively searching for promising and unexplored locations near its current position. The explorer iteratively evaluates various potential options in the vicinity to identify the most suitable one based on its fitness. The proposed algorithm utilizes the following equations to compute the updated position of the agent ($X_P(I+1)$), by adjusting the auxiliary vectors Aux_v and CX_v. The updated position of the agent is represented as follows:

$$X_P(I+1) = X_P^*(I) - Aux_v \cdot |CX_v \cdot (X_P^*(I) - X_P(I))|, \quad (1)$$

where $X_P(I)$ signifies an agent from the population at the I^{th} iteration, and $X_P^*(I)$ denotes the position of the best solution (leader). Specifically, $Aux_v = 2\alpha R_{nd1} - \alpha$ and $CX_v = 2R_{nd2}$ during iterations, where α linearly decreases from 2 to 0 throughout the iterations. The values of R_{nd1} and R_{nd2} experience random variations within the range $[0, 1]$.

To foster diversity and enhance exploration, Eq. (2) is employed, involving the selection of three random search agents (paddlings denoted with (Pdl)) named X_P^{Pdl1}, X_P^{Pdl2}, and X_P^{Pdl3}. This approach ensures that agents are not excessively influenced by a single leader position, promoting more extensive exploration. Subsequently, the current search agent's position update is determined below when $|Aux_v| \geq 1$.

$$X_P(I+1) = F_{w1} * X_P^{\text{Pdl1}} + F_{w2} * S * (X_P^{\text{Pdl2}} - X_P^{\text{Pdl3}}) + F_{w3} * (1-S) * (X_P(I) - X_P^{\text{Pdl1}}), \quad (2)$$

where F_{w1}, F_{w2}, and F_{w3} denote the weight factors associated with each paddling. These weight factors vary within the interval [0, 2]. The parameter S experiences an exponential decrease, and its calculation is defined by:

$$S = 1 - \frac{I}{I_{\text{Max}}}, \quad (3)$$

where I_{Max} denotes the maximum number of iterations. The second updating procedure, which includes diminishing both α and Aux_v vector values, is applied when $R_{\text{nd3}} \geq 0.5$.

$$X_P(I+1) = F_{w4} \cdot |X_P^*(I) - X_P(I)| \cdot e^{bl} \cos(2\pi l) + [2 \cdot F_{w1} \cdot (R_{\text{nd4}} + R_{\text{nd5}})] * X_P^*(I), \quad (4)$$

where the constant parameter b remains unchanged, and l denotes a random value within the range [-1, 1]. F_{w4} varies within the interval [0, 2], while the random variables, R_{nd4} and R_{nd5}, undergo updates within the range [0, 1].

B. Exploitation Phase: Attacking

The exploitation team is responsible for refining the current solutions. At the end of each cycle, the proposed HGGOC algorithm identifies individuals with the highest fitness and recognizes their contributions. To achieve its exploitation goals, the algorithm employs two distinct strategies, as outlined below.

1) Progressing toward the optimal solution: The motion of three solutions (sentries denoted as (sty), specifically X_P^{sty1}, X_P^{sty2}, and X_P^{sty3}, serves as guidance for other individuals ($X_P^{(\text{non_sty})}$) to adjust their positions towards the estimated location of the prey. The process of updating positions is illustrated as follows:

$$X_{P1} = X_P^{\text{sty1}} - Aux_v \cdot |CX_v \cdot X_P^{\text{sty1}} - X_P(I)|, \quad (5)$$

$$X_{P2} = X_P^{\text{sty2}} - Aux_v \cdot |CX_v \cdot X_P^{\text{sty2}} - X_P(I)|, \quad (6)$$

$$X_{P3} = X_P^{\text{sty3}} - Aux_v \cdot |CX_v \cdot X_P^{\text{sty3}} - X_P(I)|. \quad (7)$$

The updated positions for the population $X_P(I+1)$ can be described as the mean of the three solutions:

$$X_P(I+1) = \frac{X_{P1} + X_{P2} + X_{P3}}{3}. \quad (8)$$

2) Exploring the vicinity of the optimal solution: It involves identifying the most promising option situated near the best response (leader) during flight. Consequently, certain individuals undertake a search for improvements by examining areas close to the optimal response, denoted as PX_{flock}. The mathematical formulation of this process is represented by:

$$X_P(I+1) = X_P(I) + F_{w1}(1-S)CX_v(X_P(I) - X_P^{\text{flock}}), \quad (9)$$

which encompasses identifying the most favorable choice located in proximity to the optimal response (leader) during flight. As a result, specific individuals engage in a quest for enhancements by scrutinizing areas close to the optimal response, indicated as X_P^{flock}.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:27:21 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2024.3510812

C. The Lévy Flight mechanism with Golden Sine Strategy

In the optimization phase of the standard native GGO algorithm, geese determine their attack position based on a fixed behavioral track post-migration. However, the inherent singularity of the action model causes geese to persistently fly in a set angle and direction during movement. Consequently, this leads to a failure in locating the optimal position, and the elevated population density significantly hampers the ability to break free from local extreme values. To address this issue, a later stage of the algorithm incorporates the golden sine strategy guided by the Lévy flight mechanism [25]. The proposed HGGOC introduces the capability to traverse all values of the sine function (all points on the identity element) by leveraging the relationship between the sine function and the unit circle. During this process, including the golden section number serves to narrow the search space and guide seagull individuals toward the optimal solution.

Simultaneously, the Lévy flight mechanism is introduced to provide guidance. By incorporating uncertainty into both the direction and step size of the Lévy flight, the algorithm disrupts the population's positions, expands the search range, and enhances its ability to escape local extrema. The mathematical model of Lévy flight is expressed as:

$$\text{Lévy}(\gamma) = 0.01 * \frac{\rho}{\lceil \sigma^{1/\gamma} \rceil}, \tag{10}$$

where ρ and σ are typically distributed and can be selected based on the $\rho \sim N(0, \vartheta_\rho^2)$ and $\sigma \sim N(0, \vartheta_\sigma^2)$, respectively.

The values of ϑ_ρ^2 and ϑ_σ^2 are calculated by:

$$\vartheta_\rho^2 = \left(\frac{\Gamma(1+\gamma)\sin\left(\frac{\pi\gamma}{2}\right)}{\Gamma\left(\frac{1+\gamma}{2}\right)\gamma 2^{\left(\frac{\gamma-1}{2}\right)}} \right)^{\frac{1}{\gamma}}, \quad \vartheta_\sigma^2 = 1, \tag{11}$$

where $\Gamma(\gamma)$ represents the gamma function, with γ being a constant that affects the trajectory of the Lévy flight step. In this paper, we have selected γ to be 1.5. The mathematical formulation of updating individual positions through the golden sine strategy guided by the Lévy flight mechanism is illustrated in Eq. (12) from the exploitation phase.

$$X_P(I+1) = X_P(I)\sin(R_{\text{nd1}}) + \text{Lévy}(\gamma)\sin(R_{\text{nd1}}) \\ F_{w1}(1-S)CX_v(X_P(I) - X_P^{\text{flock}}). \tag{12}$$

The proposed HGGOC algorithm initiates by generating a diverse set of individuals randomly, as outlined in Algorithm 1. Each individual presents a potential solution that can be considered as a candidate to address the problem. The HGGOC population is denoted as X_{P_i}, where $i = 1, 2, \ldots, n$, and n represents the number of gaggles. The Dynamic group behavior of the proposed HGGOC algorithm segregates all individuals into an exploration group (G_{n_1}) and an exploitation group (G_{n_2}). The quantity of solutions in each group dynamically adjusts during each iteration based on the best solution. The exploration group involves n_1 agents, while the exploitation group consists of n_2 agents. The proposed HGGOC algorithm initializes the groups with an equal distribution of 50% exploration and 50% exploitation. Subsequently, the number of agents in the exploration group (G_{n_1}) decreases, while the number of agents in the exploitation group (G_{n_2})

Algorithm 1 : HGGOC

Input: Initial population X_{p_i}, where $i = 1, 2, \cdots, n$, n, $I_{\max}$
Output: Best Search Agent as CH

Start Procedure
Initialize the values of $I = 1$, $S_1 = 1$, $\gamma = 1.5$, $\rho = 2$
Select values of R_{nd1}, R_{nd2}, R_{nd3}, R_{nd4}, R_{nd5}, F_{w1}, F_{w2}, F_{w3}, F_{w4}
Compute the values of $\text{CX}_{\text{vector}}$ and $\text{Aux}_{\text{vector}}$
while $I \leq I_{\max}$ **do**
  for $i = 1$ to n **do**
    Calculate each search agent's fitness using Eq. (17)
    Update the exploration groups (Gn_1) and (Gn_2)
  end for
  for $i = 1$ to n_1 **do**
    if $I \mod 2 == 0$ **then**
      if $R_{nda} < 0.5$ **then**
        if $|\text{Aux}_v| < 1$ **then**
          Modify the present location of the search agent using Eq. (1)
        else
          Select paddling search agents position X_P^{Pdl1}, X_P^{Pdl2}, and X_P^{Pdl3}
          Update the value of S using Eq. (3)
          Modify the present location of the search agent using Eq. (2)
        end if
      else
        Modify the present location of the search agent using Eq. (4)
      end if
    else
      Select search agents position X_P^{sty1}, X_P^{sty2}, and X_P^{sty3}
      Lévy Flight with Golden Sine Strategy is applied using Eq. (10)
      Update the individual's position using Eq. (12)
    end if
  end for
  for $i = 1$ to n_2 **do**
    if $I \mod 2 == 0$ **then**
      Use Eq. (5), Eq. (6), and Eq. (7) to compute X_{P1}, X_{P2}, and X_{P3}, respectively
      Update the individual's position using Eq. (8)
    else
      Select search agents position X_P^{sty1}, X_P^{sty2}, and X_P^{sty3}
      Apply Lévy Flight mechanism with Golden Sine Strategy using Eq. (10)
      Modify the present location of the search agent using Eq. (12)
    end if
  end for
  Compute fitness value using Eq. (17)
  Update all parameters for the next iteration
  $I = I + 1$
  Adapt solutions beyond the boundaries of the search space
  if Pbest is the same as the previous two iterations **then**
    Keep posted on solutions within exploration group (G_n) & (G_{n_1})
  end if
end while
Return best search agent
End procedure

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:27:21 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2024.3510812

increases. However, if the objective function value of the best solution remains unchanged for three consecutive iterations, the algorithm starts increasing the number of agents in the exploration group (G_{n_1}), in order to discover alternative best solutions and prevent potential local optima.

D. Fitness Evaluation

The main objective of the proposed work is to minimize energy consumption in 6G-enabled IoMT networks, thereby fostering sustainable communication within these systems. A crucial element in achieving this is the role of the CH, which is instrumental in managing the network's energy usage. CH is selected through an optimized solution provided by the HG-GOC algorithm. We have developed a comprehensive fitness function incorporating various critical parameters to determine the CH effectively. Each of these parameters plays a key role in the CH selection process. This detailed approach ensures an effective strategy for energy management in 6G-enabled IoMT networks. In this work, we incorporated four different fitness parameters, i.e., residual energy, node and sink proximity, remaining energy levels of nodes, and cluster's energy. This multifaceted approach ensures a more comprehensive and efficient selection process.

1) Residual Energy: This fitness parameter is very relevant for the cluster head (CH) because CH consumes more energy than other cluster member nodes due to its role in collecting and transmitting data to the sink. Therefore, choosing a CH with the highest possible energy level is vital. As the network operates, the nodes' energy levels naturally deplete over time, altering their residual energy after each round. To assess this, the first fitness factor is denoted as F_1:

$$F_1 = \sum_{K=1}^{N} \frac{\varepsilon_{\text{resid}}(K)}{\varepsilon_{\text{init}}(K)}, \tag{13}$$

which evaluates the ratio of the node's residual energy (E_{resid}) to its initial energy (E_{init}).

2) Node and Sink Proximity: In this work, the communication distance of sensor nodes, either with other nodes or the sink, is a critical factor. The proximity of a node to the sink is a key consideration, particularly for selecting the CH. It's preferable to select CH, a node that is closest to the sink, to minimize energy consumption. This fitness parameter evaluates the distance of a candidate node from the sink ($D_{\text{SNode-Sink}}$) relative to the average distance of all nodes from the sink ($D_{\text{Avg(SNodes-Sink)}}$), ensuring energy-efficient operation of the network.

$$F_2 = \sum_{K=1}^{N} \frac{D_{\text{SNode-Sink}}(K)}{D_{\text{Avg(SNodes-Sink)}}(K)}. \tag{14}$$

3) Remaining Energy Levels of Nodes: Once network operation begins, the network's energy starts to deplete, leading to an increase in the number of dead nodes over time. This situation makes the selection of the CH from the remaining active nodes a critical concern. In our approach, we define a fitness parameter, as shown below:

$$F_3 = \frac{1}{N} \sum_{K=1}^{N} \varepsilon_{\text{resid}}(K). \tag{15}$$

which assesses the network's remaining energy. For optimal CH selection, this parameter must remain as high as possible, ensuring that the most energy-efficient nodes are chosen to maintain network functionality and longevity.

4) Cluster's Energy: As nodes within a cluster consume energy, the overall energy level of the cluster decreases. To achieve effective load balancing across the network, the selection of the CH should consider the current energy status of the cluster where the candidate node resides. Initially, this factor is set to 1 in the first round since the cluster formation occurs after the CH is selected. However, from the second round onwards, the energy level of the cluster becomes a significant factor in CH selection. We calculate the fifth fitness parameter, denoted as F_4:

$$F_4 = \sum_{K=1}^{N_{\text{CH}}} \varepsilon_{\text{resid}}(K). \tag{16}$$

It needs to be high to facilitate the selection of a suitable candidate node as CH, thereby ensuring efficient energy management and balanced load distribution within the network.

In our approach, we transform a multi-objective fitness function into a single-objective one by integrating various fitness parameters. As detailed in our methodology, this is achieved by multiplying these parameters with linear weight functions. This integration simplifies the evaluation process while still considering the diverse aspects critical to the system's performance.

$$F = \alpha F_1 + \beta F_2 + \lambda F_3 + \delta F_4, \tag{17}$$

where $\alpha + \beta + \lambda + \delta = 1$. It presents the weighted sum of various factors, each multiplied by corresponding parameters. This approach involves evaluating each node based on its fitness function value. The node that exhibits the highest fitness value through this calculation is then selected as the CH.

E. Complexity

The computational complexity of the proposed HGGOC algorithm can be analyzed in various stages. The initialization phase involves generating a diverse population, with each individual presenting a potential solution. This initial step incurs a complexity of $\mathcal{O}(n)$, where n is the population size. During each iteration, the algorithm dynamically adjusts the group sizes for exploration and exploitation, contributing to the complexity of $\mathcal{O}(n)$ for each phase. The exploration phase, driven by the movement of search agents based on auxiliary vectors and random parameters, has a complexity of $\mathcal{O}(n)$ due to the necessity of updating each agent's position. The exploitation phase, which refines the solutions through the interaction of multiple agents, adds a complexity of $\mathcal{O}(n^2)$, particularly when computing the new positions based on the behavior of sentries and the Lévy flight mechanism. The fitness evaluation step involves assessing each individual's solution, which also incurs a complexity of $\mathcal{O}(n)$. Given that these steps are repeated for a maximum number of iterations I_{max}, the overall complexity of the algorithm can be expressed as $\mathcal{O}(I_{\text{max}} \cdot n^2)$. This complexity reflects the comprehensive nature of the algorithm, balancing exploration and exploitation while

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:27:21 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2024.3510812

dynamically adjusting the search process to avoid local optima and achieve optimal clustering.

The algorithm needs to maintain a population of n individuals, where each individual represents a potential solution in the search space. Assuming each solution requires m units of space (where m depends on the dimensionality of the problem or the number of variables in the solution), the space required for storing the population is $\mathcal{O}(n \cdot m)$. Therefore, the overall space complexity of the proposed HGGOC algorithm is $\mathcal{O}(n \cdot m)$. This linear space complexity reflects the algorithm's efficient management of resources while handling the optimization process.

V. Blockchain Integration

In the proposed approach blockchain technology is integrated with 6G-enabled IoT environments to significantly enhance data security and integrity [26]. The system model has a core blockchain layer that consists of high computing nodes that form a permissioned blockchain network. This implies that all participating nodes are pre-authenticated and have established trust within the network. The nodes employ a Proof of Authority (PoA) mechanism, which is a lightweight consensus protocol in which a designated set of trusted nodes, known as authorities, validate transactions and generate new blocks. When the cluster heads transmit the data to the core blockchain layer, a blockchain node performs validation, approves the transaction and generates a new block. The new block generated is disseminated to other blockchain nodes in the network, ensuring secure and efficient data management. A block in this blockchain consists of two main parts: the header and the transaction data. The header contains essential information like the block's version, generation timestamp, nonce, difficulty target, Merkle root, and the preceding block's hash. Significantly, the current block's hash incorporates the previous block's hash, ensuring a secure and tamper-evident chain of blocks. Any alteration in a block's content results in a cascading change in the hashes of subsequent blocks.

The validation of each block involves multiple checks. The verifier first ensures the existence and validity of the previous block, and that the current block's timestamp is greater than its predecessor's. Other parameters, such as the transaction root and difficulty, are also verified for validity. Each transaction within the block is individually validated, and any transaction failure leads to an error message. The block is considered validated if all transactions are validated successfully and the final state aligns with the root state. The formula for calculating a block's hash is given by:

$$\text{Hash}(B_n) = \text{Hash}(\text{Hash}(B_{n-1}) + B_{n\text{Data}}), \qquad (18)$$

where B_n represents the block for which the hash is being calculated, $\text{Hash}(B_{n-1})$ is the previous block's hash, and $B_{n\text{Data}}$ denotes the current block data.

Using the hash of the previous block in the current block's hash calculation greatly enhances the blockchain's resistance to attacks. Once a significant number of blocks have been added to the chain, altering data in any block becomes highly impractical. This is because any change would cause a mismatch in the hashes of all subsequent blocks, thereby compromising the integrity of the entire blockchain and effectively securing it against tampering.

In this system, cluster heads aggregate data from IoMT sensor devices, reducing the transaction volume reaching the blockchain and easing the processing load on nodes. A permissioned blockchain using PoA consensus, with a small set of pre-approved validators, enables faster transaction validation, lowers latency, and enhances throughput compared to traditional methods like PoW or PoS. By limiting the number of validator nodes, network congestion and energy consumption are reduced, making the system scalable and efficient for IoT devices.

TABLE II
Network Simulation Parameters

Simulation Parameters	Values
Number of 6G-enabled IoMT devices	100
n_{nrm}, n_{int} & n_{adv} nodes energy	1.0 J, 1.50 J, & 2.0 J
Target area (m^2)	100 × 100
n_{nr}, n_{in}&n_{ad} node density ratio	70%, 20% & 10% of all nodes
Number of clusters and CN	Dynamic
Sink Location	50×50
d_T	$75m$
Data packet's size	2000 bits
Initial search whales (agents)	50
ω, ψ	$2, 1$
k, k_0	$0.2, 0.3$
Initial population Size	50
I_{Max}	40

VI. Performance Evaluation

The performance of the proposed method is rigorously evaluated using essential performance metrics, including stability, network lifetime, remaining network energy, cluster headcount, and throughput analysis. MATLAB has been utilized for these evaluations, adhering to the simulation parameters used in [9], [12], [14]. In the simulations, 100 nodes are randomly deployed in a 100 × 100 m^2 area with a total energy of 70 Joules for different levels of heterogeneity. In homogeneous networks, each node has 0.7 J energy. In 2-level heterogeneous networks, normal nodes have 0.5 J, and advanced nodes have 1.0 J, with 60 and 40 nodes, respectively. In 3-level heterogeneous networks, normal nodes have 0.5 J, intermediate nodes have 1.0 J, and advanced nodes have 1.5 J, with 50, 30, and 20 nodes, respectively. The details of the energy and node categorization are given in Section III. The performance of the proposed method is benchmarked against several established protocols, including ICITS [8], OptiGA-ASR-DDC [9], OptimizedGA-MS-In&Out-DDC-ASR [12], and OptiGeA-MoMS-In&Out-DDC [14]. This study aims to validate the efficacy and adaptability of the proposed approach in improving the performance of 6G-enabled IoMT networks. Table II shows the key simulation parameters for the network.

A. Stability Analysis

Fig. 2 illustrates a comparative stability analysis among several existing methods such as ICITS [8], OptiGA-ASR-DDC

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:27:21 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2024.3510812

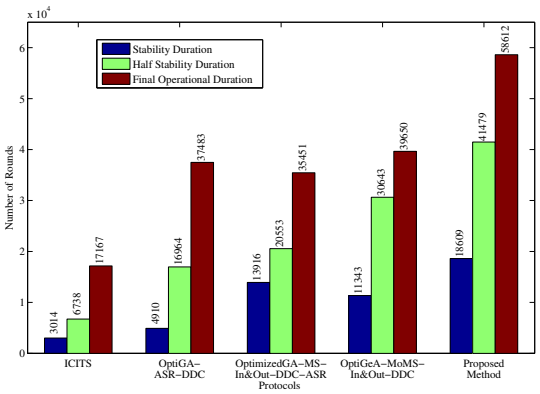


Fig. 2. Comparative Stability Analysis of Existing and Proposed Methods Across Rounds.

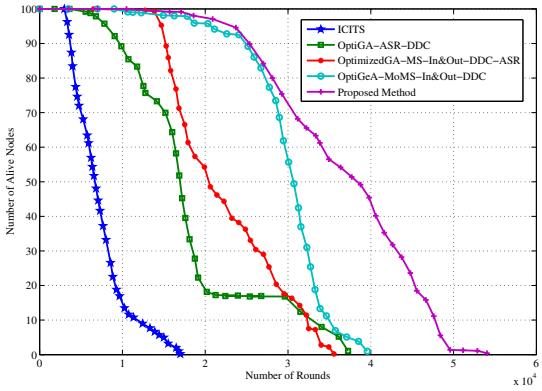


Fig. 3. Comparative Analysis of Network Lifetime of Existing and Proposed Methods Across Rounds.

[9], OptimizedGA-MS-In&Out-DDC-ASR [12], OptiGeA-MoMS-In&Out-DDC [14] and the proposed method, with the analysis conducted over a number of rounds. The existing ICITS method shows a stability period ending 3,014 rounds, while OptiGA-ASR-DDC maintains stability for 4,910 rounds. The OptimizedGA-MS-In&Out-DDC-ASR method achieves a longer stability period of 13,916 rounds, and OptiGeA-MoMS-In&Out-DDC sustains for 11,343 rounds. In stark contrast, the proposed method, HGGOC significantly outstrips these figures, boasting a stability period of 18,609 rounds. This remarkable performance of the proposed method represents substantial improvements in stability periods by 517.41% over ICITS, 279.00% over OptiGA-ASR-DDC, 33.72% over OptimizedGA-MS-In&Out-DDC-ASR, and 64.05% over OptiGeA-MoMS-In&Out-DDC.

B. Network Lifetime Analysis

Fig. 3 presents a comparative analysis focused on the number of active nodes across various rounds, encompassing both established protocols such as ICITS [8], OptiGA-ASR-DDC [9], OptimizedGA-MS-In&Out-DDC-ASR [12], and OptiGeA-MoMS-In&Out-DDC [14] and the proposed method. This comparison is particularly considered in terms of node longevity. The proposed method, HGGOC notably outperforms the established ones in maintaining active nodes over time. It shows an increase in node lifetime by 241.42% compared to ICITS, 56.36% over OptiGA-ASR-DDC, 65.33% compared to OptimizedGA-MS-In&Out-DDC-ASR, and 47.82% over OptiGeA-MoMS-In&Out-DDC. This significant improvement in the longevity of active nodes is primarily due to the optimized detection range incorporated in our protocol. The optimized distance plays a pivotal role in the formation of smaller clusters. This strategic formation effectively reduces the load on individual sensor nodes, thereby conserving energy. This aspect is especially beneficial for cluster heads located closer to the sink, as it lessens their energy expenditure. Consequently, the protocol enhances the overall energy efficiency and extends the operational lifespan of the network nodes, an essential factor for the sustainability and effectiveness of IoMT networks.

C. Network Remaining Energy Analysis

Fig. 4 provides a detailed depiction of the energy consumption metrics across different rounds within the IoMT sensor system. This system begins to expend energy as soon as information transmission is initiated, making it essential to track the behavior of the system's remaining energy after this transmission begins. As demonstrated in the figure, the proposed protocol shows a marked improvement over existing protocols like ICITS [8], OptiGA-ASR-DDC [9], OptimizedGA-MS-In&Out-DDC-ASR [12], and OptiGeA-MoMS-In&Out-DDC [14] in terms of sustaining energy for the maximum number of network activity rounds. This enhanced performance can be attributed to advancements in managing the distance factor, achieved through the strategic use of multiple sinks. This setup significantly reduces the energy required for data transmission. The protocol effectively minimizes energy drain by distributing the energy load more evenly, particularly on the cluster heads. The deployment of a single sink at the center of the target region plays a crucial role in enhancing efficiency. This sink is strategically positioned to gather information from cluster heads located across the network. These sinks are positioned to gather information from cluster heads across the network, even those sparsely situated. This arrangement prevents data packets from traveling long distances, conserving a considerable amount of energy. This approach represents a significant leap in enhancing the sustainability and effectiveness of sensor networks, particularly in scenarios requiring prolonged operation and high energy efficiency.

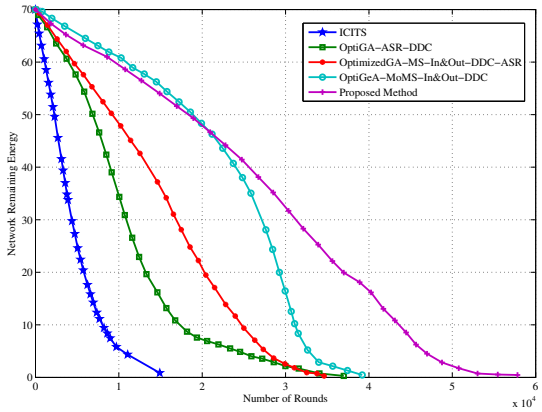


Fig. 4. Comparative Analysis of Residual Energy of Existing and Proposed Methods Across Rounds.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:27:21 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2024.3510812

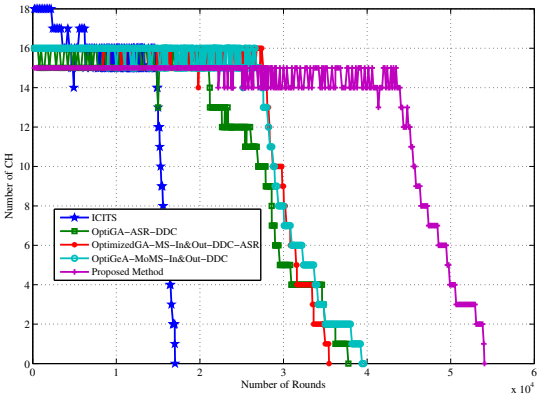


Fig. 5. Comparative Analysis of Cluster Head Count per Round for Existing and Proposed Methods.

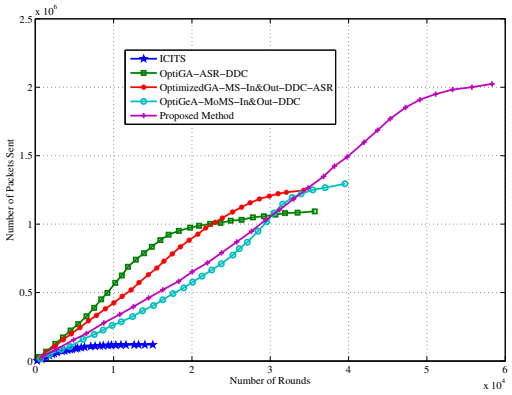


Fig. 6. Comparative Analysis of Packets Sent to Sink per Round for Existing and Proposed Methods.

D. Cluster Head Count Analysis

Fig. 5 vividly illustrates the dynamics of cluster headcount over various rounds, highlighting the influence of the fittest SN based on the efficient fitness evaluation process on cluster formation over the operational period. The protocol is designed to maintain an optimal number of cluster heads, which are generated through the proposed HGGOC approach. This optimal number of CHs is a key metric in ensuring the efficient structuring of clusters as determined by the fittest SN. Initially, the protocol starts with moving towards the optimal solution. As the process evolves, it progressively moves towards a more efficient structural arrangement. This involves strategically positioning the most relevant cluster heads based on their appropriateness fitness values The network begins its operation with this initially evolved clustering configuration and continues functioning until the stability period is reached, which is indicated by the depletion of the first node. As the network's energy gradually depletes, the selection process adapts, seeking to prolong the network's overall lifetime by adjusting the cluster head nodes.

In the early stages of operation, we observe that existing methods display a wide range of cluster head counts, from 18 to 0, with a sharp decrease. However, the proposed protocol manages to maintain a more consistent cluster headcount, typically ranging from 15 to 14. This difference is primarily due to the elimination of unnecessary cluster formations, especially those near the base stations. Moreover, this cluster headcount metric also sheds light on the functionality of the optimization process in the protocol. It illustrates how the cluster head count dynamically adjusts based on factors such as node fitness and remaining energy levels. As the operation of the system progresses, the cluster head count demonstrates a gradual decrease, reflecting the ongoing optimization process within the network. This dynamic adjustment is key to maintaining the efficiency and longevity of the network.

E. Throughput Analysis

In the network design, minimizing cluster structuring near the base station and optimizing the transmission distance factor have led to a notable improvement in the network's quality of service (QoS). The four key fitness parameters used for determining cluster heads collectively enhance the network's lifespan and increase its throughput. The capacity to transmit a higher volume of data packets is directly linked to the prolonged operational duration of the 6G-enabled IoMT devices. Fig. 6 depicts the number of data packets sent over various rounds, highlighting a significant improvement in both throughput and lifespan with our proposed method. Specifically, our approach achieves the transmission of 2.02 million packets, a substantial increase compared to the 0.11 million, 1.08 million, 1.24 million, and 1.27 million packets transmitted by existing methods. One of the key advantages of the proposed approach is the significant energy savings realized by independent 6G-enabled IoMT devices. By minimizing unnecessary traffic near the base station, the strategy conserves energy, which is then translated into increased throughput. This demonstrates the effectiveness of our method in enhancing network efficiency and sustainability.

VII. Conclusions and Future Work

In this paper, we concentrated on advancing secure and sustainable communication, specifically focusing on energy-efficient communication for 6G-enabled IoMT devices. To address this, we introduced a novel HGGOC algorithm and developed a fitness function specifically tailored for CH selection. By leveraging the robust features of blockchain, our approach ensures secure and transparent data management across expansive 6G-enabled IoMT networks. This integration mitigates potential vulnerabilities and establishes a solid foundation for accountable and auditable network operations. The effectiveness of the proposed work was evaluated using several key metrics designed to verify the protocol's efficiency and performance. The findings from these evaluations have been quite promising, particularly in terms of enhancing the network lifespans of 6G-enabled IoMT devices. This improvement is significant as it contributes directly to the sustainability and effectiveness of IoMT networks in a 6G environment.

Future work will focus on addressing the mobility aspects of IoMT devices, motivated by the expectation that 6G systems will support higher mobility. Additionally, we will explore the scalability of the HGGOC algorithm in large-scale IoMT deployments, including heterogeneous networks with diverse device types and varying energy profiles. The impact of different security concerns and blockchain configurations on

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:27:21 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2024.3510812

network performance and energy consumption could also be examined.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (31400), and partly supported by the National Natural Science Foundation of China (62071327).

REFERENCES

[1] S. R. Pokhrel, S. Verma, S. Garg, A. K. Sharma, and J. Choi, "An efficient clustering framework for massive sensor networking in industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 4917–4924, 2020.

[2] Y. Chen, Z. Ren, J. Xu, Y. Zeng, D. W. K. Ng, and S. Cui, "Integrated sensing, communication, and powering (iscap): Towards multi-functional 6g wireless networks," *arXiv preprint arXiv:2401.03516*, 2024.

[3] A. Mukherjee, D. De, N. Dey, R. G. Crespo, and E. Herrera-Viedma, "Disastdrone: A disaster aware consumer internet of drone things system in ultra-low latent 6g network," *IEEE Transactions on Consumer Electronics*, vol. 69, no. 1, pp. 38–48, 2022.

[4] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6g: Ai empowered wireless networks," *IEEE communications magazine*, vol. 57, no. 8, pp. 84–90, 2019.

[5] X. Chen, P. Wang, Y. Yang, and M. Liu, "Resource-constraint deep forest based intrusion detection method in internet of things for consumer electronic," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 2, pp. 4976–4987, 2024.

[6] J. K. Samriya *et al.*, "Blockchain and reinforcement neural network for trusted cloud-enabled iot network," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 2311–2322, 2024.

[7] S. S. Gill and R. Buyya, "Transforming research with quantum computing," *Journal of Economy and Technology*, vol. 3, pp. 1–10, 2024.

[8] S. Verma, S. Zeadally, S. Kaur, and A. K. Sharma, "Intelligent and secure clustering in wireless sensor network (wsn)-based intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 13473–13481, 2021.

[9] S. Singh, A. S. Nandan, G. Sikka, A. Malik, and N. Kumar, "A genetic-algorithm-based dynamic transmission of data for communicable disease in iomt environment," *IEEE Internet of Things Journal*, vol. 11, no. 1, pp. 1427–1438, 2023.

[10] S. P. Singh, P. Singh, D. Koundal, and M. Diwakar, "A new multi/many-objective optimization approach in wireless sensor networks for e-commerce applications," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 2, pp. 4848–4857, 2024.

[11] P. Prabaharan, "Enhanced energy proficient clustering (eepc) algorithm for minimizes the energy depletion in wireless sensor network," *Journal of Intelligent & Fuzzy Systems*, vol. 46, p. 1995–2008, Jan. 2024.

[12] A. S. Nandan, S. Singh, A. Malik, and R. Kumar, "A green data collection & transmission method for iot-based wsn in disaster management," *IEEE Sensors Journal*, vol. 21, no. 22, pp. 25912–25921, 2021.

[13] D. Gupta, J. V. N. Ramesh, M. K. Kumar, F. Y. Alghayadh, S. B. Dodda, T. A. Ahanger, Y. Ilkhamova, and S. R. Karumuri, "Optimizing cluster head selection for e-commerce-enabled wireless sensor networks," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 1640–1647, 2024.

[14] S. Singh, A. S. Nandan, G. Sikka, A. Malik, and A. Vidyarthi, "A secure energy-efficient routing protocol for disease data transmission using iomt," *Computers and Electrical Engineering*, vol. 101, p. 108113, 2022.

[15] B. M. Sahoo, T. Amgoth, and H. M. Pandey, "Particle swarm optimization based energy efficient clustering and sink mobility in heterogeneous wireless sensor network," *Ad Hoc Networks*, vol. 106, p. 102237, 2020.

[16] T. A. Alghamdi, "Energy efficient protocol in wireless sensor network: optimized cluster head selection model," *Telecommunication Systems*, vol. 74, no. 3, pp. 331–345, 2020.

[17] V. Prakash, D. Singh, S. Pandey, S. Singh, and P. K. Singh, "Energy-optimization route and cluster head selection using m-pso and ga in wireless sensor networks," *Wireless Personal Communications*, pp. 1–26, 2024.

[18] W. Osamy, B. Alwasel, A. Salim, A. M. Khedr, and A. Aziz, "Lbas: Load-balancing aware clustering scheme for iot-based heterogeneous wireless sensor networks," *IEEE Sensors Journal*, vol. 24, no. 9, pp. 15472–15490, 2024.

[19] D. Loganathan, M. Balasubramani, R. Sabitha, and S. Karthik, "Improved load-balanced clustering for energy-aware routing (ilbc-ear) in wsns.," *Computer Systems Science & Engineering*, vol. 44, no. 1, 2023.

[20] I. B. Prasad, S. S. Yadav, N. Kumar, V. Pal, *et al.*, "Bat: Barabasi–albert with topsis scale-free topology evolution for load-balanced wsns," *IEEE Sensors Journal*, vol. 23, no. 15, pp. 17627–17637, 2023.

[21] H. Ali, U. U. Tariq, M. Hussain, L. Lu, J. Panneerselvam, and X. Zhai, "Arsh-fati: A novel metaheuristic for cluster head selection in wireless sensor networks," *IEEE Systems Journal*, vol. 15, no. 2, pp. 2386–2397, 2020.

[22] T. M. Behera, S. K. Mohapatra, U. C. Samal, M. S. Khan, M. Daneshmand, and A. H. Gandomi, "Residual energy-based cluster-head selection in wsns for iot application," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5132–5139, 2019.

[23] M. Shahid, M. Tariq, Z. Iqbal, H. M. Albarakati, N. Fatima, M. A. Khan, and M. Shabaz, "Link-quality-based energy-efficient routing protocol for wsn in iot," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 4645–4653, 2024.

[24] E.-S. M. El-kenawy, N. Khodadadi, S. Mirjalili, A. A. Abdelhamid, M. M. Eid, and A. Ibrahim, "Greylag goose optimization: Nature-inspired optimization algorithm," *Expert Systems with Applications*, vol. 238, p. 122147, 2024.

[25] W. Xie, J.-S. Wang, and Y. Tao, "Improved black hole algorithm based on golden sine operator and levy flight operator," *IEEE Access*, vol. 7, pp. 161459–161486, 2019.

[26] L. K. Ramasamy, F. K. KP, A. L. Imoize, J. O. Ogbebor, S. Kadry, and S. Rho, "Blockchain-based wireless sensor networks for malicious node detection: A survey," *IEEE Access*, vol. 9, pp. 128765–128785, 2021.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:27:21 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Physica Scripta

INCLUSIVE PUBLISHING TRUSTED SCIENCE

PURPOSE-LED PUBLISHING™

ACCEPTED MANUSCRIPT

Performance evaluation of gate engineered ferroelectric MIMOS for analog/electrical IC applications

To cite this article before publication: Yash Pathak *et al* 2024 *Phys. Scr.* in press https://doi.org/10.1088/1402-4896/ad9adf

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© 2024 IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.

As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript will be available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 14.139.251.98 on 09/12/2024 at 06:16

Performance Evaluation of Gate Engineered Ferroelectric MIMOS for Analog/Electrical IC Applications

Yash Pathak[1], Kajal Verma[1], Bansi Dhar Malhotra[2] and Rishu Chaujar[1*]

[1]Department of Applied Physics, Delhi Technological University, 110042, New Delhi, India.
[2]Department of Biotechnology, Delhi Technological University, 110042, New Delhi, India.

*Corresponding author(s). E-mail(s): chaujar.rishu@dtu.ac.in;
Contributing authors: yash_2k19phdap505@dtu.ac.in;
kajalverma_2k21phdap507@dtu.ac.in; bansi.malhotra@dce.ac.in;

Abstract

This work examines various materials for gate electrode for enhancing the performance of M-Fe-MIMOS (Metal Ferroelectric Metal Insulator Metal Oxide Semiconductor Field Effect Transistor (MOSFET)). The device is analysed for different analog, electrical and RF parameters with gate materials such as chromium, tungsten and palladium. The gate stacked attached to MOSFET device consists ferroelectric layer sandwiched with insulator for focusing on reliability and stability of M-Fe-MIMOS. The palladium gate material for M-Fe-MIMOS(spacer) shows better analog parameters such as improvement in leakage current by 106 times that results 107 times higher switching ratio as compare to chromium. Also with higher threshold voltage by 428% and lower sub-threshold swing by 340% as compare to chromium, M-Fe-MIMOS(spacer) shows better immunity towards various noise distortions. The electrical properties of the device are analysed in terms of electric field , electric potential and energy band structure.All the analysed results indicate that palladium as a gate material for the M-Fe-MIMOS shows most improved electrical performance and can further be used for various applications in nanoelectronic devices and integrated circuit (IC) design.

Accepted Manuscript

Keywords: Electric field, SS, Oxide reliability, Technology Computer Aided Device (TCAD), Electrode

1 Introduction

The power and thermal management of integrated circuits (ICs) are more vital aftercare for transistor industry. Although with engineering technology and material science advancement transistor generates the low amount of heat. However, the ICs have millions of transistors on chip which generate huge amount of heat. Thus, the reduction of supply voltage (V_{dd}) is vital technique to solve the problem. Power factor depends upon square of supply voltage ($P_{dyn} \alpha V_{dd}^2$). The scaling of supply voltage for conventional MOSFET reached 60mV/dec.(fundamental limit) and cannot be lowred using methods that is known as Boltzmann Tyranny [1]. The negative capacitance might work to provide high on current and subthreshold swing lower than 60mV/dec. Recently the phenomena of negative capacitance has been demonstrated in different systems experimentally such as (i) superlattice (ii) ferroelectric dielectric bilayer (iii) isolated ferroelectric film [2, 3].

Due to its lower supply voltage, tunnel FETs have also been cited by many research organizations in the semiconductor world as the ideal replacement for traditional MOSFETs for future low-power applications. In order to fulfill the needs of ultra-low power ICs, the idea of NC has been adopted and used in MOS devices[4, 5]. According to the literature available , NCFET technology has the ability to address the drawbacks of traditional MOS technology. Additionally, it offers superior characteristics particularly SS in compare to other devices like TFET and IMOS. Because NCFET devices operate at lower supply voltage than IMOS and do not suffer from low on-state drain current [6, 7].

Channel length scaling does not cause much degration M-Fe-MIMOS due to negative capacitance phenomena well accommodated in the ferroelectric layer of device. Furthermore, the final model merely applies the Landau parameters to the outside-of-plane polarisation component, which is unreasonable for materials with large ferroelectric thicknesses[8], α= 1.0 $\times$ 10^{+7}, β= 8.9 $\times$ 10^{+8} and γ =0 [9]. The L-K and Miller models are often used to elucidate the polarisation (P) and voltage across ferroelectric (FE) materials. However, L-K has received greater recognition. This section has focused on the L-K model. The Landau–Khalatnikov (L–K) equation is a dynamic formulation of the Landau–Devonshire theory as shown in eq. 1. The L–K equation is a very useful tool for comprehending the ferroelectric switching features in the development of NCFET technology. The landau theory of non linear dielectric is given below in eq. 1:

$$G = \alpha P^2 + \beta P^4 + \gamma P^6 - EP, \alpha = \alpha_0[T - T^0] \quad (1)$$

Accepted Manuscript

Here, G is free energy density, α, β, γ are landau coefficients, P is polarisation, E is electric field[10, 11].

In our earlier studies, we have done extensive investigation to analyse how the M-Fe-MIMOS with long channels work [12, 13], giving the degradation of the device in terms of low switching speed, high leakage current, high power consumption. M-Fe-MIMOS(spacer) provides a greater ON-state flow compared to M-Fe-MIMOS, however we have found that it is not true for all ferroelectric materials. Additionally, the text still lacks for clear explanation of the real conduct of the two structures and a general performance analysis.

The energy required to transfer an electron from the Fermi level to the vacuum is known as the work function. In FETs, the work function of the gate material influences the threshold voltage, thereby improving channel control. A double metal NCFET device uses two distinct metals for the gate electrode. This structure allows for more accurate fine-tuning of the work function and optimization of the threshold voltage for enhanced performance [14]. Optimizing the performance of double metal NCFETs for both analog and digital applications necessitates an understanding of the variations in work function. By carefully selecting materials and designing the gate stack, it is possible to significantly enhance transconductance, threshold voltage, subthreshold slope, and overall device efficiency [15, 16].

This article describes an improved field effect transistor with a ferroelectric layer that operates at room temperature (T=300K) with several performance advantages over other MFMIMOS transistors, including low input power consumption, reduced drain induced barrier lowering, reduced subthreshold swing, reduced leakage current, and better control over gate channel device[17]. Visual TCAD simulator was used to derive the device performance matrix. The I_{off} in this M-Fe-MIMOS (spacer) is lying in the realm of 10^{-11}A, means less leakage and more working life of device, which is the requirement for better performance. At off-state (i.e., at $V_{gs} = 0$), the M-Fe-MIMOS has improved both Ion and depletion channel width. Through this investigation, we learn more about the device for better analog execution and power amplification performance, when M-Fe-MIMOS(spacer) and M-Fe-MIMOS are compared analogically and virtually.

2 Device structure

Fig 1. shows schematic diagram of the proposed device, metal ferroelectric metal insulator mosfet with spacer (M-Fe-MIMOS (spacer)) and conventional device, M-Fe-MIMOS. The length of bulk silicon (L_b), Drain/Source ($L_{d/s}$), Spacer1/Spacer2 (L_{sp}) are 60nm, 10nm, 8nm respectively. The thickness of Drain/Source ($T_{d/s}$), Bulk Silicon (T_{bs}), oxide layer (T_o), ferroelectric layer, metal1, metal2, insulator layer are 3nm, 50nm, 1nm, 1nm, 2nm, 1.7nm, 2.3nm respectively. The total gate thickness is denoted by T_{gt}. The material of bulk silicon, source/drain/gate, metal1, oxide layer, ferroelectric layer is Silicon

Accepted Manuscript

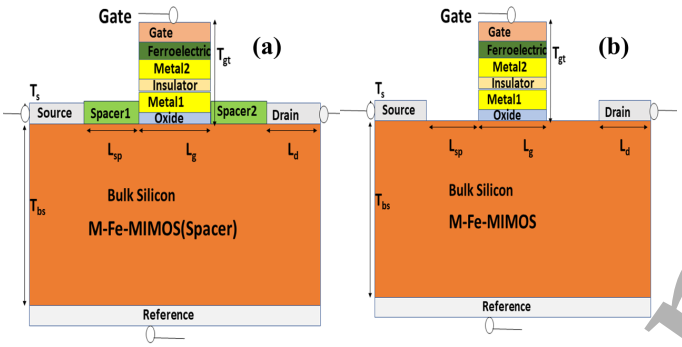


Fig. 1: Schematic design for M-Fe-MIMOS(spacer) and M-Fe-MIMOS.

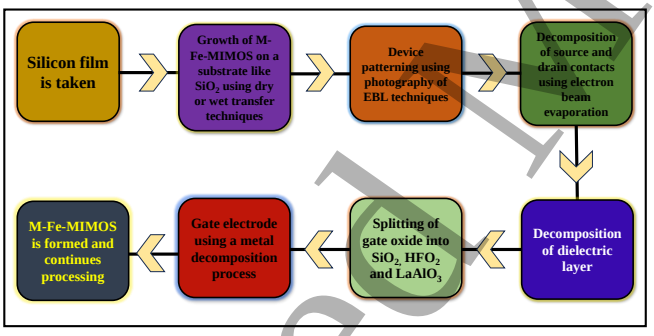


Fig. 2: Fabrication process of M-Fe-MIMOS [18].

(Si), Aluminum (Al), Au, Silicon dioxide (SiO_2), HfO_2FE, respectively. M-Fe-MIMOS (spacer) and M-Fe-MIMOS have similar architecture except 2 layers of spacers. The material of spacer1/spacer2 is NpolySi. The uniform doping concentration of bulk silicon is 1×10^{16} cm^{-3} with acceptor type. The gaussian doping concentration of drain/source ($N_{d/s}$) and threshold voltage (N_v) is 1×10^{21} cm^{-3} with donor type, 1×10^{18} cm^{-3} with acceptor type correspondingly as shown in table 1. The metal1 work function is maintained at 5.3eV.

The equation of drain current equation of MOSFET is exhibited in eq. 2.

$$I_{\mathrm{d}} = \frac{\mu \times C_{\mathrm{ox}} \times W}{L} \times [(V_{\mathrm{g}} - V_{\mathrm{t}}) - \frac{(V_{\mathrm{ds}}) \times (V_{\mathrm{ds}})}{2}] \quad (2)$$

Accepted Manuscript

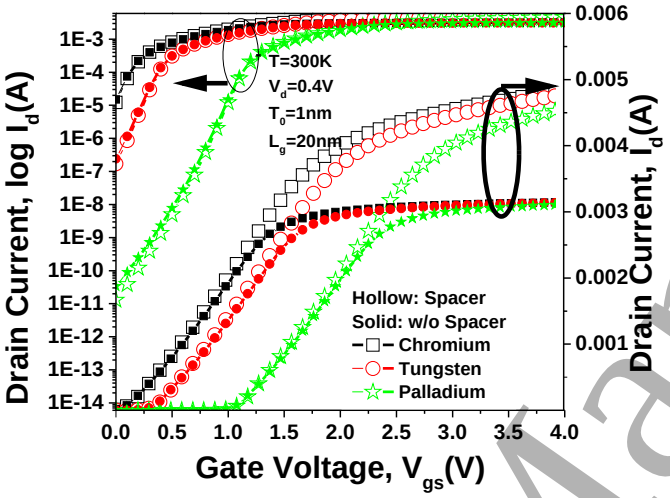


Fig. 3: The Plot of I_d vs V_{gs} for 3 work function of M-Fe-MIMOS at V_d=0.4V.

Table 1: M-Fe-MIMOS device parameter for TCAD simulation.

Parameter	M-Fe-MIMOS (Spacer)
Temperature	300k
Channel length (L_g)	20nm
Oxide thickness (T_o)	1nm
Length of drain/ source ($L_{d/s}$)	12nm
Thickness of drain/-source ($T_{d/s}$)	3nm
Doping concentration of drainsource	$1 \times 10^{21}\ cm^{-3}$
Total width of gate (T_{gt})	130nm

Here, the threshold voltage is V_t, and width, length of the channel are W, L respectively, μ is mobility of n type, C_{ox} is capacitance of oxide, V_g, V_t, V_{ds} are the gate, threshold, drain voltage [19, 20].

3 Simulation study of M-Fe-MIMOS(spacer)

Cogenda Visual TCAD(Technology Computer Aided Design) system is used for entire validation. In this study, the variation of work function for palladium, chromium, tungsten has been analyzed in gate material. The voltage of drain (V_d) and temperature (T) is secured at 0.4V and 300K. The range of gate voltage (V_{gs}) is different from 0 to 4V. he L-K and Miller models

Accepted Manuscript

Table 2: The various material with its work function.

Material	Work function ϕ (eV)
Chromium	4.3
Tungsten	4.55
Paladium	5.3

are often used to elucidate the polar- isation (P) and voltage across ferroelectric (FE) materials [10, 11]. The field dependent mobility model (FLDMOB) is used for computing the mobility, concentration-dependent mobility (CONMOB) [21, 22], polarization model, Shockley-Read–Hall (SRH) recombination model[23, 24]. We have explored the various work function (ϕ) such as 4.3eV for aluminum, 4.55eV for Chromium, 5.3eV for palladium as shown table 2.

The fabrication process is shown in fig. 2. The visual TCAD simulator has reported the feasibility of fabricating M-Fe-MIMOS. Firstly, the silicon film is thinned. Next, the Al_2O_3/HfO_2 layer is deposited on the silicon interfacing layer via atomic layer deposition (ALD) [18]. Finally, the metal gate is formed on the gate dielectric via electron beam evaporation process. The dopants of the source and drain region are implanted by spike annealing activation. Finally, the electron beam is evaporated on the surface of the source and drain, resulting in the formation of M-Fe-MIMOS.

4 Result and discussion

4.1 Effect of electrical performance on M-Fe-MIMOS

We have explored the electrical parameter of M-Fe-MIMOS such as electric field distribution over the channel, max. electric field, potential across channel, quasi fermi potential, band structure for all gate electrodes. We studied the electric field distribution over channel means the electric field is in vertical direction. The drain voltage is fixed at 0.4V with variation in gate voltage. The electric field distribution across channel exhibits almost homogeneous distribution, when device is in linear region. The device is in saturation region, when the electric field took M-shape. The electric field (E) at drain side is higher for palladium as depicted in figure 14 than source area, which reflect the better control at channel in linear region[17]. The curve of E near source is better than channel due to the electron collected under oxide layer to make the reverse bais, leads the potential difference higher thus, increases the electric field (E). The peak of E is higher at drain side for better voltage stress as shown in figure 5. The plot of potential vs position along channel are shown in fig. 6 for various gate layer of M-Fe-MIMOS at V_{gs}=V_d=0.4V. The potential and electric field of palladium is higher than other gate electrode at drain side as shown in fig. 6 (b), when device is in biased or active condition. The max. electric field and potential for palladium gate electrode are improved by 141% and 0.002% as compared to chromium gate electrode. The potential is higher

Accepted Manuscript

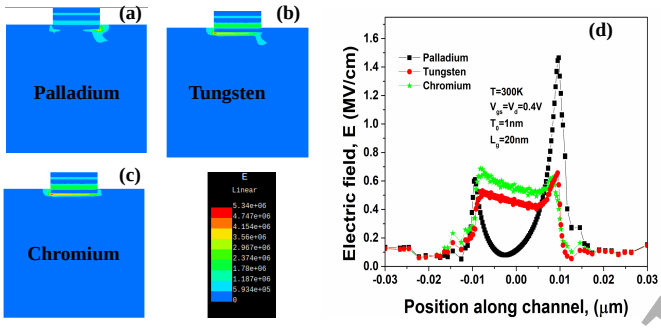


Fig. 4: The contour plot of electric field (a) palladium (b) tungsten (c) chromium (d) the graph of electric field vs position along channel for various gate electrode of M-Fe-MIMOS(spacer) at V_d=0.4V, V_{gs}=0.4V, T=300k.

at drain side, which reveals better flow of current and smooth amplification of signal in ON state (conducting).

The maximum and "M" shaped electric field profile in M-Fe-MIMOS (spacer) devices for the palladium gate electrode arises from the combined effects of gate-channel interactions, channel potential variations, ferroelectric material mobility, short channel effects, and design/material heterogeneities. These factors create a complex electric field distribution that enhances device performance by optimizing control over the channel and improving transconductance [25–28].

Figure 7 reflects the quasi fermi potential vs position along channel for different gate electrode at V_{gs}=V_d=0.4V and T=300K. The quasi fermi potential of palladium is higher than others electrode, when FET is in non equilibrium and biased condition. The quasi fermi potential for hole and electron may shift from equilibrium position on applying the external bias. The ntype FET is observed for rising the electrons by applying quasi fermi level [29, 30]. Figure 8 exhibits the energy band structure vs position along channel at V_{gs}=V_d=0.4V and T=300K for various gate electrode. The band gap of palladium is 1.1eV, which is lower than chromium 1.12eV. Valence band gap (E_v) and conduction band gap (E_c) are extracted along channel as depicted in fig. 8. The device performance and static parameter is improved for palladium.

4.1.1 Analog effect

4.2 Analog and switching parameter

Figure 3 reflects the drain current, I_d vs gate voltage, V_{gs} at constant V_d=0.4 V for various work function of M-Fe-MIMOS (spacer) and M-Fe-MIMOS. The drain current for M-Fe-MIMOS (spacer) with palladium is higher than others such as chromium, tungsten. The M-Fe-MIMOS (spacer) represents decreased

Accepted Manuscript

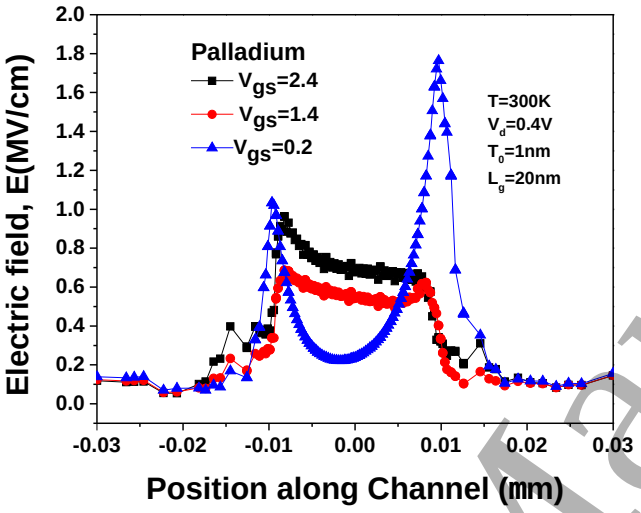


Fig. 5: The impact of electric field vs position along channel for various gate voltage of M-Fe-MIMOS(spacer) at V_d=0.4V, T=300k.

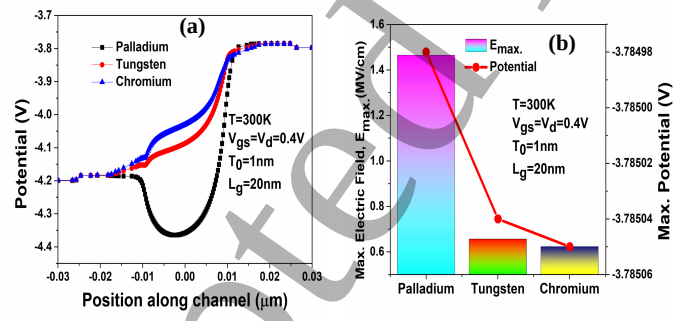


Fig. 6: The curve for various gate voltage of M-Fe-MIMOS(spacer) (a) potential (b) maximum electric field and maximum potential vs position along channel at V_d=0.4V, T=300k.

leakage current than M-Fe-MIMOS in linear drain current graph, that reveals enhanced the gate effect to control the channel.

Figure 9 shows the gate voltage vs transconductance on left side of graph and device efficiency on right side of graph at V_d=0.4 V. Transconductance (g_m) is ratio of drain current and gate voltage. The g_m of M-Fe-MIMOS (spacer) for palladium is improved than chromium, tungsten. Also, the g_m of M-Fe-MIMOS (spacer) is higher than M-Fe-MIMOS, which reveals upgraded electron mobility, better gate controllability, increase average carrier velocity,

Accepted Manuscript

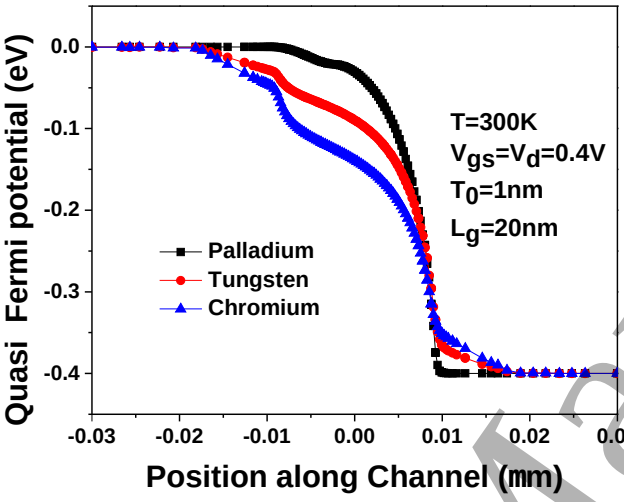


Fig. 7: The effect of quasi fermi potential vs position along channel for various gate voltage of M-Fe-MIMOS(spacer) at V_d=0.4V, T=300k.

reduced short channel effect (SCEs). TGF is gain generated by power loss. Transconductance Generation factor (TGF) of M-Fe-MIMOS (spacer) for palladium is enhanced as compare to other work functions (ϕ) [31–33], which reflects improved device efficiency tht is close to ideally $50V^{-1}$. The TGF is higher that shows the more g_m [34]. The high transconductance and device efficiency of M-Fe-MIMOS arise from their strong gate control over the channel, high carrier mobility, and material/structural enhancements. These factors collectively enable NCFET to achieve significant changes in drain current with small changes in gate voltage, making them highly efficient and effective for a wide range of electronic applications [35, 36].

For digital circuit, the operation of transistor takes place in linear region, while for analog circuit, the operation of transistor is in saturation region (V_d=V_{gs}=0.4V) [37, 38]. The electric field distribution is studied by tuning V_{gs} for different operating situation. The breakdown time of channel is lower at higher electric field. The breakdown of time dependent equation (t_b) is given below in eq. 3:

$$t_{\mathrm{b}} = Aexp(-\gamma E_{\mathrm{o}})\frac{E_{\mathrm{a}}}{K_{\mathrm{b}}T} \tag{3}$$

Here, A is electric field acceration, γ is proportionality constant, T is absolute temperature, E_a is thermal activation energy, E_o is oxide electric field, K_b is Boltzmann constant [17, 39].

Figure 10 shows the gate voltage vs leakage current (I_{off}) on left side of graph and switching ratio (I_{on}/I_{off}) on right side of graph at V_d=0.4 V for various work function.The I_{off} of M-Fe-MIMOS (spacer) for palladium is lower than chromium, tungsten. The I_{off} of M-Fe-MIMOS (spacer) is also decreased

Accepted Manuscript

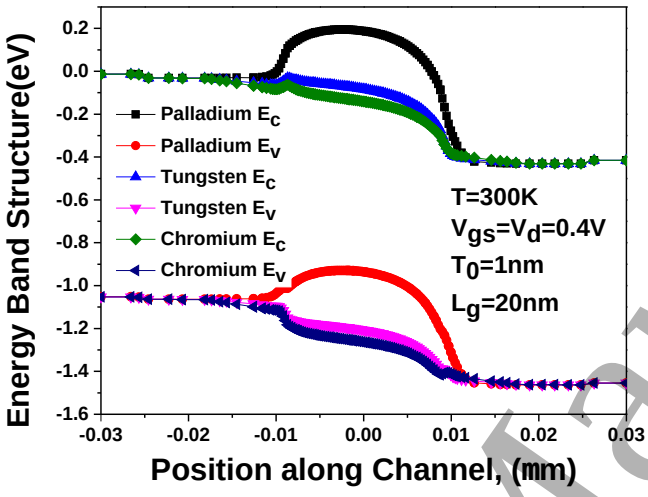


Fig. 8: The impact of energy band structure (conduction band (E_c) and valence band (E_v)) of palladium, tungsten, chromium vs position along channel for M-Fe-MIMOS(spacer) at V_d=0.4V, T=300k.

as compare to M-Fe-MIMOS, that exhibits the lower leakage current. The Y axis has two sides of the plot. The left part of the arrow represents the leakage current and the right part of the arrow represents the switching ratio. The I_{on}/I_{off} of M-Fe-MIMOS (spacer) for palladium is better than chromium, tungsten. The I_{on}/I_{off} of M-Fe-MIMOS (spacer) is higher than M-Fe-MIMOS, that affect the switching speed of device and lower leakage current as shown in table 3.

Figure 11 shows the gate voltage vs threshold voltage (V_t) on left side of plot and subthreshold swing (SS) on right side of plot at T=300K, V_d=0.4 V for different work function. The threshold voltage (V_t) of M-Fe-MIMOS (spacer) for palladium is better than chromium, tungsten. The V_t of M-Fe-MIMOS (spacer) is also higher than M-Fe-MIMOS, which reveals better potential shield at drain side better potential shield, increased gate controllability over channel. The subthreshold swing (SS) of M-Fe-MIMOS (spacer) for palladium is less than chromium, tungsten. The SS of M-Fe-MIMOS also reflects decrement as compare to M-Fe-MIMOS, which shows lower short channel effect i.e. drain barrier induced lowering (DIBL).

$$g_{\mathrm{m}} = \frac{\partial I_{\mathrm{ds}}}{\partial V_{\mathrm{g}}} \tag{4}$$

$$g_{\mathrm{d}} = \frac{\partial I_{\mathrm{d}}}{\partial V_{\mathrm{ds}}} \tag{5}$$

$$V_{\mathrm{e}} = \frac{I_{\mathrm{ds}}}{g_{\mathrm{d}}} \tag{6}$$

Accepted Manuscript

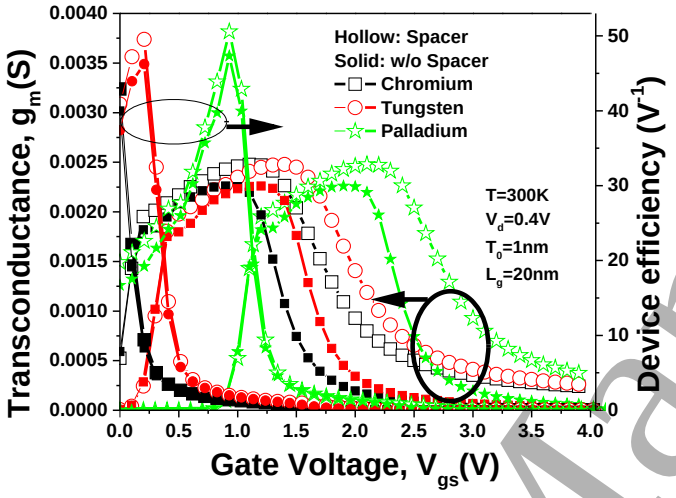


Fig. 9: The graph of transconductance and device efficiency vs V_{gs} for 3 work function of M-Fe-MIMOS at V_d=0.4V.

Table 3: M-Fe-MIMOS comparison for different work function materials such as palladium, chromium, and tungsten.

Parameter	M-Fe-MIMOS (spacer) Chromium (Cr)	M-Fe-MIMOS (spacer) Tungsten (T)	M-Fe-MIMOS (spacer) Palladium (Pd)	M-Fe-MIMOS Chromium (Cr)	M-Fe-MIMOS Tungsten (T)	M-Fe-MIMOS Palladium (Pd)
I_{off}(nA)	12200	169.0	0.013	14800	240.0	0.0338
$I_{on}/I_{off}(10^6)$	0.000394	0.0281	348	0.000212	0.013	91.8
$V_{th}(V)$	-0.28	0.201	0.9504	-0.324	0.188	0.938
SS $(mV/dec.)$	775.5	257.1	176.4	800.24	279	187.3
TGF (V^{-1})	42.89	49.5	50.6	39.9	46.2	47.3

$$A_v = \frac{g_m}{g_d} \tag{7}$$

Figure 12 exhibits the gate voltage vs intrinsic gain (A_v) plot at T=300K, V_d=0.4 V for different work function. The intrinsic gain (A_v) of M-Fe-MIMOS (spacer) for palladium is better than chromium, tungsten. The A_v of M-Fe-MIMOS (spacer) is higher than M-Fe-MIMOS, which indicates lower output conductance and higher transconductance. It is ratio of transconductance and output conductance[40, 41].

Figure 13 exhibits the drain voltage vs drain current (I_d) of plot at T=300K, V_{gs}=0.4 V for different work function. The drain current (I_d) of M-Fe-MIMOS (spacer) for palladium is better than chromium, tungsten. The drain current (I_d) of M-Fe-MIMOS (spacer) is less than M-Fe-MIMOS, which shows the lower

Accepted Manuscript

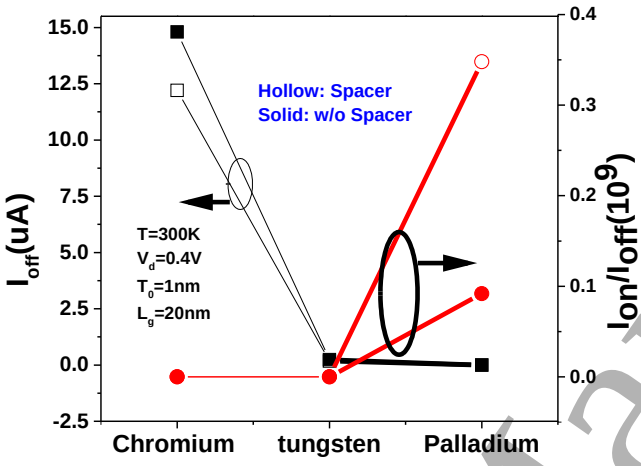


Fig. 10: The Plot of gate voltage vs leakage current and switching ratio for different work function of M-Fe-MIMOS(spacer) at V_d=0.4V, T=300k.

output conductance (g_d), lower DIBL as well as channel length modulation (CLM), thus, upgraded the control on gate.

The experiment data is taken from reference [42]. Fig. 14 shows the drain current and gate voltage curves at a drain voltage of 0.1V, a ferroelectric thickness of 5nm and a channel length of 30um for a negative capacitance field effect transistor (NCFET). To verify the physical model, experimental data extraction was taken from reference [42]. The approximate match between the simulated and experimental data validates the simulation model.

5 Conclusion

This work addresses the comparative study of M-Fe-MIMOS (spacer) and M-Fe-MIMOS for different gate electrodes, and reveals that the palladium with work function 5.3eV reflects improved result than others like chromium, tungsten. The palladium reveals 10^{-6} times lower the leakage current as compare to chromium, higher switching ratio, decayed the subthreshold swing. Also, some analog parameters found in M-Fe-MIMOS like transconductance, TGF, g_d. Theoretically, we concluded that palladium proved to be a potential material for fabrication of M-Fe-MIMOS with palladium as gate electrode. The increasing work function (ϕ) indicates improved performance of the device. The electric field across channel has also been studied for various work function thus, electric field of palladium is enhanced than other gate electrodes. The M-Fe-MIMOS (spacer) design has already seen significant advances in terms of accuracy, and component count reduction leading to downsizing, lower cost, and greater efficiency. Consequently, the findings of this research

Accepted Manuscript

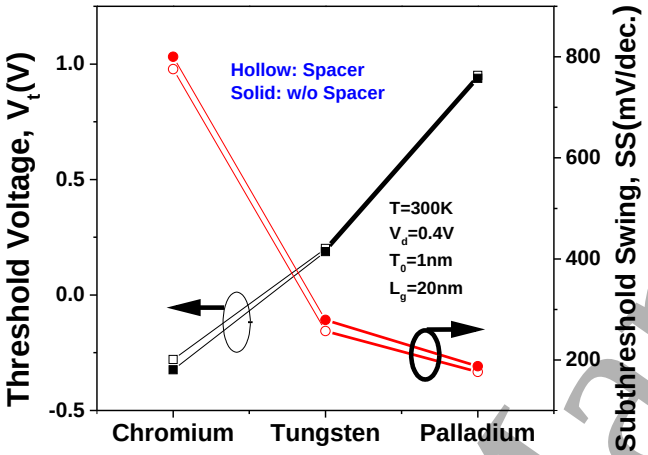


Fig. 11: The Curve of gate voltage vs threshold voltage on the left side and switching ratio on the right side for different work function of M-Fe-MIMOS(spacer) at V_d=0.4V, T=300k.

can assist engineers in designing such nanoelectronic devices that meet their requirements.

Acknowledgment

The authors thank to Vinod Dham Centre of Excellence for Semiconductor and Microelectronics (VDCOE4SM), Nano Bioelectronics and Microelectronics Lab Lab, Delhi Technological University for giving essential facilities.

Declarations

A. Funding Statement

Yash thanks the C.S.I.R. for the CSIR-JRF award (08/133(0050)/2020-EMR-I).

B. Conflict of Interests

The authors declare that they are not aware of any disagreements regarding personal ties or interests that might have impacted the work represented in this publication.

Accepted Manuscript

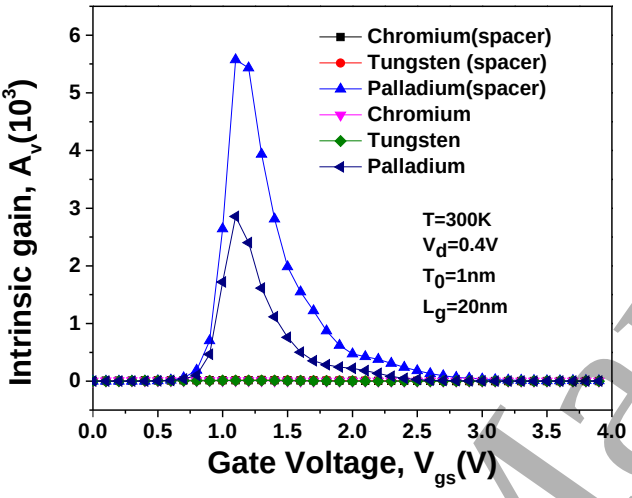


Fig. 12: The Curve of intrinsic gain vs gate voltage for various work function of M-Fe-MIMOS(spacer) at V_d=0.4V, T=300k.

C. Availability of Data & Material

The creators mentioned above possess the necessary information related to this examination work, and they are committed to revealing it in the event that they are asked to do so in the future.

D. Authors' contribution

The comprehension and design of the study were improved by all writers.

E. Compliance with Ethical Standard

The authors have located all of the moral standards and intend to adhere to them going forward.

F. Consent to participate & for publication

Since that the relevant study paper was produced in the "no-life science magazine." Therefore, at this time, it is "Non Applicable." However, the authors have provided the agencies with all journal policies and their permission for additional processing.

Accepted Manuscript

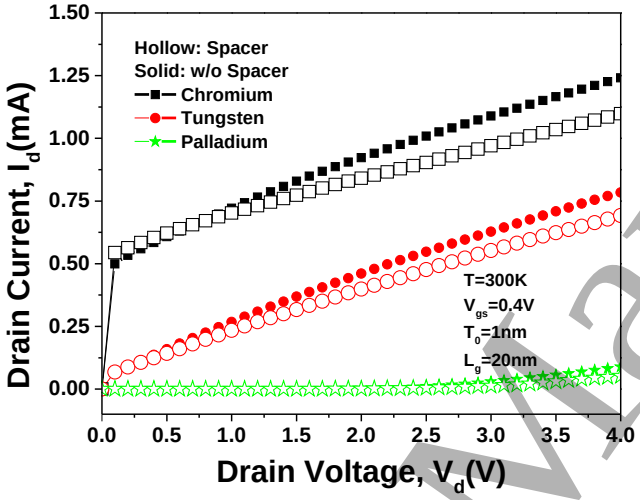


Fig. 13: The Curve of drain voltage vs drain current for different work function of M-Fe-MIMOS(spacer) at V_d=0.4V, T=300k.

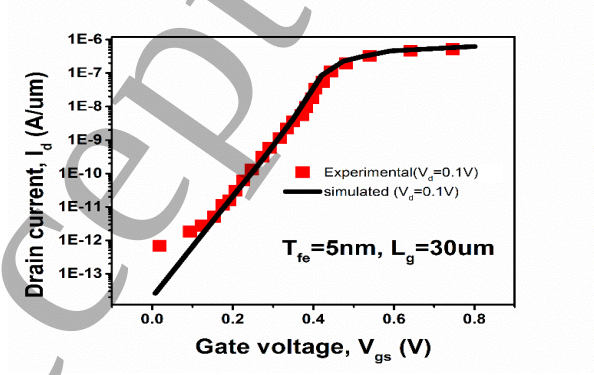


Fig. 14: Calibration result of experimental data vs simulated data [42].

Accepted Manuscript

References

[1] Rahi, S.B., Tayal, S., Upadhyay, A.K.: A review on emerging negative capacitance field effect transistor for low power electronics. Microelectronics Journal **116**, 105242 (2021)

[2] Lin, C.-I., Khan, A.I., Salahuddin, S., Hu, C.: Effects of the variation of ferroelectric properties on negative capacitance fet characteristics. IEEE transactions on electron devices **63**(5), 2197–2199 (2016)

[3] Mann, R., Pathak, Y., Chaujar, R.: Temperature analysis on short channel effects of modified ncfet: A simulation study. In: 2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), pp. 1–5 (2022). IEEE

[4] Appleby, D.J., Ponon, N.K., Kwa, K.S., Zou, B., Petrov, P.K., Wang, T., Alford, N.M., O'Neill, A.: Experimental observation of negative capacitance in ferroelectrics at room temperature. Nano letters **14**(7), 3864–3868 (2014)

[5] Kumar, S., Chatterjee, A.K., Pandey, R.: Performance analysis of gate electrode work function variations in double-gate junctionless fet. Silicon **13**, 3447–3459 (2021)

[6] Huang, X., Caron, M.: A type-based equalization technique for frequency response distortion compensation. In: 2012 IEEE Radio and Wireless Symposium, pp. 187–190 (2012). https://doi.org/10.1109/RWS.2012.6175291

[7] Choi, K.M., Choi, W.Y.: Work-function variation effects of tunneling field-effect transistors (tfets). IEEE Electron Device Letters **34**(8), 942–944 (2013)

[8] Pahwa, G., Dutta, T., Agarwal, A., Chauhan, Y.S.: Physical insights on negative capacitance transistors in nonhysteresis and hysteresis regimes: Mfmis versus mfis structures. IEEE Transactions on Electron Devices **65**(3), 867–873 (2018). https://doi.org/10.1109/TED.2018.2794499

[9] Awadhiya, B., Kondekar, P.N., Yadav, S., Upadhyay, P.: Insight into threshold voltage and drain induced barrier lowering in negative capacitance field effect transistor. Transactions on Electrical and Electronic Materials **22**(3), 267–273 (2021)

[10] Saha, A.K., Datta, S., Gupta, S.K.: "negative capacitance" in resistor-ferroelectric and ferroelectric-dielectric networks: apparent or intrinsic? Journal of Applied Physics **123**(10) (2018)

Accepted Manuscript

[11] Song, T.K.: Landau-khalatnikov simulations for ferroelectric switching in ferroelectric random access memory application. Journal of the Korean Physical Society **46**(1), 5 (2005)

[12] Pahwa, G., Agarwal, A., Chauhan, Y.S.: Numerical investigation of short-channel effects in negative capacitance mfis and mfmis transistors: Subthreshold behavior. IEEE Transactions on Electron Devices **65**(11), 5130–5136 (2018)

[13] Singh, P., Yadav, D.S.: Impact of work function variation for enhanced electrostatic control with suppressed ambipolar behavior for dual gate l-tfet. Current Applied Physics **44**, 90–101 (2022)

[14] Rassekh, A., Sallese, J.-M., Jazaeri, F., Fathipour, M., Ionescu, A.M.: Negative capacitance double-gate junctionless fets: a charge-based modeling investigation of swing, overdrive and short channel effect. IEEE Journal of the Electron Devices Society **8**, 939–947 (2020)

[15] Li, Y., Hwang, C.-H., Li, T.-Y., Han, M.-H.: Process-variation effect, metal-gate work-function fluctuation, and random-dopant fluctuation in emerging cmos technologies. IEEE Transactions on Electron Devices **57**(2), 437–447 (2009)

[16] Lee, Y., Nam, H., Park, J.-D., Shin, C.: Study of work-function variation for high-backslash kappa metal-gate ge-source tunnel field-effect transistors. IEEE Transactions on Electron Devices **62**(7), 2143–2147 (2015)

[17] Liu, B., Sun, H., Huan, C., Jia, R., Cai, Y., Ke, Q.: Investigating the reliability of a negative capacitance field effect transistor regarding the electric field across the oxide layer. Journal of Electronic Materials **52**(5), 3180–3187 (2023)

[18] Choi, S.-N., Moon, S.-E., Yoon, S.-M.: Impact of oxide gate electrode for ferroelectric field-effect transistors with metal-ferroelectric-metal-insulator-semiconductor gate stack using undoped hfo2 thin films prepared by atomic layer deposition. Nanotechnology **32**(8), 085709 (2020)

[19] Shur, M., Fjeldly, T.A., Ytterdal, T., Lee, K.: Unified mosfet model. Solid-state electronics **35**(12), 1795–1802 (1992)

[20] Anshul, A., Chaujar, R.: Semi-empirical dft based investigation of electronic and quantum transport properties of novel gs-agnr (n) fet. IEEE Transactions on Nanotechnology (2024)

[21] Agrwal, P., Kumar, A.: Performance assessment of ingaas–soi–finfet

Accepted Manuscript

for enhancing switching capability using high-k dielectric. Memories-Materials, Devices, Circuits and Systems **8**, 100117 (2024)

[22] Kumar, A., Gupta, N., Jain, A., Gupta, R., Choudhary, B., Kumar, K., Goyal, A.K., Massoud, Y.: Digital-logic assessment of junctionless twin gate trench channel (jl-tgtc) mosfet for memory circuit applications. Memories-Materials, Devices, Circuits and Systems **6**, 100087 (2023)

[23] Pathak, Y., Verma, K., Malhotra, B.D., Chaujar, R.: Tcad analysis of linearity performance on modified ferroelectric layer in fet device with spacer. Advanced Nanoscale MOSFET Architectures: Current Trends and Future Perspectives, 113–123 (2024)

[24] Raj, A., Sharma, S.K.: Investigation on palladium gate electrode-based soi junctionless fet for hydrogen gas sensing. Microelectronics Journal **151**, 106312 (2024)

[25] Song, Q., Yang, S., Tang, G., Han, C., Zhang, Y., Tang, X., Zhang, Y., Zhang, Y.: 4h-sic trench mosfet with l-shaped gate. ieee electron device letters **37**(4), 463–466 (2016)

[26] Dunga, M.V., Xi, X., He, J., Liu, W., Cao, K.M., Jin, X., Ou, J.J., Chan, M., Niknejad, A.M., Hu, C.: Bsim4. 6.0 mosfet model. University of California, Berkeley (2006)

[27] Pathak, Y., Malhotra, B.D., Chaujar, R.: Tcad analysis and simulation of double metal negative capacitance fet (dm ncfet). In: 2021 Devices for Integrated Circuit (DevIC), pp. 224–228 (2021). IEEE

[28] Sudarsanan, A., Venkateswarlu, S., Nayak, K.: Superior work function variability performance of horizontally stacked nanosheet fets for sub-7-nm technology and beyond. In: 2020 4th IEEE Electron Devices Technology & Manufacturing Conference (EDTM), pp. 1–4 (2020). IEEE

[29] Ponse, F., Masselink, W., Morkoç, H.: Quasi-fermi level bending in modfet's and its effect on fet transfer characteristics. IEEE Transactions on Electron Devices **32**(6), 1017–1023 (1985)

[30] Chanda, M., De, S., Sarkar, C.K.: Modeling of characteristic parameters for nano-scale junctionless double gate mosfet considering quantum mechanical effect. Journal of Computational Electronics **14**, 262–269 (2015)

[31] Pathak, Y., Malhotra, B.D., Chaujar, R.: TCAD analysis and simulation of double metal negative capacitance FET (DM NCFET). in 2021 Devices for Integrated Circuit (DevIC). IEEE (2021)

Accepted Manuscript

[32] Bag, A., Pandey, R., Kashyap, S., Madan, J., Ramanujam, J.: The influence of top electrode work function on the performance of methylammonium lead iodide based perovskite solar cells having various electron transport layers. Chemical Physics Letters **806**, 140009 (2022)

[33] Pathak, Y., Malhotra, B.D., Chaujar, R.: A numerical study of analog parameter of negative capacitance field effect transistor with spacer. In: 2021 7th International Conference on Signal Processing and Communication (ICSC), pp. 277–281 (2021). IEEE

[34] Kumar, B., Sharma, M., Chaujar, R.: Gate electrode work function engineered jam-gs-gaa finfet for analog/rf applications: Performance estimation and optimization. Microelectronics Journal **135**, 105766 (2023)

[35] Sai-Halasz, G.A., Wordeman, M.R., Kern, D., Rishton, S., Ganin, E.: High transconductance and velocity overshoot in nmos devices at the 0.1-mu m gate-length level. IEEE Electron Device Letters **9**(9), 464–466 (1988)

[36] Do, Q.-T., Blekker, K., Regolin, I., Prost, W., Tegude, F.: High transconductance misfet with a single inas nanowire channel. IEEE Electron Device Letters **28**(8), 682–684 (2007)

[37] Razavi, B.: Design of analog cmos integrated circuits. University of California (2005)

[38] Schneider, M.C., Galup-Montoro, C.: Cmos analog design using all-region mosfet modeling. Cambridge University Press (2010)

[39] McPherson, J., Baglee, D.: Acceleration factors for thin gate oxide stressing. In: 23rd International Reliability Physics Symposium, pp. 1–5 (1985). IEEE

[40] Khan, A.I., Radhakrishna, U., Salahuddin, S., Antoniadis, D.: Work function engineering for performance improvement in leaky negative capacitance fets. IEEE Electron Device Letters **38**(9), 1335–1338 (2017)

[41] Hu, V.P.-H., Chiu, P.-C., Lu, Y.-C.: Impact of work function variation, line-edge roughness, and ferroelectric properties variation on negative capacitance fets. IEEE Journal of the Electron Devices Society **7**, 295–302 (2019)

[42] Lee, M., Chen, P.-G., Liu, C., Chu, K., Cheng, C.-C., Xie, M.-J., Liu, S.-N., Lee, J.-W., Huang, S.-J., Liao, M.-H., *et al.*: Prospects for ferroelectric hfzrox fets with experimentally cet= 0.98 nm, ssfor= 42mv/dec, ssrev= 28mv/dec, switch-off¡ 0.2 v, and hysteresis-free strategies. In: 2015 IEEE International Electron Devices Meeting (IEDM), pp. 22–5 (2015). IEEE

Accepted Manuscript

arXiv:2412.18344v1 [math.DS] 24 Dec 2024

Predator Prey Scavenger Model using Holling's Functional Response of Type III and Physics-Informed Deep Neural Networks

Aneesh Panchal*[1], Kirti Beniwal†[2], and Vivek Kumar‡[2]

[1]Department of Computational and Data Sciences, Indian Institute of Science, Bangalore, Karnataka−560012, INDIA
[2]Department of Applied Mathematics, Delhi Technological University, Delhi−110042, INDIA

Abstract

Nonlinear mathematical models introduce the relation between various physical and biological interactions present in nature. One of the most famous models is the Lotka-Volterra model which defined the interaction between predator and prey species present in nature. However, predators, scavengers, and prey populations coexist in a natural system where scavengers can additionally rely on the dead bodies of predators present in the system. Keeping this in mind, the formulation and simulation of the predator prey scavenger model is introduced in this paper. For the predation response, respective prey species are assumed to have Holling's functional response of type III. The proposed model is tested for various simulations and is found to be showing satisfactory results in different scenarios. After simulations, the American forest dataset is taken for parameter estimation which imitates the real-world case. For parameter estimation, a physics-informed deep neural network is used with the Adam backpropagation method which prevents the avalanche effect in trainable parameters updation. For neural networks, mean square error and physics-informed informed error are considered. After the neural network, the hence-found parameters are fine-tuned using the Broyden–Fletcher–Goldfarb–Shanno algorithm. Finally, the hence-found parameters using a natural dataset are tested for stability using Jacobian stability analysis. Future research work includes minimization of error induced by parameters, bifurcation analysis, and sensitivity analysis of the parameters.

Keywords- Stability analysis; Predator prey scavenger model; Physics-informed deep neural network; Broyden–Fletcher–Goldfarb–Shanno optimization.

1 Introduction

Mathematical modeling and differential equations increase our understanding of population dynamics and ecology over time. Differential equations [1] are very useful for defining various natural phenomena such as simple harmonic motion (SHM), Newton's law of cooling, chemical equations, radioactive decay, and many more. But the study of differential equations increased drastically when the study of environment meets mathematics that is population dynamics [2], mathematical ecology [3], epidemiology [4] etc. Population dynamics [5, 6, 7] is the study of the

*Email: aneeshpanchal840850@gmail.com
†Email: kirtibnwl1912@gmail.com
‡Email: vivekkumar.ag@gmail.com

variation of the population over time according to different natural parameters. Several continuous variables can be studied jointly to generate a system of ordinary differential equations. It is generally necessary to estimate and fine-tune specific parameters to make the dynamical system valid, providing reliable mathematical explanations for certain phenomena or predictions. In most cases, estimating the parameters involved in dynamical systems is, in fact, a challenging optimization problem that needs special consideration due to some unknown natural parameters. The study of population dynamics got a boost when Italian mathematician *Vito Volterra* [8] introduced a differential equation model for the population dynamics of two different interacting species viz. predator and prey in nature. Predator-prey modeling using the Lotka-Volterra classical model and Holling's functional response has been a great instrument in advancing the understanding of ecological systems. While the Lotka-Volterra model provides a fundamental basis for studying population interactions, Holling's functional response enhances its applicability by considering the complexities of predator-feeding behavior. Integrating these models contributes to a more comprehensive and realistic representation of predator-prey dynamics in ecological modeling. Further advancements in this field are crucial for addressing contemporary ecological challenges and improving our ability to manage and conserve diverse ecosystems.

In exploring the dynamics of predator-prey models, several key research papers provide valuable insights. One such study by Sarwardi et.al. [2], dives into the behavior of a two-predator model with prey refugees. By considering the adaptive strategy of prey-seeking refugees, the research provides insights on how refugee behavior affects the stability of the system. Another study by Kant and Kumar [4], focuses on the stability analysis of a predator-prey system where prey populations migrate, and both species are susceptible to disease. This research paper deepens our understanding of migration and disease integration with predator-prey population variations and its impact on the stability of the system. Also, a study by Hadžiabdić et.al. [9] extends the classical Lotka-Volterra model by adding two predators and common prey, providing insights into the dynamics and stability analysis of multiple predators with a same prey population.

Further research by Wu et.al. [10] on the delayed predator-prey system, consider factors such as fear effect, herd behavior, and disease in the susceptible prey. The inclusion of time delays and additional ecological aspects offers a more realistic representation of predator-prey interactions, which increases our understanding of the natural behavior of predator and prey. Research by Arora and Kumar [11], introduces a delayed version of a prey-predator system with a modified Holling-Tanner response function. This modification aims to capture more accurate predator-prey interactions.

Estimating the parameters involved in dynamical systems is a challenging optimization problem that needs special consideration due to certain unknown natural parameters. Hence, the parameter estimation problem remains challenging despite the variety of methods available to estimate them. Artificial neural networks (ANN) [12] can be used to find the unknown parameters of the dynamical systems. "ANNs are information-processing systems with characteristics similar to biological neural networks" [13]. A neural network consists of several layers, the first layer is called the input layer, the last one is the output layer, and the layers between are called hidden layers. Every layer has several neurons each of which consists of activation function, weight value, and bias value [14]. A deep neural network (DNN) is an ANN with more than one hidden layer. DNNs are useful for complex optimization problems due to the large number of trainable parameters. Additionally, Raissi et al. [15] proposed a method called physics-informed neural network (PINN) for solving partial differential equations. In PINN, a loss function is formed using training data points, initial or boundary conditions, and residual of the differential equation. The problem is then converted into an optimization problem which can then be optimized using gradient-based optimizers like gradient descent and stochastic-based optimizers. In this study, a complex physics-informed deep neural network algorithm is employed to estimate the parameters of the proposed model using the natural population dataset.

1.1 Motivation and Contribution

In the classical Lotka-Volterra model, only the searching time taken by the predators is taken into consideration but in addition to that, prey require some handling time which naturally means time required to catch and kill the prey after the search. Also in terrestrial ecology, there are scavenger species populations that can consume the dead bodies of other animals. So, this research paper aims to create and analyze a three-population predator-prey scavenger mathematical model in which respective prey response to predation is considered as Holling's functional response of type III and scavenger predator (dead bodies) interaction is considered as Holling's functional response of type I as dead bodies don't require any handling time.

The objective of this research paper is to formulate a complex non-linear dynamical mathematical model that will help us understand the interactions between predator, prey, and scavengers in ecology. The analysis of the dynamical model thus created and some examples to validate the model are also required. By developing a predator-prey scavenger model, we aim to contribute to the advancement of ecological modeling and population dynamics by adding the scavenger population in existing models along with the parameter estimation of the proposed model using the given dataset [16]. The complex integration of DNN, PINN, and fine-tuning optimization helps efficiently estimate the parameters. Finally, the last objective of this paper is to analyze the stability of the model using the obtained parameters.

This paper is organized as follows: Section 2 covers the preliminaries and basic concepts required for the understanding of the subsequent sections. Section 3 covers predator-prey scavenger model formulation and Section 4 covers the mathematical analysis of the model. Section 5 consists of the numerical examples and respective simulations for the same. Section 6 consists, of problem formulation subsection 6.1, physics-informed deep neural network model subsection 6.2, and error analysis along with results in subsection 6.3. Section 7 consists of stability analysis of thus found parameters using given dataset [16]. Finally, section 8 concludes the paper along with expected future work.

2 Preliminaries

This section consists of basic terminologies and optimization algorithm concepts required for the foundation of the objective of this study which is the creation of a predator-prey scavenger model and estimating natural parameters using a physics-informed deep neural network. These terminologies help to understand the subsequent section in this research paper.

2.1 Holling's Functional Responses

Functional response is also very important in studying population dynamics. Functional response is majorly used in the work of Holling [17, 18, 19, 20]. Holling's different functional response defines the interaction between prey and predator using simple mathematical functions. Holling's functional responses as given in Fig. 1 are of three types,

1. **Type I:** $y = ax$
   It is a simple function used in the classical Lotka-Volterra model and is generally used nowadays but has some natural limitations. It is linear, which shows prey don't learn anything at any prey density in response to the predator's attacks. Also, the consumption time of the predator is taken as zero, which is not natural.

2. **Type II:** $y = \frac{ax}{1+bx}$
   It is a simple non-linear mathematical function that shows prey learns as prey density increases and slope decreases. Here consumption time of the predator is also taken into consideration which is defined by b. But in type II, learning is considered from the lowest

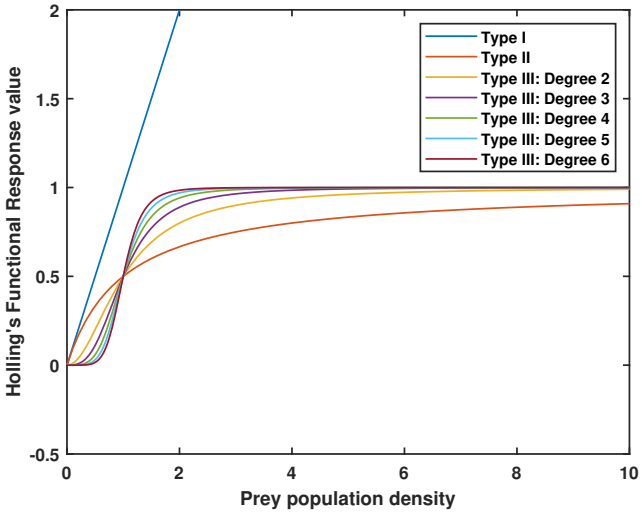


Fig. 1: Holling's functional responses.

prey density onwards. It implies prey starts to learn without any previous experience which is again not natural.

3. **Type III:** $y = \frac{ax^n}{1+bx^n}$ where, $n \geq 2$
   It is the mathematical function that is the general form of type II and is significantly less taken into consideration [18, 19, 20] in the work of population dynamics due to its complexity. It is non-linear, which shows prey initially don't know how to defend themselves from predators, so the slope of the graph increases up to a specific prey density. Then prey learns how to escape from the predator, and hence slope gradually decreases. Here n is defined by the learning rate of the prey species.

2.2 Adam Algorithm

Adam algorithm (Adaptive Moment Estimation) [21], is an adaptive learning rate algorithm to improve the efficiency of the backpropagation algorithm in deep neural networks. It changes the weights and biases of the neural net based on the history of change of gradient and this helps in the prevention of avalanche effect in parameter updation. Steps involved in adam are given in algorithm 1.

Algorithm 1: ADAM ALGORITHM

Input: $\beta_1, \beta_2, g, \theta, num_steps$
Output: θ

$m_0 = 0$
$v_0 = 0$
for $t \leftarrow 1$ ***to*** num_steps **do**
  $m_t = \beta_1 m_{t-1} + (1-\beta_1)\, g_t$
  $v_t = \beta_2 v_{t-1} + (1-\beta_2)\, g_t^2$
  $m_t^* = \frac{m_t}{1-\beta_1^t}$
  $v_t^* = \frac{v_t}{1-\beta_2^t}$
  $\theta = \theta - \frac{\alpha m_t^*}{\sqrt{v_t^* + \epsilon}}$
return θ

In algorithm 1, α represents the step size, m_t is momentum at step t, g_t is gradient at step t, β_1 is decay rate of momentum, β_2 is decay rate of squared gradient, ϵ is small parameter to prevent division from zero, and θ is parameter vector.

2.3 Broyden-Fletcher-Goldfarb-Shanno Optimization Algorithm

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) [22, 23, 24, 25] optimization algorithm is an iterative optimization method used to find the minimum of a differentiable. For easy understanding of the algorithm let's assume, We want to minimize a differentiable, scalar-valued function $f : \mathcal{R}^n \to \mathcal{R}$ with respect to input vector x. Initialize x_0 (initial guess) and B_0 (initial inverse Hessian matrix) and iteratively update x and B at each iteration k according to algorithm 2.

Algorithm 2: BFGS Optimization Algorithm

Input: $x_0, B_0, f, max_iterations$
Output: x

$x_1 = x_0$
$B_1 = B_0$
for $k \leftarrow 1$ ***to*** $max_iterations$ **do**
  $p_k = -B_k \nabla f(x_k)$
  Find α_k s.t. $\min f(x_k + \alpha_k p_k)$
  $x_{k+1} = x_k + \alpha_k p_k$
  $\nabla f_{k+1} = \nabla F(x_{k+1})$
  $\Delta x_k = x_{k+1} - x_k$
  $\Delta \nabla f_k = \nabla f_{k+1} - \nabla f_k$
  $B_{k+1} = B_k + \frac{\Delta \nabla f_k \Delta \nabla f_k^T}{\Delta \nabla f_k^T \Delta x_k} - \frac{B_k \Delta x_k \Delta x_k^T B_k^T}{\Delta x_k^T B_k \Delta x_k}$
return x_k

The algorithm 2 converges to a point x^* that represents the local minimum of the function $f(x)$. The BFGS optimization algorithm is widely used due to its efficiency and effectiveness in finding local minima [26].

3 Model Formulation

Consider three species populations, namely predator, prey, and scavengers. For modeling this system, some assumptions are required,

1. The prey population (denoted by $x(t)$) can only show logistic growth and can be eaten by predators and scavengers. The only reasons for the death of the prey population are considered competition due to scarcity of resources and predation by predators and scavengers.

2. Predator populations (denoted by $y(t)$) are considered apex predators, that is predators can eat both the prey population and scavenger population.

3. Scavenger populations (denoted by $z(t)$) can eat the prey populations and the dead bodies of the predator populations.

Here, $d = m_\alpha a, g = m_\beta b$ and $f = m_\gamma i$ where, m_α, m_β and m_γ are conversion factors. The conversion factor corresponds to an increase in the predator population due to predation of single prey.

S. No.	Parameter	Biological definition
01.	r	logistic growth rate of the prey population
02.	k	carrying capacity of the prey population
03.	a	discovery rate of prey with respect to predator
04.	a_0	handling time for prey with respect to predator
05.	b	discovery rate of prey with respect to scavenger
06.	b_0	handling time for prey with respect to scavenger
07.	i	discovery rate of scavenger with respect to predator
08.	i_0	handling time for scavenger with respect to predator
09.	d	rate of change of predator in presence of prey
10.	f	rate of change of predator in presence of scavenger
11.	g	rate of change of scavenger in presence of prey
12.	h	scavenge factor of scavengers
13.	e	natural death rate of predator
14.	j	natural death rate of scavenger

Table 1: Definitions of parameters used

A Mathematical model to take a real situation, all the parameters must have positive values. Biological definitions of the parameters used in the model formulation are given in Table 1. Discovery rate of prey with respect to the predator can be consider as the success rate of the predator in capturing the prey. Handling time for prey with respect to predator is the time required by the predator to catch and kill the prey after searching for it.

Now, the proposed mathematical predator prey scavenger model (Fig. 2) is,

$$\begin{aligned}
\frac{dx}{dt} &= rx\left(1-\frac{x}{k}\right) - \frac{ax^2y}{1+a_0x^2} - \frac{bx^2z}{1+b_0x^2} \\
\frac{dy}{dt} &= \frac{dx^2y}{1+a_0x^2} + \frac{fz^2y}{1+i_0z^2} - ey \\
\frac{dz}{dt} &= \frac{gx^2z}{1+b_0x^2} + hyz - \frac{iyz^2}{1+i_0z^2} - jz
\end{aligned} \tag{3.1}$$

with initial conditions as $x(0) > 0$, $y(0) > 0$ and $z(0) > 0$.

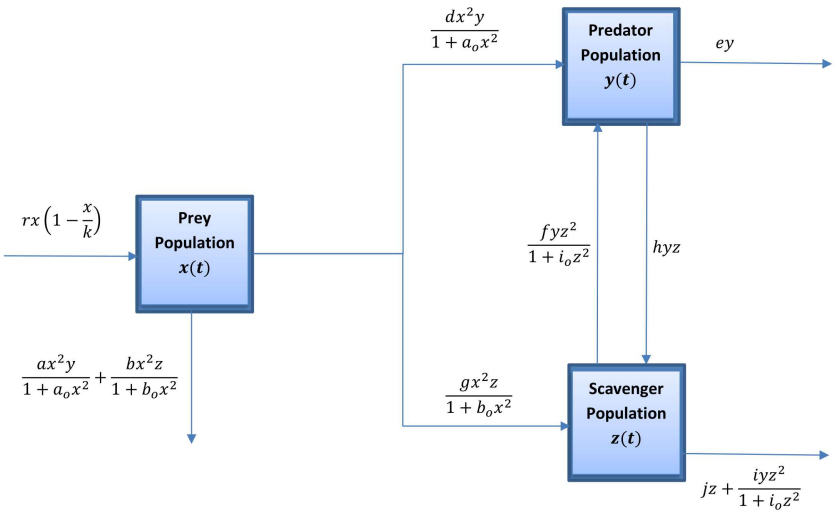


Fig. 2: Predator prey scavenger interaction flowchart

In the proposed mathematical model, prey-predator, prey-scavenger, and predator-scavenger

interactions (predation interaction only) are taken as Holling's functional response of type III as prey and scavengers require some handling time with respect to their respective predators. Also, they start to learn to defend themselves as the respective prey density increases. But scavenger-predators (dead) interaction is taken as Holling's functional response of type I as dead bodies of predator species don't require any handling time which is obvious.

In the proposed model, Holling's functional response of type III with $n = 2$ is considered as it provides the learning curve of the respective prey species more realistic because it gradually increases the slope and then gradually decreases it also. Due to gradual learning of brain with experience, $n = 2$ curve seems to be realistic. A higher degree $n > 2$ is not reasonable for the natural systems as the learning curve did not seems realistic because $n > 2$ shows that prey species learns at a much faster rate which is not realistic.

4 Mathematical Analysis of Model

In this section, analysis of different systems is done, which are extracted from the predator prey scavenger model (3.1). There are many systems which can be derived, given in Table 2. Out of systems given in Table 2, prey only, predator only and scavenger only are trivial systems.

S. No.	Model	Consideration	Remarks
01.	Prey only model	$y(t) = 0$ and $z(t) = 0$	exist in nature
02.	Predator only model	$x(t) = 0$ and $z(t) = 0$	can't exist in nature
03.	Scavenger only model	$x(t) = 0$ and $y(t) = 0$	can't exist in nature
04.	Predator scavenger model	$x(t) = 0$	exist in nature
05.	Scavenger prey model	$y(t) = 0$	exist in nature
06.	Predator prey model	$z(t) = 0$	exist in nature
07.	Predator prey scavenger model	-	exist in nature

Table 2: Different systems derived from predator prey scavenger model

4.1 Predator Scavenger Model

Scavengers can only feed on some proportion of dead bodies of the predators, and predators can feed on scavengers only. As population of scavengers decreases due to predation, predator population also decreases, resulting in a much faster decrease in scavenger population. Hence, predator scavengers can't exist in nature without prey which can be seen from the analysis of the predator scavengers model,

$$\begin{aligned} \frac{dy}{dt} &= \frac{fz^2y}{1 + i_0z^2} - ey \\ \frac{dz}{dt} &= hyz - \frac{iyz^2}{1 + i_0z^2} - jz \end{aligned} \tag{4.1}$$

The steady state points are, $(0, 0)$ and $\left(z_0, \frac{fjz_0}{hfz_0 - ie}\right)$ where, $z_0 = \left[\frac{e}{f - i_0e}\right]^{1/2}$.

1. ***Steady state*** $(0, 0)$,
   The jacobian matrix for this case is,

$$J = \begin{bmatrix} -e & 0 \\ 0 & -j \end{bmatrix}$$

   implies eigenvalues are $-e$ and $-j$, which means steady state $(0, 0)$ is stable as both eigenvalues are negative.

2. ***Steady state*** $\left(z_0, \frac{fjz_0}{hfz_0-ie}\right)$,
The existence conditions for this steady state point are $f - i_0e > 0$ and $hfz_0 - ie > 0$. The jacobian matrix for predator scavenger model for this steady point case is,

$$J_{yz} = \begin{bmatrix} 0 & \frac{2je(f-i_0e)}{hfz_0-ie} \\ hz_0 - \frac{ie}{f} & \frac{jei(2i_0e-f)}{f(hfz_0-ie)} \end{bmatrix}$$

Hence, the characteristic equation is given as,

$$m(\lambda) = \lambda^2 - \lambda\left[\frac{jei(2i_0e-f)}{f(hfz_0-ie)}\right] - \frac{2je(f-i_0e)}{f} = 0$$

From the above equation, the eigenvalues (λ_1 and λ_2) gives,

$$\lambda_1\lambda_2 = -\frac{2je(f-i_0e)}{f} < 0$$

According to Routh Hurwitz criterion [27], steady state $\left(z_0, \frac{fjz_0}{hfz_0-ie}\right)$ is unstable.

4.2 Predator Prey Model

Predator prey without scavengers can exist in nature which can be seen from the analysis of the predator prey model,

$$\begin{aligned} \frac{dx}{dt} &= rx\left(1-\frac{x}{k}\right) - \frac{ax^2y}{1+a_0x^2} \\ \frac{dy}{dt} &= \frac{dx^2y}{1+a_0x^2} - ey \end{aligned} \tag{4.2}$$

The steady state points are $(0,0)$, $(k,0)$ and $\left(x_0, \frac{drx_0}{aek}(k-x_0)\right)$ where, $x_0 = \left[\frac{e}{d-a_0e}\right]^{1/2}$.

1. ***Steady state*** $(0,0)$,
The jacobian matrix for this case is,

$$J = \begin{bmatrix} r & 0 \\ 0 & -e \end{bmatrix}$$

that is eigenvalues are r and $-e$, which means $(0,0)$ is an unstable steady state point as one of the eigenvalue is positive.

2. ***Steady state*** $(k,0)$,
The jacobian matrix for this case is,

$$J_x = \begin{bmatrix} -r & \frac{-ak^2}{1+a_0k^2} \\ 0 & \frac{dk^2}{1+a_0k^2} - e \end{bmatrix}$$

that is this steady state point is stable if $k^2 < \frac{e}{d-a_0e}$.

3. ***Steady state*** $\left(x_0, \frac{drx_0}{aek}(k-x_0)\right)$,
The existence condition for this steady state point is $d - a_0e > 0$. The jacobian matrix for this case is,

$$J_{xy} = \begin{bmatrix} -r + \frac{2ra_0e(k-x_0)}{dk} & \frac{-ae}{d} \\ \frac{2r(d-a_0e)(k-x_0)}{ak} & 0 \end{bmatrix}$$

The characteristic equation is given as,

$$m(\lambda) := \lambda^2 - \lambda\left[-r + \frac{2ra_0e(k-x_0)}{dk}\right] + \frac{2re(d-a_0e)(k-x_0)}{dk} = 0.$$

Since, population capacity > population at any instantaneous point ($k > x_0$). It is clear from above that $m(0) > 0$ and using Routh Hurwitz criterion [27], $\lambda_1\lambda_2 > 0$. Therefore, either both roots are positive or both negative. For both roots ($\lambda_1 \& \lambda_2$) to be negative, $\lambda_1\lambda_2 > 0$ and $\lambda_1 + \lambda_2 < 0$. Hence,

$$\lambda_1 + \lambda_2 = \frac{2ra_0e(k-x_0)}{dk} - r < 0.$$

So, the steady state $\left(x_0, \frac{drx_0}{aek}(k-x_0)\right)$ is stable if $\frac{2a_0e(k-x_0)}{dk} < 1$.

4.3 Scavenger Prey Model

Scavenger prey without predators can exist in nature (because scavenger prey interaction is similar to predator prey interaction) which can be seen from the analysis of the scavenger prey model,

$$\begin{aligned} \frac{dx}{dt} &= rx\left(1 - \frac{x}{k}\right) - \frac{bx^2z}{1+b_0x^2} \\ \frac{dz}{dt} &= \frac{gx^2z}{1+b_0x^2} - jz \end{aligned} \tag{4.3}$$

The steady state points are $(0,0)$, $(k,0)$ and $\left(x_0, \frac{grx_0}{bjk}(k-x_0)\right)$ where, $x_0 = \left[\frac{j}{g-b_0j}\right]^{1/2}$.

1. ***Steady state*** $(0,0)$,
   The jacobian matrix for this case is,

   $$J = \begin{bmatrix} r & 0 \\ 0 & -j \end{bmatrix}$$

   that is eigenvalues are r and $-j$ which means $(0,0)$ is an unstable steady state point as one of the eigenvalue is positive.

2. ***Steady state*** $(k,0)$,
   The jacobian matrix for this case is,

   $$J_x = \begin{bmatrix} -r & \frac{-bk^2}{1+b_0k^2} \\ 0 & \frac{gk^2}{1+b_0k^2} - j \end{bmatrix}$$

   that is this steady state point is stable if $k^2 < \frac{j}{g-b_0j}$.

3. ***Steady state*** $\left(x_0, \frac{grx_0}{bjk}(k-x_0)\right)$,
   The existence condition for this steady state point is $g - b_0j > 0$. The jacobian matrix for this case is,

   $$J_{xz} = \begin{bmatrix} -r + \frac{2rb_0j(k-x_0)}{gk} & \frac{-bj}{g} \\ \frac{2r(g-b_0j)(k-x_0)}{bk} & 0 \end{bmatrix}$$

   The characteristic equation is given as,

   $$m(\lambda) := \lambda^2 - \lambda\left[-r + \frac{2rb_0j(k-x_0)}{gk}\right] + \frac{2rj(g-b_0j)(k-x_0)}{gk} = 0.$$

Since, population capacity > population at any instantaneous point ($k > x_0$). It is clear from above that $m(0) > 0$ and using Routh Hurwitz criterion [27], $\lambda_1\lambda_2 > 0$. Therefore, either both roots are positive or both negative. For both roots ($\lambda_1 \& \lambda_2$) to be negative, $\lambda_1\lambda_2 > 0$ and $\lambda_1 + \lambda_2 < 0$. Hence,

$$\lambda_1 + \lambda_2 = \frac{2rb_0j(k - x_0)}{gk} - r < 0.$$

So, the steady state $\left(x_0, \frac{grx_0}{bjk}(k - x_0)\right)$ is stable if $\frac{2b_0j(k-x_0)}{gk} < 1$.

4.4 Predator Prey Scavenger Model

Considering all populations in the system viz. prey, predator and scavenger. Predator prey scavengers all can coexist in nature which can be seen from the analysis of the predator prey scavenger model,

$$\begin{aligned}
\frac{dx}{dt} &= rx\left(1 - \frac{x}{k}\right) - \frac{ax^2y}{1 + a_0x^2} - \frac{bx^2z}{1 + b_0x^2} \\
\frac{dy}{dt} &= \frac{dx^2y}{1 + a_0x^2} + \frac{fz^2y}{1 + i_0z^2} - ey \\
\frac{dz}{dt} &= \frac{gx^2z}{1 + b_0x^2} + hyz - \frac{iyz^2}{1 + i_0z^2} - jz
\end{aligned} \tag{4.4}$$

Let steady states for the system be denoted by $(0, 0, 0)$, $(E_x^{[2]}, 0, 0)$, $(0, E_y^{[3]}, E_z^{[3]})$, $(E_x^{[4]}, E_y^{[4]}, 0)$, $(E_x^{[5]}, 0, E_z^{[5]})$ and $(E_x^{[6]}, E_y^{[6]}, E_z^{[6]})$. It is clear from Table 2, the cases of $(0, E_y, 0)$ and $(0, 0, E_z)$ are included in $(0, 0, 0)$ case as both yields same stable steady state point $(0, 0, 0)$.

1. ***Steady state*** $(0, 0, 0)$,
   The jacobian matrix for this case is,

$$J_0 = \begin{bmatrix} r & 0 & 0 \\ 0 & -e & 0 \\ 0 & 0 & -j \end{bmatrix}$$

   that is eigenvalues are $r, -e$ and $-j$, which means $(0, 0, 0)$ is an unstable steady state point as one of the eigenvalue is positive.

2. ***Steady state*** $(E_x^{[2]}, 0, 0) = (k, 0, 0)$,
   The jacobian matrix for this case is,

$$J_x = \begin{bmatrix} -r & \frac{-ak^2}{1+a_0k^2} & \frac{-bk^2}{1+b_0k^2} \\ 0 & \frac{dk^2}{1+a_0k^2} - e & 0 \\ 0 & 0 & \frac{gk^2}{1+b_0k^2} - j \end{bmatrix}$$

   that is steady state $(E_x^{[2]}, 0, 0)$ is stable if $k^2 < \frac{e}{d - a_0e}$ and $k^2 < \frac{j}{g - jb_0}$.

3. ***Steady state*** $(0, E_y^{[3]}, E_z^{[3]}) = \left(0, \frac{fjz_0}{hfz_0 - ie}, z_0\right)$ where, $z_0 = \left[\frac{e}{f - i_0e}\right]^{1/2}$
   The existence conditions of this steady state are $f - i_0e > 0$ and $hfz_0 - ie > 0$. The jacobian matrix for this case is,

$$J_{yz} = \begin{bmatrix} r & 0 & 0 \\ 0 & 0 & \frac{2je(f - i_0e)}{hfz_0 - ie} \\ 0 & hz_0 - \frac{ie}{f} & \frac{jei(2i_0e - f)}{f(hfz_0 - ie)} \end{bmatrix}$$

which means steady state $(0, E_y^{[3]}, E_z^{[3]})$ is unstable as one of the eigenvalues is r which is positive.

4. ***Steady state*** $(E_x^{[4]}, E_y^{[4]}, 0) = \left(x_0, \frac{drx_0}{aek}(k - x_0), 0\right)$ where, $x_0 = \left[\frac{e}{d-a_0e}\right]^{1/2}$
The existence condition of this steady state is $d - a_0e > 0$. The jacobian matrix for this case is,

$$J_{xy} = \begin{bmatrix} -r + \frac{2ra_0e(k-x_0)}{dk} & \frac{-ae}{d} & \frac{-be}{d+(b_0-a_0)e} \\ \frac{2r(d-a_0e)(k-x_0)}{ak} & 0 & 0 \\ 0 & 0 & \frac{ge}{d+(b_0-a_0)e} + \frac{hdrx_0(k-x_0)}{aek} - j \end{bmatrix}$$

This steady state is stable if,

1. $\frac{2a_0e(k-x_0)}{dk} < 1$ (from Section 4.2)

2. $hdrx_0\left(1 - \frac{x_0}{k}\right)[d + (b_0 - a_0)e] + ae[ge - j(d + (b_0 - a_0)e)] < 0$

As condition 2. corresponds to one of the eigenvalues, hence must be negative for the system to be stable.

5. ***Steady state*** $(E_x^{[5]}, 0, E_z^{[5]}) = \left(x_0, 0, \frac{grx_0}{bjk}(k - x_0)\right)$ where, $x_0 = \left[\frac{j}{g-jb_0}\right]^{1/2}$
The existence condition of this steady state is $g - b_0j > 0$. The jacobian matrix for this case is,

$$J_{xz} = \begin{bmatrix} -r + \frac{2rb_0j(k-x_0)}{gk} & \frac{-aj}{g+j(a_0-b_0)} & \frac{-bj}{g} \\ 0 & \frac{dj}{g+j(a_0-b_0)} + \frac{fz_0^2}{1+i_0z_0^2} - e & 0 \\ \frac{2r(g-jb_0)(k-x_0)}{bk} & \frac{hgrx_0(k-x_0)}{bjk} - \frac{iz_0^2}{1+i_0z_0^2} & 0 \end{bmatrix}$$

where, $z_0 = \frac{grx_0}{bjk}(k - x_0)$. This steady state is stable if,

1. $\frac{2b_0j(k-x_0)}{gk} < 1$ (from Section 4.3)

2. $\frac{fz_0^2}{1+i_0z_0^2} < \frac{e(g+j(a_0-b_0))-dj}{g+j(a_0-b_0)} \quad (t(\text{say})) \Rightarrow \quad z_0^2 < \frac{t}{f-i_0t}$

As condition 2. corresponds to one of the eigenvalues, hence must be negative for the system to be stable.

6. ***Steady state*** $(E_x^{[6]}, E_y^{[6]}, E_z^{[6]}) = (x^*, y^*, z^*)$
where x^* is the only positive real root of equation (see Appendix for values of p_i),

$$\sum_{i=0}^{12} p_i X^i = 0 \tag{4.5}$$

The values of y^* and z^* are given as,

$$z^{*2} = \frac{e + (a_0e - d)x^{*^2}}{f(1 + a_0x^{*^2}) - i_0\left(e + (a_0e - d)x^{*^2}\right)}$$

$$y^* = \left(j - \frac{gx^{*^2}}{1 + b_0x^{*^2}}\right)\left(\frac{1 + i_0z^{*^2}}{h + i_0z^{*^2} - iz^*}\right)$$

The conditions for existence of this steady state are,

1. x^* must be only positive real root of equation $\sum_{i=0}^{12} p_i X^i = 0$

2. $z^* < min\left\{\frac{r(1+b_0{x^*}^2)(k-x^*)}{bx^*}, \frac{ie}{hf}\right\}$

The jacobian matrix for this case is,

$$J_{xyz} = \begin{bmatrix} r(1-\frac{2x^*}{k}) - \frac{2ax^*y^*}{(1+a_0{x^*}^2)^2} - \frac{2bx^*z^*}{(1+b_0{x^*}^2)^2} & \frac{-a{x^*}^2}{1+a_0{x^*}^2} & \frac{-b{x^*}^2}{1+b_0{x^*}^2} \\ \frac{2dx^*y^*}{(1+a_0{x^*}^2)^2} & 0 & \frac{2fz^*y^*}{(1+i_0{z^*}^2)^2} \\ \frac{2gx^*z^*}{(1+b_0{x^*}^2)^2} & hz^* - \frac{i{z^*}^2}{1+i_0{z^*}^2} & \frac{iy^*z^*}{1+i_0{z^*}^2} - \frac{2iy^*z^*}{(1+i_0{z^*}^2)^2} \end{bmatrix}$$

Considering the above matrix as,

$$J = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Let's assume some variables as,

1. $m_1 = -trace(J)$

2. $m_2 = M_{11} + M_{22} + M_{33}$ where, M_{ii} is the minor of a_{ii}^{th} element, for $i = 1, 2, 3$.

3. $m_3 = -determinant(J)$

These variables are such that the characteristic equation of matrix J is, $\lambda^3 + m_1\lambda^2 + m_2\lambda + m_3 = 0$. According to Routh Hurwitz criterion [27], steady state $(E_x^{[6]}, E_y^{[6]}, E_z^{[6]})$ stable if,

1. $m_i > 0$ where, $i = 1, 2, 3$

2. $m_1m_2 - m_3 > 0$

5 Numerical Simulations

Example 5.1. *Consider predator scavenger system with parameter values as* $f = 0.5, i_0 = 0.25, e = 0.5, h = 0.5, i = 0.5$ *and* $j = 1.5$*, with initial conditions* $y(0) = 4$ *and* $z(0) = 6$*.*

$$\frac{dy}{dt} = \frac{0.5z^2y}{1 + 0.25z^2} - 0.5y$$
$$\frac{dz}{dt} = 0.5yz - \frac{0.5yz^2}{1 + 0.25z^2} - 1.5z$$

For the predator scavenger system, it is clear from Fig. 3 that both populations stabilize at point $(0, 0)$ in the system and converge to zero. Here,

$$f - i_0e = 0.375 > 0$$

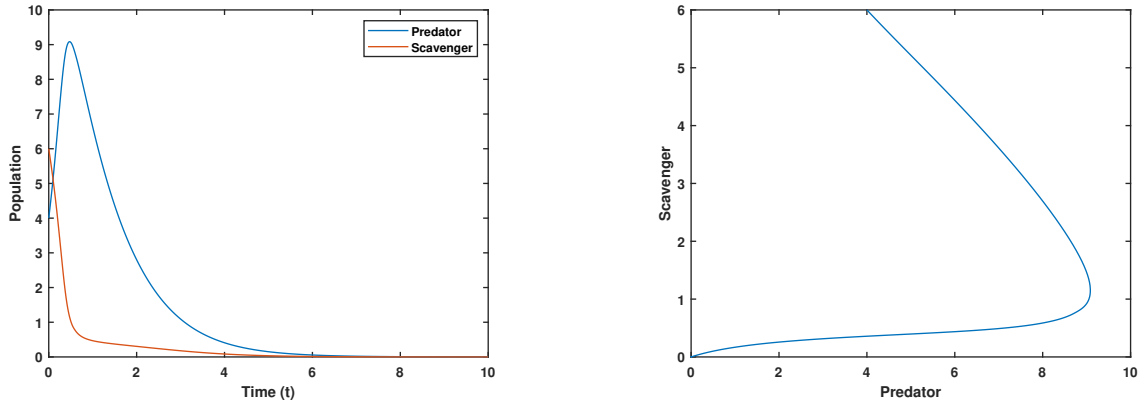
$$hf\left(\frac{e}{f - i_0e}\right)^{1/2} - ie = \frac{2 - \sqrt{3}}{4\sqrt{3}} = 0.038675 > 0$$

that means the existence condition is satisfied for the equilibrium points. Here,

$$z_0 = \left[\frac{e}{f - i_0e}\right]^{1/2} = \frac{2}{\sqrt{3}}$$

and hence, equilibrium points are $(0, 0)$ and $(1.1547, 22.3923)$,

$$\left(z_0, \frac{fjz_0}{hfz_0 - ie}\right) = \left(\frac{2}{\sqrt{3}}, 6(2 + \sqrt{3})\right) = (1.1547, 22.3923)$$



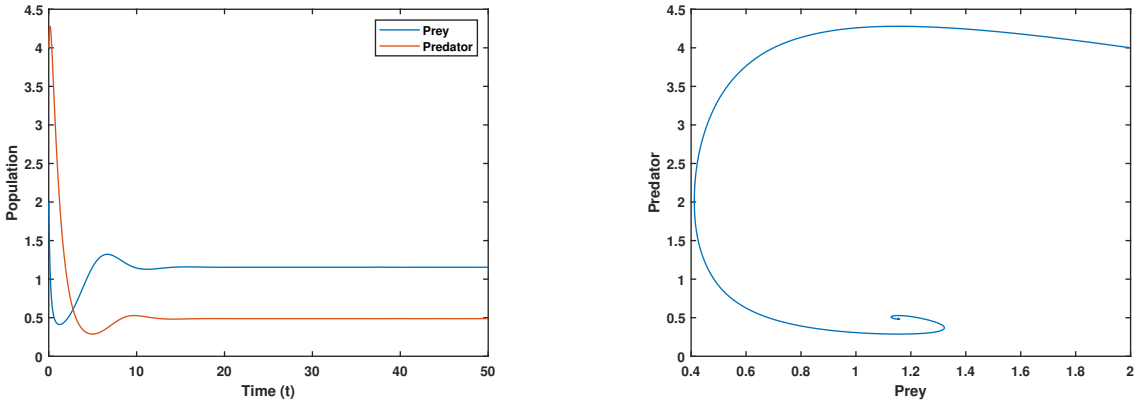
(a) Dynamical behaviour. (b) Phase portrait.

Fig. 3: Predator scavenger model for Example 5.1.

According to subsection 4.1, the equilibrium point $(0, 0)$ is stable, and the equilibrium point $(1.1547, 22.3923)$ is unstable. It is clear from the Fig. 3 that both the scavenger population as well as predator population converge to zero. The stability of $(0, 0)$ can be explained easily as the only source of food for the predators is scavengers, and the only source of food for the scavengers population are the dead bodies of the predators. Initially, there are more scavengers but their natural death rate is greater, so their population decreases and the population of predators increases. As the population of scavengers decreases, food for the predators starts to decline as well. Hence, predator population also starts falling. As a result, both populations stabilizes at $(0, 0)$ or, in other words both populations extinct.

Example 5.2. *Consider predator prey system with parameter values as* $r = 1$, $k = 2$, $a = 1$, $a_0 = 0.25$, $d = 1$ *and* $e = 1$ *with initial conditions* $x(0) = 2$ *and* $y(0) = 4$.

$$\frac{dx}{dt} = x\left(1 - \frac{x}{2}\right) - \frac{x^2 y}{1 + 0.25x^2}$$
$$\frac{dy}{dt} = \frac{x^2 y}{1 + 0.25x^2} - y$$



(a) Dynamical behaviour. (b) Phase portrait.

Fig. 4: Predator prey model for Example 5.2.

For predator prey system, it is clear from Fig. 4 that both of the populations stabilize at some specific value of the population which is neither $(0, 0)$ nor $(k, 0) = (2, 0)$. Here, $d - a_0 e = 0.75 > 0$

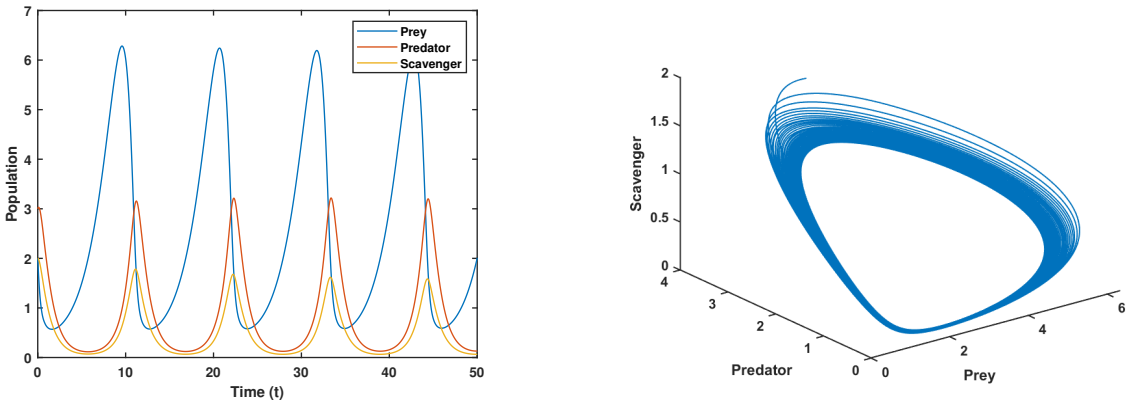
that means existence condition is satisfied for the equilibrium points. Here, $x_0 = \left[\frac{e}{d-ea_0}\right]^{1/2} = \frac{2}{\sqrt{3}}$ and hence, equilibrium points are $(0,0)$, $(2,0)$ and $(1.1547, 0.4880)$,

$$\left(x_0, \frac{drx_0}{aek}(k - x_0)\right) = (1.1547, 0.4880)$$

As discussed in the predator prey model subsection 4.2, the equilibrium point $(0,0)$ is unstable. For the equilibrium point $(2,0)$, $k^2 = 4$ and $\frac{e}{d-ea_0} = \frac{4}{3}$. As $k^2 > \frac{e}{d-ea_0}$, so equilibrium point $(2,0)$ is not stable because if this point gets stabilize then it means that only prey exists in nature and predators extinct which is not possible. For the equilibrium point $(1.1547, 0.4880)$, $\frac{2a_0e(k-x_0)}{dk} = 0.211325 < 1$ which satisfies the stability condition for the equilibrium point $(1.1547, 0.4880)$. Hence, the system stabilizes and both predator and prey populations coexist in this system.
Remarks: Similarly, the existence of solution and the stability condition for the scavenger prey model can be determined. In scavenger prey model, scavengers will show the same behaviour as the predator for the prey.

Example 5.3. *Consider predator prey scavenger system with parameter values such as* $r = 0.5$, $k = 100$, $a = 0.5$, $a_0 = 0.25$, $b = 0.5$, $b_0 = 0.25$, $d = 0.5$, $e = 1$, $f = 0.1$, $g = 0.5$, $h = 0.1$, $i = 0.1$, $i_0 = 0.25$ *and* $j = 1$ *with initial conditions* $x(0) = 4$, $y(0) = 3$ *and* $z(0) = 2$.

$$\begin{aligned}
\frac{dx}{dt} &= 0.5x\left(1 - \frac{x}{100}\right) - \frac{0.5x^2y}{1 + 0.25x^2} - \frac{0.5x^2z}{1 + 0.25x^2} \\
\frac{dy}{dt} &= \frac{0.5x^2y}{1 + 0.25x^2} + \frac{0.1z^2y}{1 + 0.25z^2} - y \\
\frac{dz}{dt} &= \frac{0.5x^2z}{1 + 0.25x^2} + 0.1yz - \frac{0.1yz^2}{1 + 0.25z^2} - z
\end{aligned}$$



(a) Dynamical behaviour. (b) Phase portrait.

Fig. 5: Predator prey scavenger model for Example 5.3.

For predator prey scavenger system, it is clear from Fig. 5 that all of the population don't stabilize at any value of the population.
Equilibrium point $(0,0,0)$: It is clear that $(0,0,0)$ is an unstable equilibrium point because if this point becomes stable, then the system collapse. As described in subsection 4.4, $(0, E_y, 0)$ and $(0,0,E_z)$ cases are also included in this $(0,0,0)$ case as both yields equilibrium point $(0,0,0)$ and both are not stable as well.
Equilibrium point $(E_x^{[2]}, 0, 0)$: For equilibrium point $(100,0,0)$,

$$\frac{e}{d - ea_0} = 4 < 100^2 = k^2 \quad \text{and} \quad \frac{j}{g - jb_0} = 4 < 100^2 = k^2$$

Hence, the equilibrium point $(100, 0, 0)$ is not stable.

Equilibrium point $(0, E_y^{[3]}, E_z^{[3]})$: As, $f - i_0 e = -0.15 < 0$ didn't satisfy the existence condition for the equilibrium point $\left(0, \frac{fjz_0}{hfz_0 - ie}, z_0\right)$ because equilibrium point turns out to have a negative or, complex numbers which has no significance in nature. Hence, equilibrium point $(0, E_y^{[3]}, E_z^{[3]})$ does not exist in the system.

Equilibrium point $(E_x^{[4]}, E_y^{[4]}, 0)$: As $d - a_0 e = 0.25 > 0$ implies the existence condition for the equilibrium point,

$$\left(\left[\frac{e}{d - a_0 e}\right]^{1/2}, \frac{drx_0(k - x_0)}{aek}, 0\right) = (2, 0.98, 0)$$

For this equilibrium point,

$$\frac{2a_0 e(k - x_0)}{dk} = 0.98 < 1$$

$$hrdx_0\left(1 - \frac{x_0}{k}\right)(d + (b_0 - a_0)e) + ae(ge - j(d + (b_0 - a_0)e)) = 0.0254 > 0$$

Hence, the stability conditions are not met which means predator and prey may coexist in an unstable system.

Equilibrium point $(E_x^{[5]}, 0, E_z^{[5]})$: As $g - b_0 j = 0.25 > 0$ implies the existence condition for the equilibrium point,

$$\left(\left[\frac{j}{g - b_0 j}\right]^{1/2}, 0, \frac{grx_0(k - x_0)}{bjk}\right) = (2, 0, 0.98)$$

For this equilibrium point,

$$\frac{2b_0 j(k - x_0)}{gk} = 0.98 < 1$$

$$t = \frac{e(g + j(a_0 - b_0)) - dj}{g + j(a_0 - b_0)} = 0$$

and hence $z_0^2 = 0.9604 > \frac{t}{f - i_0 t} = 0$. Therefore, the stability condition doesn't satisfy which means scavenger and prey may coexist in an unstable system.

Equilibrium point $(E_x^{[6]}, E_y^{[6]}, E_z^{[6]})$: Value of $E_x^{[6]}$ is given by the positive real root of the equation given by,

$X^8 - 200X^7 + 9770.73X^6 + 23631.746X^5 - 56491.68254X^4 - 130234.92X^3 + 110210X^2 + 154819X - 121904.7619 = 0$

which means equilibrium point is $(1.951629, 0.470157934, 0.510619637)$. Here, $x^* = 1.951629$, which is only positive real root of equation above and $z^* = 0.510619637 < \min\{10, 98.0778\}$, that means existence condition is satisfied for the equilibrium point in which the coexistence of predator, prey and scavenger are present. Putting these values, the jacobian matrix corresponding to this equilibrium point turns out to be,

$$J_{xyz} = \begin{bmatrix} -0.022 & -0.98 & -0.98 \\ 0.24 & 0 & 0.042 \\ 0.26 & 0.027 & -0.02 \end{bmatrix}$$

For this matrix characteristic polynomial becomes,

$$\lambda^3 + 0.041538\lambda^2 + 0.4892535\lambda + 0.02166$$

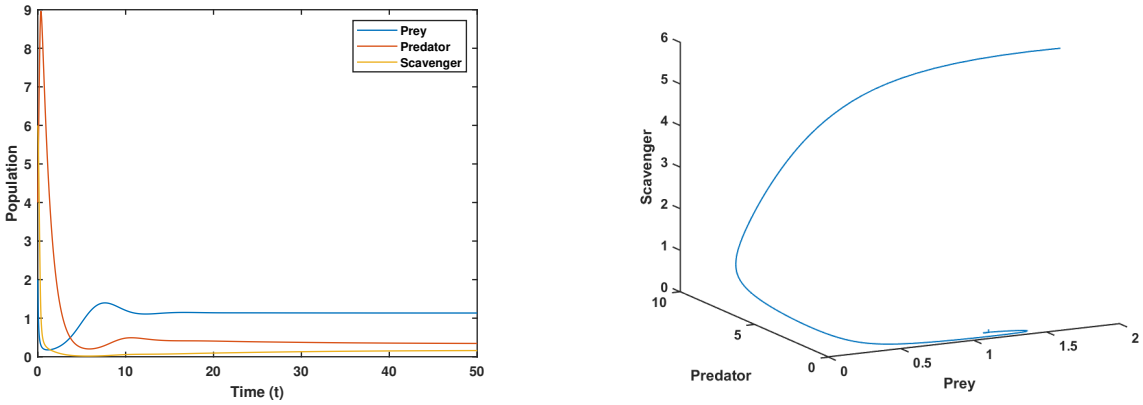
For this characteristic polynomial, $m_1 = 0.041538, m_2 = 0.4892535, m_3 = 0.02166$ which means $m_1, m_2, m_3 > 0$ and $m_1 m_2 - m_3 = -0.001337 < 0$. That means the equilibrium point $(1.95, 0.47, 0.51)$ is unstable, but all three populations can coexist in the system.

As it is clear from Fig. 5 that all the populations coexist in the system as when the population of prey decreases, the food for scavengers and predators decreases. Hence, the population of predators and scavengers also decreases. But as the predator and scavengers can also rely on each other, their population decreases slower due to interrelation. As prey population decreases, scavenger population decreases, hence the predator population. As the predator population decreases the scavenger population starts decreasing much faster. As the population of predators and scavengers starts decreasing, prey population starts increasing.
Remarks: Results for Example 5.3 are as follows,

1. Predator prey can coexist in the system in the absence of scavengers (Unstable).
2. Scavenger prey can coexist in the system in the absence of predators (Unstable).
3. Predator scavengers can't coexist in the system in the absence of prey.
4. Predator prey scavengers can all coexist in the system (Unstable).

Example 5.4. *Consider predator prey scavenger system with parameter values such as* $r = 1$, $k = 2$, $a = 1$, $a_0 = 0.25$, $b = 1$, $b_0 = 0.25$, $d = 1$, $e = 1$, $f = 1$, $g = 1$, $h = 0.25$, $i = 1$, $i_0 = 0.25$ *and* $j = 1$, *with initial conditions* $x(0) = 4$, $y(0) = 3$ *and* $z(0) = 2$.

$$\begin{aligned}
\frac{dx}{dt} &= x\left(1 - \frac{x}{2}\right) - \frac{x^2 y}{1 + 0.25x^2} - \frac{x^2 z}{1 + 0.25x^2} \\
\frac{dy}{dt} &= \frac{x^2 y}{1 + 0.25x^2} + \frac{z^2 y}{1 + 0.25z^2} - y \\
\frac{dz}{dt} &= \frac{x^2 z}{1 + 0.25x^2} + 0.25yz - \frac{yz^2}{1 + 0.25z^2} - z
\end{aligned}$$



(a) Dynamical behaviour. (b) Phase portrait.

Fig. 6: Predator prey scavenger model for Example 5.4.

For predator prey scavenger system, it is clear from Fig. 6 that all of the population gets stabilize at specific population values.
Equilibrium point $(0, 0, 0)$: It is clear that $(0, 0, 0)$ is an unstable equilibrium point because if this point becomes stable, then the system collapse. As described in subsection 4.4, $(0, E_y, 0)$ and $(0, 0, E_z)$ cases are also included in this $(0, 0, 0)$ case as both yields equilibrium point $(0, 0, 0)$ and both are not stable as well.
Equilibrium point $(E_x^{[2]}, 0, 0)$: For equilibrium point $(2, 0, 0)$,

$$\frac{e}{d - ea_0} = \frac{4}{3} < 2^2 = k^2 \quad \text{and} \quad \frac{j}{g - jb_0} = \frac{4}{3} < 2^2 = k^2$$

Hence, equilibrium point $(2, 0, 0)$ is not stable.

Equilibrium point $(0, E_y^{[3]}, E_z^{[3]})$: As, $f - i_0 e = 0.75 > 0$ and

$$hf\left[\frac{e}{f - i_0 e}\right]^{1/2} - ie = \frac{1 - 2\sqrt{3}}{2\sqrt{3}} = -0.7113248654 < 0$$

implies the existence conditions are not satisfied for the equilibrium point $\left(0, \frac{fjz_0}{hfz_0 - ie}, z_0\right)$ as the equilibrium point turns out to have a negative number, which has no significance in nature. Hence, equilibrium point $(0, E_y^{[3]}, E_z^{[3]})$ does not exist in the system.

Equilibrium point $(E_x^{[4]}, E_y^{[4]}, 0)$: As, $d - a_0 e = 0.75 > 0$ implies the existence condition for the equilibrium point,

$$\left(\left[\frac{e}{d - a_0 e}\right]^{1/2}, \frac{drx_0(k - x_0)}{aek}, 0\right) = \left(\frac{2}{\sqrt{3}}, \frac{2(\sqrt{3} - 1)}{3}, 0\right) = (1.1547, 0.4880, 0)$$

For this equilibrium point,

$$\frac{2a_0 e(k - x_0)}{dk} = 0.211325 < 1$$

$$hrdx_0\left(1 - \frac{x_0}{k}\right)[d + (b_0 - a_0)e] + ae[ge - j[d + (b_0 - a_0)e]] = 0.122 > 0$$

Hence, the stability condition are not met which means predator and prey may coexist in an unstable system.

Equilibrium point $(E_x^{[5]}, 0, E_z^{[5]})$: As, $g - b_0 j = 0.75 > 0$ implies the existence condition for the equilibrium point

$$\left(\left[\frac{j}{g - b_0 j}\right]^{1/2}, 0, \frac{grx_0(k - x_0)}{bjk}\right) = \left(\frac{2}{\sqrt{3}}, 0, \frac{2(\sqrt{3} - 1)}{3}\right) = (1.1547, 0, 0.4880)$$

For this equilibrium point,

$$\frac{2b_0 j(k - x_0)}{gk} = 0.211325 < 1$$

$$t = \frac{e(g + j(a_0 - b_0)) - dj}{g + j(a_0 - b_0)} = 0$$

and hence $z_0^2 = 0.238177 > \frac{t}{f - i_0 t} = 0$. Therefore, the stability condition doesn't satisfy that means scavenger and prey may coexist in an unstable system.

Equilibrium point $(E_x^{[6]}, E_y^{[6]}, E_z^{[6]})$: Value of $E_x^{[6]}$ is given by the positive real root of the equation given by,

$X^8 - 4X^7 + 3.926X^6 - 4.221845X^5 + 7.906X^4 + 6.77647X^3 - 5.163X^2 + 3.87227X - 17.19832 = 0$ which means equilibrium point is $(1.1331137, 0.33726859, 0.168040476)$. Here, $x^* = 1.1331137$, which is only positive real root of the equation above and $z^* = 0.168040476 < \min\{4, 1.01062\}$, that means the existence condition is satisfied for the equilibrium point in which the coexistence of predator, prey and scavenger are present. Putting these values, the jacobian matrix corresponding to this equilibrium point turns out to be,

$$J_{xyz} = \begin{bmatrix} -0.79 & -0.97 & -0.97 \\ 0.44 & 0 & 0.11 \\ 0.22 & 0.014 & -0.055 \end{bmatrix}$$

For this matrix characteristic polynomial becomes,

$$\lambda^3 + 0.8448437\lambda^2 + 0.68008\lambda + 0.052045$$

For this characteristic polynomial, $m_1 = 0.845, m_2 = 0.68, m_3 = 0.052$ which means $m_1, m_2, m_3 > 0$ and $m_1 m_2 - m_3 = 0.5225 > 0$. That means the equilibrium point $(1.133, 0.337, 0.168)$ is stable. Hence, all three populations can coexist in the system. As it is clear from Fig. 6 that all the populations coexist in the system. When the population of prey decreases, the food for scavengers and predators decreases, and hence the population of predators and scavengers also decreases. As the prey population decreases, the scavenger population also decreases and, hence the predator population. As the predator population decreases the scavenger population starts decreasing much faster. As the population of predators and scavengers starts decreasing, prey population starts increasing. Hence, the system gets stabilize, and all the population (prey, predator and scavengers) coexist in this system.
Remarks: Results for Example 5.4 are as follows,

1. Predator prey can coexist in the system in the absence of scavengers (Unstable).
2. Scavenger prey can coexist in the system in the absence of predators (Unstable).
3. Predator scavengers can't coexist in the system in the absence of prey.
4. Predator prey scavenger can all coexist in the system (Stable).

6 Parameter Estimation

Parameter estimation require proper dataset for the formulation and preparation of the model for the same. In this study population of various species from the American wildlife forest dataset [16]. Two step model for parameter estimation is done in this study with first step consists of physics-informed deep neural network using adam backpropagation defined in section 2.2 and second step consist of fine-tune optimization of parameters using BFGS optimization algorihtm defined in section 2.3.

6.1 Problem Formulation

Fig. 7 represents the dataset used in this study. Dataset consists of population of various species across 40 years (1970-2009). Hence, initial assumptions for categorization of population species in the dataset and parameters of the proposed model (3.1) are as follows,

1. Apex predators are assumed as predators.
2. Scavengers are rotten dead body eater species.
3. Rest of the species are taken as prey.
4. All the parameters must have positive values.

After applying these assumptions to the dataset, the dataset is then normalized using min-max scaling over independent parameter time (years) and numerical population of different species (predator, prey and scavenger).

Consider a dynamical system of n ordinary differential equation with m unknown parameters,

$$\begin{aligned} \frac{dX(t)}{dt} &= f(t, X(t), p), \qquad t \in [t_0, t_{end}] \subset \mathcal{R}, \\ X(t_0) &= X_0, \qquad X_0 \in \mathcal{R}^n, \end{aligned} \tag{6.1}$$

where $X(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T \in \mathcal{R}^n$ is population vector, $p = [p_1, p_2, \ldots, p_m] \in \mathcal{R}^m$ is parameter vector, and f is a vector function defined by,

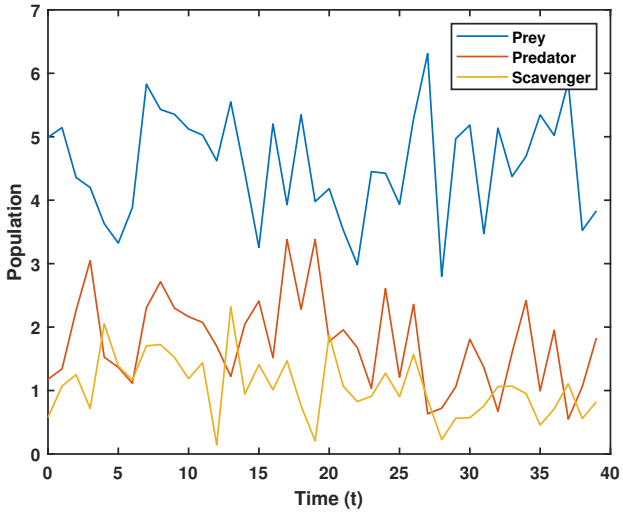


Fig. 7: American forest dataset [16].

$$f(t, X(t), p) = \begin{bmatrix} f_1(t, x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m) \\ f_2(t, x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m) \\ \cdots \\ f_n(t, x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m) \end{bmatrix}$$

Now in proposed model (3.1), number of population variables is $n = 3$ and number of parameters is $m = 14$. Here, the problem becomes finding the unknown vector p using other factors such as, physics-informed error (PIE), mean square error (MSE), and given data which will be explained in next subsections.

6.2 Physics-Informed Deep Neural Network Model

Fig. 8 represents the physics-informed deep neural network model used for estimating the parameters of a dynamical system (3.1). First, initialize the parameters of the dynamical system normally. Then random parameters are input into the neural network with the Adam backpropagation algorithm, and the objective function consists of MSE and PIE. Initialize the optimization process of neural network and update the weights and biases up to specific epoch limits using the dataset provided. The hence-found parameters are then input in the BFGS optimization algorithm for fine-tuning the parameter values using the objective function as MSE. The hence-found parameter values are the final predicted values for the parameters using the dataset.

The physics-informed neural network (PINN) uses physical laws of differential equations into consideration of the loss function. Initialize the parameters $p_{init} = [p_1, p_2, \ldots, p_{14}] \in \mathcal{R}^{14}$ using lognormal distribution that is,

$$\log(p_i) \sim \mathcal{N}(\mu, \sigma^2) \qquad \text{where, } i \in [1, 2, \ldots, 14], \qquad \text{with } \mu = 0, \text{ and } \sigma = 1.$$

Considering a neural network with a backpropagation algorithm is adapted as Adam algorithm defined in subsection 2.2, the required neural network can be presented as,

$$NN^{\Theta} : \mathcal{R}^{14} \to \mathcal{R}^{14} \qquad \text{where, trainable parameters, } \Theta = \{W, b\}$$

where, $\{W, b\}$ is a collection of trainable parameters, namely weights, and biases. The neural network defined above is trained to minimize the combined error (Mean Square Error

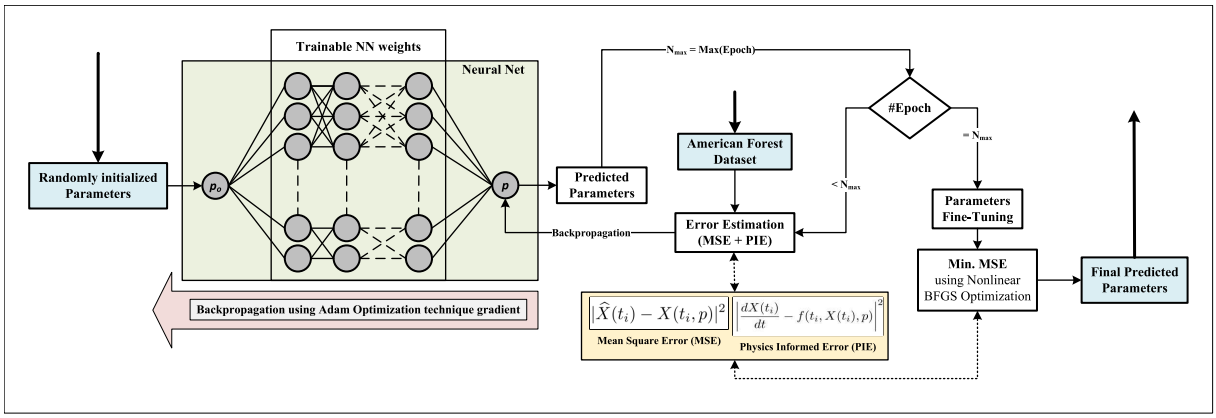


Fig. 8: Flowchart for physics-informed deep neural network model.

and Physics-Informed Error) and predicts the parameters. Now, the objective function or, loss function (for minimization) for which PINN optimization is required can be written as,

$$arg\min(MSE + PIE) \tag{6.2}$$

where MSE is the mean square error term and PIE is the physics-informed error term. Let p be the predicted parameter vector, $T_d = 40$ be the total data rows available, $\widehat{X}(t_i)$ be the true population values at time t_i and $X(t_i, p)$ be the predicted population value at time t_i using predicted parameters p. Now, the overall error term can be represented as,

$$E(p) = \frac{1}{|T_d|} \sum_{i \in T_d} \left(\left| \widehat{X}(t_i) - X(t_i, p) \right|^2 + \left| \frac{dX(t_i)}{dt} - f(t_i, X(t_i), p) \right|^2 \right) \tag{6.3}$$

where, $\left|\widehat{X}(t_i) - X(t_i, p)\right|^2$ is the mean square error and $\left|\frac{dX(t_i)}{dt} - f(t_i, X(t_i), p)\right|^2$ is the physics-informed error. Combined error is normalized using the total number of data values T_d.

The neural network structure used in this research is as follows, the input layer requires the input shape of vector to be $(14,)$, then 3 hidden layers are implemented, each with 32 neurons along with swish activation function. There are 14 neurons in the output layer with Rectified Linear Unit (ReLU) activation function and the output shape of the predicted parameter vector is the same as input vector $(14,)$. The maximum epoch limit for the neural network is set to 100 epochs.

After the neural network, the predicted parameters are then fine-tuned using the BFGS optimization algorithm defined in subsection 2.3 with objective function as,

$$arg\min(MSE) \tag{6.4}$$

where MSE is defined the same as in equation (6.3). That is, MSE $= \left|\widehat{X}(t_i) - X(t_i, p)\right|^2$. The maximum iteration limit for the BFGS optimization algorithm is 200 and threshold movement is also defined for BFGS optimization in physics physics-informed deep neural network model given by Fig. 8. Resulted parameters are final predicted parameters obtained from the dataset.

6.3 Error Analysis and Results

Fig. 9 represents the total error (MSE + PIE) propagation over iterations. Physics-informed neural network with Adam backpropagation runs for 100 epochs with initial total error 3.128 which propagates through epochs and final error after neural net optimizes to 0.644. After the neural net, the parameters are input into the BFGS optimization algorithm for fine-tuning. Now,

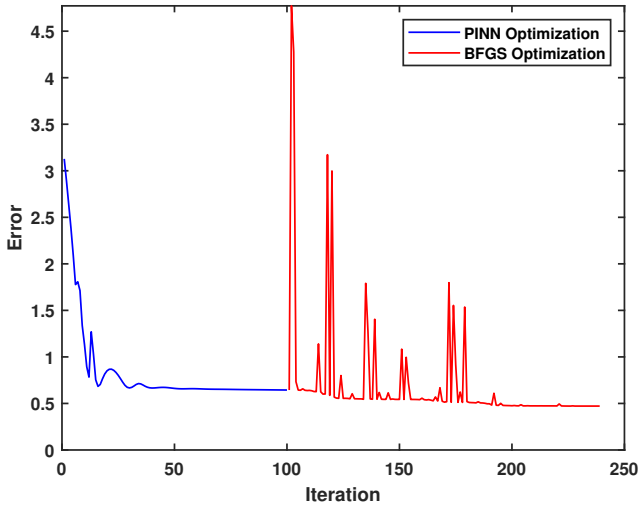


Fig. 9: Error analysis: Total error propagation.

the initial total error for BFGS 0.644 optimizes to 0.472. The final parameters are predicted parameters for model (3.1) using a natural dataset.

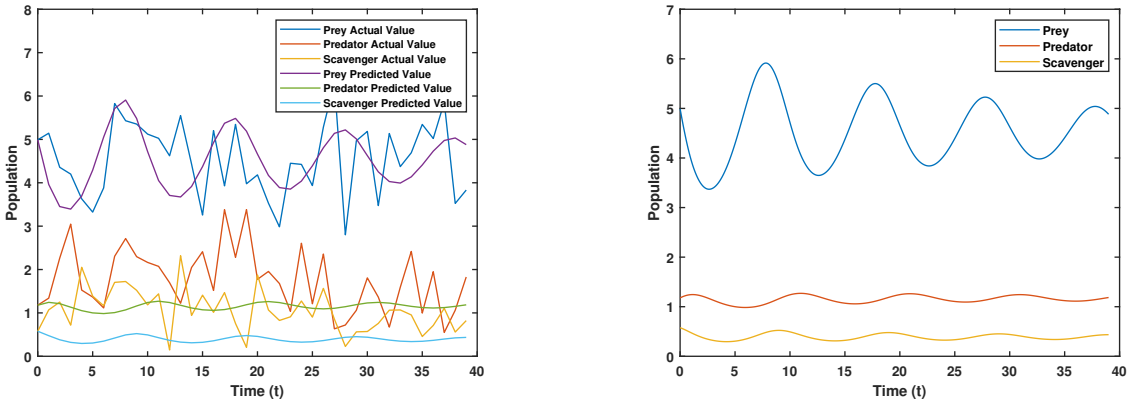
After physics-informed deep neural network, the final predicted parameter for model (3.1) is given by Table 3. These parameter values define the underlying dynamics of the model (3.1). Value r defines that the population of prey increases with a decreasing rate in nature where total resources for prey is k, which is very large as compared to population values. Values a_0, b_0, and i_0 are handling time required by respective predators for their respective prey. It can be easily concluded that handling time is the least required by scavengers for handling prey populations. Values d, f and g are the rate of change of population of predators and scavengers and h is the scavenge factor for the scavengers. From here it can be concluded that scavengers rely on the dead bodies of predators more than the prey population.

S. No.	Parameter	Numerical values
01.	r	0.9701107742719246
02.	k	573.2545487545212544
03.	a	0.7668876233328743
04.	a_0	0.4686878655732233
05.	b	0.6893067418603573
06.	b_0	0.053266947840986595
07.	d	0.42441058569930494
08.	f	0.46630691773424437
09.	e	0.888598932493589
10.	g	0.08334616995047056
11.	h	0.16502232050920586
12.	i	1.05992612257741696
13.	i_0	0.105259076974745925
14.	j	0.5320956432008955

Table 3: Resulting values of parameters

Fig. 10a represents the comparison graph of the normalized natural dataset with the model (3.1) using final predicted parameters. Fig. 10b represents the plot for the predator-prey scavenger model using the parameters given in Table 3. The comparison graph shown in Fig. 10a

shows that predicted parameters using a physics-informed deep neural network model explained in section 6.2 captures most of the features of the natural dataset and mimics the data appropriately. Further, these predicted parameters can be more optimized and fine-tuned when the large dataset is available as the present dataset consists of only 40 data point that is, population values.



(a) Comparison Graph.

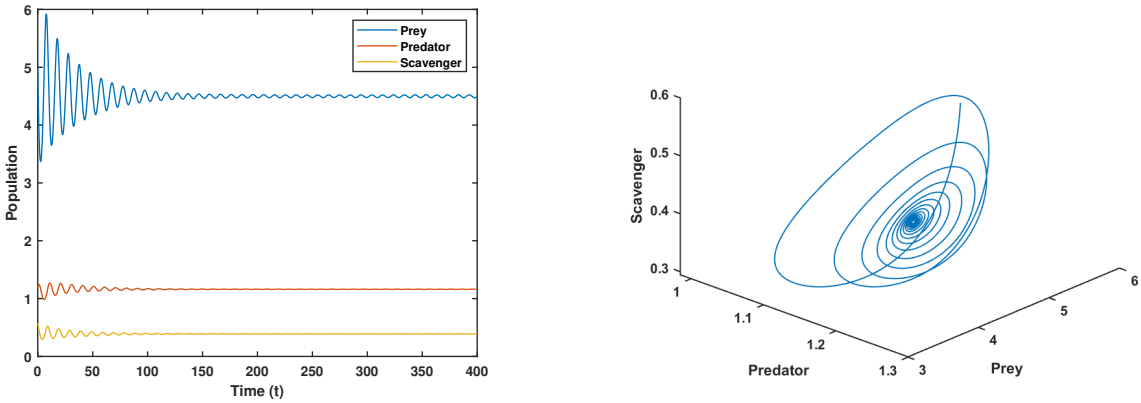
(b) Plot for predicted parameters.

Fig. 10: Parameter accuracy comparison analysis.

7 Stability Analysis of Parameters

Using parameters obtained from Table 3 with initial values as $x(0) = 4.991$, $y(0) = 1.178$ and $z(0) = 0.577$, the predator prey scavenger model (3.1) takes the form,

$$\frac{dx}{dt} = 0.97011x\left(1 - \frac{x}{573.25455}\right) - \frac{0.76689x^2y}{1 + 0.46869x^2} - \frac{0.68930x^2z}{1 + 0.05327x^2}$$
$$\frac{dy}{dt} = \frac{0.42441x^2y}{1 + 0.46869x^2} + \frac{0.46630z^2y}{1 + 0.10526z^2} - 0.88860y$$
$$\frac{dz}{dt} = \frac{0.08335x^2z}{1 + 0.05327x^2} + 0.16502yz - \frac{1.05992yz^2}{1 + 0.10526z^2} - 0.53210z$$



(a) Dynamical behaviour.

(b) Phase portrait.

Fig. 11: Predator prey scavenger model using parameters obtained in section 6.

For predator prey scavenger system using parameters obtained in Table 3, it is clear from Fig. 11 that all of the population oscillates about specific population values.

Equilibrium point $(0,0,0)$: It is clear that $(0,0,0)$ is an unstable equilibrium point because if this point becomes stable, then the system collapse. As described in subsection 4.4, $(0,E_y,0)$ and $(0,0,E_z)$ cases are also included in this $(0,0,0)$ case as both yields equilibrium point $(0,0,0)$ and both are not stable as well.

Equilibrium point $(E_x^{[2]},0,0)$: For equilibrium point $(573.25455,0,0)$,

$$\frac{e}{d-ea_0} = 111.984 < (573.25455)^2 = k^2 \quad \text{and} \quad \frac{j}{g-jb_0} = 9.673966 < (573.25455)^2 = k^2$$

Hence, equilibrium point $(573.25455,0,0)$ is not stable.

Equilibrium point $(0,E_y^{[3]},E_z^{[3]})$: As, $f - i_0 e = 0.3727738 > 0$ and

$$hf\left[\frac{e}{f-i_0e}\right]^{1/2} - ie = -0.8230414 < 0$$

implies the existence conditions are not satisfied for the equilibrium point $\left(0, \frac{fjz_0}{hfz_0-ie}, z_0\right)$ as the equilibrium point turns out to have a negative number, which has no significance in nature. Hence, equilibrium point $(0,E_y^{[3]},E_z^{[3]})$ does not exist in the system.

Equilibrium point $(E_x^{[4]},E_y^{[4]},0)$: As, $d - a_0 e = 0.007935 > 0$ implies the existence condition for the equilibrium point,

$$\left(\left[\frac{e}{d-a_0e}\right]^{1/2}, \frac{drx_0(k-x_0)}{aek}, 0\right) = (10.582252, 6.2756, 0)$$

For this equilibrium point,

$$\frac{2a_0e(k-x_0)}{dk} = 1.9263771 > 1$$

Hence, the stability condition are not met which means predator and prey may coexist in an unstable system.

Equilibrium point $(E_x^{[5]},0,E_z^{[5]})$: As, $g - b_0 j = 0.055003 > 0$ implies the existence condition for the equilibrium point

$$\left(\left[\frac{j}{g-b_0j}\right]^{1/2}, 0, \frac{grx_0(k-x_0)}{bjk}\right) = (3.1103, 0, 0.6819357)$$

For this equilibrium point,

$$\frac{2b_0j(k-x_0)}{gk} = 0.67643973 < 1$$

$$t = \frac{e(g+j(a_0-b_0)) - dj}{g+j(a_0-b_0)} = 0.1467$$

and hence $z_0^2 = 0.465 > \frac{t}{f-i_0t} = 0.32537$. Therefore, the stability condition doesn't satisfy that means scavenger and prey may coexist in an unstable system.

Equilibrium point $(E_x^{[6]},E_y^{[6]},E_z^{[6]})$: Value of $E_x^{[6]}$ is given by the positive real root of the equation (4.5) using parameter values given in Table 3, which means equilibrium point is $(4.4984538, 1.161178, 0.38895175)$. Here, $x^* = 4.4984538$, which is only positive real root of the equation above and $z^* = 0.38895175 < \min\{12.239589, 369.74242836\}$, that means the existence condition is satisfied for the equilibrium point in which the coexistence of predator, prey and scavenger are present. Putting these values, the jacobian matrix corresponding to this equilibrium point turns out to be,

$$J_{xyz} = \begin{bmatrix} 0.34 & -1.5 & -6.7 \\ 0.04 & 0 & 0.41 \\ 0.068 & -0.094 & -0.46 \end{bmatrix}$$

For this matrix characteristic polynomial becomes,

$$\lambda^3 + 0.1178643784\lambda^2 + 0.39684137\lambda + 0.02975781$$

For this characteristic polynomial, $m_1 = 0.1178643784, m_2 = 0.39684137, m_3 = 0.02975781$ which means $m_1, m_2, m_3 > 0$ and $m_1 m_2 - m_3 = 0.01701565148 > 0$. That means the equilibrium point $(4.4984538, 1.161178, 0.38895175)$ is stable. Hence, all three populations can coexist in the system. As it is clear from Fig. 11 all the populations coexist in the system. When the population of prey decreases, the food for scavengers and predators decreases, and hence the population of predators and scavengers also decreases. As the prey population decreases, the scavenger population also decreases and, hence the predator population. As the predator population decreases the scavenger population starts decreasing much faster. As the population of predators and scavengers starts decreasing, the prey population starts increasing. Hence, the system gets stabilized, and all the population (prey, predator, and scavengers) coexist in this system.

Remarks: Results for hence found real parameters are as follows,

1. Predator-prey can coexist in the system in the absence of scavengers (Unstable).
2. Scavenger prey can coexist in the system in the absence of predators (Unstable).
3. Predator scavengers can't coexist in the system in the absence of prey.
4. Predator prey scavenger can all coexist in the system (Stable).

8 Conclusion and Discussion

In conclusion, this research aims to address the dynamic behavior of predator prey scavenger interactions in natural ecosystems through the development and analysis of a mathematical model. By extending classical Lotka-Volterra dynamics to incorporate scavenger behavior, this study aims to enhance the understanding of complex interactions between different species present in the ecosystem. The proposed model, based on principles of population dynamics and Holling's functional response of type III, provides a framework for simulating the interactions between predators, prey, and scavenger populations under different scenarios. Through numerical simulations, the model demonstrates robustness and versatility, offering insights into non-linear dynamics present in ecological systems. Real-world validation is achieved through parameter estimation using a natural population dataset provided by the Forest Service, U.S. Department. Leveraging computational techniques, physics-informed neural networks using Adam backpropagation and BFGS optimization algorithm, the model's parameters are predicted which approximately reflect observed ecological patterns. Stability analysis, conducted via Jacobian matrix analysis, ensures the reliability of the model under different conditions. By examining the stability of steady states, this analysis provides crucial insights into the long-term behavior of different populations namely, predator, prey, and scavenger present in the ecosystem.

The proposed predator prey scavenger model using Holling's functional response of type III of order $n = 2$ is used for respective predator-prey interactions. Two-step parameter estimation from the natural dataset is also done in this paper using physics-informed deep neural network and BFGS optimization for further fine-tuning the estimated parameters. Numerical simulations and stability analysis of the proposed predator-prey scavenger model are also done for different scenarios. Looking ahead, this research sets the stage for further exploration and refinement of the model, by minimizing parameter-induced errors, conducting bifurcation analysis, and sensitivity analysis of the parameters to enhance predictive accuracy. Incorporating scavenger population and scavenge factor into mathematical modeling, this research aims to bridge the gap between theoretical modeling and actual population. Hence, this study also contributes to the advancement of ecological science and the sustainable management of natural resources.

Ethics declarations

Conflict of interest

The authors declare no conflict of interest.

Ethical approval

This article does not contain any studies with human participants and/or animals performed by the author.

Data availability

Wildlife population in U.S. dataset, is available to download from Forest Service, U.S. Dept. of Agriculture (USDA) repository at https://doi.org/10.2737/RMRS-GTR-296 [16]

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] M. L. Abell, J. Braselton, Modern differential equations: Theory, applications, technology (2001).

[2] S. Sarwardi, P. K. Mandal, S. Ray, Dynamical behaviour of a two-predator model with prey refuge, Journal of biological physics 39 (4) (2013) 701–722.

[3] R. Dilao, T. Domingos, A general approach to the modelling of trophic chains, Ecological Modelling 132 (3) (2000) 191–202.

[4] S. Kant, V. Kumar, Stability analysis of predator–prey system with migrating prey and disease infection in both species, Applied Mathematical Modelling 42 (2017) 509–539.

[5] H. Shim, P. Fishwick, Visualization and interaction design for ecosystem modeling (2008).

[6] T. D. Schowalter, Insect ecology: an ecosystem approach, Academic press, 2022.

[7] A. Hastings, L. Gross, Encyclopedia of theoretical ecology, no. 4, Univ of California Press, 2012.

[8] M.-C. Anisiu, Lotka, volterra and their model, Didáctica mathematica 32 (01) (2014).

[9] V. Hadžiabdić, M. Mehuljić, J. Bektešević, Lotka-volterra model with two predators and their prey, Tem Journal 6 (1) (2017) 2217–8309.

[10] S.-X. Wu, X.-Y. Meng, et al., Dynamics of a delayed predator-prey system with fear effect, herd behavior and disease in the susceptible prey, AIMS Mathematics 6 (4) (2021) 3654–3685.

[11] C. Arora, V. Kumar, A delayed version of prey predator system with modified holling-tanner response, in: AIP Conference Proceedings, Vol. 2336, AIP Publishing, 2021.

[12] I. A. Basheer, M. Hajmeer, Artificial neural networks: fundamentals, computing, design, and application, Journal of microbiological methods 43 (1) (2000) 3–31.

[13] N. Yadav, A. Yadav, M. Kumar, et al., An introduction to neural network methods for differential equations, Vol. 1, Springer, 2015.

[14] W. S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, The bulletin of mathematical biophysics 5 (1943) 115–133.

[15] M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational physics 378 (2019) 686–707.

[16] C. H. Flather, M. S. Knowles, M. F. Jones, C. Schilli, Wildlife population and harvest trends in the united states: a technical document supporting the forest service 2010 rpa assessment, General Technical Report RMRS-GTR-296, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO (2013).

[17] J. Dawes, M. Souza, A derivation of holling's type i, ii and iii functional responses in predator–prey systems, Journal of theoretical biology 327 (2013) 11–22.

[18] Y. Zhang, X. Meng, Dynamic study of a stochastic holling iii predator-prey system with a prey refuge, IFAC-PapersOnLine 55 (3) (2022) 73–78.

[19] Y. Huang, F. Chen, L. Zhong, Stability analysis of a prey–predator model with holling type iii response function incorporating a prey refuge, Applied Mathematics and Computation 182 (1) (2006) 672–683.

[20] R. Banerjee, P. Das, D. Mukherjee, Stability and permanence of a discrete-time two-prey one-predator system with holling type-iii functional response, Chaos, Solitons & Fractals 117 (2018) 240–248.

[21] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[22] C. G. Broyden, The convergence of a class of double-rank minimization algorithms 1. general considerations, IMA Journal of Applied Mathematics 6 (1) (1970) 76–90.

[23] R. Fletcher, A new approach to variable metric algorithms, The computer journal 13 (3) (1970) 317–322.

[24] D. Goldfarb, A family of variable-metric methods derived by variational means, Mathematics of computation 24 (109) (1970) 23–26.

[25] D. F. Shanno, Conditioning of quasi-newton methods for function minimization, Mathematics of computation 24 (111) (1970) 647–656.

[26] J. Nocedal, S. J. Wright, Numerical optimization, Springer, 1999.

[27] E. X. DeJesus, C. Kaufman, Routh-hurwitz criterion in the examination of eigenvalues of a system of nonlinear ordinary differential equations, Physical Review A 35 (12) (1987) 5288.

Appendix: Coefficients of Equation (4.5)

$$\mathbf{p_0} = e^3i^2{i_0}^2k^2r^2 - 2e^2fi^2{i_0}^2k^2r^2 + ef^2h^2{i_0}^2k^2r^2 + ef^2i^2k^2r^2 - f^3h^2k^2r^2$$

$$\mathbf{p_1} = -2aef^2hi_0jk^2r - 2e^3i^2{i_0}^2kr^2 + 2af^3hjk^2r + 4e^2fi^2i_0kr^2 - 2ef^2h^2i_0kr^2 - 2ef^2i^2kr^2 + 2f^3h^2kr^2$$

$$\begin{aligned}\mathbf{p_2} = {} & 3a_0e^3i^2{i_0}^2k^2r^2 + 2b_0e^3i^2{i_0}^2k^2r^2 - 6a_0e^2fi^2i_0k^2r^2 + 3a_0ef^2h^2i_0k^2r^2 - 4b_0e^2fi^2i_0k^2r^2 + 2b_0ef^2h^2i_0k^2r^2 - \\ & 3de^2i^2{i_0}^2k^2r^2 + a^2ef^2{i_0}j^2k^2 - 2abe^2fiii_0jk^2 + 3a_0ef^2i^2k^2r^2 - 3a_0f^3h^2k^2r^2 + b^2e^3i^2i_0k^2 + 2b_0ef^2i^2k^2r^2 - \\ & 2b_0f^3h^2k^2r^2 + 4defi^2i_0k^2r^2 - df^2h^2i_0k^2r^2 - a^2f^3j^2k^2 + 2abef^2ijk^2 + 2aef^2hi_0jkr - b^2e^2fi^2k^2 + b^2ef^2h^2k^2 - \\ & df^2i^2k^2r^2 + e^3i^2{i_0}^2r^2 - 2af^3hjkr - 2e^2fi^2i_0r^2 + ef^2h^2i_0r^2 + ef^2i^2r^2 - f^3h^2r^2\end{aligned}$$

$$\begin{aligned}\mathbf{p_3} = {} & -4aa_0ef^2hi_0jk^2r - 4ab_0ef^2hi_0jk^2r - 6a_0e^3i^2{i_0}^2kr^2 - 4b_0e^3i^2{i_0}^2kr^2 + 4aa_0f^3hjk^2r + 4ab_0f^3hjk^2r + 2adf^2hi_0jk^2r + \\ & 2aef^2ghi_0k^2r + 12a_0e^2fi^2i_0kr^2 - 6a_0ef^2h^2i_0kr^2 + 8b_0e^2fi^2i_0kr^2 - 4b_0ef^2h^2i_0kr^2 + 6de^2i^2{i_0}^2kr^2 - 2af^3ghk^2r - \\ & 6a_0ef^2i^2kr^2 + 6a_0f^3h^2kr^2 - 4b_0ef^2i^2kr^2 + 4b_0f^3h^2kr^2 - 8defi^2i_0kr^2 + 2df^2h^2i_0kr^2 + 2df^2i^2kr^2\end{aligned}$$

$$\begin{aligned}\mathbf{p_4} = {} & 3a_0{}^2e^3i^2{i_0}^2k^2r^2 + 6a_0b_0e^3i^2{i_0}^2k^2r^2 + b_0{}^2e^3i^2{i_0}^2k^2r^2 - 6a_0{}^2e^2fi^2i_0k^2r^2 + 3a_0{}^2ef^2h^2i_0k^2r^2 - 12a_0b_0e^2fi^2i_0k^2r^2 + \\ & 6a_0b_0ef^2h^2i_0k^2r^2 - 6a_0de^2i^2{i_0}^2k^2r^2 - 2b_0{}^2e^2fi^2i_0k^2r^2 + b_0{}^2ef^2h^2i_0k^2r^2 - 6b_0de^2i^2{i_0}^2k^2r^2 + a^2a_0ef^2{i_0}j^2k^2 + \\ & 2a^2b_0ef^2{i_0}j^2k^2 - 4aa_0be^2fiii_0jk^2 - 2abb_0e^2fiii_0jk^2 + 3a_0{}^2ef^2i^2k^2r^2 - 3a_0{}^2f^3h^2k^2r^2 + 3a_0b^2e^3i^2i_0k^2 + 6a_0b_0ef^2i^2k^2r^2 - \\ & 6a_0b_0f^3h^2k^2r^2 + 8a_0defi^2i_0k^2r^2 - 2a_0df^2h^2i_0k^2r^2 + b_0{}^2ef^2i^2k^2r^2 - b_0{}^2f^3h^2k^2r^2 + 8b_0defi^2i_0k^2r^2 - 2b_0df^2h^2i_0k^2r^2 + \\ & 3d^2ei^2{i_0}^2k^2r^2 - a^2a_0f^3j^2k^2 - 2a^2b_0f^3j^2k^2 - a^2df^2i_0j^2k^2 - 2a^2ef^2gi_0jk^2 + 4aa_0bef^2ijk^2 + 4aa_0ef^2hi_0jkr + \\ & 2abb_0ef^2ijk^2 + 4abdefiii_0jk^2 + 2abe^2fgiii_0k^2 + 4ab_0ef^2hi_0jkr - 3a_0b^2e^2fi^2k^2 + 3a_0b^2ef^2h^2k^2 - 2a_0df^2i^2k^2r^2 + \\ & 3a_0e^3i^2{i_0}^2r^2 - 3b^2de^2i^2i_0k^2 - 2b_0df^2i^2k^2r^2 + 2b_0e^3i^2{i_0}^2r^2 - 2d^2fi^2i_0k^2r^2 + 2a^2f^3gjk^2 - 4aa_0f^3hjkr - \\ & 2abdf^2ijk^2 - 2abef^2gik^2 - 4ab_0f^3hjkr - 2adf^2hi_0jkr - 2aef^2ghi_0kr - 6a_0e^2fi^2i_0r^2 + 3a_0ef^2h^2i_0r^2 + \\ & 2b^2defi^2k^2 - b^2df^2h^2k^2 - 4b_0e^2fi^2i_0r^2 + 2b_0ef^2h^2i_0r^2 - 3de^2i^2{i_0}^2r^2 + 2af^3ghkr + 3a_0ef^2i^2r^2 - 3a_0f^3h^2r^2 + \\ & 2b_0ef^2i^2r^2 - 2b_0f^3h^2r^2 + 4defi^2i_0r^2 - df^2h^2i_0r^2 - df^2i^2r^2\end{aligned}$$

$$\begin{aligned}\mathbf{p_5} = {} & -2aa_0{}^2ef^2hi_0jk^2r - 8aa_0b_0ef^2hi_0jk^2r - 2ab_0{}^2ef^2hi_0jk^2r - 6a_0{}^2e^3i^2{i_0}^2kr^2 - 12a_0b_0e^3i^2{i_0}^2kr^2 - 2b_0{}^2e^3i^2{i_0}^2kr^2 + \\ & 2aa_0{}^2f^3hjk^2r + 8aa_0b_0f^3hjk^2r + 2aa_0df^2hi_0jk^2r + 4aa_0ef^2ghi_0k^2r + 2ab_0{}^2f^3hjk^2r + 4ab_0df^2hi_0jk^2r + \\ & 2ab_0ef^2ghi_0k^2r + 12a_0{}^2e^2fi^2i_0kr^2 - 6a_0{}^2ef^2h^2i_0kr^2 + 24a_0b_0e^2fi^2i_0kr^2 - 12a_0b_0ef^2h^2i_0kr^2 + 12a_0de^2i^2{i_0}^2kr^2 + \\ & 4b_0{}^2e^2fi^2i_0kr^2 - 2b_0{}^2ef^2h^2i_0kr^2 + 12b_0de^2i^2{i_0}^2kr^2 - 4aa_0f^3ghk^2r - 2ab_0f^3ghk^2r - 2adf^2ghi_0k^2r - 6a_0{}^2ef^2i^2kr^2 + \\ & 6a_0{}^2f^3h^2kr^2 - 12a_0b_0ef^2i^2kr^2 + 12a_0b_0f^3h^2kr^2 - 16a_0defi^2i_0kr^2 + 4a_0df^2h^2i_0kr^2 - 2b_0{}^2ef^2i^2kr^2 + 2b_0{}^2f^3h^2kr^2 - \\ & 16b_0defi^2i_0kr^2 + 4b_0df^2h^2i_0kr^2 - 6d^2ei^2{i_0}^2kr^2 + 4a_0df^2i^2kr^2 + 4b_0df^2i^2kr^2 + 4d^2fi^2i_0kr^2\end{aligned}$$

$$\begin{aligned}\mathbf{p_6} = {} & a_0{}^3e^3i^2{i_0}^2k^2r^2 + 6a_0{}^2b_0e^3i^2{i_0}^2k^2r^2 + 3a_0b_0{}^2e^3i^2{i_0}^2k^2r^2 - 2a_0{}^3e^2fi^2i_0k^2r^2 + a_0{}^3ef^2h^2i_0k^2r^2 - 12a_0{}^2b_0e^2fi^2i_0k^2r^2 + \\ & 6a_0{}^2b_0ef^2h^2i_0k^2r^2 - 3a_0{}^2de^2i^2{i_0}^2k^2r^2 - 6a_0b_0{}^2e^2fi^2i_0k^2r^2 + 3a_0b_0{}^2ef^2h^2i_0k^2r^2 - 12a_0b_0de^2i^2{i_0}^2k^2r^2 - \\ & 3b_0{}^2de^2i^2{i_0}^2k^2r^2 + 2a^2a_0b_0ef^2{i_0}j^2k^2 + a^2b_0{}^2ef^2{i_0}j^2k^2 - 2aa_0{}^2be^2fiii_0jk^2 - 4aa_0bb_0e^2fiii_0jk^2 + a_0{}^3ef^2i^2k^2r^2 - \\ & a_0{}^3f^3h^2k^2r^2 + 3a_0{}^2b^2e^3i^2i_0k^2 + 6a_0{}^2b_0ef^2i^2k^2r^2 - 6a_0{}^2b_0f^3h^2k^2r^2 + 4a_0{}^2defi^2i_0k^2r^2 - a_0{}^2df^2h^2i_0k^2r^2 + \\ & 3a_0b_0{}^2ef^2i^2k^2r^2 - 3a_0b_0{}^2f^3h^2k^2r^2 + 16a_0b_0defi^2i_0k^2r^2 - 4a_0b_0df^2h^2i_0k^2r^2 + 3a_0d^2ei^2{i_0}^2k^2r^2 + 4b_0{}^2defi^2i_0k^2r^2 - \\ & b_0{}^2df^2h^2i_0k^2r^2 + 6b_0d^2ei^2{i_0}^2k^2r^2 - 2a^2a_0b_0f^3j^2k^2 - 2a^2a_0ef^2gi_0jk^2 - a^2b_0{}^2f^3j^2k^2 - 2a^2b_0df^2i_0j^2k^2 - \\ & 2a^2b_0ef^2gi_0jk^2 + 2aa_0{}^2bef^2ijk^2 + 2aa_0{}^2ef^2hi_0jkr + 4aa_0bb_0ef^2ijk^2 + 4aa_0bdefiii_0jk^2 + 4aa_0be^2fgii_0k^2 + \\ & 8aa_0b_0ef^2hi_0jkr + 4abb_0defiii_0jk^2 + 2ab_0{}^2ef^2hi_0jkr - 3a_0{}^2b^2e^2fi^2k^2 + 3a_0{}^2b^2ef^2h^2k^2 - a_0{}^2df^2i^2k^2r^2 + \\ & 3a_0{}^2e^3i^2{i_0}^2r^2 - 6a_0b^2de^2i^2i_0k^2 - 4a_0b_0df^2i^2k^2r^2 + 6a_0b_0e^3i^2{i_0}^2r^2 - 2a_0d^2fi^2i_0k^2r^2 - b_0{}^2df^2i^2k^2r^2 + b_0{}^2e^3i^2{i_0}^2r^2 - \\ & 4b_0d^2fi^2i_0k^2r^2 - d^3i^2{i_0}^2k^2r^2 + 2a^2a_0f^3gjk^2 + 2a^2b_0f^3gjk^2 + 2a^2df^2gi_0jk^2 + a^2ef^2g^2i_0k^2 - 2aa_0{}^2f^3hjkr - \\ & 2aa_0bdf^2ijk^2 - 4aa_0bef^2gik^2 - 8aa_0b_0f^3hjkr - 2aa_0df^2hi_0jkr - 4aa_0ef^2ghi_0kr - 2abb_0df^2ijk^2 - 2abd^2fiii_0jk^2 - \\ & 4abdefgii_0k^2 - 2ab_0{}^2f^3hjkr - 4ab_0df^2hi_0jkr - 2ab_0ef^2ghi_0kr - 6a_0{}^2e^2fi^2i_0r^2 + 3a_0{}^2ef^2h^2i_0r^2 + 4a_0b^2defi^2k^2 - \\ & 2a_0b^2df^2h^2k^2 - 12a_0b_0e^2fi^2i_0r^2 + 6a_0b_0ef^2h^2i_0r^2 - 6a_0de^2i^2{i_0}^2r^2 + 3b^2d^2ei^2i_0k^2 - 2b_0{}^2e^2fi^2i_0r^2 + b_0{}^2ef^2h^2i_0r^2 - \\ & 6b_0de^2i^2{i_0}^2r^2 - a^2f^3g^2k^2 + 4aa_0f^3ghkr + 2abdf^2gik^2 + 2ab_0f^3ghkr + 2adf^2ghi_0kr + 3a_0{}^2ef^2i^2r^2 - 3a_0{}^2f^3h^2r^2 + \\ & 6a_0b_0ef^2i^2r^2 - 6a_0b_0f^3h^2r^2 + 8a_0defi^2i_0r^2 - 2a_0df^2h^2i_0r^2 - b^2d^2fi^2k^2 + b_0{}^2ef^2i^2r^2 - b_0{}^2f^3h^2r^2 + 8b_0defi^2i_0r^2 - \\ & 2b_0df^2h^2i_0r^2 + 3d^2ei^2{i_0}^2r^2 - 2a_0df^2i^2r^2 - 2b_0df^2i^2r^2 - 2d^2fi^2i_0r^2\end{aligned}$$

$$\begin{aligned}\mathbf{p_7} = {} & -4aa_0{}^2b_0ef^2hi_0jk^2r - 4aa_0b_0{}^2ef^2hi_0jk^2r - 2a_0{}^3e^3i^2{i_0}^2kr^2 - 12a_0{}^2b_0e^3i^2{i_0}^2kr^2 - 6a_0b_0{}^2e^3i^2{i_0}^2kr^2 + 4aa_0{}^2b_0f^3hjk^2r + \\ & 2aa_0{}^2ef^2ghi_0k^2r + 4aa_0b_0{}^2f^3hjk^2r + 4aa_0b_0df^2hi_0jk^2r + 4aa_0b_0ef^2ghi_0k^2r + 2ab_0{}^2df^2hi_0jk^2r + 4a_0{}^3e^2fi^2i_0kr^2 - \\ & 2a_0{}^3ef^2h^2i_0kr^2 + 24a_0{}^2b_0e^2fi^2i_0kr^2 - 12a_0{}^2b_0ef^2h^2i_0kr^2 + 6a_0{}^2de^2i^2{i_0}^2kr^2 + 12a_0b_0{}^2e^2fi^2i_0kr^2 - 6a_0b_0{}^2ef^2h^2i_0kr^2 + \\ & 24a_0b_0de^2i^2{i_0}^2kr^2 + 6b_0{}^2de^2i^2{i_0}^2kr^2 - 2aa_0{}^2f^3ghk^2r - 4aa_0b_0f^3ghk^2r - 2aa_0df^2ghi_0k^2r - 2ab_0df^2ghi_0k^2r - \\ & 2a_0{}^3ef^2i^2kr^2 + 2a_0{}^3f^3h^2kr^2 - 12a_0{}^2b_0ef^2i^2kr^2 + 12a_0{}^2b_0f^3h^2kr^2 - 8a_0{}^2defi^2i_0kr^2 + 2a_0{}^2df^2h^2i_0kr^2 - \\ & 6a_0b_0{}^2ef^2i^2kr^2 + 6a_0b_0{}^2f^3h^2kr^2 - 32a_0b_0defi^2i_0kr^2 + 8a_0b_0df^2h^2i_0kr^2 - 6a_0d^2ei^2{i_0}^2kr^2 - 8b_0{}^2defi^2i_0kr^2 + \\ & 2b_0{}^2df^2h^2i_0kr^2 - 12b_0d^2ei^2{i_0}^2kr^2 + 2a_0{}^2df^2i^2kr^2 + 8a_0b_0df^2i^2kr^2 + 4a_0d^2fi^2i_0kr^2 + 2b_0{}^2df^2i^2kr^2 + 8b_0d^2fi^2i_0kr^2 + \\ & 2d^3i^2{i_0}^2kr^2\end{aligned}$$

$$\begin{aligned}\mathbf{p_8} = {} & 2a_0{}^3b_0e^3i^2{i_0}^2k^2r^2 + 3a_0{}^2b_0{}^2e^3i^2{i_0}^2k^2r^2 - 4a_0{}^3b_0e^2fi^2i_0k^2r^2 + 2a_0{}^3b_0ef^2h^2i_0k^2r^2 - 6a_0{}^2b_0{}^2e^2fi^2i_0k^2r^2 + \\ & 3a_0{}^2b_0{}^2ef^2h^2i_0k^2r^2 - 6a_0{}^2b_0de^2i^2{i_0}^2k^2r^2 - 6a_0b_0{}^2de^2i^2{i_0}^2k^2r^2 + a^2a_0b_0{}^2ef^2{i_0}j^2k^2 - 2aa_0{}^2bb_0e^2fiii_0jk^2 + \\ & a_0{}^3b^2e^3i^2i_0k^2 + 2a_0{}^3b_0ef^2i^2k^2r^2 - 2a_0{}^3b_0f^3h^2k^2r^2 + 3a_0{}^2b_0{}^2ef^2i^2k^2r^2 - 3a_0{}^2b_0{}^2f^3h^2k^2r^2 + 8a_0{}^2b_0defi^2i_0k^2r^2 - \\ & 2a_0{}^2b_0df^2h^2i_0k^2r^2 + 8a_0b_0{}^2defi^2i_0k^2r^2 - 2a_0b_0{}^2df^2h^2i_0k^2r^2 + 6a_0b_0d^2ei^2{i_0}^2k^2r^2 + 3b_0{}^2d^2ei^2{i_0}^2k^2r^2 - a^2a_0b_0{}^2f^3j^2k^2 - \\ & 2a^2a_0b_0ef^2gi_0jk^2 - a^2b_0{}^2df^2i_0j^2k^2 + 2aa_0{}^2bb_0ef^2ijk^2 + 2aa_0{}^2be^2fgii_0k^2 + 4aa_0{}^2b_0ef^2hi_0jkr + 4aa_0bb_0defiii_0jk^2 + \\ & 4aa_0b_0{}^2ef^2hi_0jkr - a_0{}^3b^2e^2fi^2k^2 + a_0{}^3b^2ef^2h^2k^2 + a_0{}^3e^3i^2{i_0}^2r^2 - 3a_0{}^2b^2de^2i^2i_0k^2 - 2a_0{}^2b_0df^2i^2k^2r^2 + \\ & 6a_0{}^2b_0e^3i^2{i_0}^2r^2 - 2a_0b_0{}^2df^2i^2k^2r^2 + 3a_0b_0{}^2e^3i^2{i_0}^2r^2 - 4a_0b_0d^2fi^2i_0k^2r^2 - 2b_0{}^2d^2fi^2i_0k^2r^2 - 2b_0d^3i^2{i_0}^2k^2r^2 + \\ & 2a^2a_0b_0f^3gjk^2 + a^2a_0ef^2g^2i_0k^2 + 2a^2b_0df^2gi_0jk^2 - 2aa_0{}^2bef^2gik^2 - 4aa_0{}^2b_0f^3hjkr - 2aa_0{}^2ef^2ghi_0kr - \\ & 2aa_0bb_0df^2ijk^2 - 4aa_0bdefgiii_0k^2 - 4aa_0b_0{}^2f^3hjkr - 4aa_0b_0df^2hi_0jkr - 4aa_0b_0ef^2ghi_0kr - 2abb_0d^2fiii_0jk^2 - \\ & 2ab_0{}^2df^2hi_0jkr - 2a_0{}^3e^2fi^2i_0r^2 + a_0{}^3ef^2h^2i_0r^2 + 2a_0{}^2b^2defi^2k^2 - a_0{}^2b^2df^2h^2k^2 - 12a_0{}^2b_0e^2fi^2i_0r^2 + 6a_0{}^2b_0ef^2h^2i_0r^2 -\end{aligned}$$

$3a_0{}^2de^2i^2i_0{}^2r^2+3a_0b^2d^2ei^2i_0k^2-6a_0b_0{}^2e^2fi^2i_0r^2+3a_0b_0{}^2ef^2h^2i_0r^2-12a_0b_0de^2i^2i_0{}^2r^2-3b_0{}^2de^2i^2i_0{}^2r^2-a^2a_0f^3g^2k^2-a^2df^2g^2i_0k^2+2aa_0{}^2f^3ghkr+2aa_0bdf^2gik^2+4aa_0b_0f^3ghkr+2aa_0df^2ghi_0kr+2abd^2fgii_0k^2+2ab_0df^2ghi_0kr+a_0{}^3ef^2i^2r^2-a_0{}^3f^3h^2r^2+6a_0{}^2b_0ef^2i^2r^2-6a_0{}^2b_0f^3h^2r^2+4a_0{}^2defi^2i_0r^2-a_0{}^2df^2h^2i_0r^2-a_0b^2d^2fi^2k^2+3a_0b_0{}^2ef^2i^2r^2-3a_0b_0{}^2f^3h^2r^2+16a_0b_0defi^2i_0r^2-4a_0b_0df^2h^2i_0r^2+3a_0d^2ei^2i_0{}^2r^2-b^2d^3i^2i_0k^2+4b_0{}^2defi^2i_0r^2-b_0{}^2df^2h^2i_0r^2+6b_0d^2ei^2i_0{}^2r^2-a_0{}^2df^2i^2r^2-4a_0b_0df^2i^2r^2-2a_0d^2fi^2i_0r^2-b_0{}^2df^2i^2r^2-4b_0d^2fi^2i_0r^2-d^3i^2i_0{}^2r^2$

$\mathbf{p_9} = -2aa_0{}^2b_0{}^2ef^2hi_0jk^2r-4a_0{}^3b_0e^3i^2i_0{}^2kr^2-6a_0{}^2b_0{}^2e^3i^2i_0{}^2kr^2+2aa_0{}^2b_0{}^2f^3hjk^2r+2aa_0{}^2b_0ef^2ghi_0k^2r+2aa_0b_0{}^2df^2hi_0jk^2r+8a_0{}^3b_0e^2fi^2i_0kr^2-4a_0{}^3b_0ef^2h^2i_0kr^2+12a_0{}^2b_0{}^2e^2fi^2i_0kr^2-6a_0{}^2b_0{}^2ef^2h^2i_0kr^2+12a_0{}^2b_0de^2i^2i_0{}^2kr^2+12a_0b_0{}^2de^2i^2i_0{}^2kr^2-2aa_0{}^2b_0f^3ghk^2r-2aa_0b_0df^2ghi_0k^2r-4a_0{}^3b_0ef^2i^2kr^2+4a_0{}^3b_0f^3h^2kr^2-6a_0{}^2b_0{}^2ef^2i^2kr^2+6a_0{}^2b_0{}^2f^3h^2kr^2-16a_0{}^2b_0defi^2i_0kr^2+4a_0{}^2b_0df^2h^2i_0kr^2-16a_0b_0{}^2defi^2i_0kr^2+4a_0b_0{}^2df^2h^2i_0kr^2-12a_0b_0d^2ei^2i_0{}^2kr^2-6b_0{}^2d^2ei^2i_0{}^2kr^2+4a_0{}^2b_0df^2i^2kr^2+4a_0b_0{}^2df^2i^2kr^2+8a_0b_0d^2fi^2i_0kr^2+4b_0{}^2d^2fi^2i_0kr^2+4b_0d^3i^2i_0{}^2kr^2$

$\mathbf{p_{10}} = a_0{}^3b_0{}^2e^3i^2i_0{}^2k^2r^2-2a_0{}^3b_0{}^2e^2fi^2i_0k^2r^2+a_0{}^3b_0{}^2ef^2h^2i_0k^2r^2-3a_0{}^2b_0{}^2de^2i^2i_0{}^2k^2r^2+a_0{}^3b_0{}^2ef^2i^2k^2r^2-a_0{}^3b_0{}^2f^3h^2k^2r^2+4a_0{}^2b_0{}^2defi^2i_0k^2r^2-a_0{}^2b_0{}^2df^2h^2i_0k^2r^2+3a_0b_0{}^2d^2ei^2i_0{}^2k^2r^2+2aa_0{}^2b_0{}^2ef^2hi_0jkr+2a_0{}^3b_0e^3i^2i_0{}^2r^2-a_0{}^2b_0{}^2df^2i^2k^2r^2+3a_0{}^2b_0{}^2e^3i^2i_0{}^2r^2-2a_0b_0{}^2d^2fi^2i_0k^2r^2-b_0{}^2d^3i^2i_0{}^2k^2r^2-2aa_0{}^2b_0{}^2f^3hjkr-2aa_0{}^2b_0ef^2ghi_0kr-2aa_0b_0{}^2df^2hi_0jkr-4a_0{}^3b_0e^2fi^2i_0r^2+2a_0{}^3b_0ef^2h^2i_0r^2-6a_0{}^2b_0{}^2e^2fi^2i_0r^2+3a_0{}^2b_0{}^2ef^2h^2i_0r^2-6a_0{}^2b_0de^2i^2i_0{}^2r^2-6a_0b_0{}^2de^2i^2i_0{}^2r^2+2aa_0{}^2b_0f^3ghkr+2aa_0b_0df^2ghi_0kr+2a_0{}^3b_0ef^2i^2r^2-2a_0{}^3b_0f^3h^2r^2+3a_0{}^2b_0{}^2ef^2i^2r^2-3a_0{}^2b_0{}^2f^3h^2r^2+8a_0{}^2b_0defi^2i_0r^2-2a_0{}^2b_0df^2h^2i_0r^2+8a_0b_0{}^2defi^2i_0r^2-2a_0b_0{}^2df^2h^2i_0r^2+6a_0b_0d^2ei^2i_0{}^2r^2+3b_0{}^2d^2ei^2i_0{}^2r^2-2a_0{}^2b_0df^2i^2r^2-2a_0b_0{}^2df^2i^2r^2-4a_0b_0d^2fi^2i_0r^2-2b_0{}^2d^2fi^2i_0r^2-2b_0d^3i^2i_0{}^2r^2$

$\mathbf{p_{11}} = -2a_0{}^3b_0{}^2e^3i^2i_0{}^2kr^2+4a_0{}^3b_0{}^2e^2fi^2i_0kr^2-2a_0{}^3b_0{}^2ef^2h^2i_0kr^2+6a_0{}^2b_0{}^2de^2i^2i_0{}^2kr^2-2a_0{}^3b_0{}^2ef^2i^2kr^2+2a_0{}^3b_0{}^2f^3h^2kr^2-8a_0{}^2b_0{}^2defi^2i_0kr^2+2a_0{}^2b_0{}^2df^2h^2i_0kr^2-6a_0b_0{}^2d^2ei^2i_0{}^2kr^2+2a_0{}^2b_0{}^2df^2i^2kr^2+4a_0b_0{}^2d^2fi^2i_0kr^2+2b_0{}^2d^3i^2i_0{}^2kr^2$

$\mathbf{p_{12}} = a_0{}^3b_0{}^2e^3i^2i_0{}^2r^2-2a_0{}^3b_0{}^2e^2fi^2i_0r^2+a_0{}^3b_0{}^2ef^2h^2i_0r^2-3a_0{}^2b_0{}^2de^2i^2i_0{}^2r^2+a_0{}^3b_0{}^2ef^2i^2r^2-a_0{}^3b_0{}^2f^3h^2r^2+4a_0{}^2b_0{}^2defi^2i_0r^2-a_0{}^2b_0{}^2df^2h^2i_0r^2+3a_0b_0{}^2d^2ei^2i_0{}^2r^2-a_0{}^2b_0{}^2df^2i^2r^2-2a_0b_0{}^2d^2fi^2i_0r^2-b_0{}^2d^3i^2i_0{}^2r^2$

Check for updates

Concurrency and Computation: Practice and Experience

WILEY

RESEARCH ARTICLE

REHAS: Robust and Efficient Hyperelliptic Curve-Based Authentication Scheme for Internet of Drones

Bhanu Pratap | Ayush Singh | Pawan Singh Mehra

Department of CSE, Delhi Technological University, New Delhi, India

Correspondence: Pawan Singh Mehra (pawansinghmehra@gmail.com)

Received: 16 April 2024 | **Revised:** 8 October 2024 | **Accepted:** 7 November 2024

Keywords: authentication | HECC | internet of drones | Scyther | security

ABSTRACT

Internet of Drones (IoD) is one of the most beneficial and has many versatile applications like Surveillance and Security, Delivery and Logistics, Environmental Monitoring, Agriculture, and so forth. The IoD network is crucial for collecting sensitive data like geo-coordinates, vehicle traffic data, and property details while surveying the various deployment locations in smart cities. The communication between users and drones can be compromised over insecure wireless channels by multiple attacks such as Man-in-the-middle-attack, Denial of Service, and so forth. Many schemes have already been propounded in the field of IoD. Still, many of them cannot address the resource constraints problem of drones, and existing protocols have higher computation and communication costs. Therefore, this paper has proposed a robust and efficient Hyper-Elliptic Curve-based authentication scheme (REHAS), which provides a session key for secure communication. Artificial Identities are generated using a hash function and random numbers. Fuzzy Extractor is used for user biometric authentication, which makes the smart device secure when lost. HECC is used with a smaller bit size of 80 bits rather than ECC of 160 bits. The security of the REHAS has been ensured using Scyther simulation. Furthermore, the resilience, safety, and robustness of REHAS are ensured by Informal security analysis. Lastly, a comparative study of the REHAS has been performed with other related Authentication and key agreement (AKA) protocols regarding communication cost, Computation cost, and security features, demonstrating that REHAS incurred less computation cost (6.7171 ms), communication overhead (1696 bits), and energy consumption (22.5 mJ) than other existing AKA schemes.

1 | Introduction

Internet of Drones (IoD) is the network that allows UAVs or drones to connect to the Internet and collect sensitive data or information from the real environment through various sensors like Radar Sensors, Global Positioning System (GPS), time-of-flight (ToF) sensors, Camera and Image Sensors, light-pulse distance sensors, Barometric Pressure Sensor, orientation sensors, magnetic-field change sensors, and so forth [1, 2]. This sensitive information and data gathered by the UAVs is sent directly to the Base station (BS), which further manages the data acquisition for the users and keeps a record of all the transactions. In the IoD network, drones can also directly communicate with the users through the BS to share sensitive information. As the communication in IoD networks happens publicly, this causes a threat/risk to the privacy of the involved entities. An intruder can intervene in the network and steal, alter, and manipulate the sensitive data that UAVs collect. Various well-known cyber attacks in the field of IoD include replay attacks, drone impersonation attacks, drone capture attacks, man-in-the-middle (MITM) attacks, stolen verifier attacks, insider attacks, session key attacks, and so forth [3].

Equally contributing authors.

© 2024 John Wiley & Sons Ltd.

To protect our network from adversary attacks such that only registered/known entities participate in the communication, an efficient and robust AKA scheme is needed to ensure the security of the entities and sensitive data on an insecure public channel. UAVs in the IoD network cannot perform heavy computation and high-cost tasks, limiting drones' ability to perform multi-factor authentication, Bio-metric authentication, digital certificates, and bilinear map pairings on the drone network. This is due to a need for more resources and battery constraints. Thus, it is necessary to develop an AKA framework that is energy efficient, fast, and secure from all threats. The rapid application of IoD frameworks has increased threats to IoD models. The data on the IoD network gets transmitted through insecure public channels. Therefore, it is crucial to protect the data against unauthorized access, tampering, malicious users, attackers, and so forth. These threats try to corrupt and modify data to obtain sensitive information. We can prevent these threats by using a robust and secure authentication protocol [4].

Drones have limited computation power and battery life, so the mutual authentication protocol should be computationally and communicationally efficient. Various public key cryptographic systems are used in the authentication, such as Hyperelliptic Curve Cryptography (HECC), Digital Signature Algorithm (DSA), Elliptic Curve Cryptography (ECC), and Rivest–Shamir–Adleman (RSA). These cryptosystems provide secure communication but with different key sizes. RSA is the least efficient in computation with a 1024-bit key size. HECC is the most efficient as it allows for an equal level of security with a relatively lesser key size, that is, 80 bit than ECC with 160 bit and RSA with 1024 bit key size [5, 6].

1.1 | Motivation

The UAVs or Drones gather sensitive data or information and send it directly to the BS via the wireless insecure channel. Moreover, the UAVs also directly communicate with users to share real-time sensitive data via public channels. This poses a threat/risk to IoD and makes them vulnerable to cyber-attacks. The privacy and security challenges in the IoD environment are discussed in detail by Yahuza et al. [7]. Iran-US R.Q-170 incident, where UAV was allegedly captured and reverse-engineered, highlights the dangers of weak security in communication channels, such as user impersonation attack [8]. The Noto Peninsula earthquake, which struck the Hokuriku region in Japan on January 1, 2024, highlighted their significance in real-world disaster scenarios [9]. Based on these events and uses of IoD networks, a robust, secure, and efficient authentication scheme is needed for IoD, which will be robust against cyber-attacks such as stolen verifier, User Impersonation, MITM, Drone Impersonation, and so forth. The scheme will ensure that only registered entities participate in the network and communicate securely. Moreover, the scheme should be computationally, communicationally and energy efficient due to the drone's computation and battery life limitations.

1.2 | Contributions

In the IoD, drones and users communicate via insecure wireless channels and share the collected data through IoT sensors [1]. Therefore, the wireless channel must be secure and robust to ensure perfect secrecy and data integrity. Considering these requirements, we have proposed a robust and efficient AKA scheme (REHAS) for the IoD, which has incurred low communication and computational overhead and is also resilient against various cyber attacks.

In the proposed scheme REHAS, HECC is used for generating public and private key pairs, utilizing only 80 bits rather than ECC using 160 bits, which makes our scheme lightweight as well as secure. Fuzzy extractor is used for biometric authentication of users, which makes the smart device secure when an adversary knows the ID and Password of the user. When the smart device of the user is lost or stolen, that is, in the revocation and reissue phase, our scheme verifies the user is a legitimate user or an adversary, unlike other proposed schemes that issue new credentials without verifying the user's authenticity. Our proposed scheme is scalable as unique sessions are created each time between the User and Drone through the BS. As a result, it reduces the risk of replay attacks and improves overall security.

REHAS can efficiently handle authentication failures or invalid login requests. If invalid credentials are provided by the user during the login phase, the smart device computes $C_{ui}^{*m} = h(C_{ui} \| R_2 \| \sigma_{ui}^{*})$ and verifies with C_{ui}^{m}, which is already computed and stored by the smart device during the registration phase. If it is not verified, the smart device immediately aborts the login request. As a result, it restricts the adversary from accessing the user's network. We have also addressed the trade-off between the performance and security of REHAS using the SHA-1 hash function. Although it is less secure than SHA-3, but it is best suited for drones due to its minimal key size and faster operations compared to SHA-3, which leads to lower computation and communication costs with reduced energy consumption. With the use of HECC, the security of the protocol is maintained, and the performance is not hindered. Our authentication protocol is designed to solve the novel challenge of scalability in the IoD during large-scale deployment. Our protocol supports the integration of as many drones as possible by adding more modules to the BS. Each sub-unit of the BS can be programmed differently according to modern needs in areas such as calamity response, natural disaster management, and military operations. Our proposed scheme also addresses the resource constraints of drones and effectively handles authentication failures or invalid requests. REHAS efficiently mitigates the smart device stolen attack, preventing an attacker from issuing fresh credentials without proper authentication.

The proposed REHAS includes the following as contributions:

- A Robust and Efficient Hyperelliptic Curve-Based AKA Scheme for IoD has been proposed in this paper. The REHAS utilizes a Fuzzy Extractor, HECC, XOR operation, and Collision-resistant one-way Hash Function (SHA-1). The proposed scheme establishes a secure wireless connection between the User and the Drone.
- Considering the resource constraints in drones, The REHAS employs a Hyper Elliptic curve with a maximum key size of 80 rather than using 160 bits Elliptic curve key size.

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- Hyper Elliptic curve multiplication is employed to generate public key and private key pair, which also utilizes the assumption of Computational Diffie–Hellman Problem (CDH) and Discrete Logarithm Problem (DLP) in preserving the secrecy of the private key.
- A fuzzy Extractor is employed to generate the biometric traits of the user, such as a key that can be regenerated to prevent data exposure in the event of a smart device being stolen.
- Formal security verification is performed under Scyther simulation, which validates the resilience of REHAS against various cyber attacks.
- The safety and resilience of the proposed scheme are assured by the Informal security analysis for various adversarial threats.
- Lastly, The communication overhead, computational cost, and energy consumption of REHAS were computed, and comparative analysis was performed with other related AKA schemes, which demonstrated the superiority of REHAS in terms of security features and lower costs.
- Our proposed REHAS scheme supports dynamic drone addition, allowing for large-scale IoD deployment. It assumes that BSs have sufficient capacity to manage this scale and ensures that each connection has a unique session key, preventing interference between multiple sessions. Additionally, the scheme is lightweight, with low communication overhead (1696 bits) and low energy consumption (22.5 mJ), making it well-suited for efficient scaling in large IoD networks without significantly increasing overall overhead.

1.3 | Paper Organization

The paper is structured as follows. Section 2, analyses the other AKA protocols in IoD. Section 3 describes System models, that is, Network and Adversary models. In Section 4, The preliminaries are discussed, that is, functionality and security requirements, HECC, Fuzzy Extractor and Complexity assumptions, In Section 5, The operative stages of the REHAS are illustrated in detail. In Section 6, both formal and informal security analysis of REHAS has been performed. Section 7 presents a comparative assessment of REHAS. In Section 8, the paper concludes.

2 | Related Work

In recent years, providing a secure and robust communication system in the IoD network has been widely researched. The drones communicate with each other and the users over the wireless insecure channel and share their sensitive data. As a result, this exposes the network to cyber-attacks. Such internet of drone privacy and security issues are discussed in detail by Yahuza et al. [7]. In 2014, Turkanović et al. [10] propounded an AKA scheme for heterogeneous WSNs. However, according to Challa et al. [11], the scheme recommended by Turkanović et al. [10] is defenseless to some security attacks like user impersonation attacks, offline password guessing attacks, stolen smart device attacks, and replay attacks. Porambage et al. [12] propounded a two-phase authentication system for WSN, but according to Challa et al. [11], this system is exposed to MITM, stolen lost device revocation, and user impersonation attacks. Berini et al. [13], propounded a lightweight AKA protocol with HECC, and XOR operation, Berini et al. [13] claims that the authentication scheme as earlier mentioned [11] is sensitive to Modification attack and GSS impersonation attack and has higher communication overhead and computation cost. Ever [14] designed a secure framework in IoD using ECC, Symmetric public key encryption, One-way hash functions, and Bilinear pairings, but [11] claims that Ever [14] is defenseless to ESL attacks, un-traceability, data integrity, GSS impersonation attacks, and so forth, and has high computation costs. Hussain et al. [15] propounded a lightweight AKA scheme for smart city surveillance using a one-way hash function and symmetric public key encryption and verified the scheme using BAN logic. However, according to some authors, his scheme is defenseless to impersonation attacks and insider attacks. Rana et al. [16] suggested a lightweight AKA scheme for IoD based on XOR operations and hash functions, but it doesn't preserve un-traceability and user anonymity. Additionally, it is not resilient against ESL attacks, biometric/password change attacks, and insider attacks.

Many of the authentication schemes are lightweight and computationally efficient. Still, they need to be more secure about physical drone capture attacks. Many authors uses a Physical Unclonable Function (PUF) to resist drone capture attacks in their scheme. Lounis and Zulkernine [17] designed a PUF-based AKA protocol for IoT. Alladi et al. [18] propounded a PUF-based certification scheme for remote Hovering Devices. Their scheme involves the computation of two session keys. As a result, they claimed that their protocol can combat attacks like MITM, capture attacks, and replay attacks. PUF is extensively used to enhance the secrecy and robustness of cyber security protocols. Sharma et al. [19] proposed a PUF-based authentication scheme for IoD, which was resilient to various threats. However, the scheme did not address how a user could issue new credentials in the event of a lost smart device, as there was no discussion of a revocation and reissue stage for re-authenticating the user.

In 2018, Semal et al. [20] propounded an authorization scheme that used the approach of certificate key agreement generation and authentication. This ensures the identity, integrity, and authenticity of the messages transmitted within the protocol. This led the protocol to use higher communication overhead. This problem was observed by Malani et al. [21], who propounded a scheme that used a hash function and ECC encryption technique. These two techniques help the entities authenticate each other. However, it is found that this protocol is not resistant to device secrecy and device impersonation attacks.

Wazid et al. [22] suggested an AKA protocol that employed SHA-160 and Exclusive-OR (XOR) functions. This scheme has limitations regarding user impersonation attacks and fails to resist insider attacks. Jangirala et al. [23] in the year 2019 used XOR, an advanced encryption standard, SHA in his scheme. They used ROM to validate the security of the propounded protocol. This protocol used a credential-based AKA scheme, which checks the validity of the user and then creates a private and safe passage or channel between the DRN_1 and DRN_2. However, it failed to ensure the indistinguishability of encryption, intractability, and

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

anonymity features. Then, Qureshi et al. [24] propounded a verification protocol for UAVs. In that scheme, they suggested an algorithm that ensured the security of assets, the packet drop, latency, and computational overhead cost. The major disadvantage of this AKA protocol is the need for more distance among UAVs. Recently, Tanveer et al. [25] proposed SEAF-IoD, a user drone authentication framework for the IoD using PUF, Fuzzy Extractor, and XOR function. Their proposed scheme is secure but does not discuss the revocation and reissue phase, nor the smart device stolen attack. In the same year, PAF-IoD was proposed by Tanveer et al. [26] using SHA, AEAD, XOR, and Fuzzy Extractor, but there was no discussion about the revocation and reissue phase and the smart device stolen attack. Pu proposed two schemes for IoD: a lightweight and privacy-preserving AKA protocol [27] in 2022, and the Featherweight AKA scheme [28] in 2023. Both schemes do not discuss user drone authentication, and Pu et al. [27] also does not perform any formal security verification of the proposed protocol. Conclusively, Table 1 compares other existing authentication schemes in the IoD environment.

3 | System Model

In this Section, We describe the system model of the REHAS.

3.1 | Network Model

REHAS involves three entities, that is, user (Ui), drone (DRNj), and BS. Figure 1 displays the network model of the REHAS. The UAVs or drones deployed collect sensitive data in the real environment using various sensors. The user (UI) can receive sensitive data from the drone (DRNj) by registering and authenticating itself through the BS. For example, if a farmer desires to get information about the physical conditions of a particular farm, he can do so by connecting to the drones deployed in the specific region. BS here works as a trusted authority, and the communication between Drone and the user happens through BS. Both entities register to BS and validate each other through BS. After successful authorization, the DRNj and Ui calculate the session key, which is used in further communication.

TABLE 1 | Analysis of other relevant AKA protocols.

Year	AKA scheme	Highlights	Drawbacks
2014	Turkanović et al. [10]	Hash Function, Bit-wise XOR	Absence of node anonymity, traceability and prone to MITM attacks, node impersonation attacks
2016	Jiang et al. [29]	Hash Function (SHA-160), XOR, ECC	Prone to ESL attacks, smart device stolen attack denial of service attacks and no un-traceability
2017	Challa et al. [11]	Hash Function (SHA-160), XOR, ECC	Prone to smart device stolen attacks, Insider attacks
2018	Wazid et al. [22]	Hash Function (SHA-160) Bit-wise XOR	Prone to user and drone impersonation attacks, traceability, session key attack stolen verifier attack.
2019	Zhou et al. [30]	Hash Function (SHA-160), XOR	Prone to session key attack, Replay attack, User Impersonation attacks
2020	Ever [14]	Billinear pairings, ECC, symmetry key encrypt/decrypt	Prone to ESL attacks, no un-traceability and anonymity and incurred high computation and communication cost.
2020	Ali et al. [31]	Hash Function (SHA-160), XOR, AES	Prone to DoS, insider attack, BS impersonation attack and freshness of key feature, no perfect forward secrecy.
2021	Hussain et al. [15]	Symm key encryption, one way hash function	Prone to impersonation attacks, Insider attacks
2021	Rana et al. [16]	XOR, SHA	Prone to bio-metric/password attacks, ESL attack, UI, Insider attacks and absence of anonymity and un-traceability features.
2022	Tanveer et al. [32]	AEGIS, ECC, XOR, Hash Function (SHA-160)	High computational and communication cost
2022	Pu et al. [27]	PUF, hash, XOR, Chaotic system	Does not support user-drone authentication, No formal security simulation tools like Scyther used
2023	Sharma et al. [19]	PUF, XOR, Hash Function	No revocation and reissue stage, no password change phase
2023	Pu [28]	Hash, XOR, PUF	Does not support user-drone authentication
2024	Tanveer et al. [25]	PUF, Fuzzy Extractor, XOR Function	No discussion about revocation and reissue stage and smart device stolen attack
2024	Tanveer et al. [26]	SHA, AEAD, Exclusive-OR, Fuzzy Extractor	Vulnerable to smart device stolen attack, no discussion about revocation and reissue stage

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

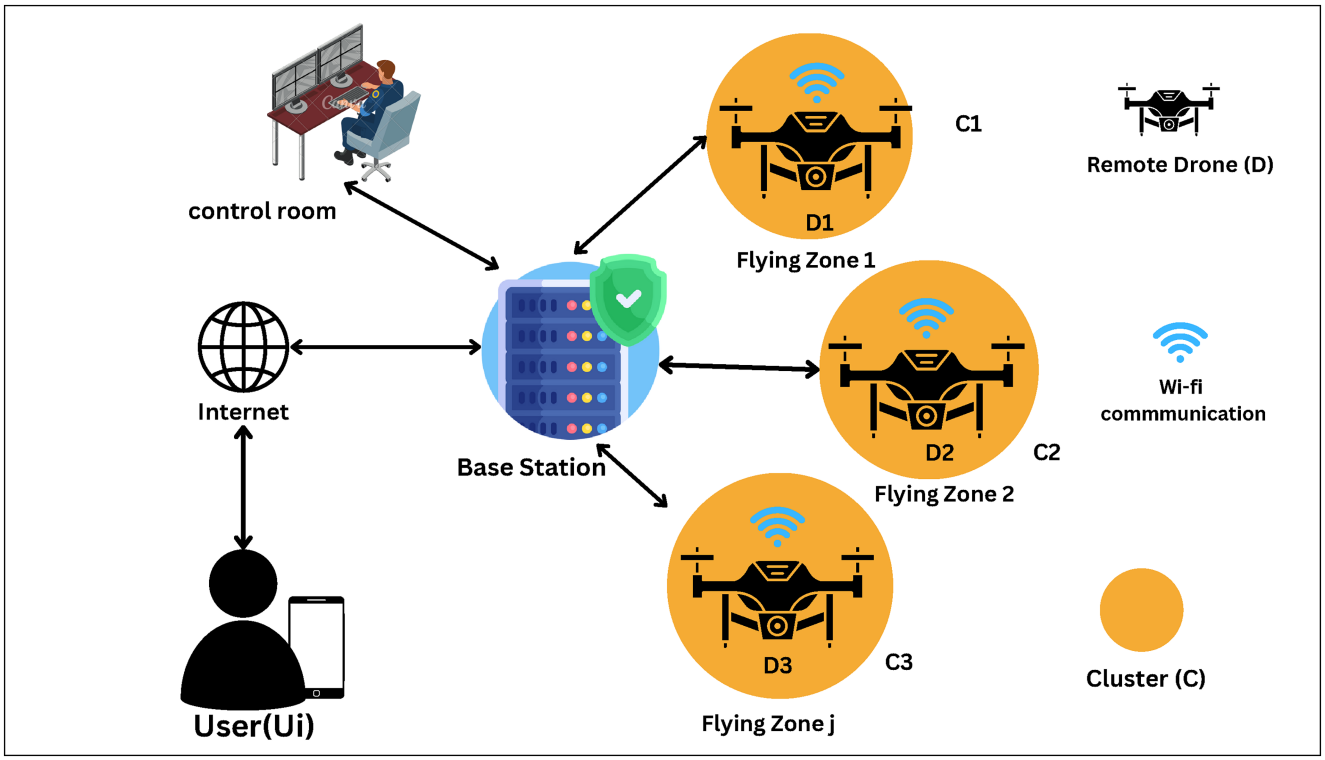


FIGURE 1 | Network model.

The control room acts as a central and trusted authority that manages communication between entities, that is, users, drones, and BSs. It also oversees and supervises all drone operations. In the event of a drone capture attack, the control room deploys a new drone, as demonstrated in the dynamic drone addition stage. Before deployment, all BSs and drones are registered with the central authority.

3.2 | Adversary Model

We also acknowledge the following threat models when designing our authentication scheme.

1. Dolev–Yao (DY) threat model

In the IoD environment, as the network entities communicate via the insecure public channel, the DY threat model [25, 33, 34] assumes that an attacker A can catch the messages exchanged on the public unsecured channel and can therefore alter, delete or insert bogus information in between the connection channel. Moreover, it is assumed in this threat model that the communication participants, that is, the UAVs, are not trustworthy in the system.

2. Canetti and Krawczyk (CK)-threat model

Under CK adversary model [25, 35] including assumptions of the DY threat model, an attacker A is capable of obtaining the session states and private credentials stored in the drone's memory. Therefore, adversary A can physically capture the drones and get the drone's private credentials by using a power analysis (PA) attack.

4 | Preliminaries

This section presents the necessary preliminary data used in the REHAS.

4.1 | Functional and Security Requirements

Various distinct functional and privacy requirements should be considered while designing a protocol in an IoD network, which are mentioned below:

- Anonymity: Real identities of the entities should be kept anonymous, that is, not to be shared on insecure public channels. So that privacy is ensured.
- Authentication: There must be mutual authentication between the entities (User, Drone, BS) to prevent impersonation attacks.
- Session key agreement: After successful authentication between the entities, that is, the drone and the user, both entities will compute a common session key, which is utilized for further safe communication. The session key should be confidential and not accessible to an attacker.
- Un-traceability: An attacker must not be able to identify any interrelation between messages transmitted by the user, drone, and BS. Therefore, Untraceability should be ensured.

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- Security verification: The proposed scheme must be formally validated using software tools like Scyther [36], ProVerif, or AVISPA. The scheme should also be robust against various cyber-attacks, which can be verified using informal security verification techniques.
- Efficiency: The scheme should be computationally, communicationally, and energy efficient.
- Robustness: The scheme should be robust against various following cyber attacks:
  1. Drone Impersonation attack (DI): Protocol should defend itself from DI attacks, that is, any attacker A cannot impersonate a registered drone and validate itself further.
  2. User Impersonation attack (UI): In the authentication protocol, adversary A should not be capable of cloning the legit user and further validating itself in the AKA process.
  3. Verifier attack: An AKA protocol should be able to defend itself from a stolen verifier attack; in other words, any adversary who can obtain the private credentials stored in the drone/BS should not be capable of calculating the session key of the non captured drones.
  4. Drone Capture attack: The scheme should be able to handle drone capture attacks, that is, if any attacker physically captures a drone and obtains the secret credentials of the drone by applying a PA attack, it should not be capable of calculating the private session key of the non captured drones.

Moreover, many other cyber attacks like MITM attacks, insider attacks, BS impersonation attacks, replay attacks and Smart device stolen attacks, and so forth. The protocol should be resilient against adversarial attacks to ensure security and privacy.

4.2 | Hyper Elliptic Curve Cryptography (HECC)

In 1988, Koblitz published an article discussing the generalized form of elliptic curves to curves with higher genus values, now known as HECC. The HECC is suitable for memory-constrained environments and small processors [37]. The HECC is well known for providing the same safety standards with a shorter key with size (i.e., 80 bits) compared to RSA and ECC with the key size 1024 bits and 160 bits respectively [38]. A HECC curve C_h for genus $g \geq 1$ over a field F is the set of solutions $(x, y) \in F \times F$ to the equation given by [37]:

$$C_h : y^2 + v(x)y = z(x) \tag{1}$$

where the $z(x)$ and $v(x)$ are polynomials. For the HECC over Fq (finite field), we will need the group order of at least $g \cdot \log_2 q \approx 2^{160}$ where g is the genus of the curve. Thus, a group order $q = 2^{80}$ is needed for genus two curves. Similarly, $q = 2^{54}$ for genus three curves and $q = 2^{40}$ for genus four curves.

The Divisor D is the formal submission of points on the hyper-elliptic curve

$$D = \sum_{p_j \in C} m_j P_j, \quad m_j \in \mathbb{Z} \tag{2}$$

Where P_j is the point on the curve and m_j represents the multiplicity of jth point P_j.

4.3 | Complexity Assumptions

4.3.1 | Assumptions of DLP

The probability of calculating a from $p = a \cdot D$ where a is arbitrary integer $a \in \{1, n-1\}$ and D is the Divisor of Hyperelliptic Curve is negligible. The problem of computing a from $p = a \cdot D$ is computationally complicated and is known as the DLP problem.

4.3.2 | Assumptions of CDH

The probability of calculating q and z from $P = qzD$, where q and z are arbitrary integers $q, z \in \{1, n-1\}$ and D is the Divisor of Hyperelliptic Curve, is negligible. The problem of computing q and z from $P = qzD$ is computationally complicated and is known as the CDH problem.

4.4 | Fuzzy Extractor (FE)

The FE is extensively used to validate users' Biometric data. FE consists of two algorithms, that is, reproduction and generation of the Bio-metric key.

4.4.1 | Gen()

Gen(.) takes the user's Bio-metrics as input and generates the reproduction parameter and user's unique Bio-metric key. Further, Gen(.) is a probabilistic algorithm.

4.4.2 | Rep()

Rep(.) takes reproduction parameter (Rp) and Bio-metric ($BIO_{u'}$) and generates user's Bio-metric key $B_{key} = \text{Rep}(BIO_{u'}, Rp)$, satisfying the necessary condition, that is, $\text{HMD}(BIO_{u'}, BIO_u) \leq E$, where HMD is the hamming distance between $BIO_{u'}$ and BIO_u and E is error tolerance.

5 | The Proposed Scheme: REHAS

In the proposed scheme REHAS, initially the entities that is, user U_i and drone *DRNj* register themselves to the BS which issues and store the secret credentials after complete registration which will be utilized for authentication later, afterward in the login and authentication phase user and drone mutually authenticate each other through the BS and then generate a unique session key which is used for secure communication. The REHAS comprises six stages: (i) The setup stage, (ii) The registration stage (drone and user registration), (iii) The login and authentication stage, (iv) The password/bio-metric change stage, (v) The dynamic drone addition stage, and (vi) revocation and reissue stage. Table 2 represents all the notations used in the REHAS.

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 | Symbol table.

Symbols	Description of the symbol
$X_{\mathrm{DRN}j}$, X_{ui}	Value corresponding to drone j and user i
ID_x	Identity of entity x
BS	Base station
$\mathrm{PW}_{\mathrm{ui}}$	Password of ith user
PB_x, PV_x	Public and private key of entity x
D	Divisor of hyper-elliptic curve
Ti	Timestamp ($1 \leq i \leq 4$)
ΔT	Threshold value for timestamp
σ_{Ui}	Bio-metric key generated by FE
τ_{ui}	Reproduction parameter generated by FE
BIO_u	Bio-metric data of user
FE	Fuzzy extractor
Gen()	FE bio-metric key generation function
Rep()	FE reproduction function
AID_x	Artificial Identity of entity x
R_1, R_2, R_3, R_4, Z_1	Arbitrary numbers
$h(.)$	Hash function
$\Vert$	Concatenation operation
$\oplus$	XOR function
$\mathrm{SS}_{\mathrm{key}}$	Session key

5.1 | Setup Stage

Before deployment, the drones and BS are registered with the control room, which acts as a central and trusted authority in our proposed scheme. The BS generates its private key and the system's public parameters during this stage. BS generates its own private credentials, specifically a private key $\mathrm{PV}_{\mathrm{BS}}$ and a public key $\mathrm{PB}_{\mathrm{BS}}$, using an asymmetric key encryption/decryption algorithm, namely HECC. These credentials are then used for the registration and authentication of users and drones.

- **Step 1** The BS picks an arbitrary number from range $\{1, \ n-1\}$ as the private key $\mathrm{PV}_{\mathrm{BS}}$.
- **Step 2** Then generates its public key $P\mathrm{B}_{\mathrm{BS}} = \mathrm{PV}_{\mathrm{BS}} \cdot D$, where $D =$ Divisor of Hyper-Elliptic curve.
- **Step 3** Finally the BS publishes the parameters $\{D, \mathrm{PB}_{\mathrm{BS}}, n = 2^{80}, h(.)\}$ publicly, where $h(.)$ is a one way hash function.

5.2 | Registration (Reg) Stage

During the Reg. stage, the BS works as a reg authority and registers all the drones and the users.

5.2.1 | Drone Registration (DR) Stage

In the DR. stage, the BS registers the drone (DRN_j) and sends the secret parameters in return, which are utilized in authentication later. Figure 2 depicts the whole drone registration process. The steps below are followed for complete registration.

- **Step 1** Firstly, the DRNj picks his Identity $\mathrm{ID}_{\mathrm{DRN}j}$ and sends to the BS over a secure channel.
- **Step 2** Then, BS picks an arbitrary number $R_1 \in n$ and calculates the following:

$$\mathrm{AID}_{\mathrm{DRN}j} = h\left(\mathrm{ID}_{\mathrm{DRN}j} \| R_1\right) \tag{3}$$

$$K_{\mathrm{DRN}j} = h\left(h\left(\mathrm{ID}_{\mathrm{DRN}j} \| \mathrm{PV}_{\mathrm{BS}}\right) \oplus h\left(\mathrm{AID}_{\mathrm{DRN}j} \| R_1\right)\right) \tag{4}$$

- **Step 3** Then BS stores $\{\mathrm{ID}_{\mathrm{DRN}j}, K_{\mathrm{DRN}j}, \mathrm{AID}_{\mathrm{DRN}j}\}$ in its database and sends $\{\mathrm{ID}_{\mathrm{DRN}j}, K_{\mathrm{DRN}j}, \mathrm{AID}_{\mathrm{DRN}j}, h(.)\}$ to DRN_j over a secure channel. DRN_j receives the message $\{\mathrm{ID}_{\mathrm{DRN}j}, K_{\mathrm{DRN}j}, \mathrm{AID}_{\mathrm{DRN}j}, h(.)\}$ from BS and stores $\{\mathrm{ID}_{\mathrm{DRN}j}, K_{\mathrm{DRN}j}, \mathrm{AID}_{\mathrm{DRN}j}, h(.)\}$ in its memory.

5.2.2 | User Registration Stage

In this stage, the BS registers the User (Ui) and sends the secret parameters to the mobile device, which will be utilized in authentication later. Figure 3 depicts the whole user registration process. The steps below are followed in this stage.

- **Step 1** The ith User Ui picks an Identity $\mathrm{ID}_{\mathrm{Ui}}$, password $\mathrm{PW}_{\mathrm{Ui}}$ and imprints his Bio-metric BIO_U in the Bio-metric sensors of the mobile device.
- **Step 2** The mobile device picks an arbitrary number $R_2 \in \mathbb{N}$, then calculates the following:

$$Gen\left(\mathrm{BIO}_U\right) = \left(\sigma_{\mathrm{Ui}}, \tau_{\mathrm{ui}}\right) \tag{5}$$

$$\mathrm{MGP}_{\mathrm{Ui}} = h\left(\mathrm{ID}_{\mathrm{Ui}} \| \sigma_{\mathrm{Ui}} \| \mathrm{PW}_{\mathrm{Ui}}\right) \tag{6}$$

Then the mobile device of Ui dispatches the parameters $\{\mathrm{ID}_{\mathrm{Ui}}, \mathrm{MGP}_{\mathrm{Ui}}, h(.)\}$ to BS over a secure conduit.

- **Step 3** After receiving $\{\mathrm{ID}_{\mathrm{Ui}}, \mathrm{MGP}_{\mathrm{Ui}}, h(.)\}$, BS picks an arbitrary number $R_3 \in n$ and then calculates the following:

$$\mathrm{AID}_{\mathrm{ui}} = h\left(R_3 \| \mathrm{ID}_{\mathrm{Ui}}\right) \tag{7}$$

$$T_{\mathrm{ui}} = h\left(\mathrm{AID}_{\mathrm{ui}} \| R_3 \| \mathrm{PV}_{\mathrm{BS}}\right) \tag{8}$$

$$\mathrm{B}_{\mathrm{ui}} = h\left(\mathrm{MGP}_{\mathrm{Ui}} \| T_{\mathrm{ui}} \| \mathrm{AID}_{\mathrm{ui}} \| R_3\right) \tag{9}$$

$$C_{\mathrm{ui}} = h\left(\mathrm{AID}_{\mathrm{ui}} \| B_{\mathrm{ui}}\right) \tag{10}$$

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

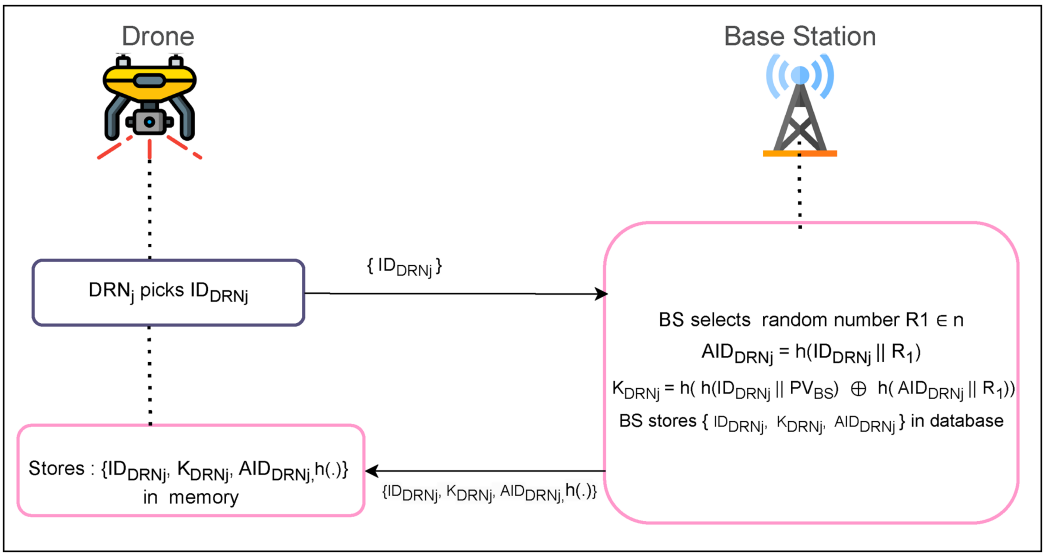


FIGURE 2 | Drone registration stage.

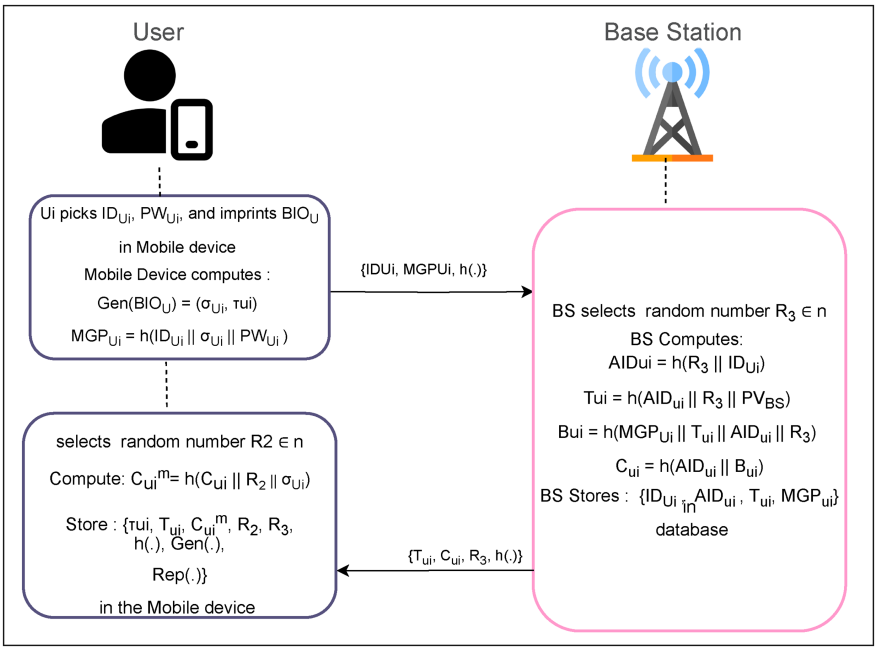


FIGURE 3 | User registration stage.

Then BS stores $\{\mathrm{ID_{Ui}}, \mathrm{AID_{ui}}, T_{\mathrm{ui}}, \mathrm{MGP_{Ui}}, R_3\}$ in its database and dispatches $\{T_{\mathrm{ui}}, C_{\mathrm{ui}}, R_3, h(.)\}$ to the mobile device over a secure channel. After obtaining $\{T_{\mathrm{ui}}, C_{\mathrm{ui}}, R_3, h(.)\}$ from BS, the mobile device computes $C_{\mathrm{ui}}^m = h(C_{\mathrm{ui}} \| R_2 \| \sigma_{\mathrm{ui}})$. Finally, the mobile device stores $\{R_2, \tau_{\mathrm{ui}}, T_{\mathrm{ui}}, C_{\mathrm{ui}}^m, R_3, h(.), Gen(.), Rep(.)\}$ in its database.

5.3 | Login and Authentication Stage

During the login and authentication stage, drone DRN_j and the user U_i authenticate each other through BS. After successful authentication, both entities compute a shared session key and communicate securely.

- **Step 1** Firstly, the User U_i enters his $\mathrm{PW_{Ui}}$ and $\mathrm{ID_{Ui}}$ into his mobile device. The U_i also imprints his Bio-metric BIO_u^* in the Bio-metric sensor of the mobile device. After this, the mobile device calculates the following:

$$\mathrm{Rep}(\mathrm{BIO}_u \tau_{\mathrm{ui}}) = \sigma_{\mathrm{Ui}}^* \tag{11}$$

$$\mathrm{MGP_{Ui}^*} = h(\mathrm{ID_{Ui}} \| \sigma_{\mathrm{Ui}}^* \| \mathrm{PW_{Ui}}) \tag{12}$$

$$\mathrm{AID_{Ui}} = h(R_3 \| \mathrm{ID_{Ui}}) \tag{13}$$

$$B_{\mathrm{ui}}^* = h(\mathrm{MGP_{ui}^*} \| T_{\mathrm{ui}} \| \mathrm{AID_{Ui}^*} \| R_3) \tag{14}$$

$$C_{\mathrm{ui}}^* = h(\mathrm{AID_{Ui}} \| B_{\mathrm{ui}}^*) \tag{15}$$

$$C_{\mathrm{ui}}^{*m} = h(C_{\mathrm{ui}}^* \| R_2 \| \sigma_{\mathrm{Ui}}^*) \tag{16}$$

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Then the mobile device verifies if $C_{ui}^{*m} = C_{ui}^{m}$. If the verification is unsuccessful, the request of user login is immediately aborted. Else, the mobile device picks a Timestamp T_1, its private key $PV_{ui} \in n$ and an arbitrary number $R_4 \in n$ and computes values mentioned below:

$$PB_{ui} = PV_{ui} \cdot D \tag{17}$$

$$F_1 = R_4 \oplus h\left(AID_{Ui} \| T_1 \| AID_{DRN_j} \| T_{ui}\right) \tag{18}$$

$$F_2 = h\left(F_1 \| R_4 \| T_{ui}\right) \tag{19}$$

The mobile device sends $\{F_1, F_2, T_1, PB_{ui}, AID_{DRN_j}, AID_{ui}\}$ to BS via a public channel.

- **Step 2** The BS receives the message $\{F_1, F_2, T_1, PB_{ui}, AID_{DRN_j}, AID_{ui}\}$ at timestamp T_2, verifies the message freshness by checking $| T_2 - T_1 | \leq \Delta T$. If it's verified, the BS validates the message from the mobile device by computing F_2^*:

$$R_4^* = F_1 \oplus h\left(AID_{Ui} \| T_1 \| AID_{DRN_j} \| T_{ui}\right) \tag{20}$$

$$F_2^* = h\left(F_1 \| R_4^* \| T_{ui}\right) \tag{21}$$

Then, the BS validates the message by verifying $F_2 = F_2^*$. The BS rejects the request and immediately aborts the session if it is not validated. Else, the BS proceeds toward the next step and computes the following:

$$S_1 = h\left(K_{DRN_j} \| T_2\right) \oplus AID_{Ui} \tag{22}$$

$$q = S_1 \oplus T_{ui} \tag{23}$$

$$S_2 = h\left(S_1 \| T_2 \| AID_{DRN_j} \| K_{DRN_j} \| T_{ui}\right) \tag{24}$$

Finally, the BS sends the authentication message $\{q, S_2, T_2, AID_{Ui}, PB_{ui}\}$ to DRN_j via a public channel.

- **Step 3** The DRN_j receives the message $\{q, S_2, T_2, AID_{Ui}, PB_{ui}\}$ at timestamp T_3, and verifies the message timeliness by checking $| T_3 - T_2 | \leq \Delta T$. If it is verified, the DRN_j validates the message from BS by computing:

$$S_1^* = h\left(K_{DRN_j} \| T_2\right) \oplus AID_{Ui} \tag{25}$$

$$T_{ui} = S_1^* \oplus q \tag{26}$$

$$S_2^* = h\left(S_1^* \| T_2 \| AID_{DRNj} \| K_{DRN_j} \| T_{ui}\right) \tag{27}$$

Then the DRN_j validates the message by verifying $S_2 = S_2^*$. The DRN_j aborts the authentication request and immediately stops the session if not validated. Else, it proceeds to the next step, picks an arbitrary number $Z_1 \in n$, its private key $PV_{DRN_j} \in n$, and computes the following:

$$PB_{DRN_j} = PV_{DRN_j} \cdot D \tag{28}$$

$$A = PV_{DRNj} \cdot PB_{ui} \tag{29}$$

$$SSkey_j = h\left(AID_{DRN_j} \| Z_1 D \| AID_{Ui} \| T_{ui} \| A \| T_3\right) \tag{30}$$

$$AUTH_{key} = h\left(SSkey_j \| T_3 \| A \| T_{ui}\right) \tag{31}$$

Finally, the DRN_j sends the following authentication message to User U_i via a public channel.

$$\left\{T_3, AUTH_{key}, Z_1 D, PB_{DRN_j}\right\}$$

- **Step 1** Upon, obtaining the message $\{T_3, AUTH_{key}, Z_1 D, PB_{DRN_j}\}$ at timestamp T_4 from DRN_j, the mobile device verifies the message timeliness by checking $| T_4 - T_3 | \leq \Delta T$. If it is verified, the U_i validates the message from DRN_j by computing:

$$A = PV_{Ui} \cdot PB_{DRN_j} \tag{32}$$

$$SSkey_i = h\left(AID_{DRN_j} \| Z_1 D \| AID_{Ui} \| T_{ui} \| A \| T_3\right) \tag{33}$$

$$AUTH_{key}^* = h\left(SSkey_i \| T_3 \| A \| T_{ui}\right) \tag{34}$$

Then the U_i validates the message by verifying $AUTH_{key}^* = AUTH_{key}$. The DRN_j discards the authentication request and immediately aborts the session if not validated. Else, the User U_i and DRN_j have finally authenticated each other. The commonly calculated session key is saved for further safe communications, and now the U_i and DRN_j can communicate and share data securely. Figure 4 represents the whole authentication stage of REHAS.

5.4 | Dynamic Drone Addition Stage

There is a probability of physical capturing of drones [35] or technical issues or low battery in drones. In this case, a new drone should be immediately deployed in that particular region. Therefore, REHAS allows dynamic drone additions by following the procedure below:-.

- **Step 1** Firstly, the DRN_j picks an ID_{DRN_j} and dispatches it to the BS via a secure channel. Then, BS picks an arbitrary number $R_1 \in n$ and computes the following:
  Figure 5 depicts the dynamic drone addition process.

$$AID_{DRN_j} = h\left(ID_{DRN_j} \| R_1\right) \tag{35}$$

$$K_{DRN_j} = h\left(h\left(ID_{DRN_j} \| PV_{BS}\right) \oplus h\left(AID_{DRN_j} \| R_1\right)\right) \tag{36}$$

- **Step 2** Then BS stores $\left\{ID_{DRN_j}, K_{DRN_j}, AID_{DRN_j}\right\}$ in its database and sends $\left\{ID_{DRN_j}, K_{DRN_j}, AID_{DRN_j}, h(.)\right\}$ to DRN_j via a secure channel.
- **Step 3** The DRN_j receives the message from BS and finally stores $\left\{ID_{DRN_j}, K_{DRN_j}, AID_{DRN_j}, h(.)\right\}$ in its memory.

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

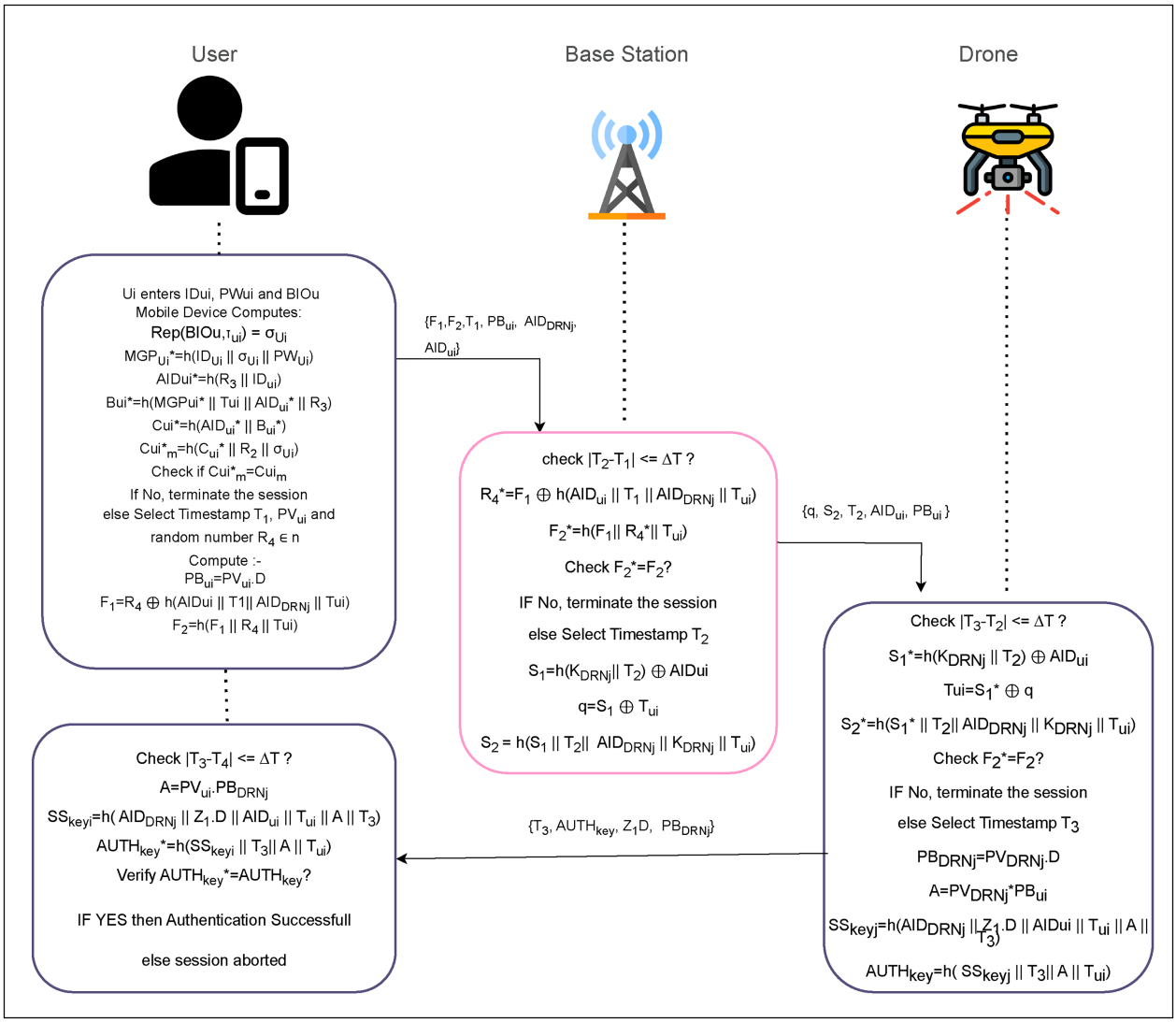


FIGURE 4 | Login and AKA stage.

5.5 | Password/ Bio-Metric Update Stage

The REHAS allows the user to change bio-metrics or password at any time. The complete procedure, step by step, is discussed below. Figure 6 depicts the Password/Bio-metric update stage.

- **Step 1** Firstly, User U_i enters his old $\mathrm{ID_{ui}}$ $\mathrm{PW_{ui}}$ and Bio-metric BIO_u in the mobile device. The mobile device validates the User U_i by computing the following:

$$\mathrm{Rep}\left(\mathrm{BIO}_u\,\tau_{\mathrm{ui}}\right) = \sigma^*_{\mathrm{Ui}} \tag{37}$$

$$\mathrm{MGP}^*_{\mathrm{Ui}} = h\left(\mathrm{ID}_{\mathrm{Ui}} \| \sigma^*_{\mathrm{Ui}} \| \mathrm{PW}_{\mathrm{Ui}}\right) \tag{38}$$

$$\mathrm{AID}_{\mathrm{Ui}} = h\left(R_3 \| \mathrm{ID}_{\mathrm{Ui}}\right) \tag{39}$$

$$B^*_{\mathrm{ui}} = h\left(\mathrm{MGP}^*_{\mathrm{ui}} \| T_{\mathrm{ui}} \| \mathrm{AID}^*_{\mathrm{Ui}} \| R_3\right) \tag{40}$$

$$C^*_{\mathrm{ui}} = h\left(\mathrm{AID}_{\mathrm{Ui}} \| B^*_{\mathrm{ui}}\right) \tag{41}$$

$$C^{*m}_{\mathrm{ui}} = h\left(C^*_{\mathrm{ui}} \| R_2 \| \sigma^*_{\mathrm{Ui}}\right) \tag{42}$$

- **Step 2** Finally, the mobile device verifies if $C^{*m}_{\mathrm{ui}} = C^m_{\mathrm{ui}}$?; if verification is unsuccessful, the mobile device aborts the login request, else asks the U_i to enter the new $\mathrm{PW_{ui}}$ or Bio-metric $\mathrm{BIO}_{\mathrm{new}_u}$. The mobile device calculates the following:

$$\mathrm{Gen}^{\mathrm{new}}_{\mathrm{BIO}_u} = \left(\sigma^{\mathrm{new}}_{\mathrm{Ui}}, \tau^{\mathrm{new}}_{\mathrm{ui}}\right) \tag{43}$$

$$\mathrm{MGP}^{\mathrm{new}}_{ui} = h\left(\mathrm{ID}_{\mathrm{ui}} \| \sigma^{\mathrm{new}}_{\mathrm{Ui}} \| \mathrm{PW}^{\mathrm{new}}_{\mathrm{ui}}\right) \tag{44}$$

$$\mathrm{AID}_{\mathrm{ui}} = h\left(R_3 \| \mathrm{ID}_{\mathrm{ui}}\right) \tag{45}$$

$$B^{\mathrm{new}}_{\mathrm{ui}} = h\left(\mathrm{MGP}^{\mathrm{new}}_{\mathrm{ui}} \| T_{\mathrm{ui}} \| \mathrm{AID}_{\mathrm{ui}} \| R_3\right) \tag{46}$$

$$C^{\mathrm{new}}_{\mathrm{ui}} = h\left(\mathrm{AID}_{\mathrm{ui}} \| B^{\mathrm{new}}_{\mathrm{ui}}\right) \tag{47}$$

$$C^{\mathrm{new}}_{\mathrm{ui}} = h\left(C^{*\mathrm{new}}_{\mathrm{ui}} \| R_2 \| \sigma^{\mathrm{new}}_{\mathrm{Ui}}\right) \tag{48}$$

- **Step 3**Finally, the mobile device changes the old C^m_{ui}, τ_{ui} with new $C^{\mathrm{new}}_{\mathrm{ui}}$ and $\tau^{\mathrm{new}}_{\mathrm{ui}}$ in its memory and sends $\mathrm{MGP}^{\mathrm{new}}_{\mathrm{ui}}$ to BS, then BS replaces the old $\mathrm{MGP_{ui}}$ with $\mathrm{MGP}^{\mathrm{new}}_{\mathrm{ui}}$ in its database.

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

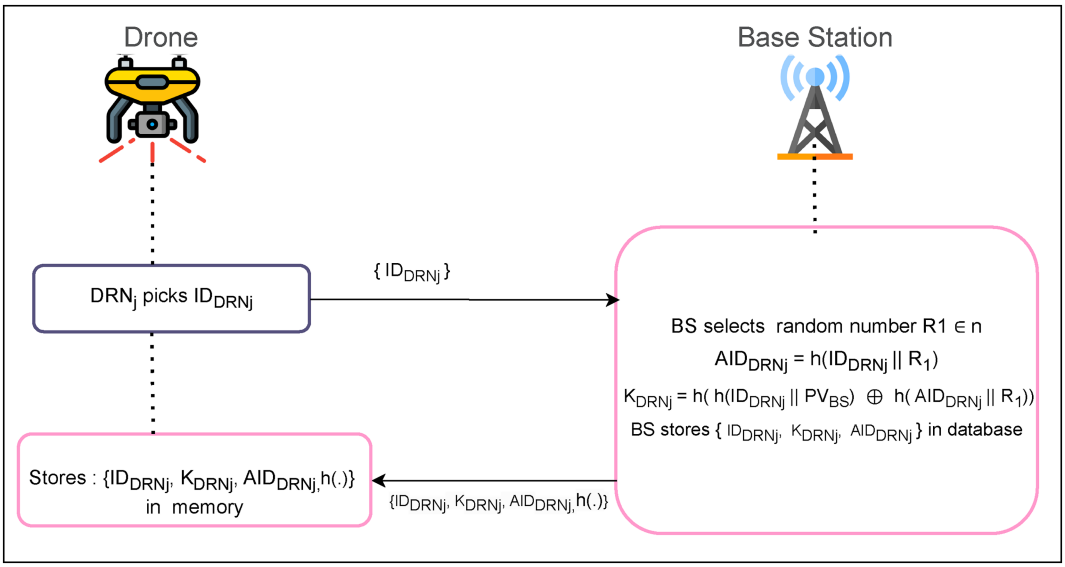


FIGURE 5 | Dynamic drone addition stage.

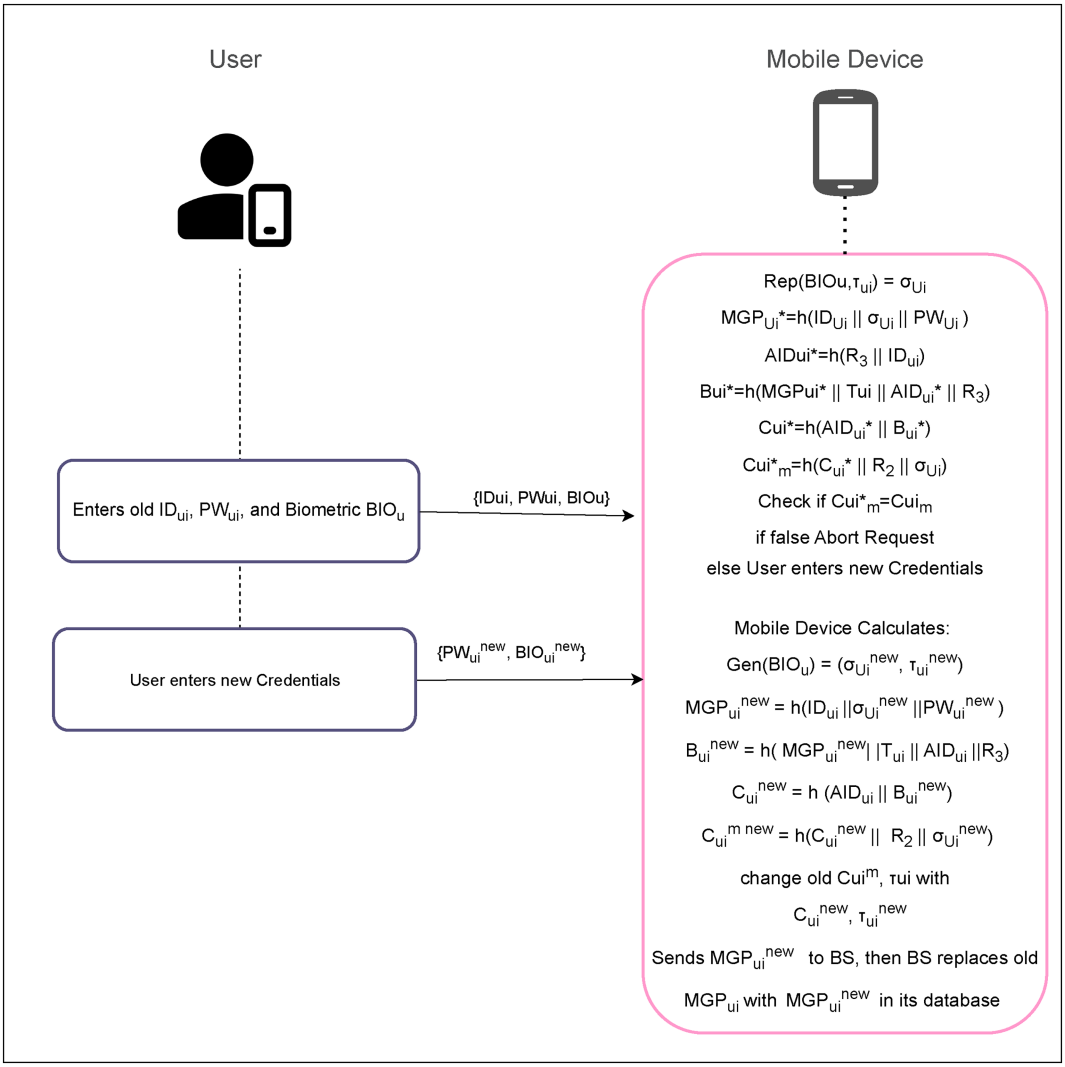


FIGURE 6 | Password/bio-metric change phase.

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

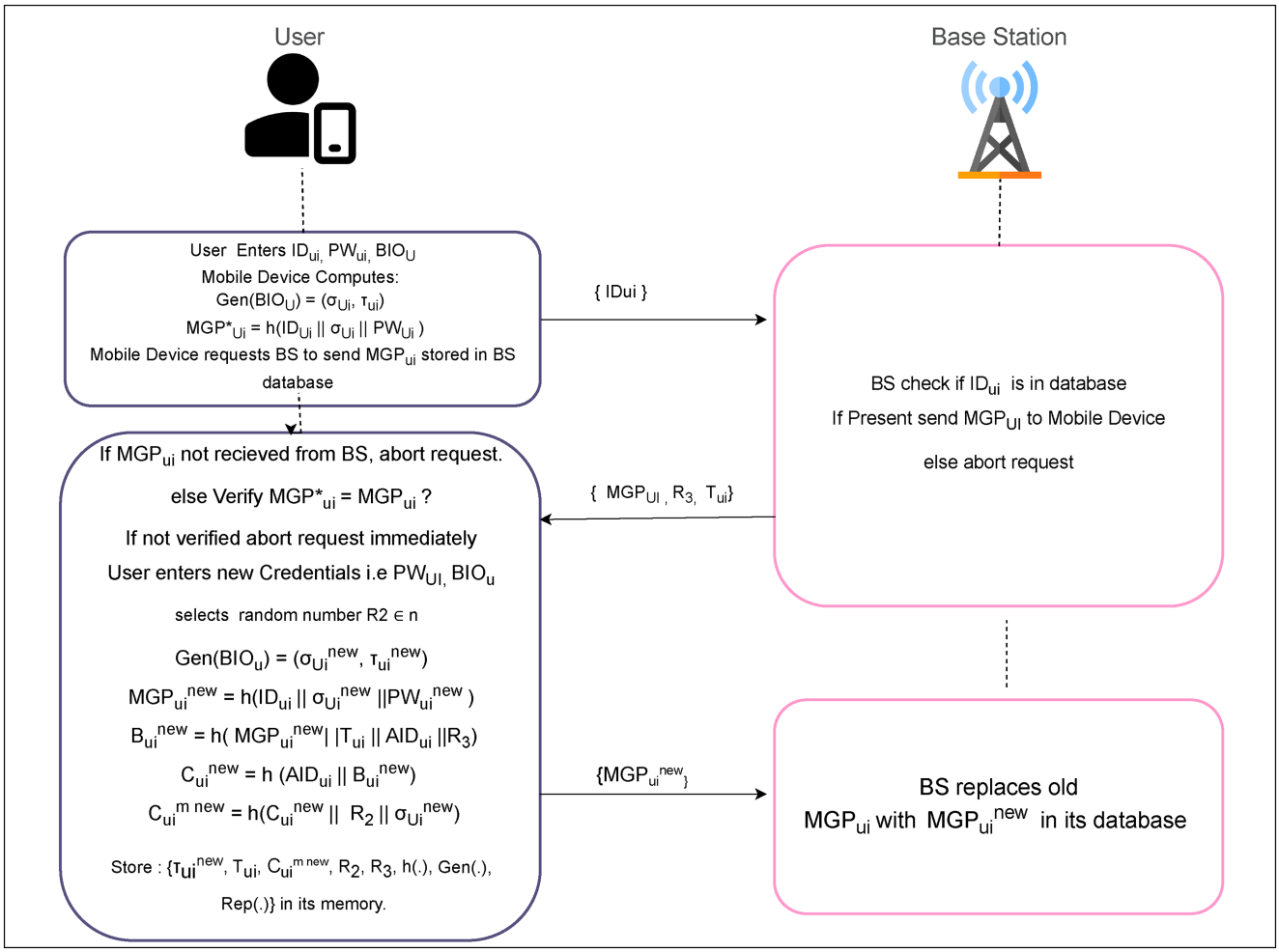


FIGURE 7 | Revocation and reissue stage.

5.6 | Revocation and Reissue Stage

If the smart device of the user is lost/stolen, new credentials are issued to the user after successful authentication, regardless of whether it is a legitimate user or an adversary issuing the new credentials. The smart device will now authenticate the user by verifying if the previously computed MGP_{ui} (stored in the BS database) matches with the freshly computed MGP^*_{ui}. If it matches, it means the actual user is reissuing the new credentials; otherwise, the request is immediately aborted. Figure 7 depicts the revocation and reissue stage.

If the mobile device of the user U_i is compromised (e.g., lost or stolen), the user can reissue new credentials by following this procedure:

- **Step 1** The user U_i enters their credentials ID_{ui}, PW_{ui}, and BIO_u. The new mobile device computes the following:

$$Gen(BIO_u) = (\sigma_{Ui}, \tau_{ui}) \tag{49}$$

$$MGP^*_{ui} = h(ID_{ui} || \sigma_{Ui} || PW_{ui}) \tag{50}$$

The mobile device sends a request to BS to obtain $\{MGP_{ui}, R_3, T_{ui}, AID_{ui}\}$ stored in the BS database.

- **Step 2** If the mobile device does not receive a message from the BS, the request is aborted because the requested ID_{ui} is invalid. Otherwise, the mobile device verifies whether $MGP^*_{ui} = MGP_{ui}$?. If the verification fails, the request is immediately aborted.
- **Step 3** If the verification is successful, the user U_i enters new credentials PW^{new}_{ui} and BIO_u. The mobile device selects a random number $R_2 \in n$ and computes:

$$Gen(BIO_u) = (\sigma^{new}_{Ui}, \tau^{new}_{ui}) \tag{51}$$

$$MGP^{new}_{ui} = h(ID_{ui} || \sigma^{new}_{Ui} || PW^{new}_{ui}) \tag{52}$$

- **Step 4** The mobile device then computes:

$$B^{new}_{ui} = h(MGP^{new}_{ui} || T_{ui} || AID_{ui} || R_3) \tag{53}$$

$$C^{new}_{ui} = h(AID_{ui} || B^{new}_{ui}) \tag{54}$$

$$C^{mnew}_{ui} = h(C^{new}_{ui} || R_2 || \sigma_{Ui}) \tag{55}$$

Finally, the mobile device stores $\{R_2, \tau^{new}_{ui}, T_{ui}, C^{mnew}_{ui}, R_3, h(.), Gen(.), Rep(.)\}$ in its database. Then the mobile device sends $\{MGP^{new}_{ui}\}$ to BS via a secure channel.

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- **Step 5** Upon receiving $\{MGP_{ui}^{new}\}$, the BS updates its database by replacing the old MGP_{ui} with MGP_{ui}^{new}.

6 | Security Analysis

This Section analyses the security and robustness of the REHAS. In subsection 6.1, Formal security analysis is performed of the REHAS under Scyther simulation [36]. In Section 6.2, an Informal security analysis of REHAS is demonstrated, which illustrates robustness against various Cyber attacks.

6.1 | Formal Security Verification Using Scyther Simulation

Similar to other verification tools like Pro-Verif, AVISPA, and so forth. Scyther is an extensively used software tool employed to verify the robustness and security of the AKA protocols. The Scyther tool considers the DY-threat model [33] and ensures the secret information is inaccessible to an adversary in the authentication process. The Scyther tool evaluated REHAS based on various claims, such as the Nisynch Claim (U, Nisynch), ensuring that the messages transmitted between entities were delivered successfully. The Alive Claim (U, Alive) claims that entity E has carried out some events. The weakagree Claim (U, Weakagree) assures that the protocol is impersonation attack-resistant. Therefore, as shown in Figure 8, REHAS has verified all the claims (as mentioned above) under Scyther simulation.

6.2 | Formal Security Verification Using Random Oracle Model (ROM)

6.2.1 | Random Oracle Model

In this segment, a formal security analysis of REHAS is performed using the ROM model. The ROM model involves an adversary A who interacts with the tth instance of an executing participant, that is, U_i, BS, DRN_j denoted by $P_{t1}U_i$, $P_{t2}BS$, and $P_{t3}DRN_j$ respectively.

Also, the ROM model presumes several queries simulating an actual attack such as Send(.), Corrupt(.), Execute(.), Test(.), Reveal(.) as mentioned in Table 3. In addition, all the participating entities, including adversary A, can access the collision-resistant one-way hash function.

6.2.2 | Session Key Semantic Security

A can execute various Test (P_x) queries on an instance of P_x with the intention of breaking the security of REHAS. Let Q represent the correct bits and Q' represent the bits guessed by A. A will win the game if $Q = Q'$ holds. The success probability is $P_b([Q = Q'])$. The advantage of A in breaching the SK security during the execution of the AKA phase of REHAS is represented in the following equation. REHAS is secure if Adv_{REHAS}^{A} is negligible under ROM.

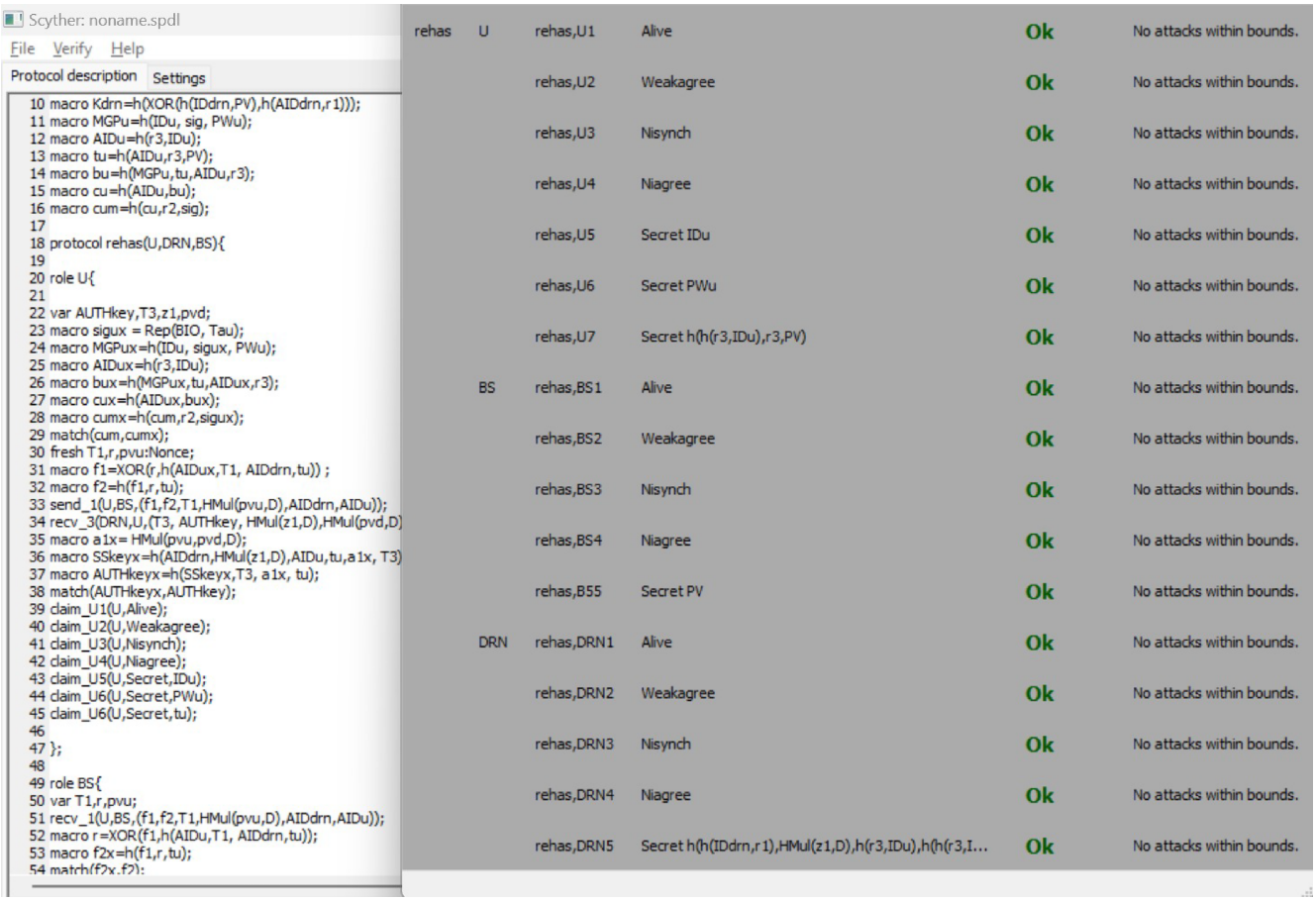


FIGURE 8 | Formal security verification using Scyther simulation.

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 | Queries in the random Oracle model.

Query	Description
Send (P_t, MSG)	Using this query, Adversary A can execute an active attack by transmitting the message MSG to P_t and receiving the response from P_t accordingly.
Reveal (P_t)	By applying this query, A can obtain the session key SK established during the AKA process between P_t and its partner.
Execute $(P_{t1}U_i, P_{t2}\text{BS}, P_{t3}\text{DRN}_j)$	This query enables A to eavesdrop all the communication messages exchanged between the network entities. It returns the message exchanged among participants.
Test (P_t)	A can check if the established SK is real or a random output by using this query. After receiving this query, C flips an unbiased coin b, and returns the actual SK if $b = 1$, else $(b = 0)$ returns a random value to A.
Corrupt $(P_{t1}U_i)$	Using this query, A can obtain the secret credentials which are stored in MD_{ui} memory by utilizing the power analysis attack.
Corrupt $(P_{t3}\text{DRN}_j)$	Using this query, A can obtain the secret credentials which are stored in DRN_j memory by utilizing the power analysis attack.

$$\text{Adv}^{A}_{\text{REHAS}} = \mid 2 \cdot \text{Prob}\left[Q = Q'\right] - 1 \mid .$$

Theorem 1. *Suppose the adversary A tries to compromise the SK security of REHAS within polynomial time T. Let Q_{rs}, Q_{rh}, L and | Hash | be the number of send queries, number of hash queries, length of Biometric Key, and size of the one-way collision-resistant hash function h(.). The advantage of A in breaking the SK security established between U_i and DRN_j within polynomial time is*:

Proof. The SK security is proved by utilizing the following three games, namely, GAME^A_i $(i = 1,2,3)$ using the queries mentioned in Table 3. The success probability of A to be the winner of GAME^A_i is:

$$\text{Adv}^{\text{REHAS}}_{A,\text{Game}_i} = \text{Prob}\left[\text{Success}^{A}_{\text{Game}_i}\right] \tag{56}$$

□

Game 1. In this game, A executes a real attack against REHAS in the ROM. At the start of the game, it is important for A to guess the bit Q. Hence we have,

$$\text{Adv}^{\text{REHAS}}_{A}(P_t) = \left|2 \cdot \text{Adv}^{\text{REHAS}}_{A,\text{Game}_i} - 1\right| \tag{57}$$

Game 2. In this game, A effectuates an eavesdropping attack on the message transferred in the proposed REHAS. A can intercept all the communication messages, that is, $M_1 = \{F_1, F_2, T_1, P_{Bui}, \text{AID}_{DRNj}, \text{AID}_{ui}\}$, $M_2 = \{q, S_2, T_2, AID_{U_i}, P_{Bui}\}$, and $M_3 = \left\{T_3, AUTH_{key}, Z_1 D, P_{B_{DRN_j}}\right\}$ by utilizing the Execute query described in Table 3 to execute the REHAS protocol. At the end of the game, A executes Test and Reveal queries to check whether SK is a valid or random key. In the proposed REHAS, the session key $SS_{key_i} = h\left(AID_{DRNj} \| Z_1 D \| AID_{ui} \| T_{ui} \| A \| T_3\right)$ contains both LTS, that is, T_{ui}, and STS, that is, (PV_{ui}, PV_{drnj}). A requires knowing both LTS and STS to construct a valid session key, but they are not accessible to A. Therefore, by only eavesdropping the communication messages, that is, M_1, M_2, M_3, A's winning probability is not enhanced at all in Game 2. As a result, Game 1 and Game 2 are indistinguishable. Hence we can say,

$$\text{Adv}^{\text{REHAS}}_{\text{Game}_1} = \text{Adv}^{\text{REHAS}}_{\text{Game}_2} \tag{58}$$

Game 3. In this game, adversary A executes query Corrupt $(P_{t1}U_i)$ to obtain the secret credentials stored in MD_{ui} memory, that is, $\{\tau_{ui}, T_{ui}, C_{m_{ui}}, R_3, R_2, h(.), Gen(.), Rep(.)\}$ by applying a power analysis attack. A cannot access the user's private credentials ID_{ui}, PW_{ui}, and σ_{U_i} due to the employed one-way hash function $h(.)$. Even if A is able to compute these private credentials, A can't construct a valid authentication message $M_1 = \{F_1, F_2, T_1, P_{Bui}, AID_{drnj}, AID_{ui}\}$ because A can't compute P_{Bui} due to the computational hardness of HECDLP. Also, A can't access the random number $R4$ used to construct F_1 and F_2.

Even if the attacker performs a hash query, no collision will happen. Therefore, Game 2 and Game 3 are indistinguishable. Consequently, due to the birthday paradox concept and HECDLP, the following results are obtained:

$$\mid \text{Adv}^{\text{REHAS}}_{A,\text{Game2}} - \text{Adv}^{\text{REHAS}}_{A,\text{Game3}} \mid \leq \frac{Q^2_{\text{Hash}}}{2 \mid \text{Hash} \mid} + \text{Adv}^{\text{HECCDH}}_{A}(P_t). \tag{59}$$

It is sufficient to predict the bit c to gain the *Game* once the *Test* query has been performed, and therefore we obtain the following output:

$$\text{Adv}^{\text{REHAS}}_{A,\text{Game3}} = \frac{1}{2}. \tag{60}$$

Using Equation (57) gives:

$$\frac{1}{2}\text{Adv}_A(P_t)^{\text{REHAS}} = \left|\text{Adv}^{\text{REHAS}}_{A,\text{Game1}} - \frac{1}{2}\right| \tag{61}$$

Simplifying the Equations (57–59), and using the triangular inequality result in the following derivation from Equation (60).

$$\frac{1}{2}\text{Adv}_A(P_t)^{\text{REHAS}} = \mid \text{Adv}^{\text{REHAS}}_{A,\text{Game1}} - \text{Adv}^{\text{REHAS}}_{A,\text{Game3}} \mid \tag{62}$$

$$= \mid \text{Adv}^{\text{REHAS}}_{A,\text{Game2}} - \text{Adv}^{\text{REHAS}}_{A,\text{Game3}} \mid \tag{63}$$

$$\leq \frac{Q^2_{\text{Hash}}}{2 \mid \text{Hash} \mid} + \text{Adv}^{\text{HECCDH}}_{A}(P_t). \tag{64}$$

The final result is obtained by multiplying both sides of the Equation (63) by "2" as follows:

$$\text{Adv}_A(P_t)^{\text{REHAS}} \leq \frac{Q^2_{\text{Hash}}}{\mid \text{Hash} \mid} + 2\text{Adv}^{\text{HECCDH}}_{A}(P_t). \tag{65}$$

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

6.3 | Informal Analysis

Analysis of REHAS for several functional and security features is performed within this subsection.

6.3.1 | Authentication

In REHAS, the mobile device validates the login request by checking if $C_{ui}^{*m} = h(C_{ui}\|R_2\|\sigma_{Ui})$ equals C_{ui}^{m}; if so, authentication succeeds, otherwise login request fails. Then BS validates the message $\{F_1, F_2, T_1, PB_{ui}, AID_{DRNj}, AID_{ui}\}$ received from the user by verifying $F_2 = F_2$? Similarly, Drone (DRN_j) validates the message $\{q, S_2, T_2, AID_{Ui}, PB_{ui}\}$ received from BS by calculating S_2^* and verifying $(S_2^* = S_2)$. If yes, then BS is validated. Otherwise, the session is immediately aborted. Finally, User (Ui) verifies the authentication message $\{T_3, AUTH_{key}, Z_1D, PB_{DRNj}\}$ received from the drone (DRN_j) by computing the $AUTH_{key}^*$ and verifies it with the $Auth_{key}$ received from a drone (DRN_j). Hence, Mutual authentication is present in the REHAS.

6.3.2 | Anonymity

In REHAS, the user (Ui) and Drone (DRN_j) original identities are anonymous. The ID_{ui} that is, identity of (Ui) is encrypted into artificial identity $AID_{ui} = h(R_3\|ID_{ui})$. Similarly, the identity of DRN_j is also encrypted as $AID_{DRNj} = h(ID_{DRNj} \| R_1)$, where R_1 and R_3 are arbitrary secret numbers. The ID_{ui} of the user and ID_{DRNj} of *Drone* are kept secret and not shared over the insecure channel. AID_u and AID_{DRN} are the artificial identities that are shared and used in the whole authentication protocol. Due to the one-way hash function, it becomes arduous for an adversary A to get original identities or secret information. Thus, anonymity is ensured in REHAS.

6.3.3 | Session Key Agreement

After the *DRNj* and the *Ui* have mutually validated, they both calculate the session key that is, $SS_{keyi} = h(\mathrm{AID}_{\mathrm{DRN}j}\| Z_1D\|\mathrm{AID}_{\mathrm{ui}}\|T_{ui}\|A\|T_3)$. This common session key is used in further secure communication. Hence, the session key agreement is ensured in the REHAS.

6.3.4 | Un-Traceability

We have used a set of arbitrary numbers such as R_4 and Z_1 and timestamps T_1, T_2, and T_3 which are specific to a session only. Hence, an adversary A will never be able to get any correlation between the messages transmitted between Drone, User, and BS and will not be able to trace back to the sender. Hence, REHAS provides un-traceability.

6.3.5 | Resistance to Insider Attacks (RIA)

In REHAS, the User U_i did not share their private credentials that is, password PW_i and Bio-metric key σ_{Ui} directly with the BS, so if an attacker is present in the BS who can access the database of BS, cannot get the secret credentials of *Ui* because of the presence of one-way collision-resistant hash function. In this way, REHAS can prevent Insider attacks.

6.3.6 | Resistance to Drone Impersonation Attacks

During the mutual authentication between DRN_j and U_i, the user validated the drone by computing $AUTH_{keyj}^* = h(SS_{keyi}\|T_3\|A\|T_{ui})$ and then verifies it with $AUTH_{keyj}$ which was computed by the drone. So, suppose an adversary tries to impersonate the actual drone. In that case, he will be unable to compute the actual $AUTH_{key}$ because he will be incapable of generating or accessing the session key $SS_{keyi} = h(AID_{DRNj}\|Z_1D\|AID_{ui}\|T_{ui}\|A\|T_3)$. Thus, REHAS is Drone Impersonation Attack-resistant.

6.3.7 | Resistance to Recognized Session-Key Specific Information Attacks

In REHAS session key is built as, $SS_{keyi} = h(AID_{DRNj}\|Z_1D\|AID_{ui}\|T_{ui}\|A\|T_3)$, which includes the arbitrary number Z_1, T_{ui}, and A. Suppose an attacker R can guess the arbitrary number. In that case, he will not be able to access T_{ui}, which is formed using the private key of BS. Thus, the adversary can't compute a valid session key.

6.3.8 | Resistance to User Impersonation Attacks

When the BS validates the user, it computes $F_2^* = h(F_1\|R_4\|T_{ui})$ and verifies it with the F_2 which is calculated by the user. Suppose an adversary attempts to pretend to be an actual user and construct valid messages that include F_2. In that case, it is computationally impractical for an adversary A to access the private credential of U_i, that is, T_{ui}. Thus, the BS can easily distinguish between the actual User and the adversary User.

6.3.9 | Resistance to BS Impersonation Attacks

DRN authenticates the message received from BS by computing $S_2^* = h(S_1\|T_2\|AID_{DRNj}\|K_{DRNj}\|T_{ui})$. If an intruder attempts to impersonate the actual BS, he needs to construct a valid message and send it to DRN_j to get validated finally. However, an attacker will not be able to calculate valid S_2 because of no access to secret credentials, that is, K_{DRNj} and T_{ui}. To compute K_{DRNj} and T_{ui} in polynomial time, an attacker has to obtain the private key of BS, which is not available to A. Thus, the REHAS is BS impersonation attack resistant.

6.3.10 | Resistance to the Session Key Attacks

The session key is formed using an arbitrary number Z_1, A and timestamp T_3, as $SS_{keyi} = h(AID_{DRNj}\|Z_1D\|AID_{ui}\|T_{ui}\|A\|T_3)$. The intruder cannot access the current session key from the previous session key due to the presence of an arbitrary number Z_1, A and timestamp T_3 which are unique for different sessions. Hence, REHAS is session key attack resistant.

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

6.3.11 | Resistance to MITM Attacks

The authentication messages between the User, Drone, and BS are shared between them on a public channel, so MITM attacks are probable. An attacker can access the authentication messages in between and manipulate/modify them so the receivers believe that they are receiving the messages from the actual sender. For example, when the user U_i sends the message $\{F_1, F_2, T_1, PB_{ui}, AID_{DRNj}, AID_{ui}\}$ to BS on a public channel if an intruder attempts to alter the contents of the message, he can't do so because of no knowledge of T_{ui}, that is, private credential information of U_i. Similarly, when the BS sends the authentication message $\{q, S_2, T_2, AID_{ui}, PB_{ui}\}$ to the drone, the adversary A does not have access to K_{DRNj} and T_{ui}. Similarly, between the drone and the user $\{T_3, AUTH_{key}, Z_1 D, PB_{DRNj}\}$, the attacker can't access the $AUTH_{key}$ which is formed using SS_{key} A and T_{ui}, which are further formed using the secret parameters of BS and U_i. Thus, the REHAS is MITM attack-resistant.

6.3.12 | Resistance Against Denial of Service (DoS) Attacks

When the user enters his $\mathrm{ID_{ui}}$, $\mathrm{PW_{ui}}$, and Bio-metric BIO_U in his mobile device, the device validates the user by computing $C^{*m}_{\mathrm{ui}} = h(C_{\mathrm{ui}} \| R_2 \| \sigma_{\mathrm{Ui}})$ and then authentication whether $C^{*m}_{\mathrm{ui}} = C^{m}_{\mathrm{ui}}$. If a user provides an invalid password, identity, or Bio-metric, the mobile device will not be able to verify the user, will not send any authentication request further to BS, and will immediately abort the login request. Thus, REHAS is DoS attack-resistant.

6.3.13 | Resistance to Replay Attacks

At each authentication step, the received message is checked to determine whether it is replayed by an attacker or not. If the gap between the timestamp of sending message T_1 and the receiving of message T_2 is smaller than or equal to $\Delta T\left(|T_1 - T_2| \leq T\right)$, that is, the maximum time threshold for receiving messages, further authentication continues, or else the session is immediately aborted. This prevents any alteration or modification to the authentication messages. Therefore, REHAS is replay attack-resistant.

6.3.14 | Drone Capture Attacks

In the IoD, it is not feasible to observe all the drones all the time. So, there is a possibility an adversary physically captures some drones and extracts all the crucial information from the Drone's database. However, through the extracted data from the attacked or compromised drones, the adversary cannot get the sensitive information of the non-compromised drones because the sensitive information for all the drones is generated with unique attributes. Moreover, the adversary cannot calculate the SS_{key} of the captured drones by the extracted information because the adversary will be unable to compute A due to the assumption of the CDH problem.

6.3.15 | Resistant to Robbed/Lost Mobile Device Attacks

If the User's U_i mobile is misplaced or robbed, with the help of power analysis attacks the secret information from the mobile's database can be retrieved easily. The private information of U_i is not directly stored in the database due to the use of the one-way hash function. In this way, the attacker can never get any of the $\mathrm{ID_{ui}}$, $\mathrm{PW_{ui}}$, BIO_u. Additionally, an adversary will not be able to issue new credentials from the user's smart device without first authenticating with the user's existing credentials. Based on the above reasons, it is ensured that REHAS is stolen smart device attack-resistant.

7 | Performance Evaluation

This section is on efficiency evaluation of the REHAS and then the detailed comparison is performed with other proposed AKA schemes, that is, Challa et al. [11], Tanveer et al. [39], Jangirala et al. [40], Sutrala et al. [41], Tanveer et al. [25], Tanveer et al. [26] in terms of total computation cost, total communication overhead and total energy expenditure.

7.1 | Security Attributes Comparison

A comparative analysis with various other authentication schemes concerning the functional and security features is done within this section. The results are depicted in Table 4, demonstrating that the REHAS fulfills all the security and functional features against other schemes [11, 25–28, 39–41]. Symbols used in Table 4 are: Supported, —: Not Supported, and abbreviation *SA* denotes Security Attributes.

7.2 | Computation Cost and Energy Consumption Comparison

This subsection examines the computation cost and total energy expenditure comparison of the REHAS with other related schemes. The authentication algorithm includes encryption functions such as the Concatenation operation, XOR operation, One-way hash function, HECC multiplication operations, and comparison operation. Concatenation, comparison, and XOR operation are negligible compared to other operations. To evaluate the computation cost and compare it with other AKA protocols, we utilize the experimental data as reported in [13, 42]. Kilinc and Yanik [42] estimated the running time of various cryptographic operations by using PBC library version (0.5.12) [43] using ubuntu 12.04.1 LTS 32bit operating system, 2048 MB of RAM, Intel Pentium Dual CPU E2200 2.20 GHz processor. Table 5 depicts the execution time and hash equivalents of various cryptographic operations such as Hash Function (T_h), Fuzzy Extractor Function(Rep(), Gen()) (T_{fe}), symmetric decryption/encryption (T_{sm}), AEGIS (AEAD scheme) (Tag), ECC multiplication (T_{ecm}), HECC divisor multiplication (T_{hec}), ECC point addition (T_{eca}), PUF function (T_{pu}).

In REHAS, the computation cost for the entities User, BS, and Drone is determined as $1T_{\mathrm{fe}} + 9T_h + 2T_{\mathrm{hec}} = 4.4727$ ms,

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 | Security attributes comparison.

Features	[39]	[11]	[40]	[41]	[25]	[27]	[26]	[28]	Proposed REHAS
Authentication	✓	✓	✓	✓	✓	✓	✓	✓	✓
Anonymity	✓	✓	✓	✓	✓	✓	✓	✓	✓
Session key agreement	✓	✓	✓	✓	✓	✓	✓	✓	✓
Un-traceability	—	✓	✓	✓	✓	✓	✓	✓	✓
Insider attack	✓	✓	✓	✓	✓	—	✓	—	✓
Drone impersonation attack	✓	—	✓	✓	✓	✓	✓	✓	✓
User impersonation attack	—	✓	✓	✓	✓	—	✓	—	✓
BS impersonation attack	—	—	✓	✓	✓	✓	✓	✓	✓
Session specific temporary information	—	—	✓	—	—	—	✓	—	✓
Session key attack	✓	✓	✓	✓	✓	✓	—	—	✓
Man in the middle (MITM) attacks	✓	✓	✓	✓	✓	✓	✓	—	✓
Denial of service attacks	—	✓	✓	✓	✓	—	✓	—	✓
Replay attacks	✓	✓	✓	✓	✓	✓	✓	✓	✓
Drone capture attacks	—	✓	✓	✓	✓	✓	✓	✓	✓
Mobile device stolen attack	✓	✓	✓	✓	—	—	—	—	✓
Formal verification: Scyther tool	—	✓	✓	✓	✓	—	✓	✓	✓
Dynamic drone addition stage	✓	✓	✓	✓	—	✓	✓	✓	✓
Smart card reissue and revocation stage	✓	✓	✓	—	—	—	—	—	✓
Password/bio-metric change stage	✓	✓	✓			—		—	✓

TABLE 5 | Execution times in milliseconds and their hash equivalents.

Symbols	Time (ms)	Hash equivalent
T_h	0.0023	$1T_h$
T_{ag}	0.415	$181T_h$
T_{ecm}	2.226	$968T_h$
T_{fe}	2.226	$968T_h$
T_{hec}	1.113	$484T_h$
T_{sm}	0.0046	$2T_h$
T_{eca}	0.0288	$13T_h$
T_{pu}	0.054	$24T_h$

TABLE 6 | Computation cost comparison with various similar schemes.

AKA protocols	User	BS	Drone	Total time (ms)
Challa et al. [11]	$(5T_h + 5T_{\text{ecm}} + T_{\text{fe}})$ 13.3675	$(4T_h + 5T_{\text{ecm}})$ 11.1392	$(3T_h + 4T_{\text{ecm}})$ 8.9109	33.4176
Tanveer et al. [39]	$(6T_h + 3T_{\text{ag}} + 3T_{\text{ecm}} + T_{\text{fe}})$ 10.1628	$(2T_h + 1T_{\text{ecm}} + 3T_{\text{ag}})$ 3.4756	$(3T_h + 2T_{\text{ecm}} + 2T_{\text{ag}})$ 5.2889	18.9273
Jangirala et al. [40]	$(16T_h + 5T_{\text{ecm}} + 2T_{\text{eca}} + 1T_{\text{fe}})$ 13.4504	$(11T_h + 3T_{\text{ecm}} + 1T_{\text{eca}})$ 6.7321	$(8T_h + 3T_{\text{ecm}} + 1T_{\text{eca}})$ 6.7252	26.9077
Sutrala et al. [41]	$(16T_h + 5T_{\text{ecm}} + 2T_{\text{eca}} + 1T_{\text{fe}})$ 13.4504	$(9T_h + 3T_{\text{ecm}} + 2T_{\text{eca}})$ 6.7563	$(8T_h + 4T_{\text{ecm}} + 1T_{\text{eca}})$ 8.9512	29.1579
Tanveer et al. [25]	$(5T_{sm} + 7T_h + 2T_{fe} + T_{pu})$ 4.5451	$(4T_{sm} + 3T_h + T_{fe} + T_{pu})$ 2.3053	$(3T_{sm} + 5T_h + T_{fe} + T_{pu})$ 2.3053	10.4939
Tanveer et al. [26]	$(4T_h + 5T_{ag} + T_{fe})$ 4.3102	$(T_h + 5T_{ag})$ 2.0773	$(3T_h + 2T_{ag} + T_{fe} + T_{pu})$ 3.1169	9.4544
Proposed REHAS	$(1T_{\text{fe}} + 9T_h + 2T_{\text{hec}})$ 4.4727	$(4T_h)$ 0.0092	$(2T_{\text{hec}} + 4T_h)$ 2.2352	6.7171

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

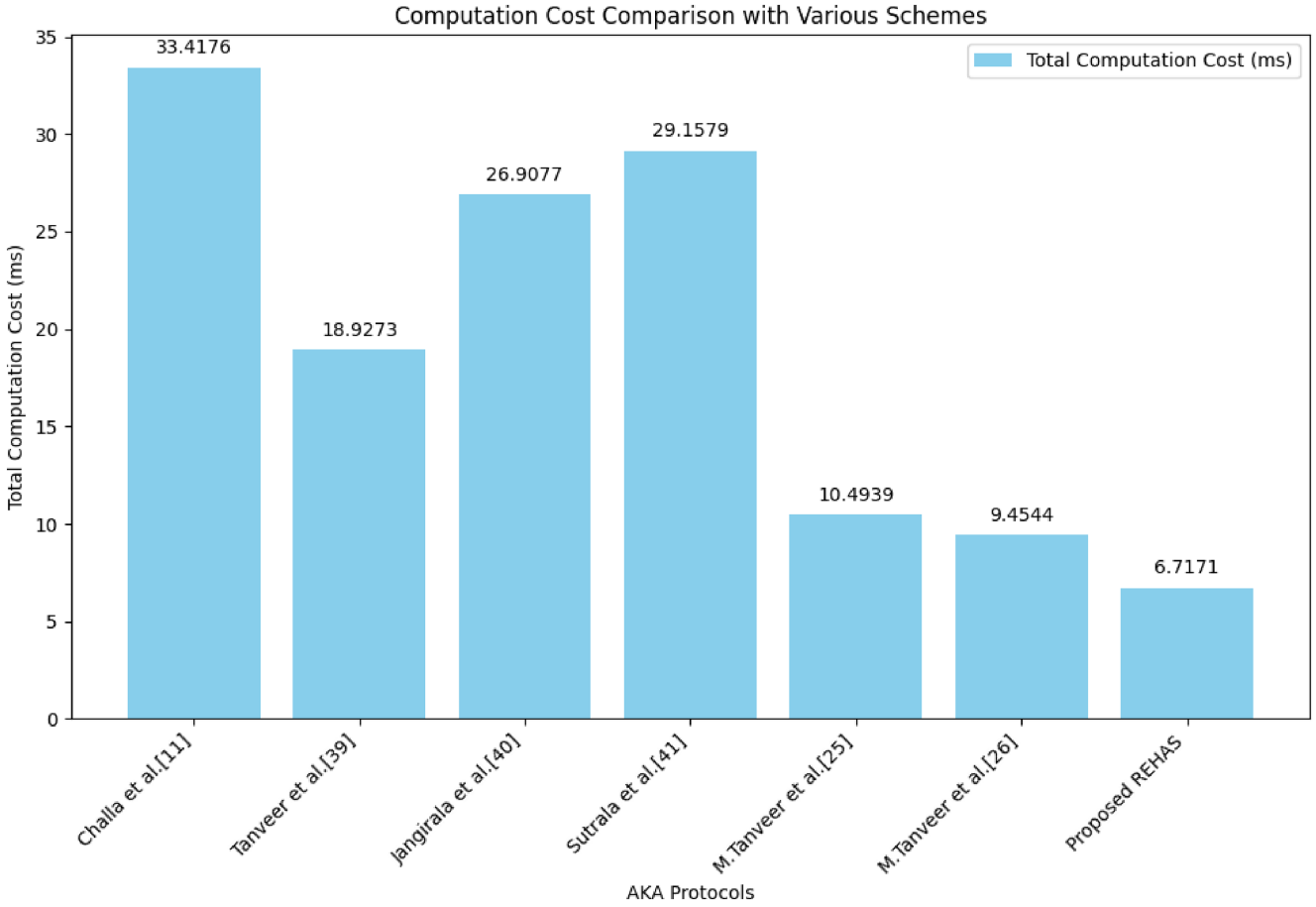


FIGURE 9 | Computation cost comparison.

$4T_h = 0.0092$ ms, and $4T_h + 2T_{\text{hec}} = 2.2352$ ms correspondingly. Following the above-calculated results, the total computational time of REHAS is 6.7171 ms. In Table 6, we analyze the computational expense of other related frameworks with the computational expense of REHAS. This comparison is further represented in Figure 9. Thus, regarding computation cost, REHAS performs better than [11, 25, 26, 39–41]. As the Drones have limited energy, [44, 45], therefore it is crucial to evaluate the energy consumption to evaluate the efficacy of the proposed protocol. According to Mehra et al. [46] the energy expended to perform one hash operation T_h is 0.00768 mJ. The total energy consumption of REHAS is equal to $2921T_h \approx 22.5$ mJ. In Table 7, we analyze the energy consumption of other related frameworks with the energy consumption of REHAS. This comparison is further represented in Figure 10 which demonstrates REHAS consumed less energy as compared to other recent AKA protocols.

TABLE 7 | Energy consumption comparison.

AKA protocols	Total hash equivalent	Total energy consumption (mJ)
Challa et al. [11]	14,532	111.2
Tanveer et al. [39]	9683	74.5
Jangirala et al. [40]	11,703	89.9
Sutrala et al. [41]	12,682	97.4
Tanveer et al. [25]	3985	30.6
Tanveer et al. [26]	4140	31.8
Proposed REHAS	2921	22.5

7.3 | Communication Overhead Comparison

Based on communication overhead, this subsection compares the REHAS with other authentication protocols [11, 25, 26, 39–41]. Here, Total Communication overhead is the aggregate cost of all the messages passed between all the participants, that is, User, Drone, and BS. To compute total communication cost, we presumed the bit sizes of ID as 160 bits, the Timestamp as 32 bits, the arbitrary number as 160 bits, the Elliptic curve point as 160 bits, the Hyper-elliptic curve divisor as 80 bits, and the size of the hash function (SHA-1) as 160. In our authentication scheme, a total of three messages are exchanged between User, Drone, and BS, that are $\text{MSG}_1 = \{F_1, F_2, T_1, \text{PB}_{\text{ui}}, \text{AID}_{\text{DRN}j}, \text{AID}_{\text{ui}}\}$, $\text{MSG}_2 = \{q, S_2, T_2, \text{AID}_{\text{ui}}, \text{PB}_{\text{ui}}\}$ and $\text{MSG}_3 = [T_3, AUTHkey, Z_1 D, \text{PB}_{\text{DRN}j}]$ whose communication costs are $[160 + 160 + 32 + 80 + 160 + 160] = 752$ bits, $[160 + 160 + 32 + 160 + 80] = 592$ bits, and $[32 + 160 + 80 + 80] = 352$ bits respectively. Therefore, the accumulated communication overhead is $752 + 592 + 352$ bits $= 1696$ bits. Table 8 and Figure 11 depicts the comparative analysis of the communication overhead of the REHAS against other similar AKA schemes, which demonstrates that REHAS requires less communication overhead than other related existing AKA protocols [11, 25, 26, 39–41].

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

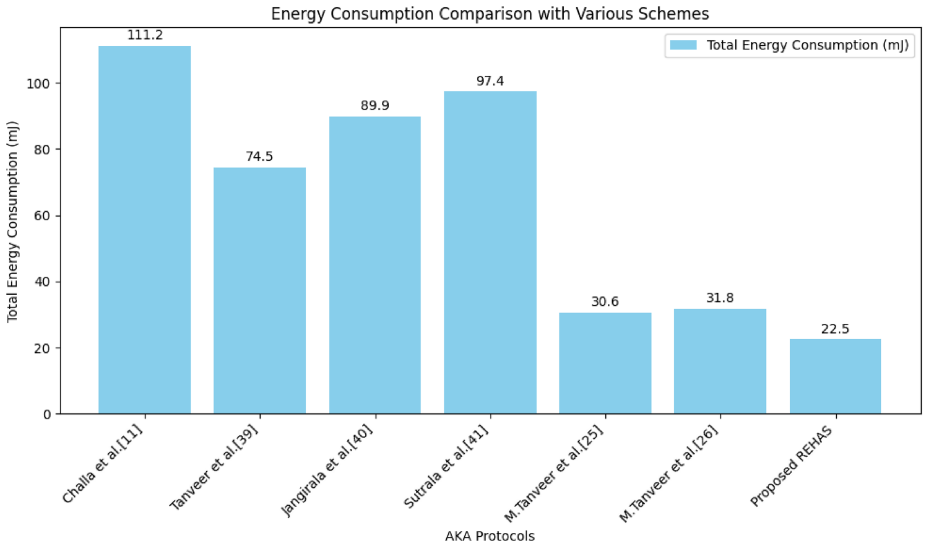


FIGURE 10 | Energy expenditure comparison.

TABLE 8 | Communication overhead comparison with various similar schemes.

Author scheme	First message	Second message	Third message	Total overhead
Challa et al. [11]	992	1024	512	2528
Tanveer et al. [39]	704	672	480	1856
Jangirala et al. [40]	1152	672	832	2656
Sutrala et al. [41]	1152	1024	1024	3200
Tanveer et al. [25]	800	800	672	2272
Tanveer et al. [26]	800	800	800	2400
Proposed REHAS	752	592	352	1696

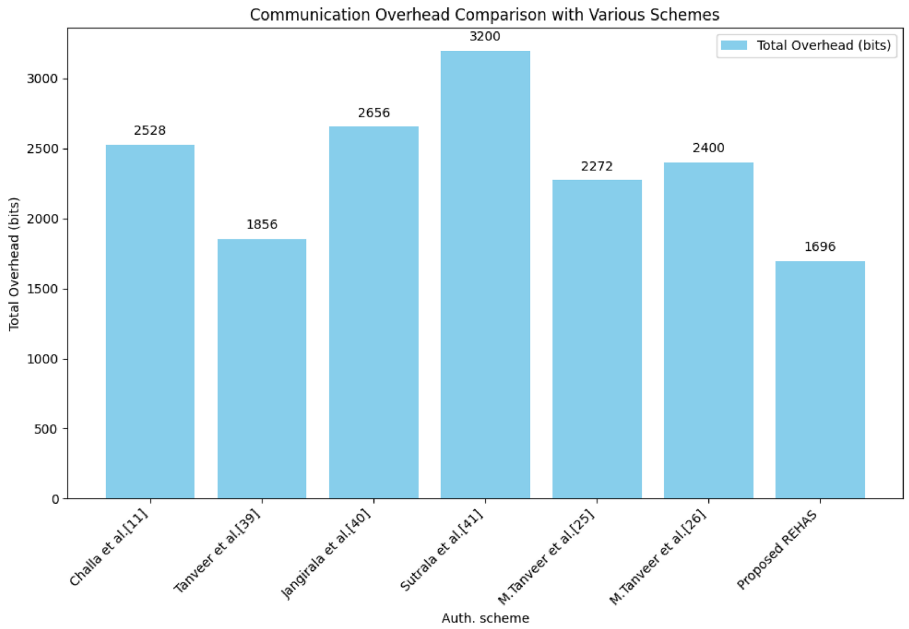


FIGURE 11 | Communication overhead comparison.

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

8 | Conclusion and Future Work

The paper proposes an efficient and robust AKA scheme REHAS in IoD using a hyper-elliptic curve. The scheme involves three participants: the User, Drone, and BS. The drone and user communicate by mutually authenticating through BS. REHAS includes functional stages such as authentication, registration, password change, revocation and reissue, and dynamic drone deployment. We also performed a formal security authentication using the software tool Scyther. Informal security analysis against various cyber attacks was also conducted, which assures the safety of secret credentials in REHAS. A comparative study in terms of computation cost, security features, and communication overhead depicts the superiority of REHAS in comparison of other AKA schemes. In the near future, we aim to integrate data storage in the REHAS using block chain technology.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

1. S. Aggarwal and N. Kumar, "Path Planning Techniques for Unmanned Aerial Vehicles: A Review, Solutions, and Challenges," *Computer Communications* 149 (2020): 270–299.

2. M. Nikooghadam, H. Amintoosi, S. H. Islam, and M. F. Moghadam, "A Provably Secure and Lightweight Authentication Scheme for Internet of Drones for Smart City Surveillance," *Journal of Systems Architecture* 115 (2021): 101955.

3. M. A. Ferrag, L. Shu, X. Yang, A. Derhab, and L. Maglaras, "Security and Privacy for Green IoT-Based Agriculture: Review, Blockchain Solutions, and Challenges," *IEEE Access* 8 (2020): 32031–32053.

4. H. S. Jennath, V. S. Anoop, and S. Asharaf, "Blockchain for Healthcare: Securing Patient Data and Enabling Trusted Artificial Intelligence. International Journal of Interactive Multimedia and Artificial," *Intelligence* 6, no. 3 (2020): 15–23.

5. S. Hussain, S. Chaudhry, O. Alomari, M. Alsharif, K. Khan, and N. Kumar, "Amassing the Security: An ECC-Based Authentication Scheme for Internet of Drones," *IEEE Systems Journal* 09, no. 15 (2021): 4431–4438.

6. P. Balaji, K. Murugan, K. Srinivasan, S. Shridevi, S. Shamsudheen, and Y. C. Hu, "Improved Authentication and Computation of Medical Data Transmission in the Secure IoT Using Hyperelliptic Curve Cryptography," *Journal of Supercomputing* 01 (2022): 78.

7. M. Yahuza, M. Y. I. Idris, I. B. Ahmedy, et al., "Internet of Drones Security and Privacy Issues: Taxonomy and Open Challenges," *IEEE Access* 9 (2021): 57243–57270.

8. L. Meng, L. Yang, W. Yang, and L. Zhang, "A Survey of GNSS Spoofing and Anti-Spoofing Technology," *Remote Sensing* 09, no. 14 (2022): 4826.

9. M. Ishiwatari, "Leveraging Drones for Effective Disaster Management: A Comprehensive Analysis of the 2024 Noto Peninsula Earthquake Case in Japan," *Progress in Disaster Science* 23 (2024): 100348.

10. M. Turkanović, B. Brumen, and M. Hölbl, "A Novel User Authentication and Key Agreement Scheme for Heterogeneous Ad Hoc Wireless Sensor Networks, Based on the Internet of Things Notion," *Ad Hoc Networks* 20 (2014): 96–112.

11. S. Challa, M. Wazid, A. K. Das, et al., "Secure Signature-Based Authenticated Key Establishment Scheme for Future IoT Applications," *IEEE Access* 5 (2017): 3028–3043.

12. P. Porambage, C. Schmitt, P. Kumar, A. Gurtov, and M. Ylianttila, "Two-Phase Authentication Protocol for Wireless Sensor Networks in Distributed IoT Applications," in *2014 IEEE Wireless Communications and Networking Conference (WCNC), Istanbul, Turkey* (2014), 2728–2733.

13. A. D. E. Berini, M. A. Ferrag, B. Farou, and H. Seridi, "HCALA: Hyperelliptic Curve-Based Anonymous Lightweight Authentication Scheme for Internet of Drones," *Pervasive and Mobile Computing* 92 (2023): 101798.

14. Y. K. Ever, "A Secure Authentication Scheme Framework for Mobile-Sinks Used in the Internet of Drones Applications," *Computer Communications* 155 (2020): 143–149.

15. S. Hussain, K. Mahmood, M. K. Khan, C. M. Chen, B. A. Alzahrani, and S. A. Chaudhry, "Designing Secure and Lightweight User Access to Drone for Smart City Surveillance," *Computer Standards & Interfaces* 80 (2022): 103566.

16. M. Rana, A. Shafiq, I. Altaf, et al., "A Secure and Lightweight Authentication Scheme for Next Generation IoT Infrastructure," *Computer Communications* 165 (2021): 85–96.

17. K. Lounis and M. Zulkernine, "T2T-MAP: A PUF-Based Thing-To-Thing Mutual Authentication Protocol for IoT," *IEEE Access* 9 (2021): 137384–137405.

18. T. Alladi, V. Chamola, and K. N. Naren, "PARTH: A Two-Stage Lightweight Mutual Authentication Protocol for UAV Surveillance Networks," *Computer Communications* 160 (2020): 81–90.

19. M. Sharma, B. Narwal, R. Anand, A. K. Mohapatra, and R. Yadav, "PSECAS: A Physical Unclonable Function Based Secure Authentication Scheme for Internet of Drones," *Computers and Electrical Engineering* 108 (2023): 108662.

20. B. Semal, K. Markantonakis, and R. N. Akram, "A Certificateless Group Authenticated Key Agreement Protocol for Secure Communication in Untrusted UAV Networks," in *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC), London, UK* (IEEE 2018), 1–8.

21. S. Malani, J. Srinivas, A. K. Das, K. Srinathan, and M. Jo, "Certificate-Based Anonymous Device Access Control Scheme for IoT Environment," *IEEE Internet of Things Journal* 6, no. 6 (2019): 9762–9773.

22. M. Wazid, A. K. Das, N. Kumar, A. Vasilakos, and J. Rodrigues, "Design and Analysis of Secure Lightweight Remote User Authentication and Key Agreement Scheme in Internet of Drones Deployment," *IEEE Internet of Things Journal* 6, no. 2 (2018): 3572–3584.

23. S. Jangirala, A. K. Das, N. Kumar, and J. Rodrigues, "TCALAS: Temporal Credential-Based Anonymous Lightweight Authentication Scheme for Internet of Drones Environment," *IEEE Transactions on Vehicular Technology* 68, no. 7 (2019): 6903–6916.

24. K. Qureshi, M. Sandila, I. Javed, T. Margaria, and L. Aslam, "Authentication Scheme for Unmanned Aerial Vehicles Based Internet of Vehicles Networks," *Egyptian Informatics Journal* 07 (2021): 23.

25. M. Tanveer, A. Aldosary, N. Kumar, and S. A. Aldossari, "SEAF-IoD: Secure and Efficient User Authentication Framework for the Internet of Drones," *Computer Networks* 247 (2024): 110449.

26. M. Tanveer, A. Aldosary, A. K. Das, S. A. Aldossari, S. A. Chaudhry, et al., "PAF-IoD: PUF-Enabled Authentication Framework for the Internet of Drones," *IEEE Transactions on Vehicular Technology* 73, no. 7 (2024): 9560–9574.

27. C. Pu, A. Wall, K. K. R. Choo, I. Ahmed, and S. Lim, "A Lightweight and Privacy-Preserving Mutual Authentication and Key Agreement Protocol for Internet of Drones Environment," *IEEE Internet of Things Journal* 9, no. 12 (2022): 9918–9933.

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

28. C. Pu, "A Featherweight Authentication and Key Agreement Scheme for Internet of Drones Applications," in *2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Toronto, Canada* (2023), 1–6.

29. Q. Jiang, J. Ma, F. Wei, Y. Tian, J. Shen, and Y. Yang, "An Untraceable Temporal-Credential-Based Two-Factor Authentication Scheme Using ECC for Wireless Sensor Networks," *Journal of Network and Computer Applications* 76 (2016): 37–48.

30. L. Zhou, X. Li, K. H. Yeh, C. Su, and W. Chiu, "Lightweight IoT-Based Authentication Scheme in Cloud Computing Circumstance," *Future Generation Computer Systems* 91 (2019): 244–251.

31. Z. Ali, S. A. Chaudhry, M. S. Ramzan, and F. Al-Turjman, "Securing Smart City Surveillance: A Lightweight Authentication Mechanism for Unmanned Vehicles," *IEEE Access* 8 (2020): 43711–43724.

32. M. Tanveer, A. Alkhayyat, A. Naushad, N. Kumar, A. G. Alharbi, et al., "RUAM-IoD: A Robust User Authentication Mechanism for the Internet of Drones," *IEEE Access* 10 (2022): 19836–19851.

33. D. Dolev and A. Yao, "On the Security of Public Key Protocols," *IEEE Transactions on Information Theory* 29, no. 2 (1983): 198–208.

34. J. Sharma and P. S. Mehra, "HCFAIUN: A Novel Hyperelliptic Curve and Fuzzy Extractor-Based Authentication for Secure Data Transmission in IoT-Based UAV Networks. Vehicular," *Communications* 49 (2024): 100834.

35. R. Canetti and H. Krawczyk, "Universally Composable Notions of Key Exchange and Secure Channels," in *Advances in Cryptology€ EUROCRYPT 2002: International Conference on the Theory and Applications of Cryptographic Techniques Amsterdam* (The Netherlands: Springer, 2002), 337–351.

36. C. J. F. Cremers, *Scyther: Semantics and Verification of Security Protocols* [Phd Thesis 1 (Research TU/e / Graduation TU/e), Mathematics and Computer Science], (Eindhoven, Netherlands: Technische Universiteit Eindhoven, 2006).

37. T. Wollinger, J. Pelzl, and C. Paar, "Cantor Versus Harley: Optimization and Analysis of Explicit Formulae for Hyperelliptic Curve Cryptosystems," *IEEE Transactions on Computers* 54, no. 7 (2005): 861–872.

38. I. Ullah, N. Ul Amin, M. Zareei, et al., "A Lightweight and Provable Secured Certificateless Signcryption Approach for Crowdsourced IIoT Applications," *Symmetry* 11, no. 11 (2019): 1386.

39. M. Tanveer, A. U. Khan, N. Kumar, and M. Hassan, "RAMP-IoD: A Robust Authenticated Key Management Protocol for the Internet of Drones," *IEEE Internet of Things Journal* 9, no. 2 (2021): 1339–1353.

40. J. Srinivas, A. K. Das, M. Wazid, and A. V. Vasilakos, "Designing Secure User Authentication Protocol for Big Data Collection in IoT-Based Intelligent Transportation System," *IEEE Internet of Things Journal* 8, no. 9 (2020): 7727–7744.

41. A. Sutrala, M. Obaidat, S. Saha, A. K. Das, M. Alazab, and Y. Park, "Authenticated Key Agreement Scheme With User Anonymity and Untraceability for 5G-Enabled Softwarized Industrial Cyber-Physical Systems," *IEEE Transactions on Intelligent Transportation Systems* 23, no. 3 (2021): 2316–2330.

42. H. H. Kilinc and T. Yanik, "A Survey of SIP Authentication and Key Agreement Schemes," *IEEE Communications Surveys & Tutorials* 16, no. 2 (2013): 1005–1023.

43. B. Lynn, *The Pairing-Based Cryptography Library* (CA, USA: Stanford, 2006).

44. J. Sharma and P. S. Mehra, "Secure Communication in IOT-Based UAV Networks: A Systematic Survey," *Internet of Things* 7, no. 23 (2023): 100883, https://doi.org/10.1016/j.iot.2023.100883.

45. J. Sharma, P. S. Mehra, D. Chawla, D. Dabas, and A. Jamshed, "Secure Communication and Authentication in IoT-Based UAV Networks," in *Network Optimization in Intelligent Internet of Things Applications* (New York, NY: Chapman and Hall/CRC, 2024), 257–273.

46. P. S. Mehra, M. N. Doja, and B. Alam, "Codeword Authenticated Key Exchange (CAKE) Light Weight Secure Routing Protocol for WSN," *International Journal of Communication Systems* 32, no. 3 (2019): e3879.

15320634, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cpe.8333 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Lepr Rev (2024) **95,** e2024067 doi:10.47276/lr.95.4.2024067

Secular trends in leprosy burden: Comparative insights from India and Brazil in the last three decades using the Global Burden of Disease study data

Shalini[a], Ravi P. Jha[b], Deepak Dhamnetiya[c], Krittika Bhattacharyya[d] & Abhishek Mehra[e]

[a]*Department of Microbiology, Maulana Azad Medical College and Associated Hospitals, New Delhi, 110002, India*
[b]*Department of Community Medicine, Dr Baba Saheb Ambedkar Medical College and Hospital, Sector-6, Rohini, Delhi, 110085, India*
[c]*Department of Community Medicine, Atal Bihari Vajpayee Institute of Medical Sciences and Dr. Ram Manohar Lohia Hospital, Connaught Place, New Delhi, 110001, India*
ORCID: https://orcid.org/0000-0002-5634-3315
[d]*Department of Statistics, University of Calcutta, West Bengal, 700019, India*
[e]*Department of Mathematics and Computing, Delhi Technological University, Rohini, Delhi, 110042, India*

Submitted 20 June 2024; Accepted 9 August 2024

Summary The objective of this study is to analyze the changes in leprosy burden in India and Brazil between 1990 and 2019. Using data from the 2019 Global Burden of Disease study, this research evaluated leprosy rates (incidence, prevalence, and DALYs) from 1990 to 2019, in India and Brazil. It investigated age and gender-specific rates in both countries, adjusting for age and providing rates with 95% uncertainty intervals. Additionally, the study illustrated the geographic distribution of prevalence across different states in India. Both India and Brazil have seen reductions in leprosy rates including incidence, prevalence, and DALYs. In India, incidence dropped from 6.04 to 1.81 cases per 100,000 between 1990 and 2019. Prevalence decreased from 73.45 to 20.00 cases per 100,000, and DALYs went down from 4.04 to 1.09 per 100,000 during the same period. In Brazil, incidence declined from 4.76 to 3.48 cases per 100,000, prevalence dropped from 26.13 to 22.20 cases, and DALYs decreased from 1.13 to 1.04 per 100,000 between 1990 and 2019. Leprosy remains a significant health

Correspondence to: Dr. Deepak Dhamnetiya, Assistant Professor (Epidemiology)
(e-mail: drdeepak.dhamnetiya@rmlh.nic.in)

© The author(s). This article is Open Access under CC BY 4.0

issue in India and Brazil. A multi-center study is essential to grasp the barriers in the existing strategies of their national leprosy eradication programs.

Keywords: Hansen's disease, incidence, India

Introduction

Leprosy (Hansen's disease) is an ancient infectious ailment caused by *Mycobacterium leprae*. It causes peripheral sensory, motor and autonomic neuropathy, which can lead to physical impairments and permanent disabilities.[1] The 44th World Health Assembly held in 1991, set the goal of global leprosy elimination as a public health problem by the year 2000 – i.e. prevalent leprosy cases reduced to less than one per 10,000 population. Of 122 leprosy endemic countries, only 108 attained the goal of elimination at the national level by the year 2000, and this number reached to 116 by the end of 2005.[2] Global elimination of leprosy has been achieved in year 2000,[3] but it continues to pose a noteworthy public health concern in both India and Brazil.[4,5]

Globally, a total of 202,256 new cases of leprosy were reported from 118 countries in 2018–19; of these, 143,787 and 29,936 cases were detected in the WHO South-East Asian Region (SEAR) and the WHO Region of the Americas (AMR), respectively. Three countries – India, Brazil and Indonesia – contribute 79% of the cases reported globally. Gender-wise distribution of the new cases shows 38.9% were females and the percentage of children was 7.4%.[6] There is uneven distribution of leprosy prevalence according to WHO regions, from 0.04 per 10,000 population in the Western Pacific Region to 0.63 per 10,000 population in Southeast Asia.[3] Since the introduction of multidrug therapy (MDT) in 1982, there has been a substantial decrease in the incidence of leprosy cases over the course of the past four decades.[7]

In India, the total number of registered cases of leprosy was nearly 4,000,000 in 1981, which decreased to 114,451 cases in 2019–20, which contributes around 80% of the cases reported from WHO SEAR; the prevalence rate reduced from >50 cases/10,000 population in 1981 to 0.57 cases/10,000 population in 2019–20. High numbers of leprosy cases were reported from Bihar, Maharashtra and Uttar Pradesh (16,595, 16,572 and 15,484, respectively). Higher prevalence rates were reported from Dadar & Nagar Haveli, Chhattisgarh and Odisha (2.61, 2.08 and 1.45 per 10,000 population, respectively).[8]

In Brazil, the total number of reported cases of leprosy were 27,863 in 2018–19, which comprises 93% of the total cases reported from WHO AMR and the prevalence rate reduced from 190 cases/10,000 population in 1985 to 1.5 cases/10,000 population in 2019–20.[6,9] In the year 2014, the rate of identification in the State of Rio de Janeiro stood at 7.36 per 100,000 individuals in the population.[10] A study conducted by the Ministry of Health in 2009 found that the States of Rondônia, MatoGrosso, Maranhão, Pará and Tocantins had high risk of leprosy transmission.[11] Leprosy cases were unevenly distributed through various regions of the country, the prevalence rates per 10,000 populations in the Midwest, the North, the Northeast, the Southeast and the South were 3.75, 3.49, 2.35, 0.61 and 0.44, respectively.[12]

Although the prevalence of leprosy has diminished in India and Brazil over the recent decades, it continues to be a significant public health challenge in both nations. The purpose of this study is to analyze the changes in leprosy burden in India and Brazil between 1990 and 2019, with the objective of monitoring the advance of leprosy eradication initiatives in these countries.

Methodology

The Global Burden of Disease (GBD) study serves as a valuable tool for comprehending the evolving health challenges faced by people around the world in the twenty-first century. Over the preceding two decades, the Institute for Health Metrics and Evaluation (IHME) has devised a methodology to quantitatively assess the impact of diseases, injuries, and risk factors, aiding health programs and policy formulation. Through GBD, consistent assessments of crucial disease burden indicators, encompassing the incidence, prevalence, and DALYs rate of leprosy, are routinely provided. This present study leveraged the GBD database from 2019 to systematically compile, analyze, and contrast the incidence, prevalence, and DALYs of leprosy, including chanages since 1990, within the contexts of India and Brazil.

Data sources

In the GBD (2019) assessment, every pertinent epidemiological measure — including incidence, prevalence, mortality, years lived with disability (YLDs), years of life lost (YLLs), and disability-adjusted life-years (DALYs) — was estimated. These estimations were computed across 23 distinct age groups and were dissaggregated by sex (males, females, and both sexes combined). Additionally, these calculations encompassed a total of 204 countries and territories, which were further organized into 21 specific regions and seven broader super-regions.To achieve this, cause-specific death rates and cause fractions are computed using the Cause of Death Ensemble model (CODEm) and spatial-temporal Gaussian process regression. The CODEm, published in detailed here facilitates the calculation of cause-specific death rates for various causes.[13–15] These rates are then adjusted to align with the total all-cause deaths, determined as part of the GBD population, fertility, and mortality estimates. Years of life lost (YLLs) are calculated by multiplying cause-specific deaths by the standard life expectancy at each age. Ensuring consistency across key metrics, such as incidence, prevalence, remission, excess mortality, and cause-specific mortality, is facilitated by the use of DisMod-MR 2.1. This Bayesian meta-regression modeling tool plays a pivotal role in harmonizing these elements for most causes. To calculate years lived with disability (YLDs) for each cause, the estimated prevalence is multiplied by disability weights assigned to mutually exclusive sequelae of diseases and injuries. Uncertainty intervals for each metric are derived by extracting the 25th and 75th ordered 1000 draw values from the posterior distribution, providing a comprehensive and robust assessment of uncertainty in the GBD 2019 estimates.[16] The key sources of data to model the cause of disability due to leprosy in India included the Global Leprosy Update-weekly Epidemiological record by WHO for various years, the Indian National sample survey on Leprosy burden, the Indian National Leprosy Eradication programme Annual Report for various years and published scientific articles. The primary data sources employed to construct the model for quantifying the disability attributed to leprosy in Brazil encompassed the Brazil Information System for Notifiable Diseases - Leprosy for multiple years, as well as the Global Leprosy Update - a weekly publication in the Epidemiological Record by the World Health Organization.[16] Within the GBD framework, leprosy is characterized as a chronic bacterial infection caused by *Mycobacterium leprae*, primarily affecting the nervous system, skin, respiratory tract, and eyes. The International Classification of Diseases, 10th Revision (ICD-10) code for Leprosy is A30.9.

Information regarding the various rates associated with Leprosy was gathered through an openly accessible online resource created by the IHME known as the Global Health Data

Exchange (GHDx) query tool (https://vizhub.healthdata.org/gbd-results/).[17] We have utilized the data gathered from GBD 2019 study and analyzed it to see the trends in the incidence, prevalence and DALYs rate over the period 1990 to 2019 for India and Brazil. Gender-specific and age-specific incidence, prevalence and DALYs rates have also been compared within India and Brazil. This study also maps out the geographical distribution of leprosy prevalence in different states within India and Brazil.

Results

Overall, the incidence, prevalence and DALY rate of leprosy have fallen in both India and Brazil. Nevertheless, the fall is more pronounced in India when compared to Brazil. Regarding the overall count of new leprosy cases, prevalent leprosy cases and total number of DALYs attributed to leprosy in India, a decrease from 1990 to 2019 was observed. However, Brazil encountered an elevation during the same period (Table 1).

Age Standardized Incidence rates of leprosy in India continuously decreased from 1990 to 2019, whereas in Brazil, the decrease in incidence rates is less during the same period. Age Standardized prevalence rates of leprosy in India gradually fall from 1990 to 2005 and then shows almost constant trend till 2017, with a further fall from 2017 to 2019, reaching 20.00 cases per 100,000 in 2019, from a level of 73.45 in 1990. In Brazil, prevalence rates of leprosy show an almost constant trend over the past three decades, except for a slight increase in the period of 2003 to 2009. Age Standardized DALYs rates for India and Brazil show similar trends as reported for prevalence rates of the respective countries (Figure 1). In India, the age-standardized incidence rates of leprosy have consistently declined from 1990 (6.04 cases per 100,000) to 2019 (1.81 cases per 100,000). On the other hand, in Brazil, the incidence rates decreased from 1990 to 2000, followed by a relatively steady trend with minor fluctuations, reaching 3.48 cases in 2019 from 4.76 in 1990.

Incidence, prevalence and DALYs rates of leprosy in India were higher than in Brazil from 1990 to 2007, almost similar to Brazil in 2008 and lower than Brazil from 2009 to 2019. In 1990, incidence, prevalence and DALYs rates per 100,000 children (0–14 years of age) in India were 0.479, 0.910 and 0.010, respectively; these indicators declined to 0.251, 0.496 and 0.005 in 2019. In Brazil, incidence, prevalence and DALYs rates 100,000 children (0–14 years of age) were 0.442, 0.865 and 0.010, respectively, in 1990; these indicators declined to 0.345, 0.679 and 0.007 in 2019 (Figure 2). In both India and Brazil, the age-standardized incidence, prevalence, and DALYs rates of leprosy have been consistently higher in males compared to females across all years spanning from 1990 to 2019. India has shown a minor reduction in the gender-based disparity in age-standardized incidence rates over the last decade. Conversely, Brazil has been unable to diminish the gender-based gap in age-standardized incidence rates over the past three decades (Figure 3). In the year 2019, the age categories of under 5, 5 to 9, and 10 to 14 exhibited lower incidence, prevalence, and DALYs rates of leprosy compared to the age group of 15 years and above, both in India and Brazil. In all age groups, the incidence rate was higher in Brazil compared to India. Concerning prevalence rates, India had a higher prevalence rate compared to Brazil in the age groups of 15 to 39 years. Conversely, in all other age groups, the prevalence rate was higher in Brazil than in India. Until the age of 54, the DALYs rate was greater in India than Brazil. However, beyond the age of 54, the DALYs rate became higher in Brazil than in India (Figure 4).

In India, female cases contribute around 39% of total leprosy cases in 1990, falling to 35.45% in 2019; in Brazil, female cases contribute 31.45% of total leprosy cases in 1990,

Table 1. Incidence, prevalence and DALYs rates of Leprosy, percentage change and annualized rates of change (95% UI) between 1990 and 2019 for India and Brazil

		# or rate (95% UI) 1990		# or rate (95% UI) 2019		% Change		ARC (95% UI)	
Measure	Metric	India	Brazil	India	Brazil	India	Brazil	India	Brazil
Incidence	**Number**	42394.09 (36351.40–50438.98)	5600.74 (4593.25–6880.05)	24856.35 (21237.72–29726.35)	8370.20 (6896.54–10205.64)	−41.37	49.45	−0.41 (−0.45 to −0.38)	0.49 (0.43 to 0.56)
	Rate	6.04 (5.14–7.21)	4.76 (3.90–5.79)	1.81 (1.55–2.15)	3.48 (2.89–4.22)	−70.07	−26.87	−0.70 (−0.72 to −0.68)	−0.27 (−0.29 to −0.25)
Prevalence	**Number**	460206.75 (386369.62–540850.98)	28452.36 (24476.48–32926.12)	258855.26 (220463.07–304861.56)	53433.62 (45641.33–62703.89)	−43.75	87.80	−0.44 (−0.48 to −0.38)	0.88 (0.80 to 0.95)
	Rate	73.45 (61.87–86.29)	26.13 (22.45–30.38)	20.00 (17.07–23.46)	22.20 (19.00–26.02)	−72.78	−15.02	−0.73 (−0.75 to −0.70)	−0.15 (−0.17 to −0.13)
DALYs	**Number**	25487.86 (16242.04–37356.91)	1213.42 (768.80–1794.94)	14170.95 (9102.40–20627.37)	2514.62 (1598.46–3715.03)	−44.40	107.23	−0.44 (−0.50 to −0.38)	1.07 (0.94 to 1.22)
	Rate	4.04 (2.62–5.89)	1.13 (0.71–1.67)	1.09 (0.70–1.57)	1.04 (0.66–1.54)	−72.93	−7.82	−0.73 (−0.75 to −0.70)	−0.08 (−0.13 to −0.02)

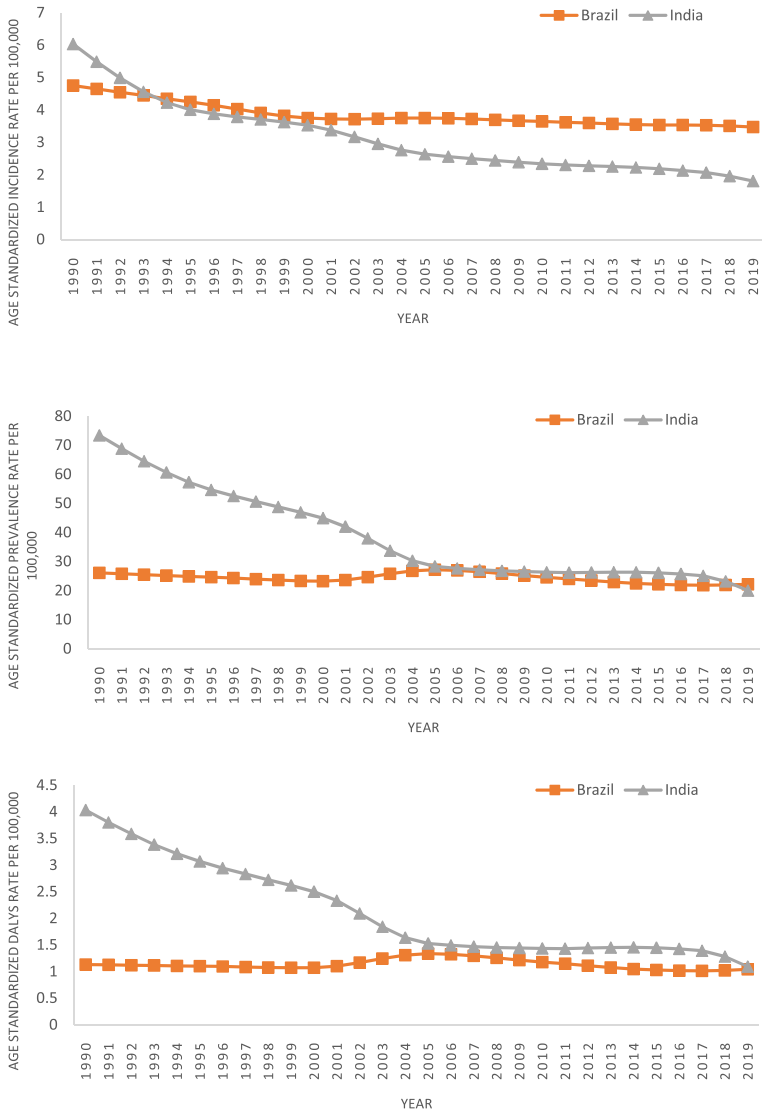


Figure 1. Age standardized Incidence, Prevalence and DALY rates of Leprosy for India and Brazil from 1990 to 2019.

increasing to 32.8% in 2019 (Figure 5). In 1990 in India, leprosy prevalent cases among 0–14 years age group population contributed 0.65% of the total leprosy cases, whereas in Brazil the figure was more than doubled (1.6%). However, in 2019, leprosy prevalent cases among 0–14 years age group population contribute 0.72% of the total leprosy cases in India, whereas they contributed 0.61% in Brazil (Figure 6). In India, sub-national distribution of leprosy prevalent case rate during 2019 shows that only Kerala, Goa, Delhi and Union territories had achieved the elimination target of <10 cases per 100,000 population for leprosy, whereas Bihar, Uttar Pradesh and Chhattisgarh reported higher prevalence rates of 32.22, 33.21 and 30.06 per

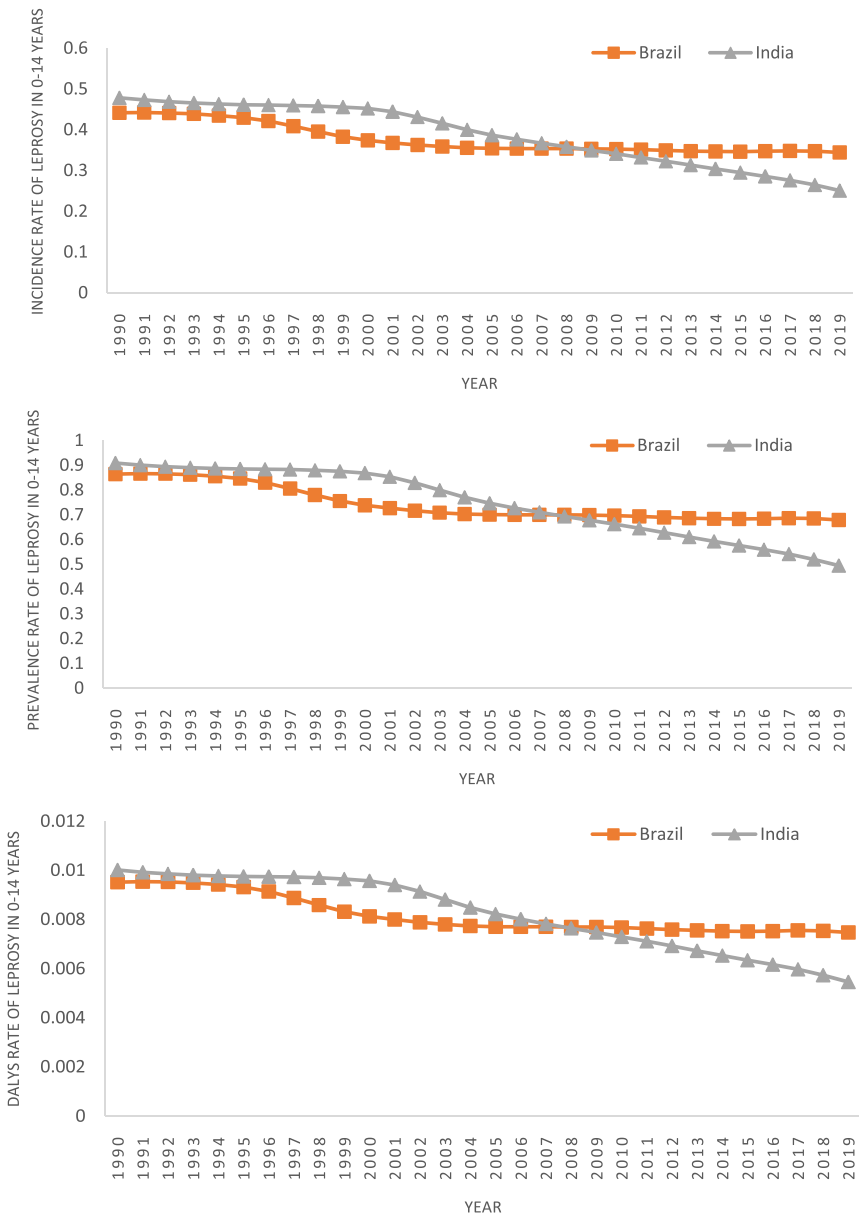


Figure 2. Incidence, Prevalence and DALY rates of Leprosy (0–14 age group) for India and Brazil from 1990 to 2019.

100,000 population, respectively (Figure 7). In Brazil, sub national distribution of leprosy prevalence rates during 2019 shows that none of their states had achieved elimination target of <10 case per 100,000 populations. The highest prevalence rate was seen in Maranhão (33.91 cases per 100,000 population), whereas the lowest prevalence rate was observed in Distrito Federal (Brazilia, 14.46 cases per 100,000 population) (Figure 8).

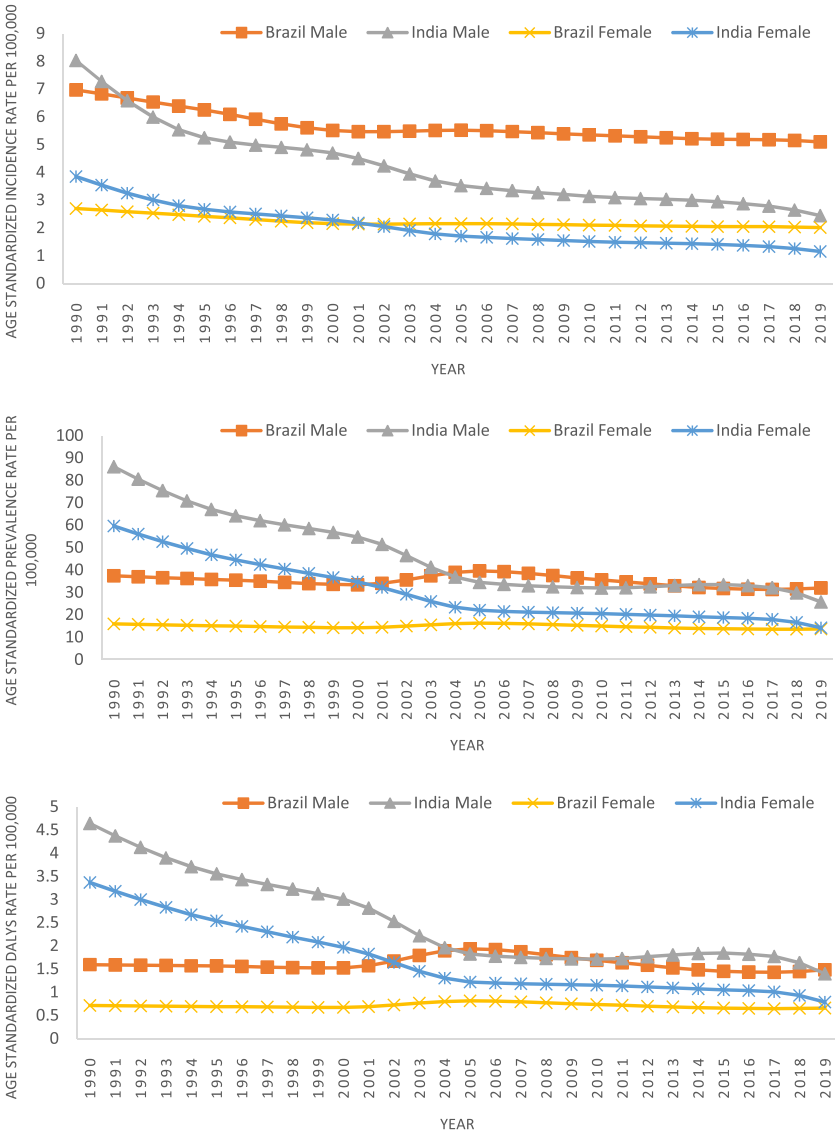


Figure 3. Gender wise trends of Incidence, Prevalence and DALY rates of Leprosy for India and Brazil; 1990 to 2019.

Discussion

Under the WHO Roadmap for Neglected Tropical Diseases 2021–2030, the World Health Organization (WHO) has introduced the Global Leprosy Strategy 2021–2030 with the objective of worldwide leprosy elimination (achieved through transmission interruption and the absence of the disease). However, due to the COVID-19 pandemic, executing plans for the global eradication of leprosy has become challenging for countries. Therefore, it is recommended that countries focus on innovative approaches such as e-learning, m-health,

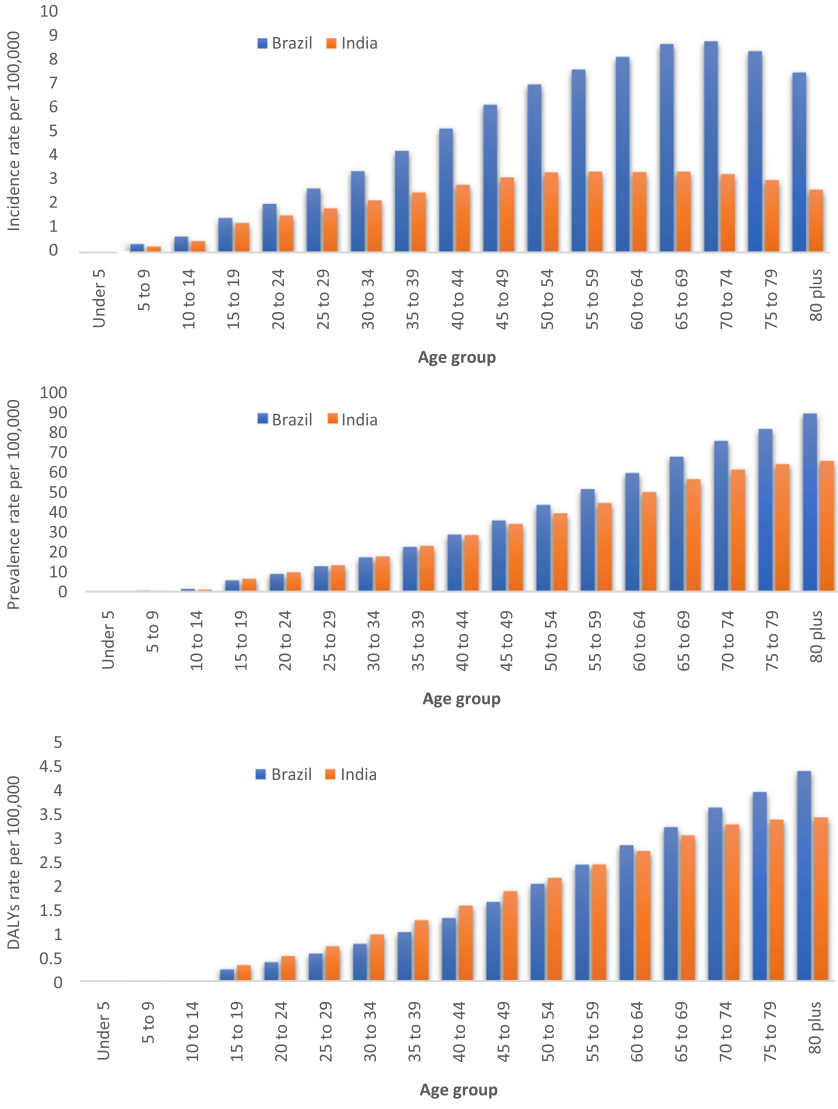


Figure 4. Age wise distribution of Incidence, Prevalence, and DALY rates of leprosy for India and Brazil in 2019.

telemedicine, and other creative methods to enhance services while safeguarding both patients and staff from potential risks.[18]

India and Brazil, both are rapidly developing countries but Brazil is an upper-middle income economy, whereas India is a lower-middle income economy. Brazil spends two and half times more of their gross domestic product (GDP) on health as compared to India.[19] India and Brazil remain, as the first and second countries for new cases of leprosy, respectively. In India, since the launch of National Leprosy Eradication Programme (NLEP) in 1983, the prevalence rate has dropped steadily and reached elimination status (<1 case/10,000 population) in 2005.[4] In Brazil, significant progress towards the elimination of leprosy has been made after the official

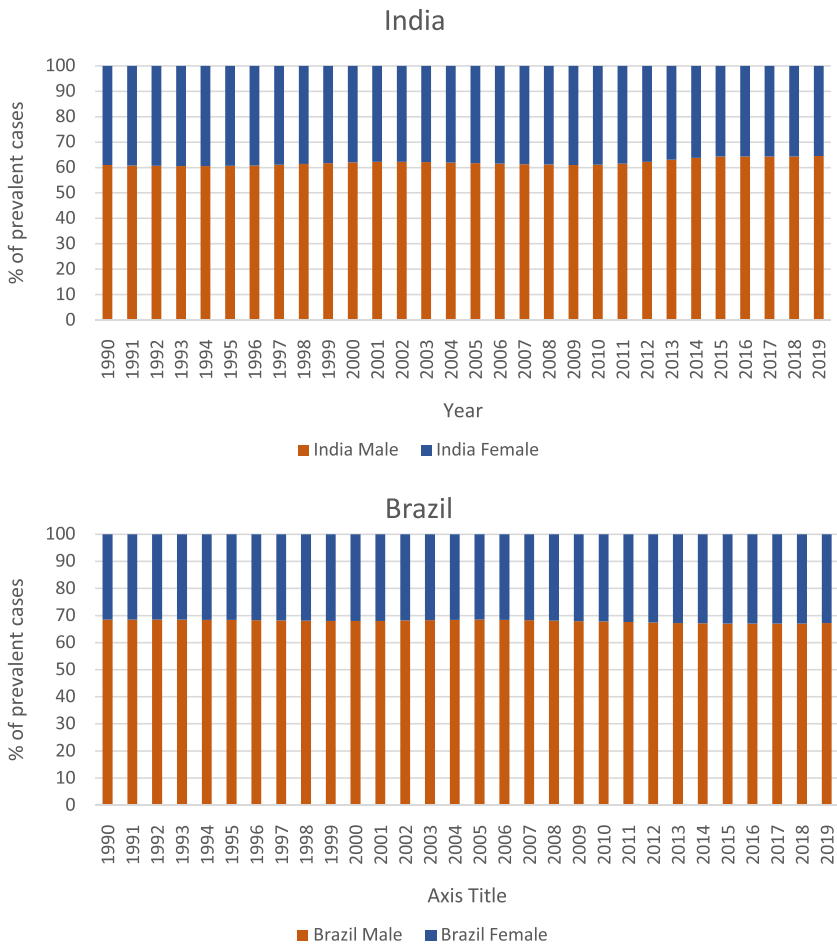


Figure 5. Gender wise percent distribution and trends of Prevalent cases of Leprosy for India and Brazil; 1990–2019.

and extensive introduction of MDT in 1993, which led to a decline in the prevalence rate to reach 4.5/10.000 in population in 1998. Despite the ongoing and dedicated efforts of the government for the past three decades, Brazil has yet to attain the goal of elimination of leprosy throughout the entire country.[20]

We have analyzed and contrasted the trends of leprosy impact in India and Brazil between 1990 and 2019, utilizing data from the Global Burden of Disease 2019 study.This study showed there is decreasing trend in the incidence, prevalence and DALYs rates of leprosy in India over past three decades. This decline may be attributed to India's constant efforts towards country wide elimination of leprosy by year 2000. Introduction of multidrug therapy (rifampicin, dapsone, and clofazimine) in mid-eighties under NLEP played a key role in reduction in the burden of leprosy; country wide coverage of MDT was achieved by the end of year 1995.[21] The cure rate of MDT was up to 98%.[22] Based on the leprosy prevalence rate whole country has been divided into high and low endemic areas for the equitable resource allocation

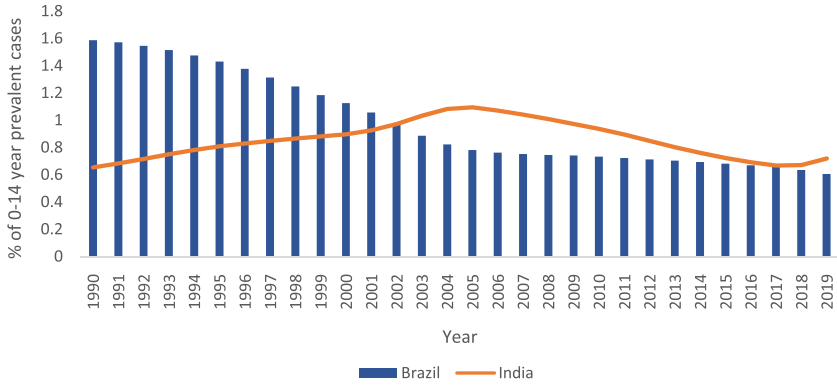


Figure 6. Percent distribution and trends of Prevalent cases among 0–14 year age group of Leprosy for India and Brazil; 1990–2019.

to control leprosy.[23] Researchers have also found poverty, illiteracy, malnutrition and poor sanitation were associated with the transmission of leprosy; hence the overall improvement in the literacy rate and sanitation might have played a role in decreasing trends of leprosy in India.[24] Starting from the late nineties, the government organized information, education and communication activities in villages, schools and communities for minimizing stigma associated with leprosy.[25] Government integrated leprosy elimination with general health services in 2005, which ensured availability of leprosy services at primary health centres, hospitals, and dispensaries.[25,26] However, certain researchers have proposed alternative explanations for the substantial reduction in leprosy prevalence during the initial years of the first decade of the 21st century. Fine (2006), for instance, raised doubts and concerns about India's criteria for prevalence and detection, suggesting that it's plausible that this nation, and possibly the global scenario as a whole, has yet to attain the prevalence rates set by the WHO for the goal of elimination.[27] Also under national leprosy eradication programme, the government revised the criteria of new cases as diagnostic confirmation by a leprologist. Relapsed cases and case with single lesions were excluded from prevalence and dropout leprosy patients were excluded from active records.[28,29] Though the stigma towards leprosy, poor knowledge and unfavorable attitude towards leprosy is high in females,[30] still leprosy cases were more in males as compared to females.[31] In our study also we have found higher proportion of leprosy cases among males vis-à-vis females, this may reflect the difficult access to healthcare services for women in India.[3] However, Kumar *et al.* found no gender wise significant difference in incidence rate of leprosy.[32] In our study highest prevalence rate of leprosy cases were reported from Bihar, Uttar Pradesh and Chhattisgarh; some other researchers also reported high prevalence of leprosy in Chhattisgarh and Subramanian *et al.* found high new case detection rate of leprosy cases in Uttar Pradesh.[4,33] However, Siddiqui *et al.* reported high prevalence of leprosy in Odisha.[34]

This study shows that there is slight decrease in the incidence, prevalence and DALYs rates of leprosy in Brazil from 1990 to 2019, but absolute number of leprosy cases was increased during the same period. A study conducted by Souza *et al.* found decreasing trends of leprosy cases in Brazil from 2001 to 2017.[35] Various researchers have ascertained that Brazil's attainment of the goal of leprosy elimination remains incomplete, and it holds the second-highest rank in terms of the total number of cases. India occupies the first position.[12,36] In Brazil,

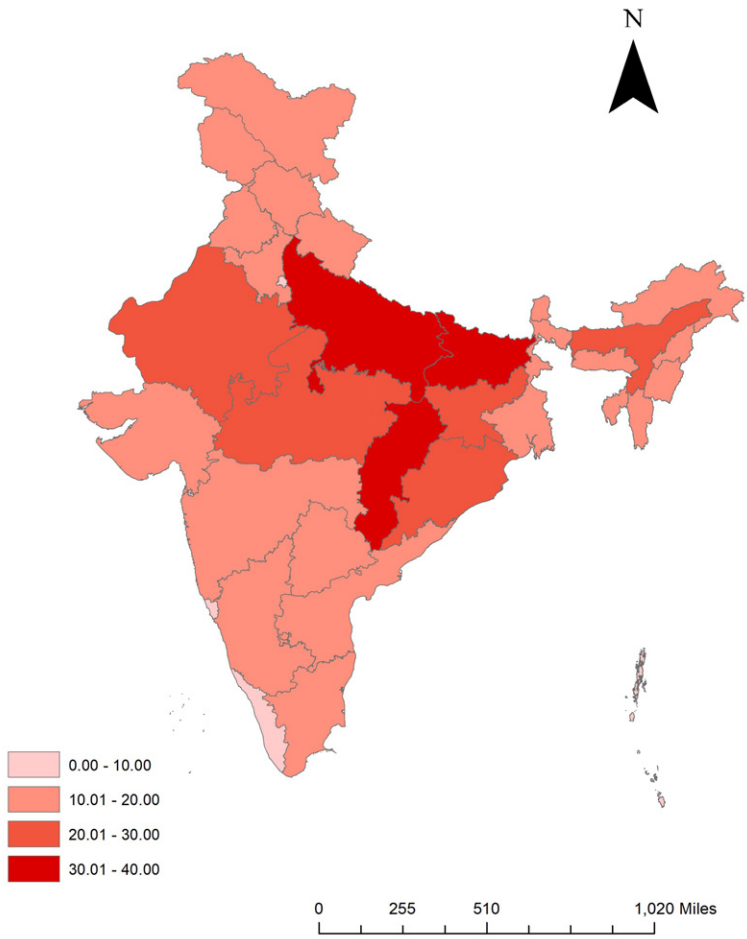


Figure 7. Sub-national distribution of age standardized prevalence rates of Leprosy in India: 2019.

the prevalence of Hansen's disease demonstrated a consistent decline, while detection rates exhibited an upward trend during the final two decades of the 20th century, likely indicative of improved patient access to primary healthcare services.[37] However, Cunha *et al.* suggested that application of revised national guidelines which excludes the defaulter patients of leprosy from active records and abolition of post treatment follow-up were some reasons for high fall leprosy cases in Brazil.[38] The National Brazilian Program altered its objective from overall elimination to a reduction in leprosy cases among children, as outlined in the country's plan for the years 2008 to 2011. However, subsequent to changes in government, the prevailing policy reverted to the goal of complete elimination.[39] A study conducted by Henry *et al.* found that social isolation, not considering their symptoms serious and visiting traditional healer's before consulting doctors were significant risk factors for the delayed diagnosis and continued transmission of leprosy in Brazil.[40] While Multidrug Therapy (MDT) is highly effective for treating leprosy,[22] it can potentially lead to gastrointestinal issues, drowsiness, dizziness, and weakness, with the side effects ranging from mild to severe.[41] A study discovered that, despite

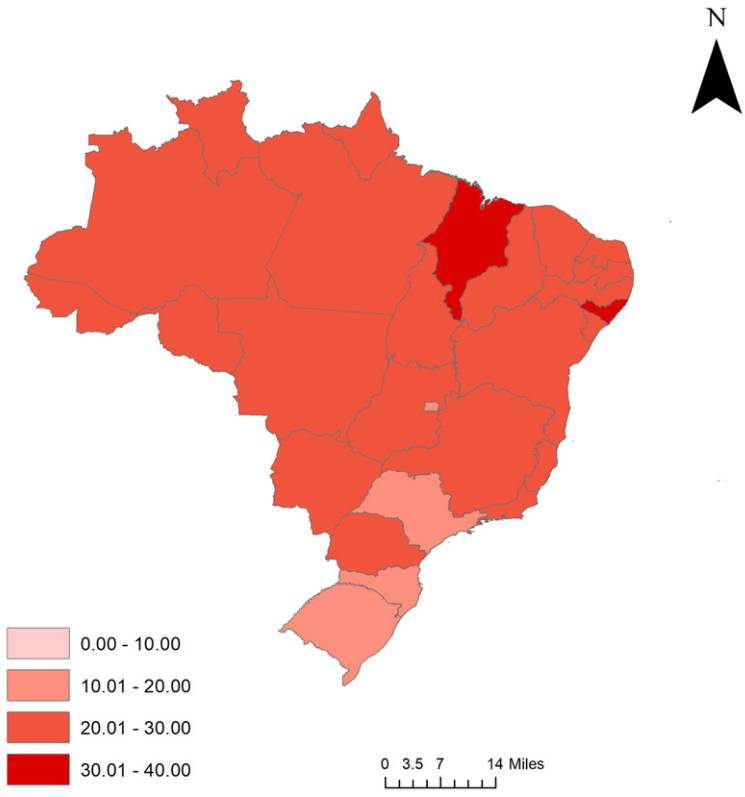


Figure 8. Sub-national distribution of age standardized prevalence rates of Leprosy in Brazil: 2019

acknowledging the risks of the disease, 62.9% of patients in Brazil exhibited low adherence to treatment.[42] In a comprehensive cohort study by de Andrade *et al.*, the default rates ranged from 6.4% in 2007 to 5.4% in 2014. Their findings indicated that individuals residing in the North region were more prone to discontinuing leprosy treatment compared to residents of the South and Southeast regions.[43] Certain researchers have linked the high infection and prevalence rates of leprosy in Brazil to factors such as poverty, illiteracy, and low socio-economic status.[44,45]

Our study has some limitations. Firstly, despite the Global Burden of Disease study employing diverse methods to address missing or incomplete data, there remains a potential for inaccuracies in the mortality data. Secondly, due to the nature of our ecological study design, the outcomes of our research are valid at the population level, but we cannot assert that they are applicable at the individual level.

Conclusions

Leprosy continues to pose a substantial health challenge in both India and Brazil. Notably, a substantial decrease in the age-standardized incidence, prevalence, and DALYs rates of leprosy has been observed in India in comparison to Brazil over the period of 1990 to 2019. Brazil has observed an increase in the absolute number of leprosy cases during the same timeframe.

To comprehensively evaluate the limitations in the current strategies of the national leprosy eradication programs in both India and Brazil, a multicenter study is essential.

Authors' contributions

S, DD and RPJ conceived the study; DD, RPJ and S designed the study protocol; DD, RPJ, KB and AM carried out the analysis and interpretation of these data. All authors drafted the manuscript; DD and RPJ critically revised the manuscript for intellectual content. All authors read and approved the final manuscript.

Guarantors

S, DD and RPJ are guarantors of the paper.

Funding

None.

Competing interests

None declared.

Data availability

The data is accessible to the public at https://vizhub.healthdata.org/gbd-results/ and can be used without any charge.

Ethical approval

Not required.

Acknowledgements

Our gratitude goes to the Institute for Health Metrics and Evaluation (IHME) for furnishing the GBD estimates.

References

1 Monteiro LD, Martins-Melo FR, Brito AL, Alencar CH, Heukelbach J. Physical disabilities at diagnosis of leprosy in a hyperendemic area of Brazil: trends and associated factors. *Lepr Rev*, 2015; **86**(3): 240–250.

2 Nsagha DS, Bamgboye EA, Assob JC, Njunda AL, Kamga HL, Zoung-Kanyi Bissek AC *et al.* Elimination of leprosy as a public health problem by 2000 AD: an epidemiological perspective. *Pan Afr Med J*, 2011; **9**: 4. doi:10.4314/pamj.v9i1.71176.

3 Reibel F, Cambau E, Aubry A. Update on the epidemiology, diagnosis, and treatment of leprosy. *Med Mal Infect*, 2015; **45**(9): 383–393. doi:10.1016/j.medmal.2015.09.002.

4 Singal A, Sonthalia S. Leprosy in post-elimination era in India: difficult journey ahead. *Indian J Dermatol*, 2013; **58**(6): 443–446. doi:10.4103/0019-5154.119952.

5 Nery JS, Pereira SM, Rasella D, Penna ML, Aquino R, Rodrigues LC *et al.* Effect of the Brazilian conditional cash transfer and primary health care programs on the new case detection rate of leprosy. *PLoS Negl Trop Dis*, 2014; **8**(11): e3357. doi:10.1371/journal.pntd.0003357.

6 World Health Organization. *Wkly Epidemiol Rec*, 2020; **95**(36): 417–440. Available at https://apps.who.int/iris/handle/10665/334140.

7 Matos AMF, Coelho ACO, Araújo LPT, Alves MJM, Baquero OS, Duthie MS *et al.* Assessing epidemiology of leprosy and socio-economic distribution of cases. *Epidemiol Infect*, 2018; **146**(14): 1750–1755. doi:10.1017/S0950268818001814.

[8] Ministry of Health and Family Welfare, Government of India 2021. National Leprosy Eradication Programme. Available at https://dghs.gov.in/WriteReadData/userfiles/file/Leprosy/State%20wise%20report-2019-20.pdf.

[9] Corrêa Rda G, Aquino DM, Caldas Ade J, Amaral DK, França FS, Mesquita ER. Epidemiological, clinical, and operational aspects of leprosy patients assisted at a referral service in the state of Maranhão, Brazil. *Rev Soc Bras Med Trop*, 2012; **45**(1): 89–94. doi:10.1590/s0037-86822012000100017.

[10] Barbieri RR, Sales AM, Hacker MA, Nery JA, Duppre NC, Machado AM *et al.* Impact of a reference center on leprosy control under a decentralized public health care policy in Brazil. *PLoS Negl Trop Dis*, 2016; **10**(10): e0005059. doi:10.1371/journal.pntd.0005059.

[11] Freitas LRS, Duarte EC, Garcia LP. Analysis of the epidemiological situation of leprosy in an endemic area in Brazil: spatial distribution in the periods 2001–2003 and 2010–2012. *Rev Bras Epidemiol*, 2017; **20**(4): 702–713. doi:10.1590/1980-5497201700040012.

[12] Lastória JC, Abreu MA. Leprosy: review of the epidemiological, clinical, and etiopathogenic aspects - part 1. *An Bras Dermatol*, 2014; **89**(2): 205–218. doi:10.1590/abd1806-4841.20142450.

[13] Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, 2012; **380**(9859): 2095–2128. doi:10.1016/S0140-6736(12)61728-0.

[14] Foreman KJ, Lozano R, Lopez AD, Murray CJ. Modeling causes of death: an integrated approach using CODEm. *Popul Health Metr*, 2012; **10**: 1. doi:10.1186/1478-7954-10-1.

[15] Murray CJ, Ezzati M, Flaxman AD, Lim S, Lozano R, Michaud C *et al.* GBD 2010: design, definitions, and metrics. *Lancet*, 2012; **380**(9859): 2063–2066. doi:10.1016/S0140-6736(12)61899-6.

[16] Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*, 2020; **396**(10258): 1204–1222. doi:10.1016/S0140-6736(20)30925-9.

[17] Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME). Available at http://ghdx.healthdata.org/gbd-results-tool.

[18] World Health Organization. Towards zero leprosy (2021). Global leprosy (Hansen's Disease) strategy 2021–2030. World Health Organization. Available at https://apps.who.int/iris/handle/10665/340774.

[19] Dhamnetiya D, Jha RP, Shalini, Bhattacharyya K. India's performance in controlling Visceral Leishmaniasis as compared to Brazil over past three decades: findings from global burden of disease study. *J Parasit Dis*, 2021; **45**(4): 877–886. doi:10.1007/s12639-021-01375-0.

[20] Nazario AP, Ferreira J, Schuler-Faccini L, Fiegenbaum M, Artigalás O, Vianna FSL. Leprosy in Southern Brazil: a twenty-year epidemiological profile. *Rev Soc Bras Med Trop*, 2017; **50**(2): 251–255. doi:10.1590/0037-8682-0229-2016.

[21] Sansarricq. *Hubert & UNDP/World Bank/WHO Special Program for Research and Training in Tropical Diseases*. 2006. Available at https://apps.who.int/iris/handle/10665/69318.

[22] Scollard DM, Adams LB, Gillis TP, Krahenbuhl JL, Truman RW, Williams DL. The continuing challenges of leprosy. *Clin Microbiol Rev*, 2006; **19**(2): 338–381. doi:10.1128/CMR.19.2.338-381.2006.

[23] Bhattacharya SN, Sehgal VN. Leprosy in India. *Clin Dermatol*, 1999; **17**(2): 159–170. Discussion 105-6. doi:10.1016/s0738-081x(99)00009-7.

[24] Freitas LR, Duarte EC, Garcia LP. Leprosy in Brazil and its association with characteristics of municipalities: ecological study, 2009–2011. *Trop Med Int Health*, 2014; **19**(10): 1216–1225. doi:10.1111/tmi.12362. Epub 2014 July 21.

[25] Arole S, Premkumar R, Arole R, Maury M, Saunderson P. Social stigma: a comparative qualitative study of integrated and vertical care approaches to leprosy. *Lepr Rev*, 2002 June; **73**(2): 186–196.

[26] Desikan KV. Elimination of leprosy & possibility of eradication - the Indian scenario. *Indian J Med Res*, 2012; **135**(1): 3–5. doi:10.4103/0971-5916.93415.

[27] Fine PE. Global leprosy statistics: a cause for pride, or frustration? *Lepr Rev*, 2006; **77**(4): 295–297.

[28] Britton WJ, Lockwood DN. Leprosy. *Lancet*, 2004; **363**(9416): 1209–1219. doi:10.1016/S0140-6736(04)15952-7.

[29] Samy AA. 'Leprosy elimination'–need for sample survey. *Lepr Rev*, 2007; **78**(2): 167–169.

[30] Singh R, Singh B, Mahato S. Community knowledge, attitude, and perceived stigma of leprosy amongst community members living in Dhanusha and Parsa districts of Southern Central Nepal. *PLoS Negl Trop Dis*, 2019; **13**(1): e0007075. doi:10.1371/journal.pntd.0007075.

[31] Kumar A, Girdhar A, Yadav VS, Girdhar BK. Some epidemiological observations on leprosy in India. *Int J Lepr Other Mycobact Dis*, 2001; **69**(3): 234–240.

[32] Kumar A, Girdhar A, Girdhar BK. Incidence of leprosy in Agra district. *Lepr Rev*, 2007; **78**(2): 131–136.

[33] Subramanian M, Showkath Ali MK, Thorat DM, Muthukumar M, Sathiskumar E, Ramadoss C *et al.* Leprosy situation in endemic states of India and prospects of elimination of the disease. *Indian J Lepr*, 2003; **75**(4): 335–345.

[34] Siddiqui MR, Velidi NR, Pati S, Rath N, Kanungo AK, Bhanjadeo AK *et al.* Integration of leprosy elimination into primary health care in orissa, India. *PLoS One*, 2009; **4**(12): e8351. doi:10.1371/journal.pone.0008351.

[35] Souza CDF, Paiva JPS, Leal TC, Urashima GDS. Leprosy in Brazil in the 21st century: analysis of epidemiological and operational indicators using inflection point regression. *An Bras Dermatol*, 2020; **95**(6): 743–747. doi:10.1016/j.abd.2019.09.031.

[36] Souza CDF, Santos FGB, Marques CS, Leal TC, Paiva JPS, Araújo EMCF. Spatial study of leprosy in Bahia, Brazil, 2001–2012: an approach based on the local empirical Bayesian model. *Epidemiol Serv Saude*, 2018; **27**(4): e2017479. doi:10.5123/S1679-49742018000400013.

[37] Penna ML, Oliveira ML, Carmo EH, Penna GO, Temporão JG. The influence of increased access to basic healthcare on the trends in Hansen's disease detection rate in Brazil from 1980 to 2006. *Rev Soc Bras Med Trop*, 2008; **41**(Suppl 2): 6–10. doi:10.1590/s0037-86822008000700003.

[38] Cunha C, Pedrosa VL, Dias LC, Braga A, Chrusciak-Talhari A, Santos M *et al.* A historical overview of leprosy epidemiology and control activities in Amazonas, Brazil. *Rev Soc Bras Med Trop*, 2015; **48**(Suppl 1): 55–62. doi:10.1590/0037-8682-0103-2013.

[39] Talhari S, Grossi MA, Oliveira ML, Gontijo B, Talhari C, Penna GO. Hansen's disease: a vanishing disease? *Mem Inst Oswaldo Cruz*, 2012; **107**(Suppl 1): 13–16. doi:10.1590/s0074-02762012000900003.

[40] Henry M, GalAn N, Teasdale K, Prado R, Amar H, Rays MS *et al.* Factors contributing to the delay in diagnosis and continued transmission of leprosy in Brazil–an explorative, quantitative, questionnaire based study. *PLoS Negl Trop Dis*, 2016; **10**(3): e0004542. doi:10.1371/journal.pntd.0004542.

[41] Lira KB, Leite JJ, Maia DC, Freitas Rde M, Feijão AR. Knowledge of the patients regarding leprosy and adherence to treatment. *Braz J Infect Dis*, 2012; **16**(5): 472–475. doi:10.1016/j.bjid.2012.04.002.

[42] Honrado ER, Tallo V, Balis AC, Chan GP, Cho SN. Noncompliance with the world health organization-multidrug therapy among leprosy patients in Cebu, Philippines: its causes and implications on the leprosy control program. *Dermatol Clin*, 2008; **26**(2): 221–229. doi:10.1016/j.det.2007.11.007.

[43] de Andrade KVF, Silva Nery J, Moreira Pescarini J, Ramond A, de Souza Teles Santos CA, Ichihara MY *et al.* Geographic and socioeconomic factors associated with leprosy treatment default: An analysis from the 100 Million Brazilian Cohort. *PLoS Negl Trop Dis*, 2019; **13**(9): e0007714. doi:10.1371/journal.pntd.0007714.

[44] Grantz KH, Chabaari W, Samuel RK, Gershom B, Blum L, Worden L *et al.* Spatial distribution of leprosy in India: an ecological study. *Infect Dis Poverty*, 2018; **7**(1): 20. doi:10.1186/s40249-018-0402-y.

[45] Nery JS, Ramond A, Pescarini JM, Alves A, Strina A, Ichihara MY *et al.* Socioeconomic determinants of leprosy new case detection in the 100 Million Brazilian Cohort: a population-based linkage study. *Lancet Glob Health*, 2019; **7**(9): e1226–e1236. doi:10.1016/S2214-109X(19)30260-8.

arXiv:2412.04819v1 [math.CV] 6 Dec 2024

Starlike Functions Associated with a Non-Convex Domain

S. Sivaprasad Kumar and Surya Giri

Abstract

We introduce and study a class of starlike functions associated with the non-convex domain

$$\mathcal{S}^*_{nc} = \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec \frac{1+z}{\cos z} =: \varphi_{nc}(z), \;\; z \in \mathbb{D} \right\}.$$

Key results include the growth and distortion theorems, initial coefficient bounds, and the sharp estimates for third-order Hankel and Hermitian-Toeplitz determinants. We also examine inclusion relations, radius problems for certain subclasses, and subordination results.

These findings enrich the theory of starlike functions associated with non-convex domains, offering new perspectives in geometric function theory.

Keywords: Univalent functions; Starlike functions; Hankel determinant; Hermitian-Toeplitz determinant.
AMS Subject Classification: 30C45, 30C50, 30C80.

1 Introduction

Let $\mathcal{A}$ denote the class of analytic functions f in the open unit disc $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ normalized by $f(0) = 0$, $f'(0) = 1$, and let $\mathcal{S} \subset \mathcal{A}$ be the subclass of functions that are univalent in $\mathbb{D}$. The well known subclasses of $\mathcal{S}$ are the classes of starlike and convex functions denoted by $\mathcal{S}^*$ and $\mathcal{C}$ respectively. A function $f \in \mathcal{S}^*$ (or $\mathcal{C}$) if and only if it satisfies $\mathrm{Re}(zf'(z)/f(z)) > 0$ (or $\mathrm{Re}(1 + zf''(z)/f'(z)) > 0$) for $z \in \mathbb{D}$. An analytic function f is said to be subordinate to other analytic function g in $\mathbb{D}$ if there exists a Schwarz function $\omega(z)$ such that $f(z) = g(\omega(z))$, $z \in \mathbb{D}$. It is denoted by $f \prec g$. In 1992, Ma and Minda [19] considered the class $\mathcal{S}^*(\varphi)$, which is defined as

$$\mathcal{S}^*(\varphi) = \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec \varphi(z) \right\}, \tag{1}$$

where $\varphi(z)$ is an analytic univalent function in $\mathbb{D}$ such that $\varphi(0) = 1$ and maps the unit disk onto the domain satisfying the following properties: (i) $\varphi(\mathbb{D})$ is starlike with respect to $\varphi(0) = 1$ (ii) $\mathrm{Re}(\varphi(z)) > 0$ (iii) and $\varphi(\mathbb{D})$ is symmetric about the real axis.

This class covers various subclasses of starlike functions, for instance, when $\varphi(z) = (1+Az)/(1+Bz)$ $(-1 \le B < A \le 1)$, the class $\mathcal{S}^*(\varphi)$ becomes the class of the Janowski starlike functions [10]. The class $\mathcal{S}^*(\alpha)$ containing starlike functions of order α can be obtained by taking $\varphi(z) = (1 + (1-2\alpha)z)/(1-z)$ $(0 \le \alpha < 1)$. For $\varphi(z) = 1 + \left(2/\pi^2\right)\left(\log\left((1+\sqrt{z})/(1-\sqrt{z})\right)\right)^2$, the class $\mathcal{S}^*(\varphi)$ reduces to the class of parabolic starlike functions introduced by Rønning [24]. In case of modified sigmoid function given by $\varphi(z) = 2/(1+e^{-z})$, we obtain the class $\mathcal{S}^*_{SG}$ introduced and studied by Goel and Kumar [9]. Geometrically, a function $f \in \mathcal{S}^*_{SG}$ if and only if $zf'(z)/f(z)$ lies in the domain $\Delta_{SG} = \{w \in \mathbb{C} : |\log(w/(2-w))| < 1\}$. Recently, Kumar and Gangania [7] considered $\varphi(z) = 1 + ze^z$, which maps the unit disk onto a cardioid domain, and introduced the subclass $\mathcal{S}^*_\varrho$.

Numerous subclasses of $\mathcal{S}^*$ have been considered by choosing an appropriate φ in (1). For instance, the interesting regions depicted by the functions e^z, $1+\sin z$ and $(1+z)^s$ were considered in place of $\varphi(z)$ by Mendiratta et al. [20], Cho et al. [5] and Liu and Liu [16], respectively. Motivated by these defined classes, we introduce the class of starlike functions associated with a non-convex shaped domain represented by $\varphi_{nc} := (1+z)/\cos z$ (see Figure 1). Let

$$\mathcal{S}^*_{nc} := \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec \frac{1+z}{\cos z} \right\}.$$

We study the geometric properties of functions belonging to the class $\mathcal{S}^*_{nc}$. In geometric function

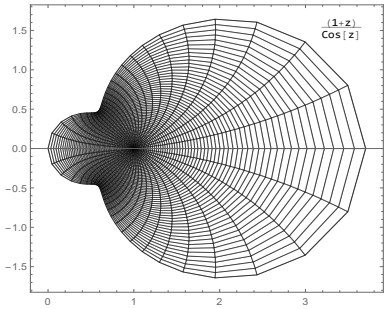


Figure 1: Image of the unit disk under the function φ_{nc}.

theory, Bieberbach conjecture for the class $\mathcal{S}$ was the famous coefficient problem, which took 66 years to solve. In between, the same problem for other subclasses of $\mathcal{S}$ were verified. Finding the bounds of Taylor's series coefficients of functions, Toeplitz, Hermitian-Toeplitz and Hankel determinants come under the coefficient problems. For more details see [8, 4, 6, 12, 18, 23]. We also derived the sharp estimate of $|a_n|$ $(n = 2, 3, 4)$, certain Toeplitz, Hermitian-Toeplitz and Hankel determinants formed over the coefficient of $f \in \mathcal{S}^*_{nc}$. For $f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in \mathcal{A}$, the m^{th} Hankel and Hermitian-Toeplitz determinant for $m \geq 1$ and $n \geq 0$ are respectively given by

$$H_m(n)(f) = \begin{vmatrix} a_n & a_{n+1} & \cdots & a_{n+m-1} \\ a_{n+1} & a_{n+2} & \cdots & a_{n+m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n+m-1} & a_{n+m} & \cdots & a_{n+2m-2} \end{vmatrix}, \tag{2}$$

$$T_{m,n}(f) = \begin{vmatrix} a_n & a_{n+1} & \cdots & a_{n+m-1} \\ \bar{a}_{n+1} & a_n & \cdots & a_{n+m-2} \\ \vdots & \vdots & \vdots & \vdots \\ \bar{a}_{n+m-1} & \bar{a}_{n+m-2} & \cdots & a_n \end{vmatrix}, \tag{3}$$

where $\bar{a}_n = \overline{a_n}$. Toeplitz matrices have constant entries along their diagonals, while Hankel matrices have constant entries along their reverse diagonals. In particular,

$$H_2(2)(f) = a_2a_4 - a_3^2, \;\; T_{3,1}(f) = 1 - 2|a_2|^2 + 2\,\mathrm{Re}(a_2^2\bar{a}_3) - |a_3|^2$$

$$T_{2,1}(f) = 1 - |a_2|^2 \text{ and } H_3(1)(f) = a_3(a_2a_4 - a_3^2) - a_4(a_4 - a_2a_3) + a_5(a_3 - a_2^2).$$

Finding the sharp bound of $|H_2(2)(f)|$ and $|H_3(1)(f)|$ for the class $\mathcal{S}$ and its subclasses has always been the focus of many researchers. Although, investigations concerning Hermitian Toeplitz are recently introduced in [3, 6]. For more work in this direction, one can refer [17, 8, 22] and the references cited therein. We obtain the sharp bounds of $|H_2(2)(f)|$, $|H_3(1)(f)|$, $T_{2,1}(f)$ and $T_{3,1}(f)$ when f belongs to the class $\mathcal{S}^*_{nc}$.

Further, we study the radius problems, inclusion relations with other subclasses of starlike functions and differential subordination implications for the same.

2 Main results

Firstly, we prove that the class $\mathcal{S}^*_{nc}$ is non-empty using the following necessary and sufficient condition for a function f to be in $\mathcal{S}^*_{nc}$.

Theorem 1. *A function $f \in \mathcal{S}^*_{nc}$ if and only if there exists an analytic function $q(z) \prec \varphi_{nc}(z) = (1+z)/\cos z$, such that*

$$f(z) = z \exp \int_0^z \frac{q(t)-1}{t} dt. \tag{4}$$

Proof. Let $f \in \mathcal{S}^*_{nc}$ and let $zf'(z)/f(z) := q(z) \prec \varphi_{nc}(z)$. Then by integrating this equation we obtain (4). If f is given by (4) with an analytic function q such that $q(z) \prec \varphi_{nc}$, then by logarithmic differentiation of (4), we obtain $zf'(z)/f(z) = q(z)$. Therefore $zf'(z)/f(z) \prec \varphi_{nc}(z)$ and $f \in \mathcal{S}^*_{nc}$. □

Define the functions $f_n(z)$ such that $f_n(0) = f_n'(0) - 1 = 0$ and

$$\frac{zf_n'(z)}{f_n(z)} = \frac{1+z^{n-1}}{\cos z^{n-1}}, \quad n = 2, 3, \cdots. \tag{5}$$

Then $f_n \in \mathcal{S}^*_{nc}$.

In particular, if we take $q(z) = \varphi_{nc}(z)$, then the corresponding function

$$\tilde{f}(z) := f_2(z) = z \exp \int_0^z \frac{1+t-\cos t}{t \cos t} dt = z + z^2 + \frac{3}{4}z^3 + \frac{7}{12}z^4 + \frac{35}{96}z^5 + \cdots \tag{6}$$

plays the role of an extremal function for many extremal problems for the class $\mathcal{S}^*_{nc}$. In [19], Ma and Minda proved some subordination results which yield that if $f \in \mathcal{S}^*_{nc}$ then $f(z)/z \prec \tilde{f}(z)/z$ and $zf'(z)/f(z) \prec z\tilde{f}'(z)/\tilde{f}(z)$ and therefore, the following result is obtained.

Theorem 2. *Let $f \in \mathcal{S}^*_{nc}$ and $\tilde{f}$ be the extremal function given by (6). Then the following holds:*

1. ***Growth theorem:*** *For $|z_0| = r < 1$, we have*

$$-\tilde{f}(-r) \leq |f(z_0)| \leq \tilde{f}(r).$$

Equality holds for some $z_0 \neq 0$ if and only if f is a rotation of $\tilde{f}$.

2. ***Rotation theorem:*** *For $|z_0| = r < 1$, we have*

$$|\arg\{f(z_0)/z_0\}| \leq \max_{|z|=r} \arg\{\tilde{f}(z)/z\}.$$

3. ***Distortion theorem:*** *For $|z_0| = r < 1$, we have*

$$\tilde{f}'(-r) \leq |f'(z_0)| \leq \tilde{f}'(r).$$

Equality holds for some $z_0 \neq 0$ if and only if f is a rotation of $\tilde{f}$

Theorem 3. *For $|z| = r$, $\min \operatorname{Re}\{\varphi_{nc}(z)\} = \varphi(-r)$ and $\max \operatorname{Re}\{\varphi_{nc}(z)\} = \varphi(r)$.*

Proof. The boundary point of $\varphi_{nc}(\mathbb{D}_{r_0})$, where $\mathbb{D}_{r_0} = \{z \in \mathbb{C} : |z| < r_0\}$, can be written in the form

$$\varphi_{nc}(r_0 e^{i\theta}) = \frac{1 + r_0 e^{i\theta}}{\cos(r_0 e^{i\theta})}.$$

The outward normal at the point $\varphi_{nc}(\zeta)$ is given by $\zeta \varphi'_{nc}(\zeta)$, where $|\zeta| = r_0$. Since we need to find the bounds of real part, it is sufficient to find the points at which, the imaginary parts of the normal, is a constant. Let us consider that $\delta(\theta) := \operatorname{Im}\{\zeta \varphi'(\zeta)\}$. A difficult computation shows that $\delta(\theta) = 0$ for $\theta = 0$ and $\theta = \pi$. It is easy to check that the maximum value is obtained at $\theta = 0$ and minimum value is obtained at $\theta = \pi$. Since r_0 is arbitrary, for $r_0 = r$ we get the desired result. □

Theorem 4. ***(Function's Bounds)*** *Let Φ_R and Φ_I denote the real and imaginary part of $\Phi_{nc}(z) = (1 + z)/\cos z$ respectively. Then we have*

1. $0 \leq \Phi_R \leq (4 \cos 1)/(1 + \cos 2)$.
2. $-\gamma_0 \leq \Phi_I \leq \gamma_0$, *where* $\gamma_0 \approx 1.6471$.
3. $|\arg \varphi_{nc}(z)| \leq \pi/2$.

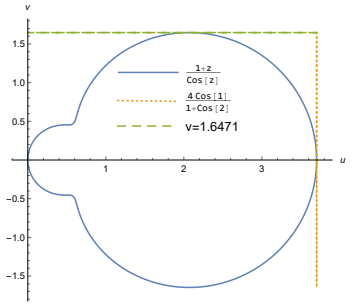


Figure 2: Bounds of φ_{nc}

2.1 Hermitian-Toeplitz determinants

Giri and Kumar [8] established the sharp bounds of $T_{2,1}(f)$ and $T_{3,1}(f)$ for f belongs to the classes $\mathcal{S}^*(\varphi)$. The following results directly follow for the class $\mathcal{S}^*_{nc}$.

Theorem 5. *If $f \in \mathcal{S}^*_{nc}$, then*

$$0 \leq \det T_{2,1}(f) \leq 1 \quad \textit{and} \quad \det T_{3,1}(f) \leq 1.$$

All the estimates are sharp.

Theorem 6. *If $f \in \mathcal{S}^*_{nc}$, then $\det T_{3,1}(f) \geq -1/15$. The bound is sharp.*

2.2 Hankel Determinants

Let $\mathcal{P}$ denote the Carathéodory class containing analytic functions $p(z) = 1 + \sum_{n=1}^{\infty} p_n z^n$ in $\mathbb{D}$ such that $\operatorname{Re} p(z) > 0$ for all $z \in \mathbb{D}$. The coefficients estimates of Carathéodory functions prominently used to deal with the coefficient problems. The following lemmas for the class $\mathcal{P}$ are used to establish the results.

Lemma 7. *[19] If $p(z) = 1 + \sum_{n=2}^{\infty} p_n z^n \in \mathcal{P}$, then for a complex number ν*

$$|p_2 - \nu p_1^2| \leq 2 \max(1, |2\nu - 1|).$$

This result is sharp for the function $p(z) = (1 + z^2)/(1 - z^2)$ and $p(z) = (1 + z)/(1 - z)$.

Lemma 8. *[27]* [5] *Let $p(z) = 1 + \sum_{n=2}^{\infty} p_n z^n \in \mathcal{P}$, then for $n, m \in \mathbb{N}$*

$$|p_{n+m} - \nu p_n p_m| \leq \begin{cases} 2, & 0 \leq \nu \leq 1 \\ 2|2\nu - 1|, & \text{elsewhere.} \end{cases}$$

The next lemma gives the representations of p_2, p_3 and p_4 in terms of p_1. The expressions for p_2 and p_3 were given in [14, 15] and for p_4 in [13].

Lemma 9. *Let $p(z) = 1 + \sum_{n=2}^{\infty} p_n z^n \in \mathcal{P}$, then*

$$\begin{aligned} 2p_2 &= p_1^2 + \gamma(4 - p_1^2), \\ 4p_3 &= p_1^3 + 2(4 - p_1^2)p_1\gamma - (4 - p_1^2)p_1\gamma^2 + 2(4 - p_1^2)(1 - |\gamma|^2)\eta \quad \text{and} \\ 8p_4 &= p_1^4 + (4 - p_1^2)\gamma(p_1^2(\gamma^2 - 3\gamma + 3) + 4\gamma) - 4(4 - p_1^2)(1 - |\gamma|^2)(p_1(\gamma - 1)\eta \\ &\quad + \bar{\gamma}\eta^2 - (1 - |\gamma|^2)\rho) \end{aligned}$$

for some complex numbers γ, η, ρ with $|\gamma| \leq 1$, $|\eta| \leq 1$ and $|\rho| \leq 1$.

Theorem 10. *If $f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in \mathcal{S}_{nc}^*$, then*

$$\sum_{n=2}^{\infty} (n^2 k_1 - 4)|a_n|^2 \leq (4 - k_1),$$

where $k_1 = \cos^2 1$.

Proof. Since $f \in \mathcal{S}_{nc}^*$, therefore by the definition of subordination

$$\frac{zf'(z)}{f(z)} = \frac{1 + \omega(z)}{\cos(\omega(z))},$$

where $\omega(z)$ is a Schwarz function. Using the identity

$$f^2(z) = (zf'(z))^2 \frac{\cos^2 w(z)}{(1 + w(z))^2},$$

we obtain

$$2\pi \sum_{n=1}^{\infty} |a_n|^2 r^{2n} = \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta = \int_0^{2\pi} |re^{i\theta} f'(re^{i\theta})|^2 \left| \frac{\cos w}{1 + w} \right|^2.$$

After some simple computation, we get

$$2\pi\sum_{n=1}^{\infty}|a_n|^2r^{2n}\geq\frac{\cos^2 1}{4}\int_0^{2\pi}|re^{i\theta}f'(re^{i\theta})|^2d\theta=2\pi\frac{\cos^2 1}{4}\sum_{n=1}^{\infty}n^2|a_n|^2r^{2n},$$

where $0<r<1$ and $a_1=1$. Thus

$$\sum_{n=2}^{\infty}\left(n^2\frac{\cos^2 1}{4}-1\right)|a_n|^2r^{2n}\leq 0.$$

On letting $r\to 1^-$, we get the required result. □

Corollary 10.1. *Let $f(z)=z+\sum_{n=2}^{\infty}a_nz^n\in\mathcal{S}^*_{nc}$ and $k_1=\cos^2 1$, then*

$$|a_n|\leq\sqrt{\frac{4-k_1}{n^2\cos^2 1-4}},\quad n=2,3,4,\cdots.$$

The following sharp bound of Fekete-Szegö functional for $f\in\mathcal{S}^*_{nc}$ can be deduced directly from [1].

Theorem 11. *Let $f\in\mathcal{S}^*_{nc}$, then*

$$|a_3-\mu a_2^2|\leq\begin{cases}-\mu+\frac{3}{4}, & \mu<\frac{1}{4}\\ \frac{1}{2}, & \frac{1}{4}\leq\mu\leq\frac{5}{4}\\ \mu-\frac{3}{4} & \mu>\frac{5}{4}.\end{cases}$$

The bound is sharp.

Theorem 12. *If the function $f(z)=z+\sum_{n=2}^{\infty}a_nz^n$ belongs to the class $\mathcal{S}^*_{nc}$, then*

$$|a_2|\leq 1,\quad |a_3|\leq\frac{3}{4},\quad |a_4|\leq\frac{7}{12},\quad |a_5|\leq\frac{1}{3}.$$

The bounds of $|a_n|$ for $n=2,3,4$ are sharp.

Proof. If $f\in\mathcal{S}^*_{nc}$, then there exists a Schwarz function $\omega(z)$ satisfying $\omega(0)=0$ and $|\omega(z)|\leq|z|$ such that

$$zf'(z)=f(z)\left(\frac{1+\omega(z)}{\cos\omega(z)}\right),\quad z\in\mathbb{D}. \tag{7}$$

By the one-to-one correspondence between the classes of Schwarz functions and the class $\mathcal{P}$, let there is $p(z)=1+\sum_{n=1}^{\infty}p_nz^n\in\mathcal{P}$ such that $\omega(z)=(p(z)-1)/(p(z)+1)$. Then by substituting the $f(z)$ and $p(z)$ in (7) and comparing the coefficients, we obtain

$$\begin{gathered}a_2=\frac{p_1}{2},\ a_3=\frac{p_1^2+4p_2}{16},\ a_4=\frac{p_1^3+4p_1p_2+16p_3}{96}\\ \text{and}\ \ a_5=\frac{-p_1^4+4p_1^2p_2+4p_1p_3+24p_4}{192}.\end{gathered} \tag{8}$$

It is well known that the coefficients of Carathéodory function $p(z)$ satisfy the inequality $|p_n|\leq 2$, $(n=1,2,3,\cdots)$, which gives $|a_2|\leq 1$. The bound of $|a_3|$ directly follow from Theorem 11 for $\mu=0$.

Further, the bound of $|a_4|$ can be obtained as a special case of Ali et al. [1, Theorem 1]. Now, the fifth coefficient

$$\begin{aligned} a_5 &= \frac{1}{192}(-p_1^4 + 4p_1^2p_2 + 4p_1p_3 + 24p_4). \\ &= \frac{1}{192}[4p_1^2(p_2 - \frac{1}{4}p_1^2) + 4(p_1p_3 + 6p_4)]. \end{aligned}$$

Applying the triangular inequality together with Lemma 7 and Lemma 8, we obtain

$$|a_5| \leq \frac{1}{192}(8|p_1|^2 + 32).$$

Using the fact that $|p_1| \leq 2$, we get the required bound of $|a_5|$.

The function $\tilde{f}$ given by (6) plays the role of an extremal function for $|a_n|$ when $n = 2, 3$ and 4. □

Theorem 13. *If $f \in \mathcal{S}^*_{nc}$, then*

$$|H_2(2)(f)| \leq \frac{1}{4}.$$

The result is sharp.

Proof. For $f \in \mathcal{S}^*_{nc}$, from (8), we can get

$$a_2a_4 - a_3^2 = \frac{1}{768}(p_1^4 - 8p_1^2p_2 - 48p_2^2 + 64p_1p_3).$$

Since the class $\mathcal{P}$ is invariant under rotation, we can take $p_1 \in [0, 2]$. Let $p := p_1$ and substituting the value of p_2 and p_3 from Lemma 9, we have

$$\begin{aligned} |a_2a_4 - a_3^2| =& \frac{1}{768}|(p^4 - 4p^2(p^2 + (4 - p^2)x) - 12(p^2 + (4 - p^2)x)^2 \\ &+ 16p(p^3 + 2p(4 - p^2)x - p(4 - p^2)x^2 + 2(4 - p^2)(1 - |x|^2)z))|. \\ =& \frac{1}{768}|(p^4 + 12p^2(4 - p^2)x + (28p^4 - 96p^2 + 192)x^2 \\ &+ 32p(4 - p^2)(1 - |x|^2)z)|. \end{aligned}$$

Now, applying the triangle inequality and replacing $|x|$ by ϱ, we obtain

$$\begin{aligned} |a_2a_4 - a_3^2| \leq& \frac{1}{768}(p^4 + 12p^2(4 - p^2)\varrho + (28p^4 + 32p^3 - 96p^2 \\ &- 128p + 192)\varrho^2 + 32p(4 - p^2)) \\ =:& G(p, \varrho) \quad \text{(say)}. \end{aligned}$$

On differentiation $G(p, \varrho)$ with respect to ϱ, we obtain

$$\frac{\partial G(p, \varrho)}{\partial \varrho} = \frac{1}{768}(12p^2(4 - p^2) + 2(192 - 128p - 96p^2 + 32p^3 + 28p^4)\varrho).$$

Since $\partial G(p, \varrho)/\partial \varrho > 0$, therefore $G(p, \varrho)$ is an increasing function and attains its maximum value at $\varrho = 1$. Hence

$$G(p, 1) =: G(p) = \frac{1}{768}(192 - 48p^2 + 17p^4).$$

A simple calculation yields that the maximum value of $G(p)$ occurs at $p = 0$. Thus,

$$|H_2(2)(f)| = |a_2a_4 - a_3^2| \leq \frac{192}{768} = \frac{1}{4}.$$

The bound is sharp for the function $f_3(z)$ given by (5). □

Theorem 14. *If $f \in \mathcal{S}^*_{nc}$, then*

$$|H_3(1)(f)| \leq \frac{1}{9}.$$

The bound is sharp.

Proof. Let $f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in \mathcal{S}^*_{nc}$. Then from (2), we have

$$H_3(1)(f) = a_3(a_2a_4 - a_3^2) - a_4(a_4 - a_2a_3) + a_5(a_3 - a_2^2).$$

Using the values of a_2, a_3, a_4 and a_5 from (8) together with Lemma 9 and $p_1 = p \in [0, 2]$, we obtain

$$H_3(1)(f) = \frac{1}{36864}\Big(\vartheta_1(p,\gamma) + \vartheta_2(p,\gamma)\eta + \vartheta_3(p,\gamma)\eta^2 + \psi(p,\gamma,\eta)\rho\Big),$$

where

$$\begin{aligned}
\vartheta_1(p,\gamma) &= 5p^6 - 26\gamma(4-p^2)p^4 - 144\gamma^2(4-p^2)p^2 + 56\gamma^2(4-p^2)p^4 + 68\gamma^2(4-p^2)^2p^2 \\
&\quad - 36\gamma^3(4-p^2)p^4 - 40\gamma^3(4-p^2)^2p^2 + 8\gamma^4(4-p^2)^2p^2, \\
\vartheta_2(p,\gamma) &= (4-p^2)(1-|\gamma|^2)(-40p^3 + 144p^3\gamma + 80p(4-p^2)\gamma - 32p(4-p^2)\gamma^2), \\
\vartheta_3(p,\gamma) &= (4-p^2)(1-|\gamma|^2)(-256(4-p^2) - 32(4-p^2)|\gamma|^2 + 144p^2\bar{\gamma}), \\
\psi(p,\gamma,\eta) &= (4-p^2)(-144p^2 + 288(4-p^2)\gamma)(1-|\gamma|^2)(1-|\eta|^2),
\end{aligned}$$

$|\gamma| \leq 1$, $|\eta| \leq 1$ and $|\rho| \leq 1$ and $p = p_1 \in [0, 2]$ by the rotationally invariant property of the class $\mathcal{P}$. Now, by taking $|\gamma| := x$, $|\eta| := y$ and keeping in mind $|\rho| \leq 1$, we get

$$|H_3(1)(f)| \leq \frac{1}{36864}\Big(|\vartheta_1(p,\gamma)| + |\vartheta_2(p,\gamma)|y + |\vartheta_3(p,\gamma)|y^2 + |\psi(p,\gamma,\eta)|\Big) =: G(p,x,y),$$

where

$$G(p,x,y) = \frac{1}{36864}\Big(g_1(p,x) + g_2(p,x)y + g_3(p,x)y^2 + g_4(p,x)\Big)$$

such that

$$\begin{aligned}
g_1(p,x) &= 5p^6 + 26p^4(4-p^2)x + 144p^2(4-p^2)x^2 + 56p^4(4-p^2)x^2 + 68p^2(4-p^2)^2x^2 \\
&\quad + 36p^4(4-p^2)x^3 + 40p^2(4-p^2)^2x^3 + 8p^2(4-p^2)^2x^4, \\
g_2(p,x) &= (4-p^2)(1-x^2)(40p^3 + 144p^3x + 80p(4-p^2)x + 32p(4-p^2)x^2), \\
g_3(p,x) &= (4-p^2)(1-x^2)(256(4-p^2) + 32(4-p^2)x^2 + 144p^2x), \\
g_4(p,x) &= (4-p^2)(1-y^2)(144p^2 + 288(4-p^2)x)(1-x^2).
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}G(p,x,y) = \frac{1}{36864}\Big(&5p^6 + 26p^4(4-p^2)x + 144p^2(4-p^2)x^2 + 56p^4(4-p^2)x^2\\ &+ 36p^4(4-p^2)x^3 + 68p^2(4-p^2)^2x^2 + 40p^2(4-p^2)^2x^3 + 8p^2(4-p^2)^2x^4\\ &+ (4-p^2)(1-x^2)(40p^3 + 144p^3x + 80p(4-p^2)x + 32p(4-p^2)x^2)y\\ &+ (4-p^2)(1-x^2)(256(4-p^2) + 144p^2x + 32(4-p^2)x^2)y^2\\ &+ (4-p^2)(144p^2 + 288(4-p^2)x)(1-x^2)(1-y^2)\Big).\end{aligned}$$

Clearly, the maximum value of $G(p,x,y)$ in the closed cuboid $\Gamma : [0,2]\times[0,1]\times[0,1]$ is the required bound of $|H_3(1)(f)|$. We establish this by finding the maximum value of $G(p,x,y)$ in the interior of six faces, on the twelve edges and in the interior of Γ.

Case I. Firstly, we proceed with interior points of Γ. Let $(p,x,y)\in(0,2)\times(0,1)\times(0,1)$. A simple computation yields

$$\begin{aligned}\frac{\partial G}{\partial y} = \frac{(4-p^2)}{2304}\Big(&9p^2(x(1-x^2)(1+2y) - 2(4-p^2)y) + 2(4-p^2)(-18(4-p^2)xy\\ &+ 8(1-x^2+2(1-x^2)y) + x^2(1-x^2+2(1-x^2)y))\Big).\end{aligned}$$

Now $\partial G/\partial y = 0$ gives

$$y_0 = \frac{p^3(-4x^2+8x+5) + 8px(2x+5)}{4(x-1)(p^2(2x-25) - 8(x-8))}.$$

The critical point y_0 lie in the interval $(0,1)$ whenever

$$p^3(4x^2-8x-5) + 4p^2(2x^2-27x+25) - 8px(2x+5) - 32(x^2-9x+8) > 0 \tag{9}$$

and

$$p^2(25-2x) > 8(8-x) \tag{10}$$

hold.

It can be simply observe that as $p\to 2$, there exists at least one $x\in(0,1)$ such that the inequality (9) is not true. Similarly, when $x\to 0$, there exists at least one $p\in(0,2)$ such that inequality (9) does not remain valid. The inequality (9) is true only for $p=2$ and $x\in(0,13/36)$, which is outside the considered domain. Hence, we deduce that $G(p,x,y)$ has no critical point.

Case II. Secondly, we consider the interior of all the six faces of the cuboid Γ. On the face $p=0$, let $G(0,x,y) = h_1(x,y)$, where

$$h_1(x,y) = \frac{1}{72}(1-x^2)(x^2y^2 - 9x(y^2-1) + 8y^2)$$

for $x,y\in(0,1)$. A simple calculation shows that

$$\frac{\partial h_1}{\partial y} = \frac{1}{72}(1-x^2)(2x^2y - 18xy + 16y) \neq 0 \quad \forall\ x,y\in(0,1).$$

Hence $h_1(x,y)$ has no critical point in $(0,1)\times(0,1)$.

On the face $p = 2$, we simply get

$$G(2, x, y) = \frac{5}{576}. \tag{11}$$

On the face $x = 0$, let $G(p, 0, y) = h_2(p, y)$, where

$$h_2(p, y) = \frac{5p^6 + 144(4 - p^2)p^2(1 - y^2) + 256(4 - p^2)^2y^2 + 40(4 - p^2)p^3y}{36864}, \tag{12}$$

$p \in (0, 2)$ and $x \in (0, 1)$. The critical points of $h_2(p, y)$ are given by the solution of $\partial h_2(p, y)/\partial p = 0$ and $\partial h_2(p, y)/\partial y = 0$. The equation $\partial h_2(p, y)/\partial y = 0$ satisfies for

$$y_0 = \frac{5p^3}{4(25p^2 - 64)}.$$

The root y_0 lies in the range $(0, 1)$ whenever $p \in (1.67136, 2)$. Also, $\partial h_2(p, y)/\partial p = 0$ implies

$$\frac{p(625p^8 - 18050p^6 + 110320p^4 - 251904p^2 + 196608)}{1536(64 - 25p^2)^2} = 0,$$

which has no solution for the range $p \in (1.67136, 2)$. Consequently, $h_2(p, y)$ has no critical point for $p \in (0, 2)$ and $x \in (0, 1)$.

On the face $x = 1$, let $G(p, 1, y) = h_3(p, y)$ where

$$h_3(p, y) = \frac{5p^6 + 144p^2(4 - p^2) + 118p^4(4 - p^2) + 116p^2(4 - p^2)^2}{36864}, \quad p \in (0, 2). \tag{13}$$

A straightforward computation reveals that the maximum of $h_3(p, y)$ is attained at $p = 2\sqrt{2(25 - \sqrt{587})/3}$ given by

$$\max h_3(p, y) = \frac{587\sqrt{587} - 14200}{324}. \tag{14}$$

On the face $y = 0$, let $G(p, x, 0) = h_4(p, x)$, where

$$\begin{aligned} h_4(p, x) = \frac{1}{36864}\Big(&5p^6 + 26p^4(4 - p^2)x + 144p^2(4 - p^2)x^2 + 56p^4(4 - p^2)x^2 \\ &+ 68p^2(4 - p^2)^2x^2 + 36p^4(4 - p^2)x^3 + 40p^2(4 - p^2)^2x^3 \\ &+ 8p^2(4 - p^2)^2x^4 + (4 - p^2)(144p^2 + 288(4 - p^2)x)(1 - x^2)\Big). \end{aligned}$$

for $x \in (0, 1)$ and $p \in (0, 2)$. A simple calculation yields that $\partial h_4/\partial p = 0$ and $\partial h_4/\partial x = 0$ have no solutions for $p \in (0, 2)$ and $x \in (0, 1)$. Hence, $h_4(p, x)$ does not have any critical point.

On the face $y = 1$, let $G(p, x, 1) = h_5(p, x)$ such that

$$\begin{aligned} h_5(p, x) = \frac{1}{36864}\Big(&5p^6 + 26p^4(4 - p^2)x + 144p^2(4 - p^2)x^2 + 56p^4(4 - p^2)x^2 \\ &+ 68p^2(4 - p^2)^2x^2 + 36p^4(4 - p^2)x^3 + 40p^2(4 - p^2)^2x^3 + 8p^2(4 - p^2)^2x^4 \\ &+ (4 - p^2)(1 - x^2)(32(4 - p^2)x^2 + 144p^2x + 256(4 - p^2)) + (4 - p^2)(1 - x^2)(144p^3x \\ &+ 40p^3 + 32(4 - p^2)px^2 + 80(4 - p^2)px)\Big). \end{aligned}$$

Proceeding as in the above case, we obtain that $h_5(p,x)$ has no critical point for $p \in (0,2)$ and $x \in (0,1)$.

Case III. Now, we find the maximum value of $G(p,x,y)$ on the eight edges of the cuboid Γ. In view of the function $G(p,x,y)$, we have

$$G(p,0,0) = 5p^6 + 144(4-p^2)p^2/36864 =: k_1(p).$$

Clearly, the maximum value of $k_1(p)$ is attained at $p = 2\sqrt{2(6-\sqrt{21})/5}$ and given by

$$\max k_1(p) = \frac{1}{300}(7\sqrt{21} - 27) = 0.0169268.$$

Again considering the function $G(p,x,y)$, we obtain $G(p,0,1) = (5p^6 + 256(4-p^2)^2 + 40(4-p^2)p^3)/36864 =: k_2(p)$ for $p \in (0,2)$. Since $k_2(p)$ is a decreasing function, therefore maximum will attain at $p = 0$. Thus

$$\max k_2(p) = \frac{1}{9}.$$

From (13), it can be observed that $G(p,1,y)$ does not depend on y, hence we directly obtain $G(p,1,0) = G(p,1,1) = h_3(p,y) =: k_3(p)$. Thus from (14), we get that $\max k_3(p) = \max h_3(p,y) = (587\sqrt{587} - 14200)/324$.

On substituting $p = 0$ in (13), we get $G(0,1,y) = 0$. Also for $p = 2$, from (11) note that $G(2,x,y)$ is independent from of all the variables x and y. Therefore the value of $G(p,x,y)$ on the edges $p = 2$, $x = 0$; $p = 2$, $x = 1$; $p = 2$, $y = 0$ and $p = 2$, $y = 1$ given by

$$G(2,0,y) = G(2,1,y) = G(2,x,0) = G(2,x,1) = \frac{5}{576}$$

respectively. From (12) at $p = 0$, we obtain $G(0,0,y) = y^2/9 =: k_4(y)$. It is easy to verify that

$$\max k_4(y) = \frac{1}{9}.$$

Again, for $p = 0$ and $y = 1$, we have $G(0,x,1) = (1-x^2)(x^2+8)/72 =: k_5(x)$. Since $k_5(x)$ is a decreasing function of x, hence

$$\max k_5(x) = \frac{1}{9}.$$

Substituting $p = y = 0$ in $G(p,x,y)$, we get $G(0,x,0) = x(1-x^2)/8 =: k_6(x)$. It is a simple exercise to check that the maximum value of $k_6(x)$ is attained at $x = 1/\sqrt{3}$ given by

$$\max k_6(x) = \frac{1}{12\sqrt{3}}.$$

In view of the above all three cases, the maximum value $1/9$ of $G(p,x,y)$ is attained at the edge $p = x = 0$ and $y = 1$, which is the required bound. The sharpness of the bound follows from the function $f_4(z)$ given by (5). □

2.3 Inclusion relations and Radius problems

In 1999, Kanas and Wiśniowska [11] introduced the class

$$k-\mathcal{ST}=\left\{f\in\mathcal{A}:\operatorname{Re}\frac{zf'(z)}{f(z)}>k\left|\frac{zf'(z)}{f(z)}-1\right|\right\}$$

of $k-$starlike functions. Geometrically, the boundary of the domain $\Omega_k=\{w\in\mathbb{C}:\operatorname{Re}w>k|w-1|$ represents an ellipse for $k>1$, a parabola for $k=1$ and a hyperbola for $k>1$. The next result shows the inclusion relation with the class $k-\mathcal{ST}$ and the following classes (see [10, 2, 26])

$$\mathcal{S}^*(M)=\left\{f\in\mathcal{A}:\left|\frac{zf'(z)}{f(z)}-M\right|<M\right\},\quad M>\frac{1}{2},$$

$$\mathcal{ST}_p(a)=\left\{f\in\mathcal{A}:\operatorname{Re}\left(\frac{zf'(z)}{f(z)}+a\right)>\left|\frac{zf'(z)}{f(z)}-a\right|,\ a>0\right\}$$

and

$$\mu(\beta)=\left\{f\in\mathcal{A}:\operatorname{Re}\left(\frac{zf'(z)}{f(z)}\right)<\beta,\ \beta>1\right\}.$$

A function f in the class $\mathcal{S}^*(M)$ is called $M-$starlike function.

Theorem 15. *$k-\mathcal{ST}\subset\mathcal{S}^*_{nc}$ for $k\geq(4\cos 1)/(4\cos 1-\cos 2-1)$.*

Proof. Let $f\in K-\mathcal{ST}$ and $\Omega_k=\{w\in\mathbb{C}:\operatorname{Re}w>k|w-1|\}$. For $k>1$, the boundary curve $\partial\Omega_k$ is an ellipse $\gamma_k:x^2=k^2(x-1)^2+k^2y^2$, which can be rewritten as

$$\frac{(x-x_0)^2}{u^2}+\frac{(y-y_0)^2}{v^2}=1,$$

where $x_0=k^2/(k^2-1)$, $y_0=0$, $u=k/(k^2-1)$ and $v=1/\sqrt{k^2-1}$. Since $u>v$, therefore for the ellipse γ_k to lie inside φ_{nc}, the range of x_0+u must be less than or equal to $(4\cos 1)/(1+\cos 2)$. A simple calculation shows that this condition holds whenever $k\geq(4\cos 1)/(4\cos 1-\cos 2-1)$. □

Theorem 16. *$\mathcal{S}^*_{nc}\subset\mathcal{ST}_p(a)$ whenever $a\geq a_0\approx 0.402301$.*

Proof. As we know that the boundary $\partial\Omega_a$ of the domain $\Omega_a=\{w\in\mathbb{C}:\operatorname{Re}w+a>|w-a|\}$ denotes a parabola. So, $\mathcal{S}^*_{nc}\subset\mathcal{ST}_p(a)$, provided $\operatorname{Re}w+a>|w-a|$, where $w=(1+z)/\cos z$. Taking $z=e^{i\theta}$, we get $T(\theta)<a$, where

$$\begin{aligned}T(\theta)=\Big((\cos(\theta)+1)\sinh(\sin(\theta))\sin(\cos(\theta))+\sin(\theta)\cos(\cos(\theta))\cosh(\sin(\theta))\Big)^2\Big/\Big(2\\(\cos(2\cos\theta)+\cosh(2\sin\theta))((\cos\theta+1)\cos(\cos\theta)\cosh(\sin\theta)\\-\sin\theta\sinh(\sin\theta)\sin(\cos\theta))\Big).\end{aligned}$$

Clearly, the graph of $T(\theta)$ in Figure 3 shows that the maximum value of $T(\theta)$ is $T(\theta_0)\approx 0.402301$, where $\theta_0\approx 0.665124$ is the root of $T'(\theta)=0$. Since $\mathcal{ST}_p(a_1)\subset\mathcal{ST}_p(a_2)$ for $a_1<a_2$, therefore $\mathcal{S}^*_{nc}\subset\mathcal{ST}_p(a)$ for $a\geq b\approx 0.402301$. □

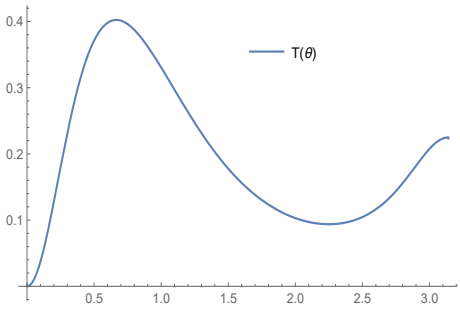
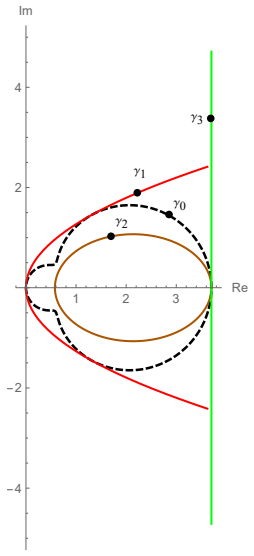


Figure 3: Graph of $T(\theta)$

The following result can be directly deduced from figure 2.

Theorem 17. *The class $\mathcal{S}^*_{nc}$ satisfies the following relations:*

1. $\mathcal{S}^*_{nc} \subset \mathcal{S}^*$.
2. $\mathcal{S}^*_{nc} \not\subset \mathcal{S}^*(\alpha)$ *for all* $\alpha \in (0,1)$.
3. $\mathcal{S}^*_{nc} \subset \mu(\beta)$ *whenever* $\beta \geq 2\sec 1$.



Legend -

$\gamma_0 : \frac{1+z}{\cos z}$

$\gamma_1 : |\omega - 0.4023| - \mathrm{Re}\,(\omega) = 0.4023$

$\gamma_2 : \mathrm{Re}\,\omega = (4\cos 1)/(4\cos 1 - \cos 2 - 1)|\omega - 1|$

$\gamma_3 : \mathrm{Re}\,\omega = 2\sec 1$

Figure 4: Inclusion graphs

Theorem 18. *Let $f \in \mathcal{S}^*_{nc}$, then the following hold:*

1. $f \in \mathcal{S}^*(\alpha)$ *in* $|z| \leq r_\alpha$, *where* r_α *is the root of* $(1-r) - \alpha\cos r = 0$.

2. $f \in \mu(\beta)$ *in* $|z| < r_\beta$, *where*

$$r_\beta = \begin{cases} r_0(\beta), & 1 < \beta < 2/(\cos 1), \\ 1, & 2/(\cos 1) \le \beta \end{cases} \tag{15}$$

and $r_0(\beta)$ *is the smallest root of*

$$1 + r = \beta \cos r.$$

The radii are sharp.

Proof. 1. Let $f \in \mathcal{S}^*_{nc}$, then

$$\frac{zf'(z)}{f(z)} \prec \frac{1+z}{\cos z}.$$

Therefore for $|z| = r < 1$, by Theorem 3, we have

$$\frac{1-r}{\cos r} \le \operatorname{Re} \frac{zf'(z)}{f(z)} \le \frac{1+r}{\cos r}.$$

Thus, $f \in \mathcal{S}^*(\alpha)$, whenever

$$\operatorname{Re} \frac{zf'(z)}{f(z)} \ge \frac{1-r}{\cos r} > \alpha.$$

This inequality is true for $r \in (0, r_\alpha)$, where r_α is the root of $(1-r) - \alpha \cos r = 0$. To see the sharpness consider the function $\tilde{f}(z)$ given by (6) as $\operatorname{Re}(z\tilde{f}'(z)/\tilde{f}(z)) = \alpha$ for $z = -r_\alpha$.

2. Similarly, $f \in \mu(\beta)$, whenever

$$\operatorname{Re} \frac{zf'(z)}{f(z)} \le \frac{1+r}{\cos r} < \beta, \quad (\beta > 1).$$

The above inequality holds for $r \in (0, r_\beta)$, where r_β is given by (15). Sharpness of the radius r_β follows from the function $\tilde{f}$ defined in (6).

□

Theorem 19. *If* $f \in \mathcal{S}^*_{nc}$, *then* f *is convex function of order* α *for* $|z| < r_c$, *where* r_c *is the root of*

$$(1-r)^2 - (r + \alpha(1-r)) \cos r - r(1-r) \sin r = 0. \tag{16}$$

Proof. Since $f \in \mathcal{S}^*_{nc}$, therefore there exists a Schwarz function $\omega(z)$ satisfying $\omega(0) = 0$ and $|\omega(z)| \le |z|$ such that

$$\frac{zf'(z)}{f(z)} = \frac{1+\omega(z)}{\cos \omega(z)},$$

A simple computation gives

$$1 + \frac{zf''(z)}{f'(z)} = \frac{z\omega'(z)}{1+\omega(z)} + \frac{z \sin \omega(z)}{\cos \omega(z)} + \frac{1+\omega(z)}{\cos \omega(z)}. \tag{17}$$

From (17), we get

$$\begin{aligned}\operatorname{Re}\left(1+\frac{zf''(z)}{f'(z)}\right) &= \operatorname{Re}\left(\frac{z\omega'(z)}{1+\omega(z)}\right)+\operatorname{Re}\left(\frac{z\sin\omega(z)}{\cos\omega(z)}\right)+\operatorname{Re}\left(\frac{1+\omega(z)}{\cos\omega(z)}\right)\\ &\geq -\frac{|z\omega'(z)|}{|1+\omega(z)|}+\operatorname{Re}\left(\frac{z\sin\omega(z)}{\cos\omega(z)}\right)+\operatorname{Re}\left(\frac{1+\omega(z)}{\cos\omega(z)}\right).\end{aligned}$$

Since $|\omega(z)|\leq|z|$. Let $\omega(z)=Re^{it}$ with $R\leq|z|=r<1$ and $t\in[0,2\pi]$. Using the Schwarz pick lemma

$$|\omega'(z)|\leq\frac{1-|\omega(z)|^2}{1-|z|^2}$$

and Theorem 3, we obtain

$$\operatorname{Re}\left(1+\frac{zf''(z)}{f'(z)}\right)\geq-\frac{r}{1-r}-\frac{r\sin r}{\cos r}+\frac{1-r}{\cos r}=:h(r).$$

Thus $f\in\mathcal{C}(\alpha)$, whenever $h(r)>\alpha$. Since $h(r)$ is a decreasing function and $h(0)=1$, therefore $h(r)>\alpha$ for $r\in(0,r_c)$, where r_c is the root of (16). □

Corollary 19.1. *The $\mathcal{C}$-radius, for the class $\mathcal{S}^*_{nc}$ is $r_c\approx 0.454$.*

Theorem 20. *If $f\in\mathcal{S}^*_{nc}$ then f is M-starlike in $|z|<r_3(M)$, where $r_3(M)$ is the root of the equation*

$$1-r-2M\cos r=0,$$

when $0<M\leq 1/2$ and $r_3(M)=1$ when $M>1/2$.

Proof. Since $f\in\mathcal{S}^*_{nc}$, there exist a Schwarz function $\omega(z)$ satisfying $|\omega(z)|<|z|$ such that

$$\frac{zf'(z)}{f(z)}=\frac{1+\omega}{\cos\omega}.$$

Thus, $f\in\mathcal{S}^*_{nc}$ is M-starlike whenever

$$\left|\frac{1+\omega}{\cos\omega}-M\right|<M.$$

By triangle inequality we can rewrite it as

$$\left|\frac{1+\omega}{\cos\omega}\right|<2M.$$

And we know that

$$\frac{1-|z|}{\cos|z|}\leq\left|\frac{1+\omega}{\cos\omega}\right|<2M.$$

From the above inequalities we get the desired result. □

2.4 Subordination Results

We need certain notations and results from the theory of differential subordination to establish our main results. Let $\mathcal{H}[a,n]$ denote the class of analytic functions f in $\mathbb{D}$ having the series expansion

$$f(z) = a + a_n z^n + a_{n+1} z^{n+1} + a_{n+2} z^{n+2} + \cdots .$$

We take $H[1,1] := \mathcal{H}$ for simplicity. Additionally, let $\mathcal{Q}$ be the class of analytic and univalent functions in $\overline{\mathbb{D}} \setminus E(q)$, where

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

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

Lemma 21. *[21, p.24] Let $q \in \mathcal{Q}$ with $q(0) = a$, and let $p(z) = a + a_n z^n + \cdots$ be analytic in $\mathbb{D}$ with $p(z) \neq a$ and $n \geq 1$. If p is not subordinate to q, then there exist points $z_0 = r_0 e^{i\theta_0} \in \mathbb{D}$ and $\zeta_0 \in \partial\mathbb{D} \setminus E(q)$ and an $m \geq n \geq 1$ for which $p(\mathbb{D}_{r_0}) \subset q(\mathbb{D})$,*

1. $p(z_0) = q(\zeta_0)$,
2. $z_0 p'(z_0) = m\zeta_0 q'(\zeta_0)$,
3. $\operatorname{Re}\{\frac{z_0 p''(z_0)}{p'(z_0)} + 1\} \geq m \operatorname{Re}\{\frac{\zeta_0 q''(\zeta_0)}{q'(\zeta_0)} + 1\}$

Lemma 22. *[21, Corollary 3.4h, p.132] Let q be univalent in $\mathbb{D}$, and let θ and ψ be analytic in a domain D containing $q(\mathbb{D})$ with $\psi(w) \neq 0$, when $w \in q(\mathbb{D})$. Set $Q(z) := zq'(z)\psi(q(z))$ and $h(z) := \theta(q(z)) + Q(z)$. Suppose that either h is convex or $Q(z)$ is starlike. In addition, assume that $\operatorname{Re}(zh'(z)/Q(z)) > 0$ $(z \in \mathbb{D})$. If p is analytic in $\mathbb{D}$ with $p(0) = q(0)$, $p(\mathbb{D}) \subset D$ and*

$$\theta(p(z)) + zp'(z)\psi(p(z)) \prec \theta(q(z)) + zq'(z)\psi(q(z)),$$

then $p \prec q$ and q is the best dominant.

Theorem 23. *If $p \in \mathcal{H}$ such that $p(0) = 1$ and*

$$\operatorname{Re}\left\{\frac{zp'(z)}{p(z)}\right\} < \frac{1}{2} + k_2, \tag{18}$$

where $k_2 = 1/\cosh 2$, then

$$p(z) \prec \frac{1+z}{\cos z}.$$

Proof. Suppose $p(z) \not\prec q(z) = \varphi_{nc}(z)$, then according to Lemma 21, there exist points $z_0 = r_0 e^{i\theta_0} \in \mathbb{D}$ and ζ_0 with $|\zeta_0| = 1$, such that we have $p(\mathbb{D}_{r_0}) \subset q(\mathbb{D})$,

$$p(z_0) = q(\zeta_0)$$

and

$$z_0 p'(z_0) = m\zeta_0 q'(\zeta_0).$$

Thus, we deduce that

$$\frac{z_0 p'(z_0)}{p(z_0)} = m\frac{\zeta_0 q'(\zeta_0)}{q(\zeta_0)}.$$

After some simple calculations we can easily get

$$\begin{aligned}\operatorname{Re}\left(\frac{z_0p'(z_0)}{p(z_0)}\right) = m\operatorname{Re}\left(\frac{\zeta_0q'(\zeta_0)}{q(\zeta_0)}\right) &= m\operatorname{Re}\left(\frac{\zeta_0\varphi_{nc}'(\zeta_0)}{\varphi_{nc}(\zeta_0)}\right)\\ &= m\left(\frac{1}{2}+\frac{1}{\cosh(2\sin\theta)}\right)\\ &\geq m\left(\frac{1}{2}+\frac{1}{\cosh 2}\right)\end{aligned}$$

for $\theta \in (0, 2\pi)$. This contradicts the hypothesis (18), and therefore, $p(z) \prec q(z)$. □

If we take $p(z) = zf'(z)/f(z)$ for $z \in \mathbb{D}$, then the following result can be readily derived.

Corollary 23.1. *If $f \in \mathcal{A}$ and*

$$\operatorname{Re}\left\{1+\frac{zf''(z)}{f'(z)}-\frac{zf'(z)}{f(z)}\right\} < \frac{1}{2}+k_2 \quad (z \in \mathbb{D}),$$

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

The next theorem provides an additional sufficient condition for a function f to belong to the class $\mathcal{S}^*_{nc}$.

Theorem 24. *If $f \in \mathcal{A}$ and satisfies*

$$\operatorname{Re}\left(1+\frac{zf''(z)}{f'(z)}\right) > \frac{1}{2}+\frac{2+\sinh 1}{\cos 1} \quad (z \in \mathbb{D}),$$

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

Proof. Suppose $f \notin \mathcal{S}^*_{nc}$ or equivlently

$$\frac{zf'(z)}{f(z)} \not\prec \varphi_{nc}(z).$$

Then according to Lemma 21, there are points $z_0 = r_0e^{i\theta_0} \in \mathbb{D}$, $\zeta_0 \in \partial\mathbb{D}$ and $m > 1$ such that

$$\frac{z_0f'(z_0)}{f(z_0)} = \varphi_{nc}(\zeta_0) \quad \text{and} \quad \left[z\left(\frac{zf'(z)}{f(z)}\right)'\right]\Big|_{z=z_0} = m\zeta_0(\varphi_{nc}(\zeta_0))'.$$

A computation gives

$$\begin{aligned}1+\frac{z_0f''(z_0)}{f'(z_0)} &= \varphi_{nc}(\zeta_0)+\frac{m\zeta_0(\varphi_{nc}(\zeta_0))'}{\varphi_{nc}(\zeta_0)}\\ &= \frac{1+\zeta_0}{\cos\zeta_0}+m\left(\frac{\zeta_0}{1+\zeta_0}+\frac{\zeta_0\sin\zeta_0}{\cos\zeta_0}\right).\end{aligned}$$

Thus, we have

$$\operatorname{Re}\left(1+\frac{z_0f''(z_0)}{f'(z_0)}\right) \leq \left|\frac{1+\zeta_0}{\cos\zeta_0}\right|+m\left(\left|\frac{\zeta_0}{1+\zeta_0}\right|+\left|\frac{\zeta_0\sin\zeta_0}{\cos\zeta_0}\right|\right). \tag{19}$$

Now, we find the maximum value of right hand side of the above inequality. Consider

$$|\cos\zeta|^2 = |\cos(e^{i\theta})|^2 = \cos^2(\cos\theta)\cosh^2(\sin\theta) + \sin^2(\cos\theta)\sin^2(\sin\theta) =: b(\theta),$$

where $-\pi \leq \theta \leq \pi$. It is easy to see that the equation $b'(\theta) = 0$ has five roots namely 0, $\pm\pi$ and $\pm\pi/2$ in $[-\pi, \pi]$. Since $b(\theta) = b(-\theta)$, it is sufficient to consider only those roots which lie in $[0, \pi]$. Now

$$\min\{b(0), b(\pi), b(\pi/2)\} = \cos^2 1.$$

Therefore, $|\cos(e^{i\theta})| \geq \cos 1$. Similarly, it can be verified that

$$|\sin e^{i\theta}| \leq \sinh 1.$$

Using these inequalities in (19), we obtain

$$\operatorname{Re}\left(1 + \frac{z_0 f''(z_0)}{f'(z_0)}\right) \leq \frac{2}{\cos 1} + m\left(\frac{1}{2} + \frac{\sinh 1}{\cos 1}\right),$$

which contradicts the hypothesis. Consequently, we have

$$\frac{zf'(z)}{f(z)} \prec \varphi_{nc}(z)$$

and $f \in \mathcal{S}^*_{nc}$. □

Now, we consider the first order differential subordination implication of the form

$$p(z) + \gamma z p'(z) \prec \frac{1+z}{\cos z} \implies p(z) \prec q(z).$$

We find the range of $\beta \in \mathbb{R}$ such that $p(z) \prec (1+Az)/(1+Bz)$, e^z, $1 + ze^z$ and $1 + \sin z$ hold.

Theorem 25. *Let p be analytic function in $\mathbb{D}$ such that $p(0) = 1$ and satisfies*

$$1 + \gamma z p'(z) \prec \frac{1+z}{\cos z}.$$

Let

$$g(z) = \int_0^z \frac{1 + t - \cos t}{t\cos t}\, dt. \tag{20}$$

Then the following hold:

1. *$p(z) \prec (1+Az)/(1+Bz)$, whenever*

$$\gamma \geq \max\left\{\frac{\gamma_2(1-B)}{A-B}, \left(\frac{1+B^2}{A-B-1-B^2}\right)\operatorname{Im} g(i)\right\},$$

   where $-1 \leq B < A \leq 1$ and $\operatorname{Im} g(i) \approx 0.862897$.

2. *$p(z) \prec e^z$ whenever $\gamma \geq (e\gamma_1)/(1-e) \approx 1.4308$,*

3. *$p(z) \prec 1 + ze^z$ whenever $\gamma \geq \gamma_2 \approx 2.45796$,*

4. $p(z) \prec 1 + \sin z$ *whenever* $\gamma \geq \gamma_2 \csc 1 \approx 1.82614$,

where $\gamma_1 = g(-1) \approx -0.904233$ *and* $\gamma_2 = g(1) \approx 1.53664$.

Proof. 1. The differential equation

$$1 + \gamma z q'(z) = \frac{1+z}{\cos z}$$

has a solution $q_\beta : \overline{\mathbb{D}} \to \mathbb{C}$ given by

$$q_\gamma(z) = 1 + \frac{1}{\gamma}\int_0^z \frac{1 + t - \cos t}{t \cos t} dt = 1 + \frac{1}{\gamma}\left(z + \frac{z^2}{4} + \frac{z^3}{6} + \frac{5z^4}{96} + \frac{z^5}{24} + \cdots\right).$$

Taking $\theta(w) = 1$ and $\psi(w) = \gamma$ in Lemma 22, the function $Q : \mathbb{D} \to \mathbb{C}$ becomes

$$Q(z) = z q'_\gamma(z)\psi(q_\gamma(z)) = \gamma z q'_\gamma(z) = \frac{1 + z - \cos z}{\cos z}.$$

A simple calculation shows that $\mathrm{Re}(zQ'(z)/Q(z)) > 0$ for $z \in \mathbb{D}$. Hence, $Q(z)$ is starlike function in $\mathbb{D}$. Since $h(z) = 1 + Q(z)$. Therefore $\mathrm{Re}(zh'(z))/Q(z) > 0$. Consequently by Lemma 22, $1 + \gamma z p'(z) \prec 1 + \gamma z q'(z)$ implies $p(z) \prec q_\gamma(z)$. Now, we only need to show that $q_\gamma(z) \prec \varphi_0(z) = (1 + Az)/(1 + Bz)$ for each parts. If $q_\gamma(z) \prec \varphi_0(z)$, then

$$\varphi_0(-1) \leq q_\gamma(-1) < q_\gamma(1) \leq \varphi_0(1) \tag{21}$$

and

$$\mathrm{Im}\, q_\gamma(i) < \mathrm{Im}\, \varphi_0(i). \tag{22}$$

A graphical representation shows that the conditions (21) and (22) are necessary as well as sufficient for $q_\gamma \prec \varphi_0(z)$. However, in other cases, the condition (21) is sufficient. Now using (21), we obtain

$$\frac{1-A}{1-B} \leq 1 + \frac{1}{\gamma} g(-1) \text{ which implies } \gamma \geq \frac{\gamma_1(1-B)}{B-A} =: \gamma_3$$

and

$$\frac{1+A}{1+B} \leq 1 + \frac{1}{\gamma} g(1) \text{ which implies } \gamma \geq \frac{\gamma_2(1-B)}{A-B} =: \gamma_4.$$

It can be easily seen that $\gamma_4 \geq \gamma_3$ for the range $-1 \leq B < A \leq 1$. From (22), we get

$$1 + \frac{1}{\gamma} g(i) \leq \frac{1 + Ai}{1 + Bi} \text{ which implies } \gamma \geq \left(\frac{1 + B^2}{A - B - 1 - B^2}\right) \mathrm{Im}\, g(i) =: \gamma_5.$$

Thus $q_\gamma(z) \prec \varphi_0(z)$ whenever $\gamma \geq \max\{\gamma_4, \gamma_5\}$.

2. For $\varphi_0(z) = e^z$, (21) reduces to

$$\frac{1}{e} \leq q_\gamma(-1) < q_\gamma(1) \leq e.$$

Now using (20) in the above inequality, we get

$$\frac{1}{e} \leq 1 + \frac{1}{\gamma} g(-1) \text{ which implies } \gamma \geq \frac{g(-1)e}{1-e} = \frac{e\gamma_1}{1-e} \approx 1.4308$$

and

$$\frac{1}{\gamma} g(1) \leq e - 1 \text{ which implies } \gamma \geq \frac{g(1)}{e-1} = \frac{\gamma_2}{e-1} \approx 0.8942.$$

Therefore $q_\gamma(z) \prec e^z$ whenever $\gamma \geq (e\gamma_1)/(1-e) \approx 1.4308$.

3. When $\varphi_0(z) = 1 + ze^z$. Then from (21), we get

$$1 - \frac{1}{e} \leq q_\gamma(-1) < q_\gamma(1) \leq 1 + e.$$

Using (20), we obtain

$$1 - \frac{1}{e} \leq 1 + \frac{1}{\gamma} g(-1) \text{ which gives } \gamma \geq -eg(-1) = -e\gamma_1 \approx 2.45796.$$

and

$$1 + \frac{g(1)}{\gamma} \leq 1 + e \text{ which gives } \gamma \geq \frac{g(1)}{e} = \frac{\gamma_2}{e} \approx 0.56529.$$

Thus $q_\gamma(z) \prec 1 + ze^z$ whenever $\gamma \geq \gamma_2 \approx 2.45796$.

4. When $\varphi_0(z) = 1 + \sin z$. Then from (21), we get

$$1 - \sin 1 \leq q_\gamma(-1) < q_\gamma(1) \leq 1 + \sin 1.$$

Now using (20), we have

$$1 - \sin 1 \leq 1 + \frac{1}{\gamma} g(-1) \text{ which gives } \gamma \geq -\gamma_1 \sin 1 \approx 0.760886.$$

and

$$1 + \frac{g(1)}{\gamma} \leq 1 + \sin 1 \text{ which gives } \gamma \geq \frac{g(1)}{\sin 1} = \frac{\gamma_2}{\sin 1} \approx 1.82614.$$

Consequently, $q_\gamma(z) \prec 1 + \sin z$ whenever $\gamma \geq \gamma_2 \csc 1 \approx 1.82614$.

□

In the following result, we consider the expression $p(z) + zp'(z)/p(z)$, where $p \in \mathcal{H}$, and the region bounded by the family of parabolas described by the equality $u = v^2/2 + (2b+1)/2$, $b < 1/2$. These parabolas are symmetric about the real axis having the vertex at $w = (2b+1)/2$. The family of domains containing point 1 inside and bounded by those parabolas may be characterised as

$$\Omega_b = \{w : \mathrm{Re}(w - b) > |w - 1 - b|\}, \tag{23}$$

or equivalently

$$\Omega_b = \{w = u + iv : 2u > v^2 + 2b + 1\}. \tag{24}$$

Theorem 26. *If $p \in \mathcal{H}$ such that*

$$p(z) + \frac{zp'(z)}{p(z)} \in \Omega_b, \tag{25}$$

then $p(z) \prec \varphi_{nc}(z)$ for $b < b_0 \approx -0.005796$.

Proof. Suppose $p(z) \not\prec \varphi_{nc}(z)$. Then according to Lemma 21, there exists points $z_0 = r_0 e^{i\theta_0} \in \mathbb{D}$, $\zeta_0 \in \partial\mathbf{D}$, $\zeta_0 \neq 1$, and $m \geq 1$ such that $p(z_0) = \varphi_{nc}(\zeta_0)$ and $z_0 p'(z_0) = m\zeta_0 \varphi'_{nc}(\zeta_0)$. On putting the value of $\varphi_{nc}(z)$, we obtain

$$\frac{z\varphi'_{nc}(z)}{\varphi_{nc}(z)} = \frac{z}{1+z} + \frac{z\sin z}{\cos z} \tag{26}$$

Note that for $\zeta_0 = e^{i\theta}$, $\theta \in [-\pi, \pi]$, we have

$$\begin{aligned} p(z_0) = \varphi_{nc}(\zeta_0) =& 2\left(\frac{(x+1)\cos x\cosh y - y\sinh y\sin x}{\cos(2x)+\cosh(2y)}\right) \\ &+ 2i\left(\frac{(x+1)\sinh y\sin x + y\cos x\cosh y}{\cos(2x)+\cosh(2y)}\right) \end{aligned}$$

and

$$\begin{aligned} \frac{z_0 p'(z_0)}{p(z_0)} &= \frac{m\zeta_0\varphi'_{nc}(\zeta_0)}{\varphi_{nc}(\zeta_0)}, \\ &= m\left(\frac{1}{2} + \frac{x\sin(2x) - y\sinh(2y)}{\cos(2x)+\cosh(2y)} + i\left(\frac{y}{(x+1)^2+y^2} + \frac{y\sin(2x)+x\sinh(2y)}{\cos(2x)+\cosh(2y)}\right)\right) \end{aligned}$$

where $x = \cos\theta$ and $y = \sin\theta$. Let us take $u := \mathrm{Re}(p(z_0) + z_0p'(z_0)/p(z_0))$ and $v := \mathrm{Im}(p(z_0) + z_0p'(z_0)/p(z_0))$, then we have

$$u = 2\left(\frac{(x+1)\cos x\cosh y - y\sinh y\sin x}{\cos(2x)+\cosh(2y)}\right) + m\left(\frac{1}{2} + \frac{x\sin(2x) - y\sinh(2y)}{\cos(2x)+\cosh(2y)}\right)$$

and

$$v = 2\left(\frac{(x+1)\sinh y\sin x + y\cos x\cosh y}{\cos(2x)+\cosh(2y)}\right) + m\left(\frac{y}{(x+1)^2+y^2} + \frac{y\sin(2x)+x\sinh(2y)}{\cos(2x)+\cosh(2y)}\right).$$

Observe, by the definition of Ω_b, that will be led to a contradiction if we show that

$$v^2 - 2u + 2b + 1 \geq 0.$$

Observe now that

$$v^2 - 2u + 2b + 1 \geq \min\{v^2 - 2u\} + 2b + 1.$$

After some calculation, we can get minimum will be ≈ -0.988408 at $\theta \approx -2.47734$. Hence

$$v^2 - 2u + 2b + 1 \geq 0,$$

if and only if $-0.988408 + 2b + 1 \geq 0$, that is, when $b \geq b_0 \approx -0.005796$ that leads to the contradiction. Thus the proof is complete. □

Corollary 26.1. *Let $f \in \mathcal{A}$ and let $1 + zf''(z)/f'(z) \in \Omega_b$, where Ω_b is given in (3.9) and $b \geq b_0 \approx -0.005796$. Then*

$$\frac{zf'(z)}{f(z)} \prec \frac{1+z}{\cos z}$$

or equivalently $f \in \mathcal{S}_P^*$.

Proof. Assume that $1 + zf''(z)/f'(z) \in \Omega_b$. Setting $p(z) = zf'(z)/f(z)$, we have $p(0) = 1$, and the above condition can be rewritten as

$$p(z) + \frac{zp'(z)}{p(z)} \in \Omega_b,$$

in $\mathbb{D}$. Now, applying Theorem 3.5, we conclude the assertion. □

3 Convolution results

Theorem 27. *A function $f \in \mathcal{S}_{nc}^*$ if and only if*

$$\frac{1}{z}\left[f(z) * \frac{z - \varphi(e^{i\theta})(z - z^2)}{(1-z)^2}\right] \neq 0, \quad \theta \in [-\pi, \pi]$$

Proof. Let $f \in \mathcal{S}_{nc}^*$ then by the definition

$$\frac{zf'(z)}{f(z)} \prec \frac{1+z}{\cos z}.$$

According to the definition of subordination, there exists a Schwarz function $\omega(z)$ with $\omega(0) = 0$ and $|\omega(z)| < 1$, such that

$$\frac{zf'(z)}{f(z)} = \frac{1+\omega(z)}{\cos(\omega(z))}.$$

Again by the definition of subordination we can get

$$\frac{zf'(z)}{f(z)} \neq \frac{1+e^{i\theta}}{\cos e^{i\theta}},$$

which is equivalent to

$$\frac{1}{z}\left[zf'(z) - f(z)\varphi(e^{i\theta})\right] \neq 0. \tag{27}$$

Now using the following basic convolution property in (4.3)

$$f(z) = f(z) * \left(\frac{z}{1-z}\right) \quad \text{and} \quad zf'(z) = f(z) * \left(\frac{z}{(1-z)^2}\right) \tag{28}$$

we can get desired result. □

If we put $f(z) = z + \sum_{n=2}^{\infty} a_n z^n$ in Theorem 27, we get the condition in terms of coefficients.

Corollary 27.1. *If $f \in \mathcal{A}$, then a necessary and sufficient condition for $f \in \mathcal{S}^*_{nc}$ is that*

$$\sum_{n=2}^{\infty}(\varphi(e^{i\theta}) - n)a_n z^{n-1} + \varphi(e^{i\theta}) \neq 1,$$

where $\varphi(z) = (1+z)/\cos z$.

Corollary 27.2. *If $f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in \mathcal{A}$ such that*

$$\sum_{n=2}^{\infty} |n - \varphi(e^{i\theta})||a_n| + M < 1, \quad z \in \mathbb{D}$$

then $f \in \mathcal{S}^*_{nc}$, where $M = 4\cos 1/(1+\cos 2)$.

Proof. Ler $\varphi(z) = (1+z)/\cos z$. Then, from Corollary 27.1, we have

$$\left|1 + \sum_{n=2}^{\infty}(n - \varphi(e^{i\theta}))a_n z^{n-1} - \varphi(e^{i\theta})\right| > 0.$$

By triangle inequality, we have

$$\begin{aligned}\left|1 + \sum_{n=2}^{\infty}(n - \varphi(e^{i\theta}))a_n z^{n-1} - \varphi(e^{i\theta})\right| &\geq 1 - \sum_{n=2}^{\infty}|(n - \varphi(e^{i\theta}))||a_n||z|^{n-1} - M, \\ &> 1 - \sum_{n=2}^{\infty}|(n - \varphi(e^{i\theta}))||a_n| - M\end{aligned}$$

$$\left|1 + \sum_{n=2}^{\infty}(n - \varphi(e^{i\theta}))a_n z^{n-1} - \varphi(e^{i\theta})\right| \geq 1 - \sum_{n=2}^{\infty}|(n - \varphi(e^{i\theta}))||a_n||z|^{n-1} - M,$$

and

$$1 - \sum_{n=2}^{\infty}|(n - \varphi(e^{i\theta}))||a_n||z|^{n-1} - M > 1 - \sum_{n=2}^{\infty}|(n - \varphi(e^{i\theta}))||a_n| - M.$$

By corollary 27.1, above inequality gives the desired result. □

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 the manuscript.

Data Availability

Not Applicable.

References

[1] R. M. Ali, V. Ravichandran and N. Seenivasagan, Coefficient bounds for p-valent functions, Appl. Math. Comput. **187** (2007), no. 1, 35–46.

[2] R. M. Ali, Subramanian, K.G., Ravichandran, V., Ahuja, O.P.: Neighborhoods of starlike and convex functions associated with parabola. J. Inequal. Appl. 346279, 9 (2008).

[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] S. Banga and S. S. Kumar, The sharp bounds of the second and third Hankel determinants for the class $\mathcal{SL}^*$, Mathematica Slovaca 70.4 (2020): 849-862.

[5] N.E. Cho, V. Kumar, S.S. Kumar, V. Ravichandran, Radius problems for starlike functions associated with the sine function. Bull. Iranian Math. Soc. 45(1), 213â€“232 (2019).

[6] K. Cudna, O. S. Kwon, A. Lecko, Y. J. Sim and B. Śmiarowska, (2020), The second and third-order Hermitian Toeplitz determinants for starlike and convex functions of order α. Bol de la Sociedad Mate Mexicana, 26(2), 361-375.

[7] S. S. Kumar and K. Gangania, A cardioid domain and starlike functions, Analysis and Mathematical Physics 11 (2021): 1-34.

[8] S. Giri and S. S. Kumar, Hermitian-Toeplitz determinants for certain univalent functions, Anal. Math. Phys. **13** (2023), no. 2, Paper No. 37, 19 pp.

[9] P. Goel , and S. S. Kumar, Certain class of starlike functions associated with modified sigmoid function, Bulletin of the Malaysian Mathematical Sciences Society 43.1 (2020): 957-991.

[10] W. Janowski, Extremal problems for a family of functions with positive real part and for some related families. Ann. Polon. Math. 23, 159-177 (1970/1971).

[11] S. Kanas and A. Wiśniowska, Conic regions and k-uniform convexity. J. Comput. Appl.Math. 105(1â€“2), 327â€“336 (1999).

[12] B. Kowalczyk, A. Lecko and D. K. Thomas, The sharp bound of the third Hankel determinant for starlike functions, Forum Mathematicum, vol. 34, no. 5, 2022, pp. 1249-1254.

[13] O. S. Kwon, A. Lecko, Y. J. Sim, On the fourth coefficient of functions in the Carathéodory class, Comput. Methods Funct. Theory 18(2), 307â€“314 (2018).

[14] R. J. Libera, E. J. Zĺotkiewicz, Early coefficients of the inverse of a regular convex function, Proc. Am. Math. Soc. 85(2), 225â€“230 (1982).

[15] R. J. Libera, E. J. Zĺotkiewicz, Coefficient bounds for the inverse of a function with derivative in $\mathcal{P}$, Proc. Am. Math. Soc. 87(2), 251â€“257 (1983).

[16] L. Liu and J. L. Liu, Properties of certain multivalent analytic functions associated with the lemniscate of Bernoulli, Axioms 10, 160 (2021).

[17] 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.

[18] A. Lecko, Y.J. Sim and B. Śmiarowska, The Sharp Bound of the Hankel Determinant of the Third Kind for Starlike Functions of Order 1/2. Complex Anal. Oper. Theory 13, 2231â€“2238 (2019).

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

[20] R. Mendiratta, S. Nagpal and V. Ravichandran, On a subclass of strongly starlike functions associated with exponential function. Bull. Malays. Math. Sci. Soc. 38(1), 365â€“386 (2015)

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

[22] M. Obradović and N. Tuneski, Hermitian Toeplitz determinants for the class $\mathcal{S}$ of univalent functions, Armen. J. Math. **13** (2021), Paper No. 4, 10 pp.

[23] N. Verma and S. S. Kumar, A Conjecture on $H_3(1)$ For Certain Starlike Functions, arXiv preprint arXiv:2208.02975 (2022) (accepted in *Mathematica Slovaca*).

[24] F. Rønning, Uniformly convex functions and a corresponding class of starlike functions, Proc. Amer. Math. Soc. **118** (1993), no. 1, 189–196.

[25] T.N. Shanmugam, Convolution and differential subordination. Int. J. Math. Math. Sci. 12(2), 333â€“340 (1989).

[26] B. A. Uralegaddi, Ganigi, M.D., Sarangi, S.M.: Univalent functions with positive coefficients. Tamkang J. Math. 25(3), 225â€“230 (1994).

[27] V. Ravichandran and S. Varma, Bound for the fifth coefficient of certain starlike functions. C. R. Math. Acad. Sci. Paris. 353(6), 505-510 (2015).

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

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Stochastic Neural Network Control for Stochastic Nonlinear Systems With Quadratic Local Asymmetric Prescribed Performance

Yu Xia, *Member, IEEE*, Ke Xiao, Jinde Cao, *Fellow, IEEE*, Radu-Emil Precup, *Senior Member, IEEE*, Yogendra Arya, *Senior Member, IEEE*, Hak-Keung Lam, *Fellow, IEEE*, and Leszek Rutkowski, *Life Fellow, IEEE*

***Abstract*—This article presents an adaptive neural network control scheme with prescribed performance for stochastic nonlinear systems. Unlike existing adaptive stochastic control schemes that primarily utilize deterministic neural networks for approximations in complex stochastic environments, we employ stochastic neural networks to approximate the stochastic nonlinear terms, effectively resolving the "memory overflow" issue. Moreover, we propose a novel prescribed performance design method, which distinguishes itself from the previous prescribed performance control schemes by integrating a quadratic characteristic capable of suppressing transient input vibrations, along with a local asymmetric characteristic that optimize both transient output overshoot and steady-state error bias. Furthermore, the proposed control scheme is implemented within a fixed-time framework to ensure that all closed-loop systems are fixed-time bounded in probability, with the tracking error consistently within the predefined performance bounds. Simulation results validate the effectiveness of the proposed control scheme.**

***Index Terms*—Fixed-time stabilization, prescribed performance, stochastic neural network (SNN), stochastic nonlinear system.**

Received 27 August 2024; revised 28 September 2024 and 10 November 2024; accepted 17 November 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 52375039 and Grant 52175022; in part by the National Key Research and Development Program of China under Grant 2022YFB4702201; in part by the Ministry of Research, Innovation and Digitization, CNCS/CCCDI—UEFISCDI, within PNCDI IV under Project ERANET-ENUAC-e-MATS; in part by the "Excellence Initiative—Research University" Program for the AGH University in Krakow as well as the ARTIQ Project under Grant UMO-2021/01/2/ST6/00004 and Grant ARTIQ/0004/2021; and in part by the Polish Ministry of Science and Higher Education funds assigned to AGH University. This article was recommended by Associate Editor B. Ribeiro. *(Corresponding authors: Ke Xiao; Radu-Emil Precup.)*

Yu Xia is with the State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, Chongqing 400044, China, and also with the State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: rainyx@sjtu.edu.cn).

Ke Xiao is with the State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, Chongqing 400044, China (e-mail: xiaoke993@cqu.edu.cn).

Jinde Cao is with the School of Mathematics, Southeast University, Nanjing 211189, China, and also with Purple Mountain Laboratories, Nanjing 211111, China (e-mail: jdcao@seu.edu.cn).

Radu-Emil Precup is with the Department of Automation and Applied Informatics, Politehnica University of Timisoara, 300223 Timisoara, Romania, and also with the Center for Fundamental and Advanced Technical Research, Romanian Academy–Timisoara Branch, 300223 Timisoara, Romania (e-mail: radu.precup@aut.upt.ro).

Yogendra Arya is with the Department of Electrical Engineering, J.C. Bose University of Science and Technology, YMCA, Faridabad 121006, India (e-mail: yarya@jcboseust.ac.in).

Hak-Keung Lam is with the Department of Engineering, King's College London, WC2R 2LS London, U.K. (e-mail: hak-keung.lam@kcl.ac.uk).

Leszek Rutkowski is with the Systems Research Institute of the Polish Academy of Sciences, 01-447 Warsaw, Poland, also with the AGH University of Krakow, 30-059 Kraków, Poland, and also with the University of Social Sciences, 90-113 Lódz, Poland (e-mail: leszek.rutkowski@ibspan.waw.pl).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2024.3502496.

Digital Object Identifier 10.1109/TCYB.2024.3502496

I. Introduction

In RECENT years, there has been a growing interest among researchers in exploring the impact of stochastic factors on control systems. This curiosity has prompted a shift toward describing the dynamic behavior of nonlinear systems using stochastic differential equations, leading to the development of numerous sophisticated stochastic control schemes. Wu et al. [1] proposed an adaptive dynamic surface control scheme incorporating neural networks to address the predefined tracking problem in uncertain stochastic nonlinear systems. Jia and He [2] developed an adaptive neural network fault-tolerant control method utilizing barrier Lyapunov functions for uncertain stochastic nonlinear systems, aiming to achieve fault-tolerant tracking control while maintaining output constraints. Song et al. [3] proposed a neuroadaptive output feedback control method for non-affine stochastic MIMO nonlinear systems, incorporating deferred constraints to mitigate the negative effects of filter errors. Wu et al. [4] proposed an adaptive neural network fixed-time control strategy for nonstrict-feedback stochastic nonlinear systems under injection and deception attacks, utilizing a self-triggered mechanism that eliminates the need for an additional observer, enhances resource efficiency, and ensures semi-global uniform boundedness of the system signals while achieving deferred output constraint control performance. Sedghi et al. [5] addressed distributed finite-time consensus control for stochastic nonlinear multiagent systems with directed communication graphs, introducing a control method that integrates backstepping, command filtering, neural networks, and auxiliary systems to handle unknown dynamics, external disturbances, and input saturation. Wang et al. [6] presented an adaptive fuzzy backstepping control strategy for stochastic high-order nonlinear systems with full-state constraints, utilizing fuzzy logic systems

2168-2267 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:12:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

and barrier Lyapunov functions. Yang et al. [7] explored the integral sliding mode control problem for a class of stochastic nonlinear systems under cyber-attack, employing the Takagi–Sugeno fuzzy model to address the stochastic nonlinear terms. Xia et al. [8] presented an event-based prescribed adaptive fuzzy control scheme for stochastic nonlinear systems, addressing initial boundary constraints on tracking errors and the limitations of conservative trigger parameter design by introducing tangential envelope functions, a barrier function, time-varying scaling transformations, and a new event-triggered mechanism. Li and Tong [9] investigated adaptive fuzzy tracking control for uncertain stochastic nonlinear systems, proposing an asymptotic tracking control scheme that eliminates the restrictive requirement of prior knowledge of unknown virtual control coefficients through the use of smooth functions, bounded estimation, and a specific composite Lyapunov function. Xia et al. [10] addressed adaptive fuzzy control for stochastic MEMS gyroscopes by proposing a control scheme that employs interval type-3 fuzzy logic to enhance precision and minimize communication load, while ensuring fixed-time convergence of tracking errors. In comparison to deterministic system control, the control schemes [1], [2], [3], [4], [5], [6], [7], [8], [9], [10] of stochastic systems primarily focus on ensuring that the system states are bounded in probability, thereby achieving stability in a statistical sense. This analytical framework is particularly valuable for applications in stochastic environments, where the dynamic behavior of systems may not be accurately captured by deterministic equations.

A challenging issue in stochastic system control is how to handle nonlinear terms containing stochastic variables. Benefiting from the rapid development of adaptive techniques, deterministic neural networks (DNNs) [1], [2], [3], [4], [5] and fuzzy logic systems [6], [7], [8], [9], [10] have been widely adopted to approximate these nonlinear terms, and data-driven and model-based control approaches are discussed in the recent survey [11]. However, using deterministic approximators to deal with stochastic nonlinear terms is unmatched. Deterministic approximators typically can only generate and memorize a single sample function, which may not capture all stochastic information when confronted with stochastic signal inputs. Moreover, since deterministic approximators need to learn each input sample, they are prone to "memory overflow" issues when applied to complex stochastic environments. In fact, researchers long ago evaluated the modeling capabilities of stochastic neural networks (SNNs) [12], [13] and probabilistic fuzzy logic systems [14], [15] for stochastic processes, confirming the superiority of stochastic (probabilistic) approximators in handling stochastic functions. However, regrettably, these outstanding achievements have yet to be applied to the control of stochastic nonlinear systems.

Another intriguing topic revolves around prescribed performance functions (PPFs), first introduced by Bechlioulis and Rovithakis [16]. As an effective class of state constraint tools, they have garnered increasing attention and application [17], [18], [19] in recent years, giving rise to various derivative techniques, such as finite-time [20], fixed-time [21], [22] PPFs, and prescribed-time [23], [24] control, as well as asymmetric designs [25], [26]. However, the concept of adding settling time to PPFs is inherently limited to the functions themselves. In the theory, the settling time of PPFs does not necessarily align with the settling time of the system states. Therefore, a more conservative design approach should focus on finite-time [27] and fixed-time [28] Lyapunov stability analysis. On the other hand, concerning transient performance analysis of PPFs, while the aforementioned approaches [25], [26] with asymmetric boundaries can effectively regulate output overshoot, these asymmetric boundaries continue to act on tracking errors even after the system reaches the steady state, leading to bias in steady-state tracking errors. An ideal design would entail maintaining asymmetric constraints for PPFs during transients but promptly adjusting to symmetric constraints upon entering steady state, thereby not affecting steady-state tracking errors. However, there is limited research on this aspect in the existing literature. Furthermore, many researchers overlook another aspect of transient behavior, namely, the issue of control input corresponding to PPFs. Typically, we need to design a relatively high convergence rate for PPFs to achieve rapid convergence of tracking errors. However, an excessively high convergence rate may cause transient input vibrations, potentially leading to safety issues. Thus, for PPFs, a novel design that can simultaneously achieve fast convergence and low transient input vibrations is meaningful.

Based on the observations above, this article aims to propose an adaptive control method that employs SNN approximation and an improved prescribed performance design. Its main merits are as follows.

1) Using SNNs for modeling and approximating uncertain terms in stochastic nonlinear systems aims to address the issues of "information loss" and memory overflow that have been observed when DNNs [1], [2], [3], [4], [5] are applied to stochastic control.
2) *A Novel Method for Prescribed Performance Design Is Proposed, Consisting of Two Main Components:* a quadratic PPF (QPPF) to regulate transient input vibrations and two adjustable envelope functions facilitating the transition from transient asymmetric constraints to steady-state symmetric constraints. This method improves upon previous PPF approaches, particularly those that either completely ignored transient input vibrations [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26] or employed global asymmetric designs [25], [26]. Therefore, this method ensures that the constrained system maintains excellent performance both in transient and steady-state conditions.

II. Preliminaries and Problem Formulation

A. Preliminaries

Consider the following stochastic nonlinear system:

$$dx(t) = f(t, x(t))dt + g(t, x(t))d\omega(t) \tag{1}$$

where $x \in R^n$ denotes the system state, $f(t, x(t)) : R^+ \times R^n \to R$ and $g(t, x(t)) : R^+ \times R^n \to R^r$ denote the unknown nonlinear

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:12:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

functions, and $f(t,0)=0$, $g(t,0)=0$, and $\omega \in R^r$ denotes the standard Brownian motion.

Definition 1 [6]: For the system (1) and any positive definite function $V(x)$, the differential operator $\mathcal{L}$ can be expressed as follows:

$$\mathcal{L}V(x)=\frac{\partial V}{\partial x}f(x)+\frac{1}{2}Tr\left\{g^T(x)\frac{\partial^2 V}{\partial x^2}g(x)\right\} \tag{2}$$

where $Tr(\cdot)$ indicates the trace of the matrix.

Lemma 1 [29]: For two functions, $\wp_1(\cdot), \wp_2(\cdot) \in k_\infty$, and some constants, $0<p<1$, $q>1$, and $\sigma_1, \sigma_2, \beta>0$, such that

$$\begin{aligned}&\mathcal{L}V(x) \le -\sigma_1 V^p(x)-\sigma_2 V^q(x)+\beta\\ &\wp_1(\|x\|) \le V(x) \le \wp_2(\|x\|).\end{aligned} \tag{3}$$

Then, the solution of the system (1) is fixed-time bounded in probability (FTBIP) with

$$\mathbb{E}[T] \le T_s=\frac{1}{\sigma_1\eta(1-p)}+\frac{1}{\sigma_2\eta(q-1)} \tag{4}$$

where $0<\eta<1$ is a constant.

Lemma 2 [30]: Consider the following filter:

$$\begin{cases}\dot{\varphi}_1=\varphi_2-r_1\Gamma^{\frac{1}{2}}\mathrm{sig}(\varphi_1-\alpha_1)^{\frac{1}{2}}\\ \qquad\quad -h_1\Lambda^{\frac{1}{2}}\mathrm{sig}(\varphi_1-\alpha_1)^{\mu_1}\\ \dot{\varphi}_2=-r_2\Gamma\mathrm{sig}(\varphi_2-\dot{\varphi}_1)\\ \qquad\quad -h_2\Lambda\mathrm{sig}(\varphi_1-\alpha_1)^{\mu_2}\end{cases} \tag{5}$$

where $\mathrm{sig}(\cdot)^\alpha=|\cdot|^\alpha\mathrm{sgn}(\cdot)$, $\Gamma, \Lambda \in R^+$, $\mu_1=\mu$, $\mu_2=2\mu-1$, and $\mu \in (1, 1+\iota)$ with $\iota \in R^+$ being a sufficiently small constant, and r_1, r_2, h_1, and h_2 are chosen under the condition that the matrices $\begin{bmatrix}-r_1 & 1\\ -r_2 & 0\end{bmatrix}$, and $\begin{bmatrix}-h_1 & 1\\ -h_2 & 0\end{bmatrix}$ are Hurwitz, as defined in (8). Then, within a fixed time T_f, the output of filter (5) satisfies $\bar{\alpha}_1=\varphi_1$, $\dot{\bar{\alpha}}_1=\dot{\varphi}_1$, and $\bar{\alpha}_1-\alpha_1 \le \bar{\eta}$, where $\bar{\eta}$ is a design parameter and T_f is represented as

$$\mathbb{E}[T] \le T_f=\frac{2\lambda_{\max}^{\frac{3}{2}}(\mathcal{M}_1)}{\lambda_{\min}(\mathcal{N}_1)}+\frac{\lambda_{\max}(\mathcal{M}_2)}{\lambda_{\min}(\mathcal{N}_2)(\mu-1)\varsigma^{\mu-1}} \tag{6}$$

where $\varsigma \le \lambda_{\min}(\mathcal{M}_2)$, and $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{N}_1$, and $\mathcal{N}_2$ satisfy

$$\begin{aligned}\mathcal{A}_1^T\mathcal{M}_1+\mathcal{M}_1\mathcal{A}_1&=-\mathcal{N}_1\\ \mathcal{A}_2^T\mathcal{M}_2+\mathcal{M}_2\mathcal{A}_2&=-\mathcal{N}_2\end{aligned} \tag{7}$$

with

$$\mathcal{A}_1=\begin{bmatrix}-r_1 & 1\\ -r_2 & 0\end{bmatrix}, \mathcal{A}_2=\begin{bmatrix}-h_1 & 1\\ -h_2 & 0\end{bmatrix}. \tag{8}$$

Lemma 3 [8]: For any $\varrho, \jmath \ge 0$ and $a_1, a_2, a_3>0$, the following holds:

$$|\varrho|^{a_1}|\jmath|^{a_2} \le \frac{a_1}{a_1+a_2}a_3|\varrho|^{a_1+a_2}+\frac{a_2}{a_1+a_2}a_3^{-\frac{a_1}{a_2}}|\jmath|^{a_1+a_2}. \tag{9}$$

B. System Model and Control Objective

Consider the following stochastic nonlinear system:

$$\begin{cases}dx_i=\left[x_{i+1}+f_i(t,\bar{x})\right]dt+g_i^T(t,\bar{x})d\omega\\ \qquad 1 \le i \le n-1\\ dx_n=\left[u(t)+f_n(t,\bar{x})\right]dt+g_n^T(t,\bar{x})d\omega\\ y=x_1\end{cases} \tag{10}$$

where $\bar{x}=[x_1, x_2, \ldots, x_n]^T \in R^n$, $y \in R$, and $u(t) \in R$ denote the system state, output, and input, respectively. $f_i(t,\bar{x}): R^+ \times R^n \to R$ and $g_i(t,\bar{x}): R^+ \times R^n \to R^r$ denote the unknown nonlinear functions, and $f_i(t,0)=0$, $g_i(t,0)=0$, $i=1,2,\ldots,n$, and $\omega \in R^r$ denotes a standard Brownian motion.

Control Objective: The control objective of this article is to develop an adaptive control scheme for the system (10), ensuring that all signals are FTBIP and maintaining tracking errors within the boundaries of predetermined performance.

Assumption 1 [20]: The tracking signal y_d and its first-order derivative $\dot{y}_d$ are bounded and continuous.

C. Stochastic Neural Network

Considering that the input of a neural network is a stochastic variable $x(t)$, inspired by [12], [13], [14], and [15], we define the following neural network $\Phi_n^T(x)$:

$$\Phi_n^T(x)=\left(\sum_{i=1}^n \psi_{i1}H_{i1}(x), \sum_{i=1}^n \psi_{i2}H_{i2}(x), \ldots, \sum_{i=1}^n \psi_{in}H_{in}(x)\right) \tag{11}$$

where $\psi_{i1}, \psi_{i2}, \ldots, \psi_{in}$ are the weight vectors and $H_{i1}, H_{i2}, \ldots, H_{in}$ represent some known smooth functions, such as radial basis functions, multilayer perceptron functions, and approximate identity functions. Integrating (11) yields

$$\int_0^{+\infty}\Phi_n^T(x)d\omega=U\Phi_n^T(x) \tag{12}$$

where U is an isometric correspondence defined in *stochastic integral* [31].

Based on the definition of *vector-valued stochastic measure* [32], if we consider ω to represent the Brownian Motion, then the function F_n involving a stochastic vector $x(t)$ can be expressed as $F_n(x)=\int_0^{+\infty}\Phi_n^T(x)d\omega$.

Analogously, the network $\Phi^T(x)$ is transformed into the function $F(x)$ by U, represented as

$$F(x)=\int_0^{+\infty}\Phi^T(x)d\omega=U\Phi^T(x). \tag{13}$$

From (11)–(13), one has

$$\begin{aligned}F_n=&\sum_{i=1}^n\psi_{i1}\int_0^{+\infty}H_{i1}(x)d\omega+\sum_{i=1}^n\psi_{i2}\int_0^{+\infty}H_{i2}(x)d\omega\\ &+\cdots+\sum_{i=1}^n\psi_{in}\int_0^{+\infty}H_{in}(x)d\omega.\end{aligned} \tag{14}$$

Rewriting (14) yields

$$F_n=\sum_{i=1}^n\bar{\psi}_i^T\bar{H}_i \tag{15}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:12:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

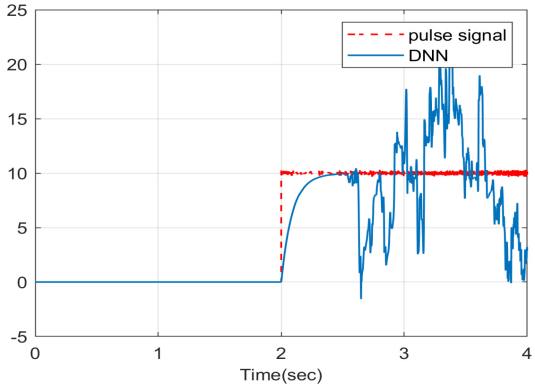


Fig. 1. Modeling capabilities of DNNs.

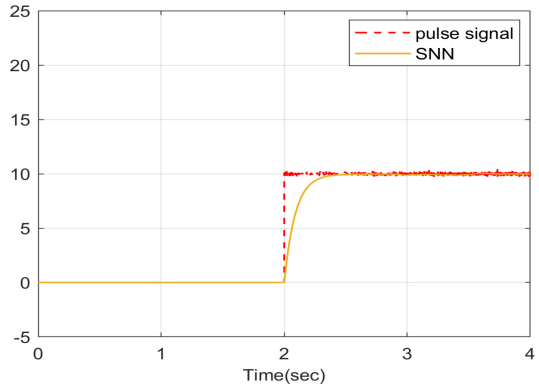


Fig. 2. Modeling capabilities of SNNs.

where $\bar{\psi}_i^T = (\psi_{i1}, \ldots, \psi_{in})$, and $\bar{H}_i = (\int_0^{+\infty} H_{i1}(x)d\omega, \ldots, \int_0^{+\infty} H_{in}(x)d\omega)^T$.

Lemma 4 [13], [15]: For an unknown nonlinear function $F(x)$ with $x(t)$ being a stochastic variable, it can be approximated by an SNN $\bar{\psi}^T\bar{H}$, i.e.,

$$F(x) = \bar{\psi}^T\bar{H}(x) + \varepsilon(x)$$
$$|\varepsilon(x)| \leq \bar{\varepsilon} \quad \forall x \in \Omega_x = \{x | \|x\| \leq \bar{x}\} \tag{16}$$

where $\varepsilon(x)$ denotes an approximation error, Ω_x denotes a compact set, and $\bar{\varepsilon}$ and $\bar{x}$ are the given constants.

Remark 1: Essentially, SNNs can maintain more information compared to deterministic ones. This is because when a neural network operates in a stochastic manner, it can generate (and consequently remember) a collection of sample functions, whereas when it operates deterministically, it can only generate (and consequently remember) a single sample function. Furthermore, unlike DNNs [1], [2], [3], [4], [5], SNNs do not require learning every input sample. As a result, SNNs are not susceptible to the memory overflow issue caused by the vast number of signals they need to remember while operating in a stochastic environment.

Remark 2: In this remark, we provide a comparison of the modeling capabilities of SNNs and DNNs in a stochastic environment. At $t = 2$s, we apply a pulse signal with a numerical value of $10 + 0.1 \times$ randn, where randn denotes Gaussian white noise with a mean of 0 and a variance of 1. To ensure fairness, both SNNs and DNNs are equipped with the same Gaussian basis functions and consist of 30 neurons each. Figs. 1–3 depict the modeling performance of these two types of neural networks regarding the pulse signal. From these figures, we can observe that DNNs can effectively model the pulse signal within a short duration after its application. However, over time, they encounter the issue of memory overflow, which undermines their learning capability. In contrast, SNNs demonstrate continuous learning ability in a stochastic environment. These results also corroborate our discussion in Remark 1.

D. Quadratic Prescribed Performance Function

To achieve the desired performance behavior with the capability of suppressing transient input vibrations, the following

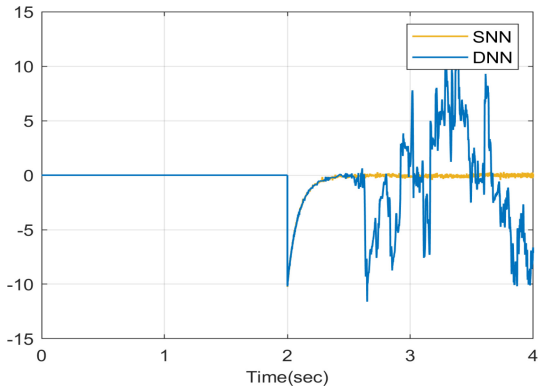


Fig. 3. Comparison of modeling errors between SNNs and DNNs.

QPPF is proposed:

$$\rho(t) = \left(\rho_0 - \rho_f\right)e^{-\nu t^2} + \rho_f \tag{17}$$

where $\nu > 0$ is the design parameter that influence the convergence rate and $\rho_0 \gg \rho_f > 0$ are the design parameters that dictate the initial and final values of the prescribed performance boundary, respectively. Moreover, the tracking error of the system (10) is defined as $e(t) = x_1 - y_d$, and satisfies the initial condition $-\rho(0) < e(0) < \rho(0)$.

Remark 3: In the existing control schemes [16], [17], [18], where the traditional PPF (TPPF) is typically designed as $\rho(t) = (\rho_0 - \rho_f)e^{-\nu t} + \rho_f$, a general control law can be expressed as

$$u = \lambda(t)^{-1}\left(-ks(t) + \frac{\lambda(t)e(t)\dot{\rho}(t)}{\rho(t)}\right) + \dot{y}_d \tag{18}$$

where k denotes a constant, and $\lambda(t)$ and $s(t)$ represent some smooth functions.

To calculate the derivative of TPPF, we can obtain $\dot{\rho}(t) = -\nu(\rho_0 - \rho_f)e^{-\nu t}$. This indicates that increasing the convergence rate ν will lead to rapid changes of $\dot{\rho}(t)$ in the initial phase. Consequently, it can be inferred from (18) that the occurrence of transient input vibrations is unavoidable at a higher convergence rate.

Remark 4: To better illustrate the advantages of QPPF, we compare it with TPPF as follows: set $\rho_0 = 0.66$ and $\rho_f = 0.06$, and consider two scenarios with $\nu = 1$ and $\nu = 10$. The comparison results are presented in Figs. 4 and 5. From these figures, it is evident that the transient behaviors

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:12:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

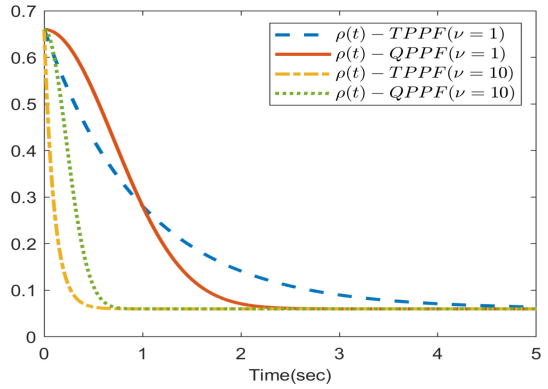


Fig. 4. Comparison of $\rho(t)$ between TPPF and QPPF.

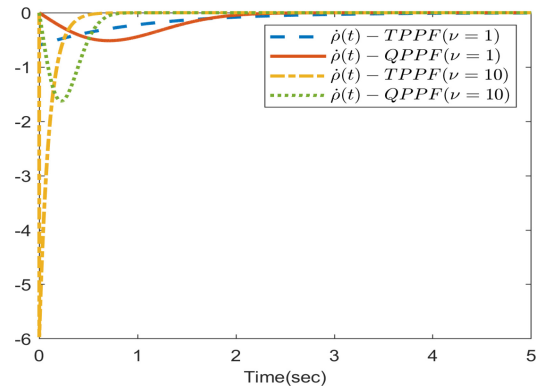


Fig. 5. Comparison of $\dot{\rho}(t)$ between TPPF and QPPF.

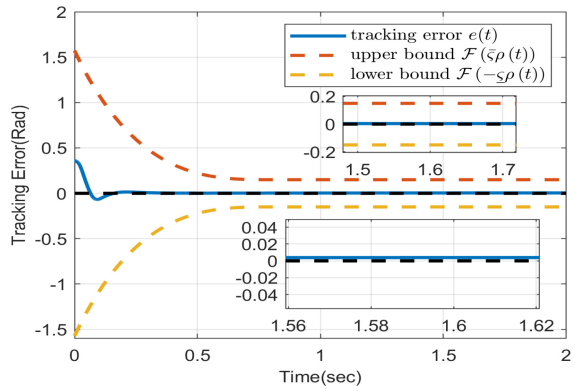


Fig. 6. Tracking error when using a symmetric design.

of $\rho(t)$ and $\dot{\rho}(t)$ for both QPPF and TPPF are similar at a lower convergence rate $\nu = 1$. However, at a higher convergence rate $\nu = 10$, QPPF shows a smaller value of $\dot{\rho}(t)$ compared to TPPF, while the transient behavior of $\rho(t)$ remains similar. Combining this observation with (18), we can conclude that QPPF exhibits reduced transient input vibrations in comparison to TPPF.

Remark 5: Although [33] can achieve similar functionality, it relies on a complex observer to implement such features, which introduce additional design parameters and filtering errors. In contrast, the QPPF is a simple improvement on TPPFs, effectively reducing design complexity.

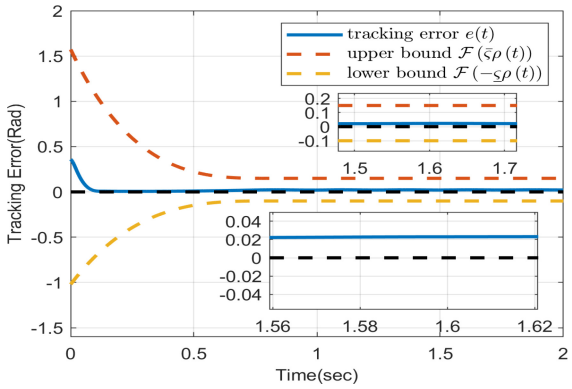


Fig. 7. Tracking error when using a global asymmetric design.

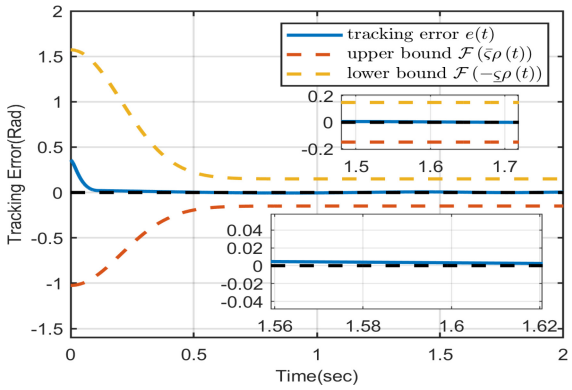


Fig. 8. Tracking error when using a local asymmetric design.

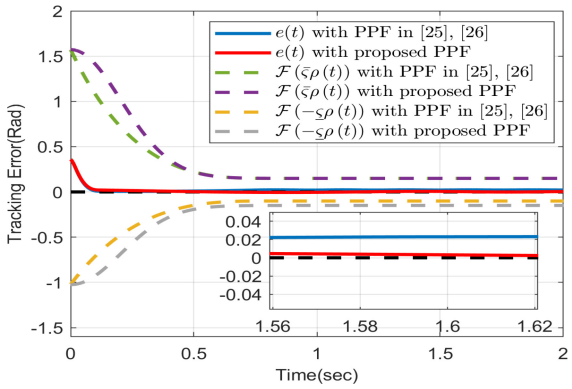


Fig. 9. Comparison of tracking errors when using global asymmetric and local asymmetric designs.

E. Asymmetric Design

To achieve controlled output overshoot, we propose two adjustable envelope curves based on the QPPF (17) as follows:

$$E_u(t) = o_u\rho(t) - \rho_u$$
$$E_l(t) = -o_l\rho(t) - \rho_l \tag{19}$$

where o_u, o_l, ρ_u, and ρ_l are the design parameters.

The new boundaries can be illustrated as follows: $E_u(t)$ serves as the upper envelope boundary, while $E_l(t)$ serves as the lower envelope boundary. Based on the previous discussions in QPPF (17), we can deduce that $o_u\rho_0 - \rho_u$ and $-o_l\rho_0 - \rho_l$ represent the initial upper and lower values for

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:12:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

prescribed bounds, respectively, and $o_u\rho_f - \rho_u$ and $-o_l\rho_f - \rho_l$ represent the steady upper and lower values for prescribed bounds, respectively.

Define the transformation function as

$$s(t) = \ln\left(\frac{\lambda(t)}{1-\lambda(t)}\right) \tag{20}$$

where $s(t)$ denotes the transformed tracking error with the normalized error $\lambda(t) = [e(t) - E_l(t)]/[E_u(t) - E_l(t)]$, and its time derivative can be deduced as

$$ds(t) = \bar{\lambda}\left[dx_1 - \dot{y}_d + \frac{E_l(t)\dot{E}_u(t) - \dot{E}_l(t)E_u(t) - e(t)\left[\dot{E}_u(t) - \dot{E}_l(t)\right]}{E_u(t) - E_l(t)}\right] \tag{21}$$

where $\bar{\lambda} = 1/\{[1-\lambda(t)][E_u(t) - E_l(t)]\lambda(t)\} > 0$, $\dot{E}_u(t) = -2\nu t[o_u(\rho_0 - \rho_f)e^{-\nu t^2}]$, and $\dot{E}_l(t) = -2\nu t[-o_l(\rho_0 - \rho_f)e^{-\nu t^2}]$.

Theorem 1: If $E_l(0) < e(0) < E_u(0)$ holds, considering a positive constant s_m such that $|s(t)| \le s_m \quad \forall t \ge 0$, the tracking error $e(t)$ will be strictly bounded within the interval $E_l(t) < e(t) < E_u(t)$.

Proof: First, the inverse transformation of (20) can be derived as

$$e^{s(t)} = \frac{\lambda(t)}{1-\lambda(t)}. \tag{22}$$

Then, from (22), we can further obtain

$$\lambda(t) = \frac{e^{s(t)}}{1+e^{s(t)}}. \tag{23}$$

Based on $|s(t)| \le s_m$, one has

$$0 < \frac{e^{-s_m}}{1+e^{-s_m}} \le \lambda(t) \le \frac{e^{s_m}}{1+e^{s_m}} < 1. \tag{24}$$

With $\lambda(t) = [e(t) - E_l(t)]/[E_u(t) - E_l(t)]$, one has

$$0 < \frac{e(t) - E_l(t)}{E_u(t) - E_l(t)} < 1. \tag{25}$$

Finally, from (25), we can obtain

$$E_l(t) < e(t) < E_u(t) \quad \forall t \ge 0. \tag{26}$$

The discussion when $e(0)$ is negative is similar to the above discussion, essentially entailing the interchange of roles between $E_u(t)$ and $E_l(t)$. ■

Remark 6: References [25], [26] also discuss the case of asymmetric constraints. In these references, a common control objective is to maintain tracking error $e(t)$ within a global asymmetric boundary $\mathcal{F}(-\underline{\varsigma}\rho(t)) < e(t) < \mathcal{F}(\bar{\varsigma}\rho(t))$. It is crucial to note that achieving this objective requires selecting different values for $\underline{\varsigma}$ and $\bar{\varsigma}$. However, while the mentioned method can achieve asymmetric constraints, such asymmetrical constraints persist throughout the system's operation. In other words, even after the system reaches a stable state, asymmetric constraints continue to exert influence, inevitably affecting the steady-state accuracy of the system. Given this observation, we endeavor to design two different envelope functions to realize local asymmetric constraints. The primary objective is to ensure that PPFs enforce asymmetric constraints during the transient phase. However, when the system reaches a stable state, it can transition into symmetric constraints. Figs. 6–9 illustrates the discussed scenario. These results are based on tracking errors of a stochastic MEMS gyroscope [20] on its driving axis. As evident from Fig. 6 ($\underline{\varsigma} = \bar{\varsigma} = 0.6$) and 7 ($\underline{\varsigma} = 0.4$ and $\bar{\varsigma} = 0.6$), although the asymmetric boundary design can enhance the output overshoot, these asymmetric constraints persist after the system is stabilized, resulting in a bias of approximately 0.02 rad in steady-state error, which is unacceptable for MEMS devices utilized in precision inertial navigation. Conversely, as depicted in Figs. 8 and 9, adopting a local asymmetric design not only improves output overshoot but also ensures that symmetric constraints are applied upon the system stabilization, thereby preserving the system's steady-state performance.

Remark 7: In this remark, we will elucidate the adjustment method for local asymmetric constraints. o_u and o_l are utilized to adjust the initial values of asymmetric boundaries. Subsequently, based on o_u, o_l, and $\rho(t)$, the steady-state values of the asymmetric boundaries can be computed. At this point, it is only necessary to adjust ρ_u, and ρ_l to transform the asymmetric boundaries into symmetric ones.

III. Main Results

A. Adaptive Neural Network Control Design

The controller design process will strictly follow the command-filtered fixed-time backstepping framework. First, define the error system as

$$z_i(t) = \begin{cases} s(t), i = 1 \\ x_i(t) - \bar{\alpha}_{i-1}(t), i = 2, 3, \ldots, n \end{cases} \tag{27}$$

where $\bar{\alpha}_{i-1}(t)$ denotes the output of the fixed-time filter (28).

To match with the fixed-time backstepping framework and avoid the issue of differential explosion, let $\alpha_{i-1}(t), i = 2, 3, \ldots, n$ pass through the following fixed-time filtering process [30]:

$$\begin{cases} \dot{\varphi}_{i1} = \varphi_{i2} - r_{i1}\Gamma^{\frac{1}{2}}\mathrm{sig}(\varphi_{i1} - \alpha_{i-1})^{\frac{1}{2}} \\ \qquad\quad - h_{i1}\Lambda^{\frac{1}{2}}\mathrm{sig}(\varphi_{i1} - \alpha_{i-1})^{\mu_{i1}} \\ \dot{\varphi}_{i2} = -r_{i2}\Gamma\mathrm{sig}(\varphi_{i2} - \dot{\varphi}_{i1}) \\ \qquad\quad - h_{i2}\Lambda\mathrm{sig}(\varphi_{i1} - \alpha_{i-1})^{\mu_{i2}} \end{cases} \tag{28}$$

where the parameters in (28) carry the same meaning as those in (5), $\bar{\alpha}_{i-1} = \varphi_{i1}$, and $\dot{\bar{\alpha}}_{i-1} = \dot{\varphi}_{i1}$.

Construct the error compensation mechanism as

$$\begin{cases} \dot{\zeta}_1 = -l_{11}\zeta_1{}^3 + \bar{\lambda}(\bar{\alpha}_1 - \alpha_1) + \bar{\lambda}\zeta_2 - \bar{\lambda}l_{12}\mathrm{sgn}(\zeta_1) \\ \dot{\zeta}_2 = -l_{21}\zeta_2{}^3 + (\bar{\alpha}_2 - \alpha_2) + \zeta_3 - \bar{\lambda}\zeta_1 - l_{22}\mathrm{sgn}(\zeta_2) \\ \dot{\zeta}_i = -l_{i1}\zeta_1{}^3 + (\bar{\alpha}_i - \alpha_i) + \zeta_{i+1} - \zeta_{i-1} - l_{i2}\mathrm{sgn}(\zeta_i) \\ \dot{\zeta}_n = -l_{n1}\zeta_n{}^3 - \zeta_{n-1} - l_{n2}\mathrm{sgn}(\zeta_n) \end{cases} \tag{29}$$

where l_{i1} and l_{i2} are the positive design parameters.

Then, the compensated error variable v_i is defined as follows:

$$v_i = z_i - \zeta_i, i = 1, 2, \ldots, n. \tag{30}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:12:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

From (10), (21) and (27)–(30), one has

$$\begin{cases} dv_1 = \big(\bar{\lambda}\,(f_1 + v_2 + \alpha_1 - \dot{y}_d + E \\ \qquad +l_{12}\mathrm{sgn}(\zeta_1)) + l_{11}{\zeta_1}^3\big)\,dt + \bar{\lambda} g_1^T dw \\ dv_2 = \big(f_2 + v_3 + \alpha_2 - \dot{\bar{\alpha}}_1 + \bar{\lambda}\zeta_1 + l_{21}{\zeta_2}^3 \\ \qquad + l_{22}\mathrm{sgn}(\zeta_2)\big)dt + g_i^T dw \\ dv_i = \big(f_i + v_{i+1} + \alpha_i - \dot{\bar{\alpha}}_{i-1} + l_{i1}{\zeta_i}^3 \\ \qquad + l_{i2}\mathrm{sgn}(\zeta_i)\big)dt + g_i^T dw \\ dv_n = \big(f_n + u - \dot{\bar{\alpha}}_{n-1} + l_{n1}{\zeta_n}^3 + \zeta_{n-1} \\ \qquad + l_{n2}\mathrm{sgn}(\zeta_n)\big)dt + g_n^T dw \end{cases} \tag{31}$$

where

$$\upsilon = \frac{E_l(t)\dot{E}_u(t) - \dot{E}_l(t)E_u(t) - e(t)\big[\dot{E}_u(t) - \dot{E}_l(t)\big]}{E_u(t) - E_l(t)}.$$

Remark 8: The single-parameter method and the minimal-learning-parameter method will be combined to reduce the adaptive computational burden in this article. The weight parameter is defined as $\Theta = \max \|\bar{\psi}_i\|^2$ for $i = 1, \ldots, n$. The estimation error is represented as $\tilde{\Theta} = \Theta - \hat{\Theta}$, where $\hat{\Theta}$ is the estimated value and Θ is the ideal value.

Step 1: The Lyapunov function is defined as

$$V_1 = \frac{1}{4}v_1^4. \tag{32}$$

Based on Definition 1, one has

$$\begin{aligned} \mathcal{L}V_1 = {} & \bar{\lambda} v_1^3 (f_1 + v_2 + \alpha_1 + \upsilon + l_{12}\mathrm{sgn}(\zeta_1) - \dot{y}_d) \\ & + l_{11}{\zeta_1}^3 v_1^3 + \frac{3}{2}\bar{\lambda}^2 v_1^2 g_1^T g_1. \end{aligned} \tag{33}$$

According to Lemma 3, one has

$$\frac{3}{2}\bar{\lambda}^2 v_1^2 g_1^T g_1 \le \frac{3}{4}\bar{\lambda}^4 v_1^4 \|g_1\|^4 + \frac{3}{4}. \tag{34}$$

Substituting (34) into (33) yields

$$\begin{aligned} \mathcal{L}V_1 \le {} & \bar{\lambda} v_1^3 (v_2 + \alpha_1 + \upsilon + l_{12}\mathrm{sgn}(\zeta_1) \\ & - \dot{y}_d + F_1(X_1)) + \frac{3}{4} \end{aligned} \tag{35}$$

where $F_1(X_1) = f_1 + 3\bar{\lambda}^3 v_1 \|g_1\|^4/4 + l_{11}{\zeta_1}^3 v_1^3/\bar{\lambda}$. Based on Lemma 4, the SNN is used to approximate $F_1(X_1)$, i.e., $F_1(X_1) = \bar{\psi}_1^T \bar{H}_1(X_1) + \varepsilon_1(X_1)$, and the approximation error of SNN satisfies $|\varepsilon_1(X_1)| \le \varepsilon_{1M}$, where $\varepsilon_{1M} \in R^+$ and $X_1 = [\bar{x}^T, v_1, \zeta_1]^T$.

According to Lemma 3 again, one has

$$\begin{aligned} \bar{\lambda} v_1^3 F_1(X_1) &\le \frac{\bar{\lambda}^2 v_1^6}{2a_{11}}\Theta \bar{H}_1^T \bar{H}_1 + \frac{3v_1^4 \bar{\lambda}^{\frac{4}{3}}}{4a_{12}} + \frac{2a_{11} + a_{12}^3 \varepsilon_{1M}^4}{4} \\ \bar{\lambda} v_1^3 l_{12}\mathrm{sgn}(\zeta_1) &\le \frac{3\bar{\lambda}^{\frac{4}{3}} v_1^4}{4a_{13}} + \frac{a_{13}^3 l_{12}^4}{4} \\ \bar{\lambda} v_1^3 v_2 &\le \frac{3\bar{\lambda}^{\frac{4}{3}} v_1^4}{4} + \frac{v_2^4}{4} \end{aligned} \tag{36}$$

where a_{11}, a_{12}, and $a_{13} \in R^+$ are the design constants.

Substituting (36) into (35) yields

$$\begin{aligned} \mathcal{L}V_1 \le \bar{\lambda} v_1^3 \Bigg(& \alpha_1 + \upsilon - \dot{y}_d + \frac{\bar{\lambda} v_1^3}{2a_{11}}\Theta \bar{H}_1^T \bar{H}_1 \\ & + \frac{3\bar{\lambda}^{\frac{1}{3}} v_1}{4} + \frac{3\bar{\lambda}^{\frac{1}{3}} v_1}{4a_{12}} + \frac{3\bar{\lambda}_1^{\frac{1}{3}} v_1}{4a_{13}} \Bigg) + \frac{v_2^4}{4} + \Xi_1 \end{aligned} \tag{37}$$

where $\Xi_1 = (3 + 2a_{11} + a_{12}^3 \varepsilon_{1M}^4 + a_{13}^3 l_{12}^4)/4$.

Design α_1 for the first subsystem as follows:

$$\begin{aligned} \alpha_1 = {} & -k_{p1} v_1^{4p-3} - k_{q1} v_1^{4q-3} - \upsilon + \dot{y}_d \\ & - \frac{\bar{\lambda} v_1^3}{2a_{11}}\hat{\Theta}\bar{H}_1^T \bar{H}_1 - \frac{\bar{\lambda}^{\frac{1}{3}} v_1}{4} - \frac{3\bar{\lambda}^{\frac{1}{3}} v_1}{4a_{12}} - \frac{3\bar{\lambda}^{\frac{1}{3}} v_1}{4a_{13}} \end{aligned} \tag{38}$$

where k_{p1} and $k_{q1} > 0$ are the designed parameters.

Substituting (38) into (37) yields

$$\mathcal{L}V_1 \le -k_{p1} v_1^{4p} - k_{q1} v_1^{4q} + \frac{\bar{\lambda}^2 v_1^6}{2a_{11}}\tilde{\Theta}\bar{H}_1^T \bar{H}_1 + \frac{v_2^4}{4} + \Xi_1. \tag{39}$$

Step i ($i = 2, \ldots, n-1$): The Lyapunov function is defined as

$$V_i = V_{i-1} + \frac{1}{4}v_i^4. \tag{40}$$

Since the design process of the virtual control laws is similar to step 1, to save space, we directly provide the control laws of α_2 and $\alpha_i (i = 3, \ldots, n-1)$

$$\begin{aligned} \alpha_2 = {} & -k_{p2} v_2^{4p-3} - k_{q2} v_2^{4q-3} + \dot{\bar{\alpha}}_1 - \bar{\lambda}\zeta_1 \\ & - \frac{v_2^3}{2a_{21}}\hat{\Theta}\bar{H}_2^T \bar{H}_2 - \frac{v_2}{4} - \frac{3v_2}{4a_{22}} - \frac{3v_2}{4a_{23}} \end{aligned} \tag{41}$$

$$\begin{aligned} \alpha_i = {} & -k_{pi} v_i^{4p-3} - k_{qi} v_i^{4q-3} + \dot{\bar{\alpha}}_{i-1} - \zeta_{i-1} \\ & - \frac{v_i^3}{2a_{i1}}\hat{\Theta}\bar{H}_i^T \bar{H}_i - \frac{v_i}{4} - \frac{3v_i}{4a_{i2}} - \frac{3v_i}{4a_{i3}} \end{aligned} \tag{42}$$

where $k_{pi}, k_{qi}, a_{i1}, a_{i2}$, and a_{i3} are the positive design constants. Furthermore, similar to Step 1, let $F_i(X_i) = f_i + 3v_i\|G_i\|^4/4 + l_{i1}{\zeta_i}^3$, then we introduce SNNs to approximate $F_i(X_i)$ i.e., $F_i(X_i) = \bar{\psi}_i^T \bar{H}_i(X_i) + \varepsilon_i(X_i)$. We assume the approximation errors of SNNs satisfy $|\varepsilon_i(X_i)| \le \varepsilon_{iM}$, where $\varepsilon_{iM} \in R^+$ and $X_i = [\bar{x}^T, v_i, \zeta_i]^T$.

From (40)–(42), one has

$$\begin{aligned} \mathcal{L}V_i \le {} & -\sum_{j=1}^{i} k_{pj} v_j^{4p} - \sum_{j=1}^{i} k_{qj} v_j^{4q} + \sum_{j=1}^{i} \Xi_j \\ & + \frac{\bar{\lambda}^2 v_1^6}{2a_{11}}\tilde{\Theta}\bar{H}_1^T \bar{H}_1 + \sum_{j=2}^{i} \frac{v_j^6}{2a_{j1}}\tilde{\Theta}\bar{H}_j^T \bar{H}_j \end{aligned} \tag{43}$$

where $\Xi_j = (3 + 2a_{j1} + a_{j2}^3 \varepsilon_{jM}^4 + a_{j3}^3 l_{j2}^4)/4$.

Step n: The Lyapunov function is defined as

$$V_n = V_{n-1} + \frac{1}{4}v_n^4. \tag{44}$$

From (31) and (44), one has

$$\begin{aligned} \mathcal{L}V_n \le {} & v_n^3 \big(u - \dot{\bar{\alpha}}_{n-1} + \zeta_{n-1} + l_{n2}\mathrm{sgn}(\zeta_n) + F_n(X_n)\big) \\ & + \sum_{j=2}^{n-1} \frac{v_i^6}{2a_{j1}}\tilde{\Theta}\bar{H}_j^T \bar{H}_j + \frac{\bar{\lambda}^2 v_1^6}{2a_{11}}\tilde{\Theta}\bar{H}_1^T \bar{H}_1 + \frac{v_n^4}{4} \\ & - \sum_{j=1}^{n-1} k_{pj} v_j^{4p} - \sum_{j=1}^{n-1} k_{qj} v_j^{4q} + \sum_{j=1}^{n-1} \Xi_j \end{aligned} \tag{45}$$

where $F_n(X_n) = 3v_n\|G_n\|^4/4 + l_{n1}{\zeta_n}^3 + f_n$ is approximated by SNN, i.e., $F_n(X_n) = \bar{\psi}_n^T \bar{H}_n(X_n) + \varepsilon_n(X_n)$. We assume the approximation error of SNN satisfies $|\varepsilon_n(X_n)| \le \varepsilon_{nM}$, where $\varepsilon_{nM} \in R^+$ and $X_n = [\bar{x}^T, v_n, \zeta_n]^T$.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:12:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Applying Lemma 3, the following inequalities hold:

$$v_n^3 F_n \le \frac{v_n^6}{2a_{n1}}\Theta\bar{H}_n^T\bar{H}_n + \frac{3v_n^4}{4a_{n2}} + \frac{2a_{41}+a_{42}^3\varepsilon_{4M}^4}{4}$$
$$v_n^3 l_{n2}\mathrm{sgn}(\zeta_n) \le \frac{3v_n^4}{4a_{n3}} + \frac{a_{n3}^3 l_{n2}^4}{4} \tag{46}$$

where a_{n1}, a_{n2}, and $a_{n3} \in R^+$ are the design constants.

Substituting (46) into (45) yields

$$\begin{aligned}\mathcal{L}V_n \le v_n^3\Big(&u - \dot{\bar{\alpha}}_{n-1} + \zeta_{n-1} + \frac{v_n^3}{2a_{n1}}\Theta\bar{H}_n^T\bar{H}_n \\ &+\frac{v_n}{4} + \frac{3v_n}{4a_{n2}} + \frac{3v_n}{4a_{n3}}\Big) \\ &+\frac{\bar{\lambda}^2 v_1^6}{2a_{11}}\tilde{\Theta}\bar{H}_1^T\bar{H}_1 + \sum_{j=2}^{n-1}\frac{v_i^6}{2a_{j1}}\tilde{\Theta}\bar{H}_j^T\bar{H}_j \\ &-\sum_{j=1}^{n-1}k_{pj}v_j^{4p} - \sum_{j=1}^{n-1}k_{qj}v_j^{4q} + \sum_{j=1}^{n}\Xi_j\end{aligned} \tag{47}$$

where $\Xi_n = (3 + 2a_{n1} + a_{n2}^3\varepsilon_{nM}^4 + a_{n3}^3 l_{n2}^4)/4$.

Design α_n for the nth subsystem as follows:

$$\begin{aligned}u = &-k_{pn}v_n^{4p-3} - k_{qn}v_n^{4q-3} + \dot{\bar{\alpha}}_{n-1} - \zeta_{n-1} \\ &-\frac{v_4}{4} - \frac{v_n^3}{2a_{n1}}\hat{\Theta}\bar{H}_n^T\bar{H}_n - \frac{3v_n}{4a_{n2}} - \frac{3v_n}{4a_{n3}}\end{aligned} \tag{48}$$

where k_{pn} and $k_{qn} > 0$ are the designed parameters.

Substituting (48) into (47) yields

$$\begin{aligned}\mathcal{L}V_n \le &-\sum_{j=1}^{n}k_{pj}v_j^{4p} - \sum_{j=1}^{n}k_{qj}v_j^{4q} + \sum_{j=1}^{n}\Xi_j \\ &+\frac{\bar{\lambda}^2 v_1^6}{2a_{11}}\tilde{\Theta}\bar{H}_1^T\bar{H}_1 + \sum_{j=2}^{n}\frac{v_i^6}{2a_{j1}}\tilde{\Theta}\bar{H}_j^T\bar{H}_j.\end{aligned} \tag{49}$$

Step $n+1$: In this step, the adaptive weight law will be deduced. First, the Lyapunov function is defined as

$$V_{n+1} = V_n + \frac{1}{2\gamma}\tilde{\Theta}^T\tilde{\Theta} \tag{50}$$

where $\gamma > 0$ is a design parameter.

From (49) and (50), one has

$$\begin{aligned}\mathcal{L}V_{n+1} \le &-\sum_{j=1}^{n}k_{pj}v_j^{4p} - \sum_{j=1}^{n}k_{qj}v_j^{4q} + \sum_{j=1}^{n}\Xi_j \\ &+\tilde{\Theta}\left(-\frac{1}{\gamma}\dot{\hat{\Theta}} + \frac{\bar{\lambda}^2 v_1^6}{2a_{11}}\bar{H}_1^T\bar{H}_1 + \sum_{j=2}^{n}\frac{v_i^6}{2a_{j1}}\bar{H}_j^T\bar{H}_j\right).\end{aligned} \tag{51}$$

Design $\dot{\hat{\Theta}}$ for the whole system as follows:

$$\begin{aligned}\dot{\hat{\Theta}} = \gamma\Bigg(&-p_{\Theta 1}\hat{\Theta} - p_{\Theta 2}\hat{\Theta}^{2q-1} \\ &+\frac{\bar{\lambda}^2 v_1^6}{2a_{11}}\bar{H}_1^T\bar{H}_1 + \sum_{j=2}^{n}\frac{v_i^6}{2a_{j1}}\bar{H}_j^T\bar{H}_j\Bigg)\end{aligned} \tag{52}$$

where $p_{\Theta 11}$ and $p_{\Theta 12} > 0$ are the design parameters.

Substituting (52) into (51) yields

$$\begin{aligned}\mathcal{L}V_{n+1} \le &-\sum_{j=1}^{n}k_{pj}v_j^{4p} - \sum_{j=1}^{n}k_{qj}v_j^{4q} + \sum_{j=1}^{n}\Xi_j \\ &+p_{\Theta 1}\tilde{\Theta}\hat{\Theta} + p_{\Theta 2}\tilde{\Theta}\hat{\Theta}^{2q-1}.\end{aligned} \tag{53}$$

Based on Lemma 3, we can obtain

$$\begin{aligned}&p_{\Theta 1}\tilde{\Theta}\hat{\Theta} \le \frac{p_{\Theta 1}}{2}\Theta^2 - \frac{p_{\Theta 1}}{2}\tilde{\Theta}^2 \\ &\left(\tilde{\Theta}^2\right)^p \le \tilde{\Theta}^2 + (1-p)p^{\frac{p}{1-p}} \\ &p_{\Theta 2}\tilde{\Theta}\hat{\Theta}^{2q-1} \le \frac{p_{\Theta 2}(2q-1)}{2q}\left(\Theta^{2q} - \tilde{\Theta}^{2q}\right).\end{aligned} \tag{54}$$

Substituting (54) into (53) yields

$$\begin{aligned}\mathcal{L}V_{n+1} \le &-\sum_{j=1}^{n}k_{pj}v_j^{4p} - \sum_{j=1}^{n}k_{qj}v_j^{4q} + \sum_{j=1}^{n}\Xi_j \\ &-\frac{p_{\Theta 1}}{2}\tilde{\Theta}^{2p} - \frac{p_{\Theta 2}(2q-1)}{2q}\tilde{\Theta}^{2q} + \frac{p_{\Theta 1}}{2}\Theta^2 \\ &+\frac{p_{\Theta 1}(1-p)p^{\frac{p}{1-p}}}{2} + \frac{p_{\Theta 2}(2q-1)}{2q}\Theta^{2q}.\end{aligned} \tag{55}$$

B. Stability Analysis

Theorem 2: For the stochastic nonlinear system (10) under Assumption 1, the virtual control signals (38), (41), (42), the actual control signal (48), and the adaptive law (52), all signals are FTBIP and the tracking error consistently remains within the prescribed performance bounds.

Proof: The Lyapunov function is defined as

$$V = \sum_{i=1}^{n}\frac{1}{4}v_i^4 + \frac{1}{2\gamma}\tilde{\Theta}^T\tilde{\Theta}. \tag{56}$$

Combining (55) and (56) yields

$$\begin{aligned}\mathcal{L}V \le &-\sum_{i=1}^{n}k_{pi}v_i^{4p} - \sum_{i=1}^{n}k_{qi}v_i^{4q} + \sum_{i=1}^{n}\Xi_i \\ &-\frac{p_{\Theta 1}}{2}\tilde{\Theta}^{2p} - \frac{p_{\Theta 2}(2q-1)}{2q}\tilde{\Theta}^{2q} + \frac{p_{\Theta 1}}{2}\Theta^2 \\ &+\frac{p_{\Theta 1}(1-p)p^{\frac{p}{1-p}}}{2} + \frac{p_{\Theta 2}(2q-1)}{2q}\Theta^{2q}.\end{aligned} \tag{57}$$

From (57), one has

$$\mathcal{L}V \le -\sigma_1\left(\sum_{i=1}^{n}v_i^{4p} + \tilde{\Theta}^{2p}\right) - \sigma_2\left(\sum_{i=1}^{n}v_i^{4q} + \tilde{\Theta}^{2q}\right) + \beta \tag{58}$$

where

$$\sigma_1 = \min\left\{4k_{pi}, p_{\Theta 1}\right\}$$
$$\sigma_2 = \min\left\{4k_{qi}, \gamma p_{\Theta 2}(2q-1)/q\right\}$$
$$\beta = \sum_{i=1}^{n}\Xi_i + \frac{p_{\Theta 1}}{2}\Theta^2 + \frac{p_{\Theta 1}(1-p)p^{\frac{p}{1-p}}}{2} + \frac{p_{\Theta 2}(2q-1)}{2q}\Theta^{2q}.$$

With appropriate constants σ_1, σ_2, and β, it can be concluded from Lemma 1 that the state v_i and the error $\tilde{\Theta}$ are FTBIP. In addition, we also need to verify the boundedness of ζ_i as we introduce the error compensation mechanism.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:12:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

The Lyapunov function is defined as

$$V_\zeta = \sum_{i=1}^{n} \frac{1}{2}\zeta_i^2. \tag{59}$$

Calculating its time derivative, one has

$$\begin{aligned} dV_\zeta =& -l_{11}{\zeta_1}^2 + \bar{\lambda}\zeta_1(\bar{\alpha}_1 - \alpha_1) + \bar{\lambda}\zeta_1\zeta_2 - \bar{\lambda}l_{12}|\zeta_1| \\ & -l_{21}{\zeta_2}^2 + \zeta_2(\bar{\alpha}_2 - \alpha_2) - \lambda_1\zeta_1\zeta_2 + \zeta_2\zeta_3 - l_{22}|\zeta_2| \\ & -l_{31}{\zeta_3}^2 + \zeta_3(\bar{\alpha}_3 - \alpha_3) - \zeta_2\zeta_3 + \zeta_3\zeta_4 - l_{32}|\zeta_3| \\ & \cdots \\ & -l_{n1}{\zeta_n}^2 - \zeta_{n-1}\zeta_n - l_{n2}|\zeta_n| \\ =& -\sum_{i=1}^{n} l_{i1}{\zeta_i}^2 + \sum_{i=1}^{n-1} \Delta_i\zeta_i(\bar{\alpha}_i - \alpha_i) - \sum_{i=1}^{n} \Delta_i l_{i2}|\zeta_i| \end{aligned} \tag{60}$$

where $\Delta_i = \bar{\lambda}, i = 1$ and $\Delta_i = 1, i = 2, \ldots, n$. Based on Lemma 2, we can obtain $\bar{\alpha}_i - \alpha_i \le \bar{\eta}_i$ within a fixed time T_{if}. Then, one has

$$\begin{aligned} \mathcal{L}V_\zeta \le& -\sum_{i=1}^{n} l_{i1}{\zeta_i}^2 - \sum_{i=1}^{n} \Delta_i(l_{i2} - \bar{\eta}_i)|\zeta_i| \\ \le& -\iota_1 V_\zeta^{\frac{1}{2}} - \iota_2 V_\zeta^2 \end{aligned} \tag{61}$$

where $\iota_1 = \sqrt{2}\min\{\Delta_i(l_{12} - \bar{\eta}_1)\}$ and $\iota_1 = (4/n)\min\{l_{i1}\}$.

From (61) and Lemma 1, we can determine that $\zeta_i, i = 1, \ldots, n$ are FIBIP with $T_\zeta = 2/\iota_1 + 1/\iota_2$. Additionally, due to $z_i = v_i + \zeta_i$, it follows that $z_i, i = 1, \ldots, n$ are also FIBIP. Furthermore, with Assumption 1 and (27), we can determine that s and $x_i, i = 2, \ldots, n$ are FIBIP as well. We also note from (20) that the boundedness of s leads to the inequality $-1 < [e(t) - E_l(t)]/[E_u(t) - E_l(t)] < 1$, which implies that $e \in \Omega_e$ is always satisfied. This guarantees that the tracking errors remain within the prescribed performance bounds. Additionally, based on Lemmas 1 and 2, we can derive that the settling time T is expressed as $T = 1/\sigma_1\eta(1-p) + 1/\sigma_2\eta(q-1) + \sum_{i=1}^{n} T_{\text{if}} + T_\zeta$. ■

IV. Simulation Results

Example: Consider a mass-spring-damper stochastic system as follows [34]:

$$\begin{cases} dx_1 = x_2 dt \\ dx_2 = \left[u - \frac{1}{m}f(x_1, x_2)\right]dt + \frac{1}{m}g(x_1, x_2)d\omega(t) \\ y = x_1 \end{cases} \tag{62}$$

where $m = 1/3$, $f(x_1, x_2) = 2x_1^2 + 0.2x_2^2\cos x_2^2$, $g(x_1, x_2) = 0.6x_1\sin(x_2) + 1.2x_2$, and $\omega(t)$ denotes an 1-D standard Brownian motion. The desired trajectory was selected as $y_d = 1.86\sin(1.5t)+0.14$. The main control parameters were selected as $k_{p1} = 10$, $k_{p2} = 60$, $k_{q1} = 50$, $k_{q2} = 30$, $l_{11} = l_{21} = 10$, $l_{12} = l_{22} = 0.001$, $a_{11} = a_{21} = 0.1$, $a_{12} = a_{22} = 0.5$, $a_{13} = a_{23} = 1$, $r_{21} = 10$, $r_{22} = 0.1$, $h_{21} = 0.5$, $h_{22} = 0.1$, $\mu_{21} = \mu_{22} = 1$, $\Gamma = \Lambda = 20$, $\gamma = 0.2$, $p_{\Theta 1} = 0.1$, $p_{\Theta 2} = 0.1$, $\nu = 1$, $\rho_0 = 1$, $\rho_f = 0.1$, $o_u = 1$, $o_l = 1$, $\rho_u = 0$, and $\rho_f = 0$. All SNNs adopt Gaussian basis functions and consist of the same 11 nodes, with centers uniformly distributed in the interval $[-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5]$, and a width of 100. Moreover, the initial conditions were selected as follows: $x_1(0) = 0.64$, $x_2(0) = 0$, $\hat{\Theta}(0) = 0$, $\varphi_{21}(0) = \varphi_{22}(0) = 0$, and $\bar{\alpha}_2(0) = \alpha_2(0) = 0$.

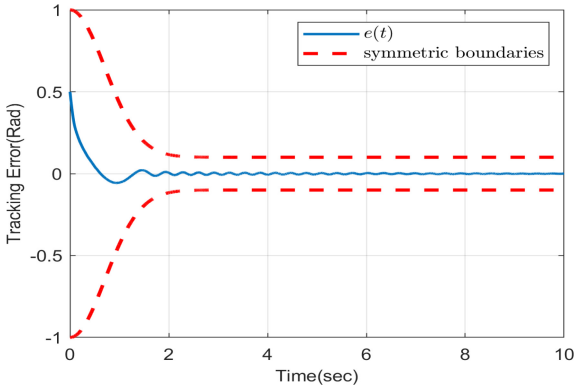


Fig. 10. Tracking error with symmetric boundaries.

Fig. 10 illustrates the prescribed performance behavior of tracking error under symmetric constraints. From Fig. 10, it is evident that the transient behavior includes a slight overshoot. To this end, by reset $o_u = 1$ and $o_l = 1$ as $o_u = 1.1$ and $o_l = 0.9$, we obtain the prescribed performance behavior of tracking error under global asymmetric constraints, which is similar to that observed in [25] and [26] (see Fig. 11). From Fig. 11, it is apparent that the overshoot has been mitigated. However, as mentioned in Remark 6, such global asymmetric constraints persist throughout the control process, meaning that after the system reaches stability, the steady-state upper bound is 0.11 rad (not 0.10 rad), and the steady-state lower bound is -0.09 rad (not -0.10 rad). This leads to an upward bias in steady-state tracking error (about 0.01 rad). To address this, we reset $\rho_u = 0$ and $\rho_f = 0$ as $\rho_u = 0.01$ and $\rho_f = 0.01$ to adjust the steady-state boundaries to symmetric constraints without affecting the initial asymmetric boundaries. The result is shown in Fig. 12, where it can be observed that the steady-state upper bound returns to 0.10 rad, and the steady-state lower bound returns to -0.10 rad. Most importantly, the bias in tracking error is resolved. Fig. 13 presents a comparison of the three scenarios, further validating our aforementioned discussion.

To demonstrate the superiority of SNNs in handling strong stochastic effects, we combine the increments of a 1-D standard Brownian motion with another process characterized by high-frequency noise (with an amplitude of 1 and a frequency spectrum following an f^{-2} distribution). Fig. 14 illustrates the variation of the adaptive law in SNNs. It can be observed that in the presence of strong stochastic effects, the adaptive laws remain bounded and do not exhibit any memory overflow phenomena similar to those observed in DNNs [1], [2], [3], [4], [5].

Fig. 15 demonstrates the control inputs of both QPPF and TPPF during the transient phase with the convergence rate set to $\nu = 1$. It can be observed that when the convergence rate is relatively low, the levels of control inputs for both methods remain essentially consistent. However, upon increasing the convergence rate to $\nu = 10$, as shown in Fig. 16, there is a noticeable increase in fluctuations within the control input generated by TPPF compared to that of QPPF. Nonetheless, following the phase of error convergence, the input levels

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:12:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

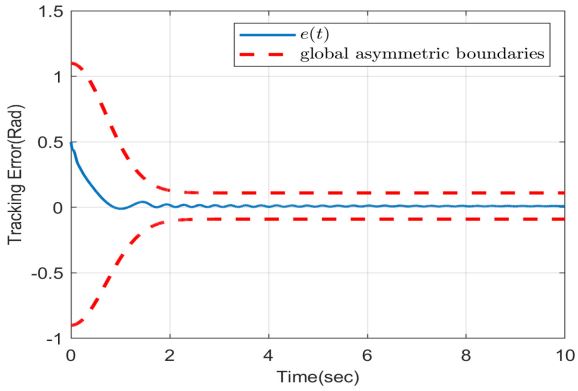


Fig. 11. Tracking error with global asymmetric boundaries.

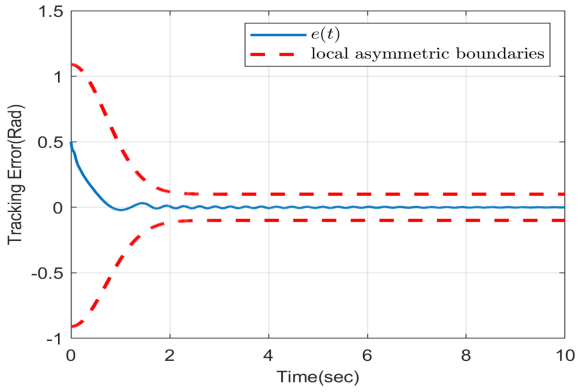


Fig. 12. Tracking error with local asymmetric boundaries.

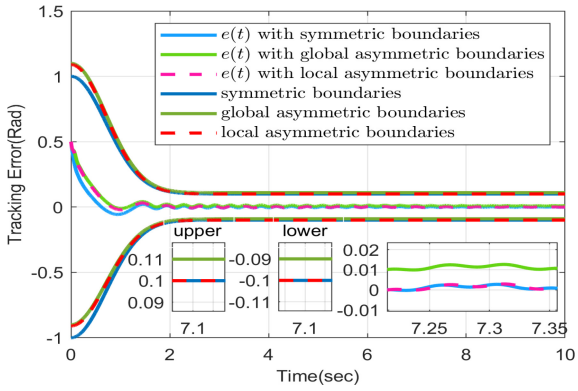


Fig. 13. Comparison of tracking errors with symmetric, global asymmetric, and local asymmetric boundaries.

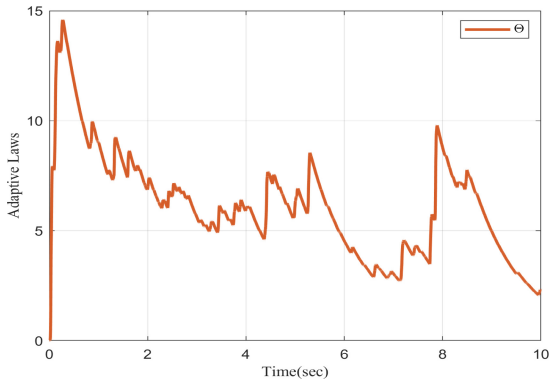


Fig. 14. Adaptive law.

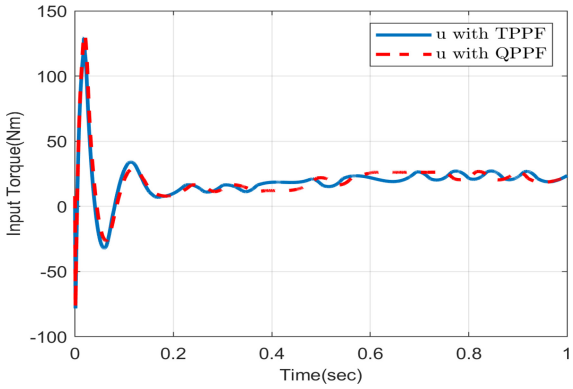


Fig. 15. Comparison of control inputs with $\nu = 1$.

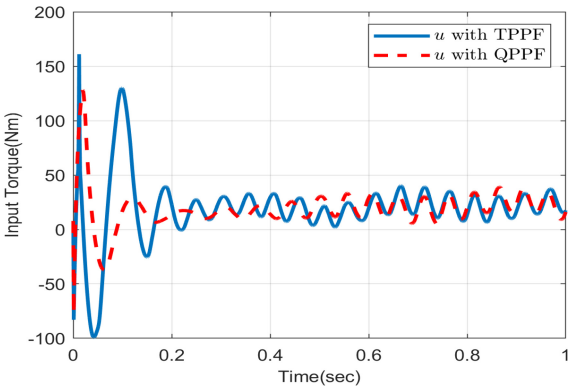


Fig. 16. Comparison of control inputs with $\nu = 10$.

for both methods stabilize and align closely, validating the observations made in Remark 4. Therefore, the QPPF-based method exhibits superior performance in suppressing transient input vibrations compared to the TPPF-based method [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26] under certain emergency conditions.

V. Conclusion

This article presents an SNN control scheme for stochastic nonlinear systems with improved prescribed performance. The SNNs are employed to approximate the stochastic nonlinear terms, mitigating the issue of memory overflow that control systems are susceptible to in stochastic environments. Additionally, the QPPF is proposed for vibration damping under fast convergence scenarios. By combining QPPF with the local asymmetric design, a smooth transition from transient asymmetric constraints to steady-state symmetric constraints is achieved, resolving the problem of steady-state tracking error bias in previous asymmetric constraint designs. Both theoretical analysis and simulation results validate the effectiveness of the proposed method.

This article has the following limitations.

1) The PPFs are designed to be strictly monotonically decreasing. However, in scenarios with highly fluctuating reference signals or periodic disturbances, these

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:12:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

strictly decreasing boundaries may be violated, potentially leading to controller failure due to insufficient input energy.

2) The study validates only SNNs as a class of low-order approximators, leaving the modeling capabilities of higher-order stochastic or probabilistic approximators in controlling stochastic nonlinear systems unexplored.

Future research will focus on two main aspects as follows.

1) Extending the PPFs from strictly monotonically decreasing to customized and self-adjusting nonmonotonic behaviors to autonomously relax boundary constraints and accommodate a wider range of application scenarios.
2) Applying probabilistic fuzzy logic systems to control stochastic nonlinear systems, and studying these in conjunction with SNNs as a foundation for high-order stochastic or probabilistic approximators.

REFERENCES

[1] J. Wu, X. Chen, Q. Zhao, J. Li, and Z.-G. Wu, "Adaptive neural dynamic surface control with prespecified tracking accuracy of uncertain stochastic nonstrict-feedback systems," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 3408–3421, May 2022.

[2] F. Jia and X. He, "Fault-tolerant control for uncertain nonstrict-feedback stochastic nonlinear systems with output constraints," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 53, no. 8, pp. 5212–5223, Aug. 2023.

[3] X. Song, P. Sun, C. K. Ahn, and S. Song, "Switching ETM-based neural adaptive output feedback control for nonaffine stochastic MIMO nonlinear systems subject to deferred constraint," *Neural Netw.*, vol. 167, pp. 668–679, Oct. 2023.

[4] J. Wu, F. He, H. Shen, S. Ding, and Z.-G. Wu, "Adaptive NN fixed-time fault-tolerant control for uncertain stochastic system with deferred output constraint via self-triggered mechanism," *IEEE Trans. Cybern.*, vol. 53, no. 9, pp. 5892–5903, Sep. 2023.

[5] F. Sedghi, M. M. Arefi, A. Abooee, and S. Yin, "Distributed adaptive-neural finite-time consensus control for stochastic nonlinear multiagent systems subject to saturated inputs," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 7704–7718, Oct. 2023.

[6] N. Wang, F. Tao, Z. Fu, and S. Song, "Adaptive fuzzy control for a class of stochastic strict feedback high-order nonlinear systems with full-state constraints," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 1, pp. 205–213, Jan. 2022.

[7] Y. Yang, B. Wen, X. Su, J. Huang, and B. Liu, "Sliding mode fuzzy control of stochastic nonlinear systems under cyber-attacks," *IEEE Trans. Cybern.*, vol. 54, no. 5, pp. 3174–3182, May 2024.

[8] Y. Xia, K. Xiao, and Z. Geng, "Event-based adaptive fuzzy control for stochastic nonlinear systems with prescribed performance," *Chaos Solitons Fractals*, vol. 180, Mar. 2024, Art. no. 114501.

[9] Y.-X. Li and S. Tong, "A bound estimation approach for adaptive fuzzy asymptotic tracking of uncertain stochastic nonlinear systems," *IEEE Trans. Cybern.*, vol. 52, no. 6, pp. 5333–5342, Jun. 2022.

[10] Y. Xia, K. Xiao, Y. Yao, Z. Geng, and Z. Lendek, "Fixed-time fuzzy vibration reduction for stochastic MEMS gyroscopes with low communication resources," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 8, pp. 4220–4233, Aug. 2024.

[11] R.-E. Precup, A.-T. Nguyen, and S. Blažič, "A survey on fuzzy control for mechatronics applications," *Int. J. Syst. Sci.*, vol. 55, no. 4, pp. 771–813, 2024.

[12] C. Turchetti, M. Conti, P. Crippa, and S. Orcioni, "On the approximation of stochastic processes by approximate identity neural networks," *IEEE Trans. Neural Netw.*, vol. 9, no. 6, pp. 1069–1085, Nov. 1998.

[13] C. Turchetti, P. Crippa, M. Pirani, and G. Biagetti, "Representation of nonlinear random transformations by non-Gaussian stochastic neural networks," *IEEE Trans. Neural Netw.*, vol. 19, no. 6, pp. 1033–1060, Jun. 2008.

[14] Z. Liu and H.-X. Li, "A probabilistic fuzzy logic system for modeling and control," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 6, pp. 848–859, Dec. 2005.

[15] H.-X. Li and Z. Liu, "A probabilistic neural-fuzzy learning system for stochastic modeling," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 4, pp. 898–908, Aug. 2008.

[16] C. P. Bechlioulis and G. A. Rovithakis, "Robust adaptive control of feedback linearizable MIMO nonlinear systems with prescribed performance," *IEEE Trans. Autom. Control*, vol. 53, no. 9, pp. 2090–2099, Oct. 2008.

[17] S. He, S.-L. Dai, Z. Zhao, T. Zou, and Y. Ma, "UDE-based distributed formation control for MSVs with collision avoidance and connectivity preservation," *IEEE Trans. Ind. Informat.*, vol. 20, no. 2, pp. 1476–1487, Feb. 2024.

[18] L. Li, K. Zhao, Z. Zhang, and Y. Song, "Dual-channel event-triggered robust adaptive control of strict-feedback system with flexible prescribed performance," *IEEE Trans. Autom. Control*, vol. 69, no. 3, pp. 1752–1759, Mar. 2024.

[19] K. Zhao and Y. Song, "Decision function-based adaptive control of uncertain systems subject to irregular output constraints," *IEEE Trans. Autom. Control*, vol. 69, no. 11, pp. 8026–8033, Nov. 2024.

[20] Y. Xia, C. Liu, Y. Tuo, and J. Li, "Command filter-based event-triggered control for stochastic MEMS gyroscopes with finite-time prescribed performance," *ISA Trans.*, vol. 148, pp. 212–223, May 2024.

[21] C. Hua, P. Ning, K. Li, and X. Guan, "Fixed-time prescribed tracking control for stochastic nonlinear systems with unknown measurement sensitivity," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 3722–3732, May 2022.

[22] X. Hu, Y.-X. Li, S. Tong, and Z. Hou, "Event-triggered adaptive fuzzy asymptotic tracking control of nonlinear pure-feedback systems with prescribed performance," *IEEE Trans. Cybern.*, vol. 53, no. 4, pp. 2380–2390, Apr. 2023.

[23] K. Zhao, Y. Song, C. L. P. Chen, and L. Chen, "Adaptive asymptotic tracking with global performance for nonlinear systems with unknown control directions," *IEEE Trans. Autom. Control*, vol. 67, no. 3, pp. 1566–1573, Mar. 2022.

[24] Y. Yao, Y. Kang, Y. Zhao, P. Li, and J. Tan, "Prescribed-time output feedback control for cyber-physical systems under output constraints and malicious attacks," *IEEE Trans. Cybern.*, vol. 54, no. 11, pp. 6518–6530, Nov. 2024.

[25] L. Li, K. Zhao, Y. Song, and F. L. Lewis, "On the uniformness of full-state error prescribed performance for strict-feedback systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 54, no. 5, pp. 3022–3031, May 2024.

[26] B. Mao, X. Wu, H. Liu, Y. Xu, and J. Lü, "Adaptive fuzzy tracking control with global prescribed-time prescribed performance for uncertain strict-feedback nonlinear systems," *IEEE Trans. Cybern.*, vol. 54, no. 9, pp. 5217–5230, Sep. 2024.

[27] S. P. Bhat and D. S. Bernstein, "Continuous finite-time stabilization of the translational and rotational double integrators," *IEEE Trans. Autom. Control*, vol. 43, no. 5, pp. 678–682, May 1998.

[28] K. Xu, H. Wang, and P. X. Liu, "Singularity-free adaptive fixed-time tracking control for MIMO nonlinear systems with dynamic uncertainties," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 71, no. 3, pp. 1356–1360, Mar. 2024.

[29] K. Xu, H. Wang, and P. X. Liu, "Adaptive fixed-time control for high-order stochastic nonlinear time-delay systems: An improved Lyapunov–Krasovskii function," *IEEE Trans. Cybern.*, vol. 54, no. 2, pp. 776–786, Feb. 2024.

[30] M. Basin, P. Yu, and Y. Shtessel, "Finite-and fixed-time differentiators utilising HOSM techniques," *IET Control Theory Appl.*, vol. 11, no. 8, pp. 1144–1152, 2017.

[31] J. L. Doob, *Stochastic Processes*. New York, NY, USA: Wiley, 1990.

[32] I. I. Gihman and A. V. Skorohod, *The Theory of Stochastic Processes*, Berlin, Germany: Springer, 1974.

[33] X. Shao, Y. Shi, and W. Zhang, "Input-and-measurement event-triggered output-feedback chattering reduction control for MEMS gyroscopes," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 9, pp. 5579–5590, Sep. 2022.

[34] Y. Xia, Z. Lendek, Z. Geng, J. Li, and J. Wang, "Adaptive fuzzy control for stochastic nonlinear systems with nonmonotonic prescribed performance and unknown control directions," *IEEE Trans. Syst., Man, Cybern., Syst.*, doi: 10.1109/TSMC.2024.3494269, to be published.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:12:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Yu Xia (Member IEEE) received the Ph.D. degree in mechanical engineering from Chongqing University, Chongqing, China, in 2024.

He is currently a Research Fellow with the State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China. His research interests include prescribed performance control, stochastic system control, and electromechanical system control.

Ke Xiao received the Ph.D. degree in mechanical engineering from Chongqing University, Chongqing, China, in 2013.

He is currently an Associate Professor with the College of Mechanical and Vehicle Engineering, Chongqing University. His research interests include special robots and electromechanical drive systems.

Jinde Cao (Fellow, IEEE) received the B.S. degree in mathematics from Anhui Normal University, Wuhu, China, in 1986, the M.S. degree in applied mathematics from Yunnan University, Kunming, China, in 1989, and the Ph.D. degree in applied mathematics from Sichuan University, Chengdu, China, in 1998.

He was a Postdoctoral Research Fellow with the Department of Automation and Computer-Aided Engineering, Chinese University of Hong Kong, Hong Kong, China, from 2001 to 2002. He is currently an Endowed Chair Professor, the Dean of the Department of Science, and the Director of the Research Center for Complex Systems and Network Sciences, Southeast University, Nanjing, China, where he is also the Director of the National Center for Applied Mathematics and the Jiangsu Provincial Key Laboratory of Networked Collective Intelligence of China. He is also an Honorable Professor with the Institute of Mathematics and Mathematical Modeling, Almaty, Kazakhstan.

Prof. Cao was a recipient of the National Innovation Award of China, the IETI Annual Scientific Award, the Obada Prize, and the Highly Cited Researcher Award in Engineering, Computer Science, and Mathematics by Clarivate Analytics. He is elected as a member of the Russian Academy of Sciences, the Academia Europaea (Academy of Europe), the Russian Academy of Engineering, the European Academy of Sciences and Arts, the Lithuanian Academy of Sciences. He is a Fellow of the African Academy of Sciences and the Pakistan Academy of Sciences.

Radu-Emil Precup (Senior Member, IEEE) received the Dipl.Ing. (Hons.) degree in automation and computers from the Traian Vuia Polytechnic Institute of Timisoara, Timisoara, Romania, in 1987, the Diploma degree in mathematics from the West University of Timisoara, Timisoara, in 1993, and the Ph.D. degree in automatic systems from the Politehnica University of Timisoara (UPT), Timisoara, in 1996.

He is currently with UPT, where he became a Professor with the Department of Automation and Applied Informatics in 2000. From 2022, he is also a Senior Researcher (CS I) and the Head of the Data Science and Engineering Laboratory, Center for Fundamental and Advanced Technical Research, Romanian Academy–Timisoara Branch, Timisoara. From 2016 to 2022, he was an Adjunct Professor with the School of Engineering, Edith Cowan University, Joondalup, WA, Australia. He is the author or co-author of more than 300 papers. His current research interests include intelligent and data-driven control systems.

Prof. Precup is a corresponding member of the Romanian Academy. He is a member of the Technical Committees on Data-Driven Control and Monitoring, and Control, Robotics and Mechatronics of the IEEE Industrial Electronics Society. He is the Editor-in-Chief of *Romanian Journal of Information Science and Technology*, the Executive Editor-in-Chief of *Information Sciences*, and an Associate Editor/Editorial Board Member of IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, *Applied Soft Computing*, *Expert Systems with Applications*, *Engineering Applications of Artificial Intelligence*, *Evolving Systems*, *Applied Artificial Intelligence*, *Healthcare Analytics*, and *Communications in Transportation Research*.

Yogendra Arya (Senior Member, IEEE) received the A.M.I.E. in electrical engineering from the Institution of Engineers, Kolkata, India, in 2008, the M.Tech. degree in electrical engineering (instrumentation and control) from Deenbandhu Chhotu Ram University of Science and Technology, Haryana, India, in 2010, and the Ph.D. degree in electrical engineering (power system) from Delhi Technological University, New Delhi, India, in 2018.

He is currently an Associate Professor with J. C. Bose University of Science and Technology, YMCA, Faridabad, India. He has published 60 SCIE journal papers. His major areas of research interests include optimization techniques, control systems, AGC/LFC in smart grid, energy storage systems, and renewable energy systems.

Dr. Arya is an Associate Editor/Editorial Board Member of *Chaos, Solitons and Fractals*, *ISA Transactions*, *Engineering Applications of Artificial Intelligence*, and *Journal of Electrical Engineering and Technology*. He has been placed among the Top 2% of Researchers in the World from 2020 to 2024 by Elsevier/Stanford University, USA. He has Google Scholar citations of 3900+ and H-index of 36. He is Fellow of IETE.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:12:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Hak-Keung Lam (Fellow, IEEE) received the B.Eng. (Hons.) and Ph.D. degrees from the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, in 1995 and 2000, respectively.

From 2000 to 2005, he was with the Department of Electronic and Information Engineering, Hong Kong Polytechnic University as Postdoctoral Fellow and a Research Fellow, respectively. He joined as a Lecturer with King's College London, London, U.K., in 2005, where he is currently a Reader. His current research interests include intelligent control, computational intelligence, and machine learning.

Dr. Lam was an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: EXPRESS BRIEFS. He is an Associate Editor of *IEEE Transactions on Fuzzy Systems*, *IET Control Theory and Applications*, *International Journal of Fuzzy Systems*, *Neurocomputing*, and *Nonlinear Dynamics*, and the Guest Editor of a number of international journals. He is on the Editorial Board of *Journal of Intelligent Learning Systems and Applications*, *Journal of Applied Mathematics*, *Mathematical Problems in Engineering*, *Modelling and Simulation in Engineering*, *Annual Review of Chaos Theory, Bifurcations and Dynamical System*, *The Open Cybernetics and Systemics Journal*, *Cogent Engineering*, and *International Journal of Sensors, Wireless Communications and Control*. He is a Co-Editor of two edited volumes: *Control of Chaotic Nonlinear Circuits* (Singapore: World Scientific, 2009) and *Computational Intelligence and Its Applications* (Singapore: World Scientific, 2012) and author/co-author of three monographs: *Stability Analysis of Fuzzy-Model-Based Control Systems* (NY, USA: Springer, 2011), *Polynomial Fuzzy Model Based Control Systems* (NY, USA: Springer, 2016), and *Analysis and Synthesis for Interval Type-2 Fuzzy-Model-Based Systems* (NY, USA: Springer, 2016). He has served as a program committee member, the international advisory board member, the invited session chair, and the publication chair for various international conferences and a reviewer for various books, international journals, and international conferences. He was named as a Highly Cited Researcher.

Leszek Rutkowski (Life Fellow, IEEE) received the M.Sc., Ph.D., and D.Sc. degrees from Wroclaw University of Technology, Wroclaw, Poland, in 1977, 1980, and 1986, respectively, and the Honoris Causa degree from the AGH University of Krakow, Krakow, Poland, in 2014.

He is a Professor with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland, and the AGH University of Krakow. He is an Honorary Professor with the Czestochowa University of Technology, Czestochowa, Poland, and also cooperates with the University of Social Sciences, Lódz, Poland. From 1987 to 1990, he held a visiting position with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK, USA. He has published more than 300 papers, including more than 50 in various series of IEEE Transactions. His research interests include machine learning, data mining, LLM, computer vision, and agent systems.

Prof. Rutkowski is a recipient of the IEEE Transactions on Neural Networks Outstanding Paper Award. He is the President and the Founder of the Polish Neural Networks Society. He organized and served as the General Chair for the International Conferences on Artificial Intelligence and Soft Computing from 1995 to 2024. He is the Editor-in-Chief of *Journal of Artificial Intelligence and Soft Computing Research*. He was elected as a Full Member (Academician) of the Polish Academy of Sciences in 2016 and a member of the Academia Europaea in 2022 and the European Academy of Sciences and Arts in 2023.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2024 at 06:12:19 UTC from IEEE Xplore. Restrictions apply.

Check for updates

View Article Online
View Journal

Sustainable Food Technology

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: C. Sharma, A. Rana, A. K. Kesharwani, D. Singh, R. Srivastava and S. N. Sharma, *Sustainable Food Technol.*, 2024, DOI: 10.1039/D4FB00182F.

This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

View Article Online
DOI: 10.1039/D4FB00182F

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainability Spotlight Statement

The study on developing nanomaterials as antimicrobial agents for crop protection aligns well with the scope of *Sustainable Food Technology*. This research focuses on sustainable agricultural practices by recycling waste cotton cloth to extract nanocrystalline cellulose (NCell) and synthesizing biocompatible titanium oxide (TiO_2) nanoparticles using Azadirachta indica leaf extract. These eco-friendly nanocomposites (CNC and GNC) demonstrated significant antimicrobial efficacy against various phytopathogens and fungi, highlighting their potential to replace toxic, non-degradable pesticides and fungicides. This innovative approach not only addresses environmental and health challenges associated with conventional agrochemicals but also contributes to the circular economy and sustainable intensification of food production.

Keywords: Waste cotton cloth, nanocrystalline cellulose, TiO_2 nanoparticles, green synthesis, nanocomposite, antimicrobial activity, sustainable agriculture, circular economy.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

Synergistic Antimicrobial Effects of Waste Cotton Cloth Extracted Cellulose with Phytofabricated TiO_2 for Potential Application in Agriculture

Chhavi Sharma[a,b], Archana Rana[a,b], Amit Kumar Kesharwani[c], Dinesh Singh*[c], Ritu Srivastava[a,b] and Shailesh Narain Sharma*[d],

[a]CSIR-National Physical Laboratory, Dr. K.S. Krishnan Marg, New Delhi-110012, India

[b]Academy of Scientific & Innovative Research (AcSIR), Ghaziabad- 201002, India

[c]Division of Plant Pathology, ICAR-Indian Agricultural Research Institute, New Delhi-110012, India

[d]Department of Applied Physics, Delhi Technological University (DTU), Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India

Corresponding authors email: shaileshnarainsharma@dtu.ac.in (Shailesh Narain Sharma)

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Abstract:

A substantial economic loss in agriculture due to plant microbial diseases has driven attention towards developing nanomaterials as antimicrobial agents for crop protection. The currently available fungicides and pesticides are highly toxic and non-degradable, causing environmental pollution and even harmful to consumers of agricultural products. However, this work attempts to develop non-toxic and biocompatible nanomaterial as a good antimicrobial agent. In this regard, a waste product, i.e., waste or used cotton cloth, has been recycled and used as a source of cellulose (natural biopolymer) extraction and for isolation of nanocrystalline cellulose (NCell) as well. Another biocompatible nanomaterial, titanium oxide (TiO_2), was synthesized using *Azadirachta indica* leaf extract to utilize the advantage of phytochemicals in the green extract. Further, the two eco-friendly nanocomposites of NCell were prepared as nanomedicines, one with commercially available chemical TiO_2 (CNC) and another with green TiO_2 (GNC). The nanocomposites (CNC and GNC), in comparison with their individual nanomaterials (NCell and TiO_2), were examined against phytopathogens: *Xanthomonas campestris* pv. *campestris* (Xcc), *Bacillus subtilis* (BS) and *Pseudomonas fluorescence* (Psfl); fungi: *Fusarium graminearium* and *Phytophthora* spp. The results illustrated the synergistic effects of NCell and TiO_2 as nanocomposites showing the stronger and longer ability to inhibit pathogen growth and, thus, proved GNC an excellent antimicrobial agent for crop protection in agriculture.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

Keywords: Waste cotton cloth, nanocrystalline cellulose, TiO_2 nanoparticles, green synthesis, nanocomposite, antimicrobial activity.

1. Introduction

Agriculture suffers significant losses globally due to microbial plant diseases caused by bacteria, fungi, and viruses. These pathogens threaten food security by reducing crop yield and quality [1]. Despite plants' natural defense mechanisms, harmful pathogens like *Pseudomonas syringae* [2]. and fungi such as *Fusarium spp.* and *Alternaria spp.* continue to cause extensive damage. These fungi often produ ce toxic mycotoxins, endangering consumers and animals [3]. The widespread use of synthetic fungicides and antimicrobials in agriculture has serious drawbacks, including toxicity, non-biodegradability, environmental pollution, and potential carcinogenic effects, necessitating the development of safer alternatives.

Nanotechnology offers innovative solutions for managing plant diseases sustainably. Metal oxide nanoparticles [4–8], including TiO_2, MgO, and ZnO, as well as graphene oxide [9], silver and gold nanoparticles [10,11], have demonstrated potent antimicrobial activity. For example, MgO nanoparticles combat soilborne pathogens [12], while TiO_2 nanoparticles inhibit microbial cells by generating reactive oxygen species (ROS). Silver nanoparticles (AgNPs) are examined as measures against the bacterium *Xanthomonas campestris* pv. *campestris* (Xcc) [13]. However, the high cost and potential environmental risks associated with silver and gold nanoparticles limit their application in agriculture. The present study addresses these challenges by developing cost-effective, eco-friendly nanocomposites using waste-derived cellulose and green-synthesized TiO_2 nanoparticles. Cellulose, a naturally abundant, biodegradable polymer, was extracted from waste cotton cloth and processed into nanocrystalline cellulose (NCC). NCC offers numerous benefits [14,15], such as a high surface area, excellent heat resistance, and functional groups (hydroxyl and carbonyl) that enhance antimicrobial properties. It has applications in various industries [16–18], including pharmaceuticals, food packaging, and textiles and also has great perceptiveness for biomedical applications like drug delivery, antibacterial wound dressing, antimicrobial films, and photobacterial nanofillers for food packaging [19–22]. The presence of a large number of chemical groups (hydroxyl or carbonyl) on its surface exhibits a high affinity towards various contaminants [23]. This property helps the cellulose-based nanomaterials used for the adsorption of charged contaminants and attack the pathogen's surface to weaken or damage its cell wall. Incorporating another biocompatible nanomaterial [24] further amplifies these

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

properties, creating nanocomposites with enhanced structural and functional capabilities to strengthen antimicrobial potency.

TiO_2 was chosen as another nanomaterial for its stability, biocompatibility, and antimicrobial potency [25–27]. Notably, its photocatalytic properties produce ROS that damage bacterial and fungal cells [28–30], making it a promising material for plant disease control [31–33]. The study employed a green synthesis method to produce TiO_2 nanoparticles using *Azadirachta indica* (neem) leaf extract [34], known for its rich phytochemical composition. Neem has been valued for centuries for its antimicrobial, anti-inflammatory, and antioxidant properties, with over 300 bioactive compounds that enhance plant defense mechanisms. The nanocomposites were prepared by combining NCC with both green-synthesized and commercially available TiO_2 nanoparticles. The materials were characterized using advanced analytical techniques, including transmission electron microscopy (TEM), X-ray diffraction (XRD), and UV-Vis spectroscopy, to assess their structural and optical properties. These nanocomposites were then evaluated for their antimicrobial efficacy against key phytopathogens and beneficial microbes.

The primary bacterial pathogen tested was *Xanthomonas campestris pv. campestris* (Xcc), responsible for black rot in Brassicaceae crops [35–37] such as cabbage, cauliflower, and broccoli. This disease causes V-shaped lesions on leaves, leading to significant yield losses. As per the reported literature[13,38–40], only silver, copper and chitosan nanoparticles have been demonstrated against Xcc. These NPs are either expensive or not eco-friendly. However, the nanomaterials explored in the present work are earth-abundant, cost-effective, non-toxic, and biocompatible. This work is a first attempt to prepare such nanomaterial with efficient antimicrobial strength against chosen antimicrobials, especially *Xanthomonas campestris pv. campestris* (Xcc). However, cellulose and TiO_2 have been explored for their antimicrobial action against different microbes but not against Xcc and other microbes explored in the present work. Moreover, only a few studies have reported antimicrobial properties of nanocomposite of cellulose and TiO_2; one such report[41] has demonstrated bacterial cellulose with CQD-TiO_2 nanocomposite action against *S. aureus* and *E. coli* bacteria only. The present work is first-hand information on the reporting of microbicidal action of Cellulose-TiO_2 nanocomposites against *Xanthomonas campestris pv. campestris* (Xcc) and other fungicides. Antifungal activity was assessed against *Fusarium graminearum* and *Phytophthora spp.*, pathogens causing severe diseases like Fusarium head blight in cereals (wheat and barley) [42,43] and water and soil-borne diseases in citrus [44,45], respectively. Both

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

fungi are highly destructive, with *F. graminearum* producing mycotoxins such as deoxynivalenol, which poses risks to human and animal health. Additionally, the compatibility of the nanocomposites with beneficial biocontrol agents, *Bacillus subtilis* and *Pseudomonas fluorescens*, was also studied. These microbes play vital roles in suppressing plant diseases and enhancing soil health. *B. subtilis* manages bacterial wilt in tomatoes caused by *Ralstonia solanacearum* [46,47], while *P. fluorescens* protects seeds and roots from fungal infections [48]. The obtained antibacterial and antifungal results of nanocomposites were analyzed and compared with each other and their individual parent component. Thus, the present work aims to utilize the synergistic effect of TiO_2 and cellulose to increase their strength to damage bacterial cell growth without affecting the growth of biocontrol agents. Hence, the potential of prepared nanocomposites has been demonstrated for plant disease management or crop protection in agriculture.

2. Materials and Methods

2.1 Materials Used

Waste cotton cloth was picked from old daily used clothes. The 10% NaOH solution and 50% hydrogen peroxide used for the extraction process were purchased from Sigma-Aldrich. Acid treatment of extracted cellulose was done using concentrated sulphuric acid (98 wt%). The solvents used for washing and other testing purposes are DI water and absolute Ethanol (99.9%). All the reactants and solvents used in the experiment were of analytical grade. The *Azadirachta indica* leaves were obtained from the ICAR-Indian Agricultural Research Institute campus. The titanium tetra isopropoxide (TTIP) with 99.99% purity was used as a precursor for the synthesis of green TiO_2, while commercially available Titanium (IV) oxide (TiO_2) with 99.7% purity was used as chemical TiO_2 (C- TiO_2). TTIP and chemical TiO_2 were obtained from Merck Chemicals Limited. The bacterial strain *Xanthomonas campestris* pv. *campestris* (Xcc) Xcc-C7, *Bacillus subtilis* DTBS-5 (BS), and *Pseudomonas fluorencens* DTPF-3 (PSFL) were obtained from Bacteriology Lab, Division of Plant Pathology, ICAR-Indian Agricultural Research Institute. The fungal pathogens *F. Graminearum* and *Phytophthora Spp* were obtained from the ITCC department of ICAR-Indian Agricultural Research Institute.

2.2 Synthesis

Cellulose extraction from cotton cloth: According to the extraction process [49], waste cotton cloth, preferably white cotton cloth (free from dyes), was washed with detergent and

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

rinsed properly to get it dried naturally in the sunlight. It was later cut into small pieces and dispersed in 10% NaOH solution at 70 °C temperature for 2 h. It was then washed with DI water and further treated with hydrogen peroxide (50%) at a temperature of 70–80 °C for 3 hrs. When the dispersion becomes off-white, then maintain the pH using NaOH and acetic acid to be 7. It was then washed with DI water repeatedly until it became a clean white solid which was further immersed in DI water with constant stirring overnight. This resulted in a stable colloid marked as Cotton cloth extracted Cellulose (CCell).

Isolation of Nanocrystalline cellulose: 0.62 mol CCell was treated by 10ml H_2SO_4, added dropwise at a temperature of 50°C for about an hour. The reaction was terminated by adding an equal amount of cold water to the heated solution. The solution was then washed with DI water thrice using centrifugation at 7000rpm. The sediment was dried at 60°C for an hour, and the obtained nanocrystalline cellulose powder was marked as NCell.

Preparation of *Azadirachta indica* leaf extract: 10 g of freshly fined chopped *Azadirachta indica* leaves were completely clean with deionized water and heated for 45 min with 50 mL of deionized water during constant stirring. The aqueous extract was filtered with Whatman filter paper no. 1 to eliminate any remaining impurities and kept at 4°C for further use.

Green synthesis of TiO_2 nanoparticles: The titanium dioxide nanoparticles were synthesized using *Azadirachta indica* leaf extract and titanium tetra isopropoxide as a precursor. According to the procedure, 4 mL of titanium tetra isopropoxide precursor was added to 50ml of green extract, and the solution was heated in a conical flask at 80°C for 3 hrs with constant stirring. The phytochemicals found in leaf extract contribute to the conversion of metal salts to metal oxide nanoparticles and are also in charge of the reaction mixture's color change of particles. The color change from green to yellowish green confirmed the formation of nanoparticles. The reaction mixture was centrifuged at 5000 rpm for 15 minutes to separate the synthesized TiO_2 nanoparticles. The obtained sediment was dried in the oven overnight at 60 ^{0}C and further calcined at 400 ^{0}C for about 3 hrs in a furnace. The dried white powder obtained was anatase form of TiO_2 nanoparticles which was stored in a glass bottle and marked as G-TiO_2. The anatase form of TiO_2 is known to be the most stable and highly non-toxic in nature with relatively low cost.

$$\underset{\text{(metal salt)}}{Ti\{OCH(CH_3)_2\}_4} \xrightarrow[\text{in DI water}]{\text{Green extract}} \underset{\text{(first product)}}{TiO_2} + 4\,(CH_3)_2CHOH$$

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Preparation of Nanocomposite: 0.2g NCell was dispersed in 40ml ethanol/DI water solution with a ratio of 1:1 and magnetically stirred. After about 15min, 0.2g of TiO_2 NPs (G-TiO_2 or C-TiO_2) were added to the suspension, and the stirring continued for another 10min. The obtained white solution was further ultrasonicated for 3 hrs followed by heating for half an hour at 70°C. Later heating was stopped, but the solution was kept on stirring overnight for homogeneous dispersion of nanoparticles. Next, the solution was centrifuged at 5000 rpm twice by ethanol and finally by DI water. The obtained sediment was dried at 50°C for an hour to get the powdered nanocomposite marked as GNC (for G- TiO_2) and CNC (for C- TiO_2).

2.3 Characterization Techniques

The synthesized nanoparticles' shape and size were analyzed by transmission electron microscope (JEM 1011 model TEM). The well-sonicated ethanol dispersion of all the prepared nanomaterials was used to prepare carbon-coated copper TEM grids for testing. The absorption profile of all the samples was analyzed by UV-Visible spectroscopy (Analytik jena SPECORD 210 PLUS Double beam UV-Visible spectrophotometer). The diffraction pattern showing the phase composition of the nanocomposites in comparison with their parent components was analyzed by using powder X-ray diffractometer (XRD) using CuK_α radiation of wavelength (λ=1.5418Å) in the range of $2\theta = 20^o$ to 80^o with a scan rate 0.02 deg/sec. Each measurement of XRD was recorded by ~50mg well dried powdered sample. The data obtained from XRD and UV-Visible spectroscopy were plotted using Origin software.

2.4 In vitro analysis of the antimicrobial property

Bacterial culture: All the bacteria strains (Xcc, BS and Psfl) were cultured in Nutrient broth (NB) (SRL, India) for 24-48 hrs at 28°C with 200 rpm in a rotatory shaker. The optical density of bacterial culture was measured using a spectrophotometer (Thermo Fisher Scientific, USA) at 600 nm wavelength and maintained at 0.5 Mcfarland unit – 1.5×10^8 cfu/ml.

Antibacterial activity: The antibacterial activity of nanomaterials (NPs and nanocomposite NC) was evaluated using the agar well diffusion method [18,50,51] against *X. campestris* pv. *campestris*, *B. subtilis* and *Pseudomonas fluorencens.* The Nutrient Agar (NA) plates were prepared by spreading 100 µl of 1.5×10^8 cfu/ml bacterial suspension evenly over solidified

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

nutrient agar. Then, a well of 5 mm was punctured in these bacterial NA plates, and 50 µl of nanomaterial to be tested was poured into the well. The prepared plates were kept at 28°C for 48 to 72 h in an incubator. The inhibition zone (IZ) was observed at 48 and 72 h of incubation. The experiment was conducted at different concentrations (2, 4 and 8 mg/ml) of nanomaterials, and the results mentioned here present the optimized concentration. Each concentration has been tested thrice, and the results were repeatedly the same with the uncertainty of ±1 mm IZ.

Antifungal activity: The poison food agar assay has been used to evaluate the nanomaterials' antifungal activity against mycopathogens – *F. Graminarium* and *Phytopthora* sp. The potato dextrose agar (PDA) medium plates were prepared and supplemented with different concentrations (4, 2, 1, 0.5, 0.25 and 0.125 mg/ml) of the samples. The small plug of 3-day mycelial growth of *F. Graminarium* and *Phytopthora* sp. was placed on the prepared PDA plates and incubated at 28°C for 5 days. The results were observed after 5 days of incubation.

3. Results and Discussions

3.1 X-ray diffraction pattern analysis

The diffraction peaks recorded by XRD spectra (shown in Fig.1) represents the crystalline phases of prepared nanocomposites (CNC and GNC) as compared with their individual components (NCell and TiO_2), reflecting their joint contributions in composites. The characteristic peaks of TiO_2 (both chemical and green synthesized) are indexed according to the JCPDS card No. 21-1272, which reflects the pure anatase phase [52]. On the other hand, the XRD spectra of nanocrystalline cellulose featured three characteristic peaks with $2\theta = 16.4^{o}$, 22.5^{o}, 34.5^{o} corresponding to (110), (200) and (004) planes, respectively [53–55]. However, the characteristic peak of NCell corresponds to the (200) plane, and XRD peaks of TiO_2 correspond to (101), (004), (200), (105), (211), (204) planes simultaneously featured in XRD spectrum of nanocomposites (CNC and GNC). This confirms the successful formation of nanocomposites featuring the contribution of both NCell and TiO_2.

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

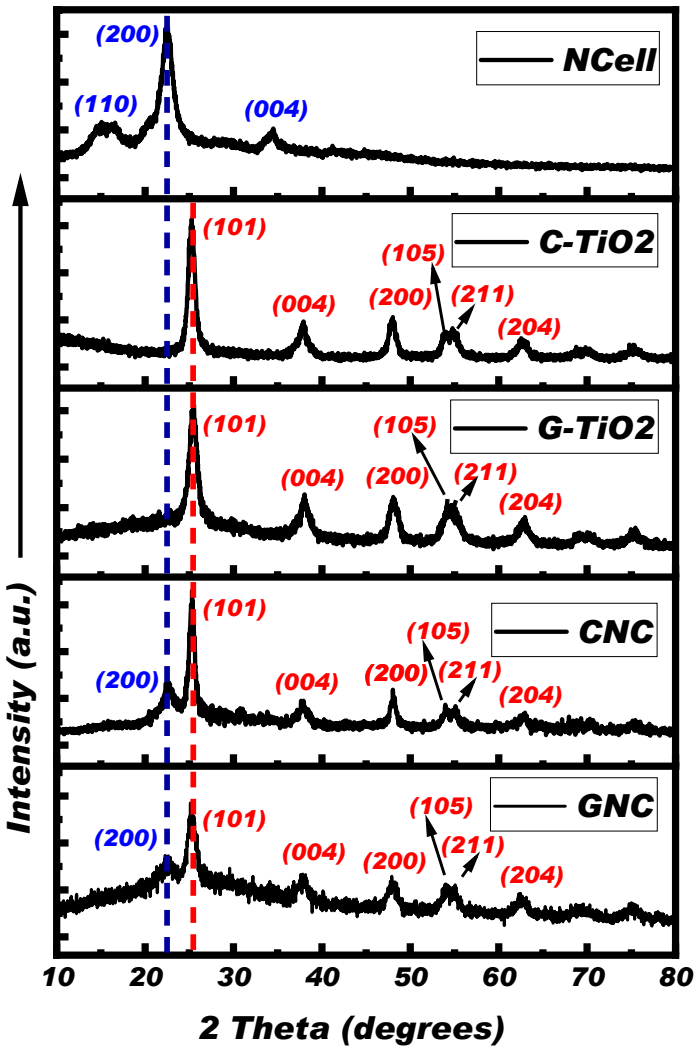


Fig. 1 X-ray diffraction pattern of nanocrystalline cellulose (NCell), C-TiO_2 (chemical TiO_2), G-TiO_2 (green synthesized TiO_2), CNC (NCell:C-TiO_2 nanocomposite) and GNC (NCell:G-TiO_2 nanocomposite)

The X-ray diffraction pattern has also been used to determine the crystallite size in a solid by using Debye Scherrer equation [56,57].

$D = K\lambda / \beta\cos\theta$

where, D = Crystalline size of the nanoparticle

K = Scherrer constant equals to 0.98

λ = Wavelength of X-ray (1.54 Å)

β = Full width at half maxima

The calculated value of crystallite size of nanomaterials has been recorded in the table1.

Sample	β	2θ	D (nm)
NCell	0.3	22.42	28.18
C-TiO_2	0.38	25.23	21.25

Sustainable Food Technology Accepted Manuscript

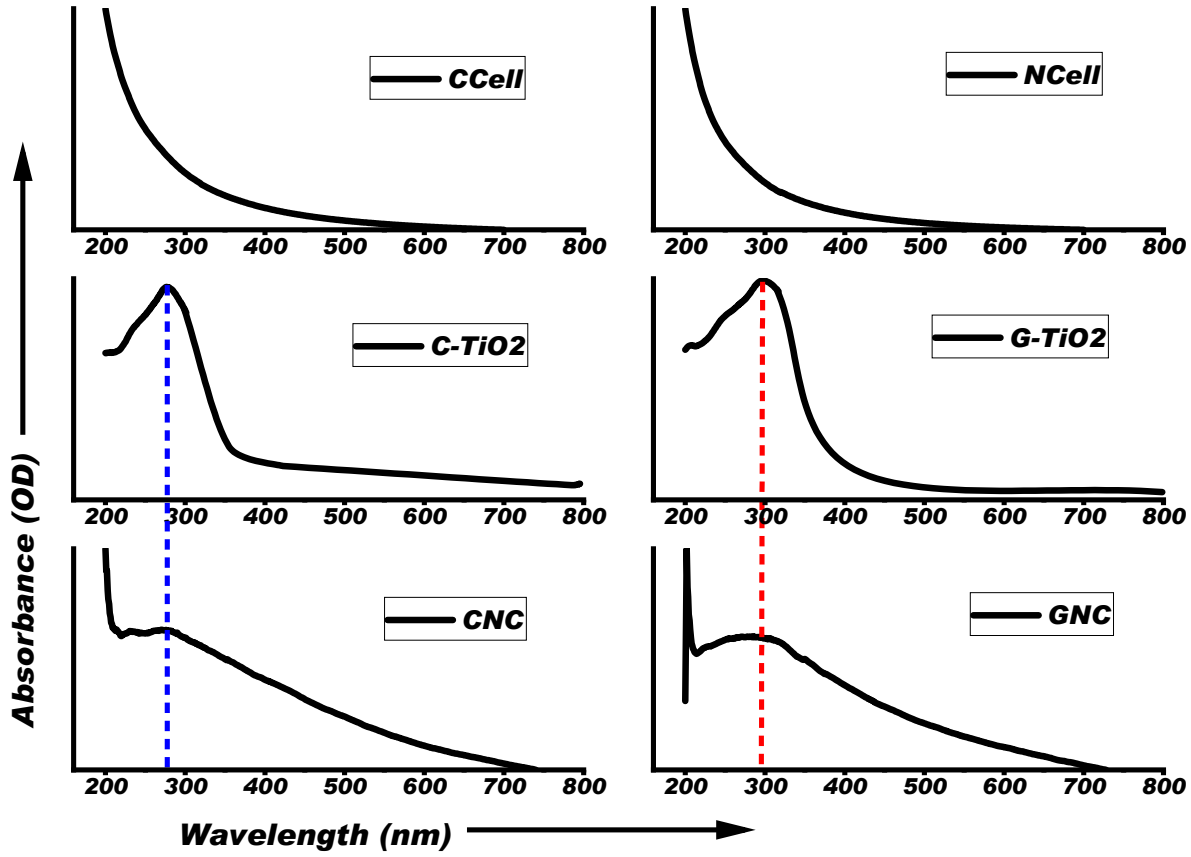
View Article Online
DOI: 10.1039/D4FB00182F

G-TiO_2	0.4	25.28	22.37
CNC	0.52	25.32	16.35
GNC	0.72	25.44	11.81

Table 1: The crystallite size of all the samples calculated by using the Debye Scherrer equation.

3.2 Optical characteristics

The absorption spectra shown in Fig. 2 have been obtained using UV-Visible spectroscopy for optical analysis. It can be clearly seen that cellulose and TiO_2 both exhibit absorption in the ultraviolet region. Considering the slope of the absorption band corresponds to the absorption edge of a material, the NCell absorption is slightly blue-shifted compared to CCell. Similarly, the C- TiO_2 absorption peak is blue-shifted than that of G- TiO_2. This blue shift in the absorption profile suggests particle size reduction. However, the absorption spectra of the nanocomposite showed the characteristic features of both NCell and TiO_2. This suggests the contribution of both individual parent components to the optical properties of nanocomposites, similar to the inference obtained from XRD spectra.



Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

Fig. 2 Absorption profile of cotton extracted cellulose (CCell), nanocrystalline cellulose (NCell), C-TiO_2 (chemical TiO_2), G-TiO_2 (green synthesized TiO_2), CNC (NCell: C-TiO_2 nanocomposite) and GNC (NCell: G-TiO_2 nanocomposite) obtained from UV-Visible spectroscopy

3.3 Structural and Morphological Analysis

The particle shape and size of prepared materials were analysed by TEM images shown in Fig. 3. image (a) depicts the cluster of long fibers of cellulose extracted from waste cotton cloth while nanocrystalline cellulose (NCell shown in image b) isolated from extracted microcrystalline cellulose shows spherical nanoparticles of size ~ 35-50 nm. Image (c) and (d) exhibits commercial TiO_2 of particle size ~ 20nm and green synthesized TiO_2 with particle size ~ 20-25 nm, respectively. Thus, the particle size of C-TiO_2 is found to be lesser as compared to G-TiO_2, commensurate well with UV-Visible absorption results. However, image (e) and (f) belongs to the nanocomposites of NCell with commercial chemical TiO_2 (CNC), and green synthesized TiO_2 (GNC), respectively. The TEM image of CNC depicts spherical nanoparticles of size ~ 16-20 nm, while that of GNC depicts ~ 7-12 nm spherical nanoparticles. Moreover, particle agglomeration is noted in the CNC image, while the GNC image depicts evenly distributed nanoparticles. Hence, TEM results suggest that the nanocrystalline cellulose nanocomposite with green synthesized TiO2 exhibits better morphology than the nanocomposite of chemical TiO_2, as GNC shows good particle distribution with lesser or no agglomeration. This indicates that the phytochemicals-rich green extract in TiO_2 material played an important role in the preparation of nanocomposite. Moreover, the results obtained from TEM commensurate well with the crystallite size calculated using X-ray diffraction pattern.

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

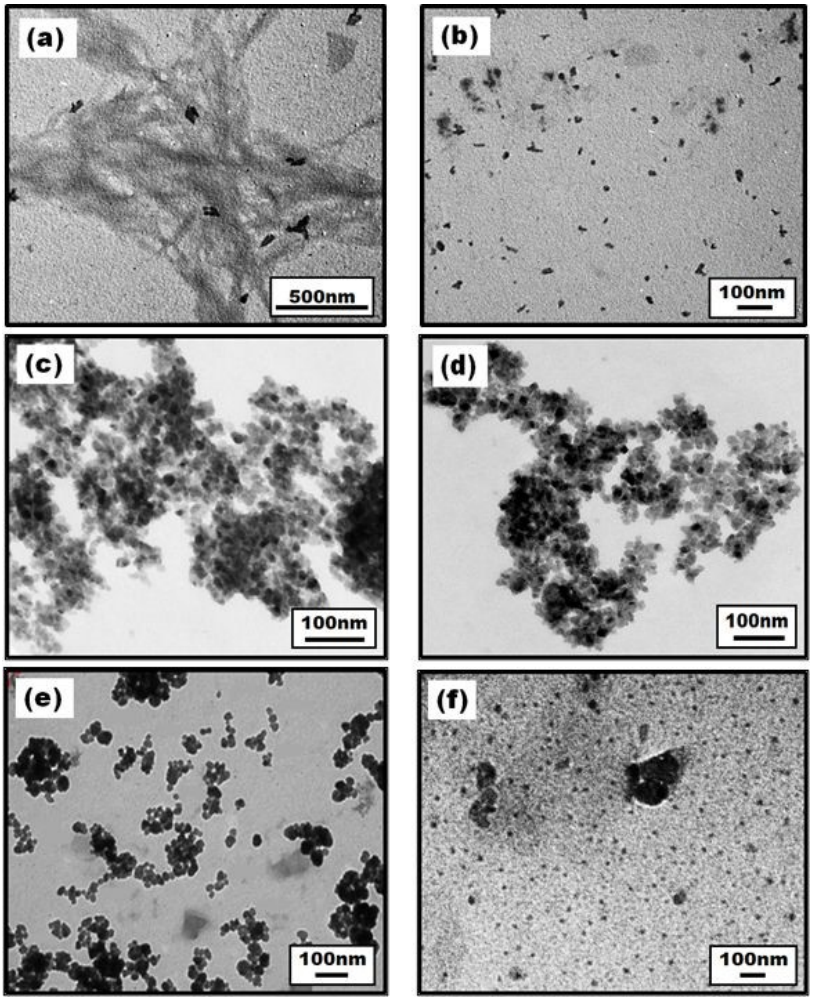


Fig. 3 Images recorded by transmission electron microscope (TEM) for (a) cotton-extracted cellulose, (b) nanocrystalline cellulose, (c) chemical TiO_2, (d) green synthesized TiO_2, (e) CNC (NCell: C-TiO_2 nanocomposite) and (f) GNC (NCell: G-TiO_2 nanocomposite)

Further, the particle size distribution has been analyzed by using Dynamic Light Scattering. Although DLS is too sensitive toward the presence of even small aggregates as I (scattering intensity) α R^6 (particle radius), it is used to determine the size distribution profile of particles in a suspension, measuring the hydrodynamic size of particles. Thus, it provide the nanoparticle size much larger than that analyzed by TEM results. Therefore, DLS data owe to hydrodynamic diameter based on ligands, ionic species, capping agents, solvents, etc. However, it can be appropriately utilized to monitor the PDI (Polydispersity index) value to determine its monodispersity/ polydispersity nature. The PDI value represents the particle size distribution, which should ideally be less than 1 to show monodispersity. Table 2. Represents the particle hydrodynamic size of all the samples along with their PDI values. It has been found that PDI value of all the samples is less than 1 owing to their monodispersive nature. Fig. 4. describes the particle size distribution of all the synthesized nanomaterials, while the

View Article Online
DOI: 10.1039/D4FB00182F

inset figure shows the plot of the DLS correlation coefficient versus time. The particle size distribution profile shows broad peak of chemical TiO_2 while the sharpest peak of phytofabricated nanocomposite i.e., GNC. Similarly, the correlation decay of GNC is steeper than all other samples, and its PDI value (0.2) is also lesser comparatively, which indicates that it is more stable or monodispersive than other nanomaterials. Thus, the DLS analysis also makes evident that the phytofabricated nanocomposite is the most stable and monodispersive nanomaterial.

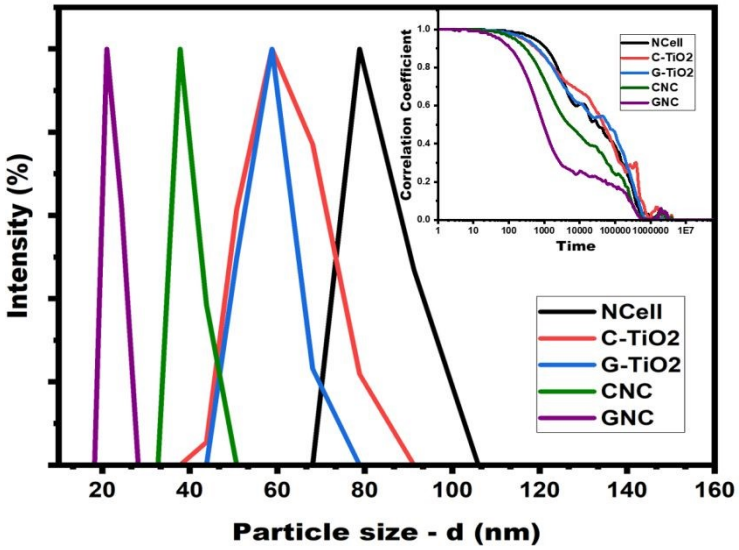


Fig. 4 DLS spectra showing the particle size distribution of (a) nanocrystalline cellulose (NCell), (b) chemical TiO_2 (C-TiO_2), (c) green synthesized TiO_2 (C-TiO_2), (d) CNC (NCell: C-TiO_2 nanocomposite) and (e) GNC (NCell: G-TiO_2 nanocomposite); inset showing correlation coefficient versus time plot of all the samples.

Sample	Hydrodynamic size (nm)	PDI value
NCell	78.8	0.584
C-TiO_2	58.8	0.9
G-TiO_2	58.8	0.4
CNC	37.8	0.327
GNC	21	0.2

Table 2: The hydrodynamic size of particles of all the samples, along with their PDI value obtained by DLS analysis.

3.4 Antibacterial study of synthesized nanomaterials:

The antibacterial properties of nanocomposites (CNC and GNC) in comparison with their parent components (NCell, chemical TiO_2 and Green TiO_2) were analyzed by agar well diffusion method at different concentrations (2, 4 and 8 mg/ml). Initially, the experiment was

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

conducted at 8mg/ml concentration, but all the nanomaterials showed no activity at this concentration. It might be the case that the density of nanoparticles is too high, providing higher chances for agglomeration, due to which the nanoparticles could no longer maintain smaller particle size. Since particle size is one of the major factors responsible for the defensive mechanism against pathogens, small particles comparatively have a high ability to rupture and penetrate the bacterial cell wall [58–60]. Keeping this in mind, the concentration of the nanomaterials has been reduced to 4mg/ml and tested against phytopathogen *X. campestris* pv. *campestris,* and the results were observed till 72 hrs, as shown in Fig. 5. Further, the concentration of all the nanomaterials was reduced to 2mg/ml and expected to provide better results but found no reactivity of nanomaterials against phytopathogen. This suggests that the density of NPs at lower concentrations (2mg/ml) is very low that the collective strength of NPs is not enough to be defensive against the pathogen. Thus, 4mg/ml concentration is found to be optimized reactive concentration against pathogen Xcc, so the results obtained from this concentration are presented here to be analyzed.

The obtained results manifest that the chemical TiO_2 (C-TiO_2) nanoparticles have shown an inhibition zone (IZ) of 10 mm till 24 hrs, and beyond that, bacterial growth was observed inside IZ. Thus, Fig. 5 depicts no activity of C-TiO_2 NPs against Xcc, while green synthesized TiO_2 depicts a 20 mm clear inhibition zone (IZ) upto 48 hrs. This suggests that the phytochemicals present in neem extract played a role in the bacterium. However, the G-TiO_2 NPs could no longer resist the pathogen attack, and bacterial cells could even cover the inhibition zone after 48 hrs. Similarly, the nanocrystalline cellulose (NCell) extracted from cotton cloth also showed 18 mm IZ till 48 hrs and depicted bacteriostatic activity (NPs still resist bacterial growth with 1 mm clear IZ) against Xcc after 72 hrs. However, the nanocomposites (CNC and GNC) showed promising results even after 72 hrs. The nanocomposite (CNC) of NCell and chemically synthesized TiO_2 manifests 30 mm IZ upto 48 hrs and even beyond that showed 24 mm IZ upto 72 hrs. Whereas the nanocomposite (GNC) of NCell and green synthesized, TiO_2 demonstrated 44 mm IZ even after 72 hrs. It significantly involves the action of phytochemicals associated with neem (*Azadirachta indica)*. Firstly, the neem extract act as a capping agent responsible for particle size reduction. Secondly, the phytochemicals present in neem also played defensive roles against pathogens, either by weakening the bacterial cell wall or slower cell growth. The reactivity of GNC against Xcc up to 72 hrs evident the synergistic effect of NCell and G-TiO_2 NPs as well as the strong defensive role of phytochemicals-rich Neem extract against phytopathogen.

View Article Online
DOI: 10.1039/D4FB00182F

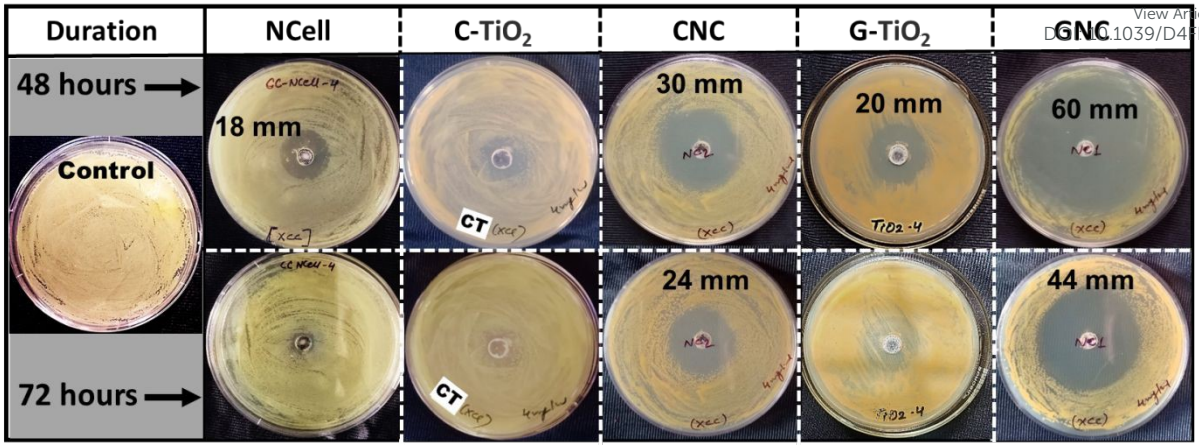


Fig. 5 Antibacterial activity of synthesized nanomaterials at 4mg/ml concentration against phytopathogen *X. campestris* pv. *campestris*

It is necessary to investigate that the nanomaterial showing activity against phytopathogen should not harm plant health. Thus, all the synthesized nanomaterials tested against pathogens *Bacillus subtilis* (BS) and *Pseudomonas fluorescens* (PSFL), which are biocontrol agents maintaining plant health. The results are shown in Fig. 6a. manifests the activity of nanomaterials against the biocontrol *B. subtilis* (BS). Although, green synthesized TiO_2 and its corresponding nanocomposite (GNC) inhibited the bacterial growth till 48 hrs with 8 mm and 20 mm IZ, respectively. In spite of that, the nanomaterials (G-TiO_2 and GNC) could no longer resist bacterial growth beyond 48 hrs and became ineffective against pathogen BS. On the other hand, Fig. 6b. also depicts similar results that *P. fluorescens* growth inhibited by green synthesized TiO_2 (GNC) with 6 mm IZ till 48 hrs, but these NPs could not be effective for another 24 hrs (no IZ after 72 hrs). This suggests that the nanomaterials which showed activity against phytopathogens Xcc are not harmful against biocontrol agents BS and PSFL. The measurement of the inhibition zone observed from antibacterial testing of all the nanomaterials against phytopathogens is noted in table 3, which indicates the potential of prepared nanocomposites.

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

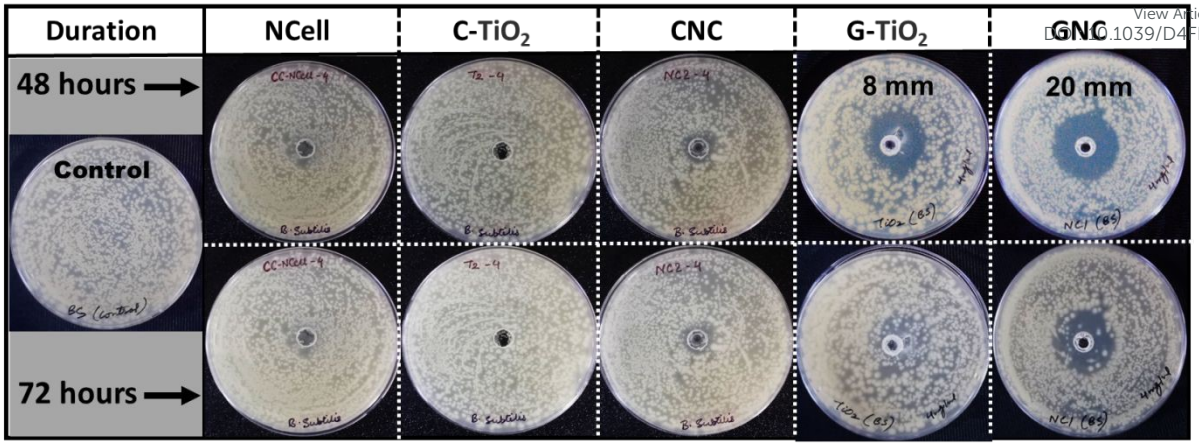


Fig. 6a Agar well diffusion assay testing nanomaterials at 4mg/ml concentration against biocontrol agent *Bacillus subtilis* (BS)

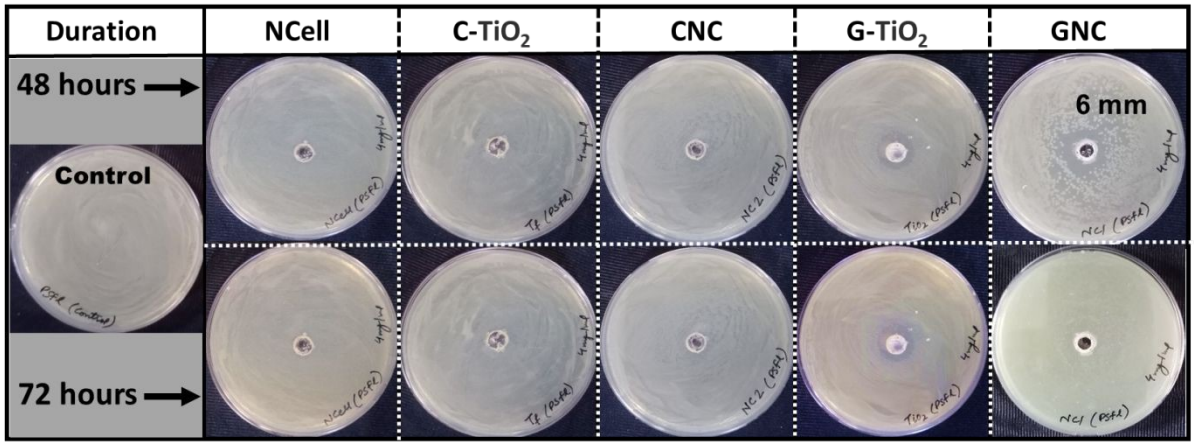


Fig. 6b Agar well diffusion assay testing nanomaterials at 4mg/ml concentration against biocontrol agent *Pseudomonas fluorescens* (PSFL)

Materials	Zone of inhibition (mm)					
	Xcc		BS		PSFL	
	48h	72h	48h	72h	48h	72h
NCell	18	NI	NI	NI	NI	NI
C-TiO_2	NI	NI	NI	NI	NI	NI
G-TiO_2	20	NI	8	NI	NI	NI
CNC	30	24	NI	NI	NI	NI
GNC	60	44	20	NI	6	NI

Table 3: Measurement of inhibition zone observed from antibacterial testing of nanomaterials against phytopathogens. NI represents No Inhibition

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

3.5 The defense mechanism of nanomaterials against pathogens:

The obtained antibacterial results were analyzed based on the observed inhibition zone. Fig. 7 attempts to explain the scenario inside the inhibition zone, or the defence mechanism of a material against pathogens, using TEM images of untreated Xcc and treated Xcc by G-TiO_2 nanomaterial as an example. In this respect, a very small area inside the inhibition zone of the G-TiO_2 plate (72 hrs) was swept out, dispersed in autoclaved DI water, and further used to prepare the TEM grid for testing. It is seen from Fig. 7 that the untreated bacterium is a rod-shaped microorganism with a fine thick cell wall and thread-like flagella. On the other hand, there are two types of defensive action during a pathogen treatment by a nanomaterial: *bactericidal* and *bacteriostatic*. In this example case, G-TiO_2 showed bactericidal action until 48 hrs showing a clear inhibition zone. This means the nanoparticles could successfully rupture the bacterial cell wall and penetrate the cell to kill it. However, after 48 hrs, the strength of nanoparticles decreases with further bacterial growth. Although at this stage (after 48 hrs), G-TiO_2 NPs could not penetrate the bacterial cell but are still effective enough to weaken or inactivate the bacterium making it unconscious. It basically suppresses the growth of bacterial cells, keeping them in a stationary phase of growth known as a bacteriostatic condition. Thus, the TEM images recorded for the sample inside the inhibition zone of 72 hrs condition of the G-TiO_2 sample plate show dead cells with ruptured cell walls and deactivated cells, also depicting bacteriostatic condition.

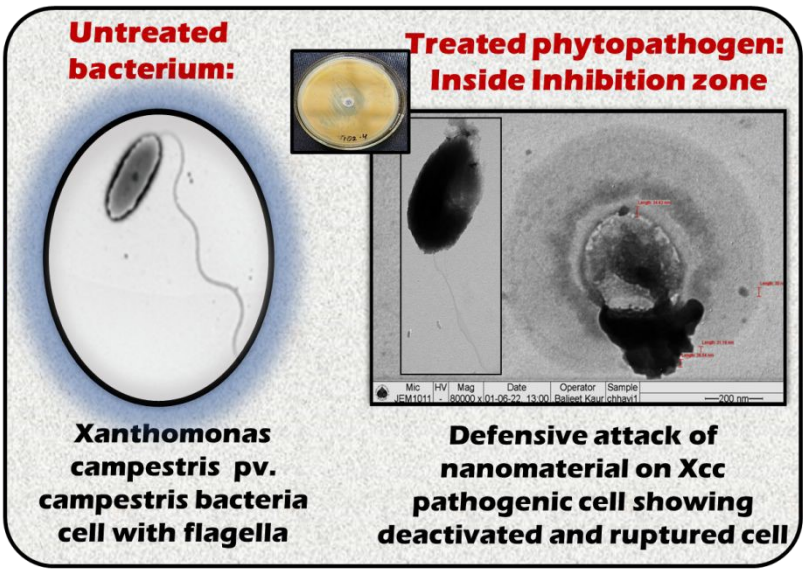


Fig. 7. Schematic of TEM images depicting the action on green synthesized TiO_2 on phytopathogen *X. campestris* pv. *campestris* (Xcc)

3.6 Antifungal studies of synthesized nanomaterials:

View Article Online
DOI: 10.1039/D4FB00182F

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

The antifungal activity of prepared nanocomposites in comparison with their parent components was analyzed using poison food agar assay against pathogens *Fusarium Graminearum* and *Phytophthora spp.,* and the results are presented in Fig. 8. Since the antibacterial studies revealed 4mg/ml to be the optimized concentration, this concentration is considered a reference for antifungal studies. However, it was found that all the nanomaterials at 4, 2, 1, and 0.5 mg/ml concentration exhibited the complete inhibition of mycopathogens *Fusarium G.* and *Phytophthora spp.* even after 7 days, the control plate showed full fungus growth within 5 days. Thus, the nanomaterial concentration was further lowered to find the minimum optimized concentration effective against fungus. Fig. 8(a). shows the antifungal results of nanomaterials at 0.25 and 0.125mg/ml concentration against *Fusarium Graminearum (FG).* It has been found that the 0.25mg/ml concentration is strongly effective against FG as all the nanomaterials completely inhibit the fungal growth. However, the results at 0.125mg/ml concentration are quite different, but the nanomaterials are still effective against FG as they showed restricted growth compared with the control plate. Moreover, the PDA plate consisting of G-TiO_2 showed no fungal growth, while only the plug of 3-day mycelial growth placed in the PDA plate of GNC is covered with fungus FG and could not grow beyond due to the defensive action of GNC NPs. Thus, this suggests that the green synthesized TiO_2 and its corresponding nanocomposite completely inhibit the fungal growth even at 0.125mg/ml concentration and thus proved their potential for preventing FHB diseases caused by *F. graminearum* in wheat and barley.

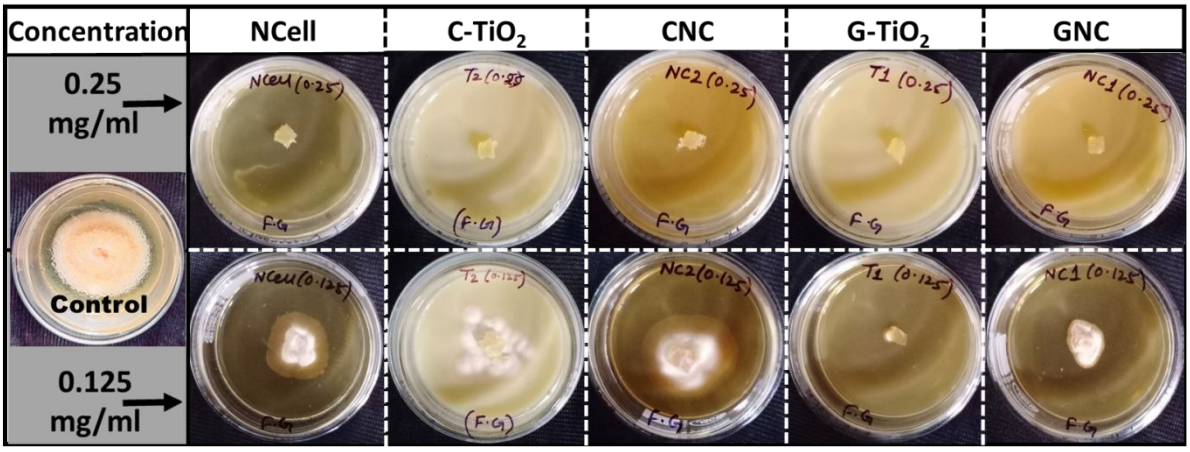


Fig. 8a. Poison food agar assay testing all the nanomaterials at 0.25 and 0.125 mg/ml concentration against *F. graminearum* under *in vitro* conditions

However, all the nanomaterials exhibited strong inhibition of fungal growth of *Phytophthora Spp* even at 0.125mg/ml concentration, as shown in Fig. 8(b). It can be seen that the control

View Article Online
DOI: 10.1039/D4FB00182F

plate showed full fungal growth within 5 days while all the nanomaterials exhibited complete inhibition and thus strongly effective against *Phytophthora Spp.* This suggests that all the prepared nanomaterials have the ability to control water and soil-borne diseases caused by *Phytophthora spp.* in citrus.

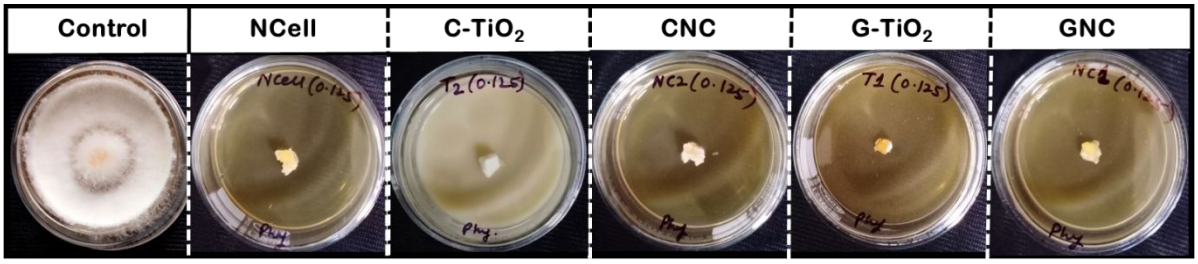


Fig. 8b. Poison food agar assay testing all the nanomaterials at 0.125 mg/ml concentration against *Phytophthora spp.* under *in vitro* conditions

Thus, the obtained antimicrobial results manifest that the synergistic effects of nanocrystalline cellulose and TiO_2 are stronger than the individual nanomaterial. Moreover, the *Azadirachta indica* leaf extract contributed well to the defense mechanism against pathogens. The phytochemicals in green extract acted as a reducing agent during synthesis and as a capping agent to stop further growth and maintain smaller particle sizes. So, the phytochemicals help to weaken the bacterial cell wall so that smaller-sized particles [61,62] easily penetrate the cytoplasm by damaging the cell wall and rupturing the membrane, causing vacuole formation due to increased NPs concentration in the cell, inhibiting cell growth, and finally open the death pathways for pathogen either by deactivating or kill the bacteria. Hence, the nanocomposite of cotton cloth extracted nanocrystalline cellulose with green synthesized TiO_2 has proved its potential to be a good antimicrobial agent for crop protection in agriculture.

4. Conclusions

This work presents two different non-toxic, biocompatible compounds to prepare environmental-friendly nanocomposite. A waste product (waste cotton cloth) has been recycled successfully used to prepare cellulose nanoparticles and tested against pathogens. Another biocompatible compound TiO_2 has been green synthesized by using *Azadirachta indica* leaf extract to utilize the effects of phytochemicals against pathogens. Moreover, a commercially available TiO_2 also has been used for comparison. So, two nanocomposites of nanocrystalline cellulose (NCell) have been prepared, one with chemical TiO_2 and another

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

with green TiO_2. All the nanomaterials (NCell, C-TiO_2, G-TiO_2, CNC and GNC) were analysed by XRD, UV-Visible spectroscopy and TEM. The results obtained from XRD and UV-Visible spectroscopy evident the contribution of NCell and TiO_2 in nanocomposites. The TEM results demonstrate the particle size of nanomaterials and prove GNC particle size is smaller than CNC. The prepared nanocomposites, in comparison with individual parent components, have been examined against phytopathogens: *Xanthomonas campestris* pv. *campestris* and two biocontrol agents - *Bacillus subtilis* and *Pseudomonas fluorescence;* fungal pathogens: *Fusarium graminearium* and *Phytophthora* spp. It is evident from antibacterial results against Xcc that only nanocomposites CNC and GNC showed bactericidal action even after 72 hrs with an inhibition zone of 24 and 44 mm, respectively. However, comparing CNC and GNC with each other, the phytochemicals present in the green extract are responsible for stronger and longer defensive actions of GNC nanocomposite against Xcc compared to CNC. Similarly, the nanocomposites have proved their efficiency against fungi *F. graminearum* and *Phytophthora Spp.* Moreover, all the prepared nanomaterials have not affected the growth of *B. subtilis* and *P. fluorescence,* which are biocontrol agents helping to maintain plant health. Thus, the results evidently demonstrated that the nanocomposites exhibited enhanced antimicrobial activity compared to their individual components, with the green-synthesized TiO_2 showing superior performance. This enhanced efficacy is attributed to the synergistic effects of TiO_2 and cellulose, which disrupt bacterial and fungal cell walls more effectively. The nanocomposites also displayed excellent biocompatibility, ensuring they do not harm beneficial microbes. The green synthesis method further reduced particle agglomeration, improved stability, and minimized environmental impact, aligning with sustainability principles. This approach highlights the potential of waste-derived nanomaterials in sustainable agriculture. Using cellulose extracted from waste cotton cloth not only reduces waste but also creates value-added products for agricultural applications. Similarly, green synthesis using neem extract leverages natural resources to produce environmentally friendly nanoparticles without hazardous chemicals. Thus, the present study demonstrates that nanocomposites of cellulose and TiO_2, particularly those synthesized using green methods, offer a promising solution for managing plant diseases. These materials provide effective antimicrobial action against harmful pathogens while preserving beneficial biocontrol agents. Their cost-effectiveness, environmental safety, and scalability make them suitable for widespread agricultural use. By integrating advanced

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

nanotechnology with sustainable practices, this work contributes to improve crop protection and ensuring global food security.

Authors contributions

Chhavi Sharma: Conceptualization, Original draft, Methodology, Investigation, data curation, and experimentation.

Archana Rana, Amit Kumar Kesharwani: Resource and Software

Dinesh Singh: Supervision for antimicrobial testings and analysis

Ritu Srivastava: Co-supervision & editing

Shailesh Narain Sharma: Supervision; Validation; Visualization; review & editing

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.

Acknowledgments

The authors also express gratitude to Director CSIR-NPL for permitting them to utilize facilities to execute the research work. C. Sharma sincerely acknowledges the Academy of Scientific and Innovative Research (AcSIR) to pursue Ph.D. in CSIR-NPL.

References:

1 S. Savary, L. Willocquet, S. J. Pethybridge, P. Esker, N. McRoberts and A. Nelson, *Nat. Ecol. Evol.*, 2019, **3**, 430–439.
2 X.-F. Xin, B. Kvitko and S. Y. He, *Nat. Rev. Microbiol.*, 2018, **16**, 316–328.
3 L. Perincherry, J. Lalak-Kańczugowska and Ł. Stępień, *Toxins*, 2019, **11**, 664.
4 M. J. Hajipour, K. M. Fromm, A. Akbar Ashkarran, D. Jimenez de Aberasturi, I. R. de Larramendi, T. Rojo, V. Serpooshan, W. J. Parak and M. Mahmoudi, *Trends Biotechnol.*, 2012, **30**, 499–511.
5 N.-Y. T. Nguyen, N. Grelling, C. L. Wetteland, R. Rosario and H. Liu, *Sci. Rep.*, 2018, **8**, 16260.
6 L. Miao, C. Wang, J. Hou, P. Wang, Y. Ao, Y. Li, N. Geng, Y. Yao, B. Lv, Y. Yang, G. You and Y. Xu, *Bioresour. Technol.*, 2016, **216**, 537–544.
7 C. López de Dicastillo, M. Guerrero Correa, F. B. Martínez, C. Streitt and M. José Galotto, in *Antimicrobial Resistance - A One Health Perspective*, eds. M. Mareș, S. Hua Erin Lim, K.-S. Lai and R.-T. Cristina, IntechOpen, 2021.
8 Y. He, S. Ingudam, S. Reed, A. Gehring, T. P. Strobaugh and P. Irwin, *J. Nanobiotechnology*, 2016, **14**, 54.
9 J. Chen, H. Peng, X. Wang, F. Shao, Z. Yuan and H. Han, *Nanoscale*, 2014, **6**, 1879–1889.
10 S. Shamaila, N. Zafar, S. Riaz, R. Sharif, J. Nazir and S. Naseem, *Nanomaterials*, 2016, **6**, 71.

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

11 K. Markowska, A. M. Grudniak and K. I. Wolska, *Acta Biochim. Pol.*, 2013, **60**, 523–530.
12 J. Chen, L. Wu, M. Lu, S. Lu, Z. Li and W. Ding, *Front. Microbiol.*, 2020, **11**, 365.
13 J. Pečenka, Z. Bytešníková, T. Kiss, E. Peňázová, M. Baránek, A. Eichmeier, D. Tekielska, L. Richtera, R. Pokluda and V. Adam, *Mater. Today Commun.*, 2021, **27**, 102284.
14 B. Thomas, M. C. Raj, A. K. B, R. M. H, J. Joy, A. Moores, G. L. Drisko and C. Sanchez, *Chem. Rev.*, 2018, **118**, 11575–11625.
15 K. Heise, E. Kontturi, Y. Allahverdiyeva, T. Tammelin, M. B. Linder, Nonappa and O. Ikkala, *Adv. Mater.*, 2021, **33**, 2004349.
16 K. J. Edgar, C. M. Buchanan, J. S. Debenham, P. A. Rundquist, B. D. Seiler, M. C. Shelton and D. Tindall, *Prog. Polym. Sci.*, 2001, **26**, 1605–1688.
17 M. Rose and R. Palkovits, *Macromol. Rapid Commun.*, 2011, **32**, 1299–1311.
18 S. Kalia, A. Dufresne, B. M. Cherian, B. S. Kaith, L. Avérous, J. Njuguna and E. Nassiopoulos, *Int. J. Polym. Sci.*, 2011, **2011**, 1–35.
19 I. Siró and D. Plackett, *Cellulose*, 2010, **17**, 459–494.
20 S. Ummartyotin and C. Pechyen, *Carbohydr. Polym.*, 2016, **142**, 133–140.
21 L. Zheng, S. Li, J. Luo and X. Wang, *Front. Bioeng. Biotechnol.*, 2020, **8**, 593768.
22 R. Koshani, J. Zhang, T. G. M. van de Ven, X. Lu and Y. Wang, *ACS Sustain. Chem. Eng.*, 2021, **9**, 10513–10523.
23 D. Georgouvelas, H. N. Abdelhamid, J. Li, U. Edlund and A. P. Mathew, *Carbohydr. Polym.*, 2021, **264**, 118044.
24 V. A. Spirescu, C. Chircov, A. M. Grumezescu, B. Ștefan Vasile and E. Andronescu, *Int. J. Mol. Sci.*, 2021, **22**, 4595.
25 V. Rodríguez-González, C. Terashima and A. Fujishima, *J. Photochem. Photobiol. C Photochem. Rev.*, 2019, **40**, 49–67.
26 X.-X. Chen, B. Cheng, Y.-X. Yang, A. Cao, J.-H. Liu, L.-J. Du, Y. Liu, Y. Zhao and H. Wang, *Small*, 2013, **9**, 1765–1774.
27 K. P. Priyanka, T. H. Sukirtha, K. M. Balakrishna and T. Varghese, *IET Nanobiotechnol.*, 2016, **10**, 81–86.
28 B. Kim, D. Kim, D. Cho and S. Cho, *Chemosphere*, 2003, **52**, 277–281.
29 J.-W. Liou and H.-H. Chang, *Arch. Immunol. Ther. Exp. (Warsz.)*, 2012, **60**, 267–275.
30 H. A. Foster, I. B. Ditta, S. Varghese and A. Steele, *Appl. Microbiol. Biotechnol.*, 2011, **90**, 1847–1868.
31 M. R. Khan and Z. A. Siddiqui, *Gesunde Pflanz.*, 2021, **73**, 445–464.
32 Q. Zhang, X. Yan, R. Shao, H. Dai and S. Li, *J. Wuhan Univ. Technol.-Mater Sci Ed*, 2014, **29**, 407–409.
33 M. A. Irshad, R. Nawaz, M. Zia ur Rehman, M. Imran, J. Ahmad, S. Ahmad, A. Inam, A. Razzaq, M. Rizwan and S. Ali, *Chemosphere*, 2020, **258**, 127352.
34 B. K. Thakur, A. Kumar and D. Kumar, *South Afr. J. Bot.*, 2019, **124**, 223–227.
35 A. K. Kesharwani, D. Singh, A. Kulshreshtha, A. S. Kashyap, A. S. Avasthi and N. Geat, *Plant Dis.*, 2023, **107**, 212.
36 D. Singh, P. S. Rathaur and J. G. Vicente, *Plant Pathol.*, 2016, **65**, 1411–1418.
37 D. Singh, A. K. Kesharwani, K. Singh, S. Jaiswal, M. A. Iquebal, N. Geat and A. S. Avasthi, *Plant Dis.*, 2022, **106**, 1502–1505.
38 S. Sreelatha, N. Kumar and S. Rajani, *Front. Microbiol.*, 2022, **13**, 1085113.
39 A. Varympopi, A. Dimopoulou, I. Theologidis, T. Karamanidou, A. Kaldeli Kerou, A. Vlachou, D. Karfaridis, D. Papafotis, D. G. Hatzinikolaou, A. Tsouknidas and N. Skandalis, *Pathogens*, 2020, **9**, 1024.
40 H. Peng and I. A. Chen, *ACS Nano*, 2019, acsnano.8b06395.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

41 S. Malmir, A. Karbalaei, M. Pourmadadi, J. Hamedi, F. Yazdian and M. Navaee, *Carbohydr. Polym.*, 2020, **234**, 115835.
42 Mohd. K. Khan, A. Pandey, T. Athar, S. Choudhary, R. Deval, S. Gezgin, M. Hamurcu, A. Topal, E. Atmaca, P. A. Santos, M. R. Omay, H. Suslu, K. Gulcan, M. Inanc, M. S. Akkaya, A. Kahraman and G. Thomas, *3 Biotech*, 2020, **10**, 172.
43 M. McMullen, G. Bergstrom, E. De Wolf, R. Dill-Macky, D. Hershman, G. Shaner and D. Van Sanford, *Plant Dis.*, 2012, **96**, 1712–1728.
44 J. Graham and E. Feichtenberger, *J. Citrus Pathol.*, DOI:10.5070/C421027203.
45 S. O. Cacciola and G. M. di S. Lio, in *Integrated Management of Diseases Caused by Fungi, Phytoplasma and Bacteria*, eds. A. Ciancio and K. G. Mukerji, Springer Netherlands, Dordrecht, 2008, pp. 61–84.
46 S. Dinesh, K. Y. D, S. Shweta, K. M. K, S. Gita, R. P. R, ey and S. Rajender, *Afr. J. Microbiol. Res.*, 2013, **7**, 5459–5470.
47 T. Stein, *Mol. Microbiol.*, 2005, **56**, 845–857.
48 G. Ganeshan and A. Manoj Kumar, *J. Plant Interact.*, 2005, **1**, 123–134.
49 C. Sharma, S. N. Sharma and R. Srivastava, *Colloid Polym. Sci.*, 2022, **300**, 1027–1036.
50 A. H. Bashal, K. D. Khalil, A. M. Abu-Dief and M. A. El-Atawy, *Int. J. Biol. Macromol.*, 2023, **253**, 126856.
51 A. M. Abu-Dief, F. M. M. Alrashedee, K. M. Emran and H. A. Al-Abdulkarim, *Inorg. Chem. Commun.*, 2022, **138**, 109251.
52 Ž. Antić, R. M. Krsmanović, M. G. Nikolić, M. Marinović-Cincović, M. Mitrić, S. Polizzi and M. D. Dramićanin, *Mater. Chem. Phys.*, 2012, **135**, 1064–1069.
53 A. D. French, *Cellulose*, 2014, **21**, 885–896.
54 D. Bondeson, A. Mathew and K. Oksman, *Cellulose*, 2006, **13**, 171–180.
55 J. Gong, J. Li, J. Xu, Z. Xiang and L. Mo, *RSC Adv.*, 2017, **7**, 33486–33493.
56 E. M. M. Ibrahim, A. M. Abu-Dief, A. Elshafaie and A. M. Ahmed, *Mater. Chem. Phys.*, 2017, **192**, 41–47.
57 L. H. Abdel Rahman, A. M. Abu-Dief, R. M. El-Khatib, S. M. Abdel-Fatah, A. M. Adam and E. M. M. Ibrahim, *Appl. Organomet. Chem.*, 2018, **32**, e4174.
58 T. T. Win, S. Khan and P. Fu, *J. Nanotechnol.*, 2020, **2020**, 1–9.
59 A. K. Madbouly, M. S. Abdel-Aziz and M. A. Abdel-Wahhab, *IET Nanobiotechnol.*, 2017, **11**, 702–708.
60 M. AlMatar, E. A. Makky, I. Var and F. Koksal, *Curr. Drug Deliv.*, 2018, **15**, 470–484.
61 C. Lee, J. Y. Kim, W. I. Lee, K. L. Nelson, J. Yoon and D. L. Sedlak, *Environ. Sci. Technol.*, 2008, **42**, 4927–4933.
62 J. Yu, W. Zhang, Y. Li, G. Wang, L. Yang, J. Jin, Q. Chen and M. Huang, *Biomed. Mater.*, 2014, **10**, 015001.

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D4FB00182F

Data availability statement

The raw/processed data required to reproduce the above findings cannot be shared at this time due to legal/ ethical reasons.

Open Access Article. Published on 07 December 2024. Downloaded on 12/16/2024 5:14:08 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

J Mater Sci: Mater Electron (2025) 36:68

Tb^{3+}/Eu^{3+} co-doped $NaLa(MoO_4)_2$ phosphors: photoluminescence and optical thermometry studies for solid-state lighting and sensor applications

Sonali Tomar[1], Vaibhav Chauhan[2], M. Jayasimhadri[3], and C. Shivakumara[1,*]

[1] *Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore 560012, India*
[2] *Atomic and Molecular Physics Division, Bhabha Atomic Research Centre, Mumbai 400085, India*
[3] *Luminescent Materials Research Lab (LMRL), Department of Applied Physics, Delhi Technological University, Delhi 110042, India*

Received: 6 June 2024
Accepted: 17 December 2024

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

ABSTRACT

Tb^{3+}/Eu^{3+} co-doped $NaLa(MoO_4)_2$ phosphors were synthesized by conventional solid-state reaction method. Rietveld refinement of the powder XRD data confirmed that all the compounds crystallized in the scheelite-type tetragonal structure with space group $I4_1/a$ (No. 88). Photoluminescence results reveal that, upon the excitation at 298 nm, Tb^{3+}-doped $NaLa(MoO_4)_2$ phosphors emit intense green light centred at 545 nm. On co-doping of Eu^{3+} ions in the $NaLa(MoO_4)_2$:Tb^{3+} (7 mol%) phosphor, the energy transfer appears from Tb^{3+} to Eu^{3+} ions, and it induces concentration-dependent colour tuning from green to red. The lifetime results confirmed that the maximum energy transfer efficiency 48.20% was observed for the composition $NaLa_{0.84}Tb_{0.07}Eu_{0.09}(MoO_4)_2$. Temperature-dependent PL of $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ at 423 K retains about 61% of emission intensity with respect to that at 298 K, and the temperature quenching is characterized by activation energy 0.38 eV. For the $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$ phosphor, an activation energy of 0.26 eV was determined for thermal quenching, with a maximum relative sensitivity of approximately 0.15% K^{-1}. The determined radiative properties, CIE, CCT values and temperature-sensing results suggest the usefulness of Tb^{3+}/Eu^{3+} co-doped $NaLa(MoO_4)_2$ phosphors in developing solid-state lighting and high-temperature sensor applications.

1 Introduction

The research on white light-emitting diodes (w-LEDs) has spurred enormous attention, revolutionizing the lighting and display field of application [1–4]. The merits of the w-LEDs encompass energy efficiency, compact size, high brightness, prolonged operational lifespan and environmental sustainability. The fabrication of the commercial w-LED requires coating a yellow Ce^{3+}-doped YAG phosphor over the ultraviolet (UV) or near UV LED chip. Alternatively, blue, green and red phosphors can be used to obtain white light.

Address correspondence to E-mail: shiva@iisc.ac.in
E-mail Addresses: sonali15466@iisc.ac.in; vaibhavchauhan.rs.phy17@itbhu.ac.in; jayasimha@dtu.ac.in

https://doi.org/10.1007/s10854-024-14109-8
Published online: 28 December 2024

The w-LEDs are applied in diverse emerging fields, such as fluorescent sensors, colour-tunable devices and encrypted information storage [5–8]. Among these applications, colour tunability from phosphor can be achieved by doping multiple rare-earth ions giving distinct emissions at the same excitation wavelength. Other factors include crystal site engineering, altering the temperature and altering the concentration of the doped rare-earth ions. Adjustable colour rendering index (CRI) and correlated colour temperature (CCT) are two benefits of phosphors with tunable emission that make them ideal for tunable light sources [9].

Over the years, rare-earth-activated phosphors have gained significant interest in producing tunable light sources via co-activation [2, 10–15]. The 4f–4f and 5d–4f electronic transitions of the rare-earth ions give rise to characteristic sharp spectral lines and broad emission band covering the entire visible spectrum. Among various rare-earth ions, the Tb^{3+} ion with its dominant $^5D_4 \rightarrow {}^7F_5$ transition has been utilized as a dominant green-emitting phosphor across multiple lighting devices [16]. Tb^{3+} can also be a suitable sensitizer to Eu^{3+} ions, an excellent red-emitting ion. Eu^{3+} mainly relies on $^5D_0 \rightarrow {}^7F_2$-induced electric dipole transition for its characteristic warm red emission [17, 18]. The colour tunability of cool to warm light is reachable by controlling the dopant concentrations of the Tb^{3+} and Eu^{3+} ions meticulously and altering the excitation wavelength.

Choosing the appropriate host for rare-earth-based phosphor is an essential consideration in various applications. Various host families, such as vanadates [19], oxides [20], sulphides [21], phosphates [22], silicates [23], garnets [24], aluminates [25], molybdates [26, 27], tungstates [28], nitrides [29] or oxynitrides [30], were studied as parent material for RE ions [31–33]. Among these phosphors, the molybdates have garnered significant attention owing to their excellent properties, including good chemical and thermal stability, cost-effective and environmentally friendly synthesis, near UV excitation and blue-green broadband emission [34–36]. An example of self-activated molybdate phosphor is $NaLa(MoO_4)_2$, which exhibits a $[MoO_4]^{2-}$ charge transfer band (CTB) resulting from O^{2-} to Mo^{6+} charge transfer. The $NaLa(MoO_4)_2$ facilitates effective energy transfer from the $^1T_2$ and $^1T_1$ energy bands of the Mo^{6+} to excited energy levels of the rare-earth ions. While previous studies, such as those by L. Peng et al. and G. Li et al., were focussed on the hydrothermal synthesis and optical properties of the Tb^{3+} and Eu^{3+} co-doped $NaLa(MoO_4)_2$ phosphors, colour tunability and thermal stability were not thoroughly explored [37, 38]. Specifically, G. Li et al. focussed on monitoring the luminescence variation by varying synthesis parameters and microstructure morphology. However, excitation-dependent alteration in the luminescence of the phosphor was not explored. Y. Zhai et al. reported the luminescent properties of the Tb^{3+} and Eu^{3+} co-doped $NaLa(MoO_4)_2$ phosphors prepared by the microwave-assisted sol–gel method [39]. However, the thermal stability of the luminescence in this co-doped phosphor was not addressed. The effectiveness of luminescence is severely reduced by thermal quenching, which also limits its applicability in lighting and display applications. Therefore, comprehensive knowledge of the thermal stability of the phosphor is essential for informed application in various contexts.

In this work, we synthesized series of Tb^{3+}/Eu^{3+} co-doped $NaLa(MoO_4)_2$ phosphors by the conventional solid-sate reaction method at relatively low temperature and shorter duration. We explore the energy transfer coupled with concentration-dependent colour-tunable properties in Tb^{3+}/Eu^{3+} pair in the $NaLa(MoO_4)_2$ host matrix for solid-state lighting applications. We investigate the temperature-dependent photoluminescence (TDPL) and important thermal quenching parameters of the prepared phosphors. In addition, the temperature-sensing property of the Tb^{3+} and Eu^{3+} co-doped $NaLa(MoO_4)_2$ phosphors was studied.

2 Experimental section

2.1 Materials synthesis

Series of $NaLa_{1-x}Tb_x(MoO_4)_2$ ($0.00 \le x \le 0.17$) phosphors and $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$ ($0.00 \le y \le 0.09$) phosphors were synthesized by the conventional solid-state reaction approach. Lanthanum oxide (La_2O_3, 99.9%, Loba Chemicals), sodium carbonate (Na_2CO_3, 99%, Merck) and molybdenum trioxide (MoO_3, 99.5%, Chemie Pvt. Ltd) were employed as starting materials in the preparation of $NaLa(MoO_4)_2$ host material. While preparing Tb^{3+}-activated $NaLa(MoO_4)_2$ phosphors, terbium oxide (Tb_2O_3, 99.9%, Sigma-Aldrich) was used as a source of Tb^{3+} dopant. The co-doped $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$ phosphors were made using a similar procedure by adding a stoichiometric

amount of Tb_2O_3 oxide and Europium oxide (Eu_2O_3, 99.9%, Sigma-Aldrich) and grinding in an agate mortar. Before preparation, the components were weighed according to a required ratio and ground in an agate mortar with lanthanum oxide, Tb_2O_3 and Eu_2O_3 (preheated at 800 °C for 6 h). The mixture was subsequently placed into a crucible, heated to 750 °C in a muffle furnace for 4 h and then cooled to room temperature. Further, all the compounds were characterizations using various experimental methods. All the glassware and agate mortar used in the preparation of the samples were thoroughly cleaned with neutral detergent followed by Aqua regia treatment. After cleaning, they were further sterilized with UV treatment to prevent contamination from airborne particles, moisture and reactive gases, ensuring material purity.

2.2 Characterizations

A PANalytical Empyrean powder X-ray diffractometer was employed to characterize the synthesized phosphors at operating 45 kV and 30 mA with CuK radiation at a scanning rate of 2°/min for 2θ range from 10° to 80°. The structural parameters for each phosphor were calculated using the Rietveld refinement method using Fullprof software. Fourier transform infrared spectroscopy (FTIR) spectra were recorded on the PerkinElmer Frontier Spectrometer to identify the functional groups. Bandgaps of the compounds were determined on the base of diffuse reflectance spectra (DRS) recorded be a PerkinElmer Lambda 750 spectrophotometer. X-ray photoelectron spectroscopy (XPS) measurements were carried out using Alkα as an X-ray source (1486.6 eV). A Jobin Yvon spectrofluorometer (Fluoromax, Horiba) with a 450 W xenon lamp as the excitation source was used to record the emission and excitation spectra. All measurements were carried out at room temperature. Temperature-dependent photoluminescence (TDPL) is measured in an Ocean optics spectrophotometer (Model: FLAME-S-XRI-ES) and household heating assembly.

3 Result and discussions

3.1 X-ray diffraction analysis

The powder XRD patterns of the $NaLa_{1-x}Tb_x(MoO_4)_2$ phosphors are presented in Fig. 1. The sharp peaks

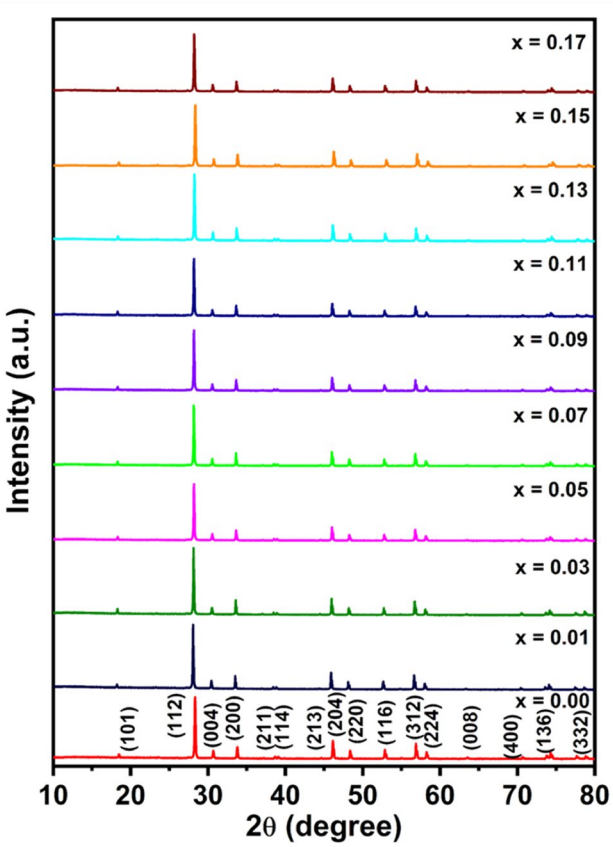


Fig. 1 XRD patterns of $NaLa_{1-x}Tb_x(MoO_4)_2$ phosphors

of the diffraction patterns revealed their high crystallinity. The diffraction peaks agree well with the reported JCPDS File No. 85–1751 [40]. Further, we did not see any extra peaks related to Tb_2O_3 or raw materials in the XRD patterns. This implies that all the samples exhibited a pure tetragonal phase, indicating that Tb^{3+} ions had been incorporated into the lattice of the $NaLa(MoO_4)_2$ host matrix. The radius percentage difference of the doped Tb^{3+} ion and the La^{3+} ion in $NaLa(MoO_4)_2$ was estimated by the following formula: [41]

$$R_r = \frac{R(CN)_h - R(CN)_d}{R(CN)_h} \times 100\%, \quad (1)$$

where R_r signifies the percentage difference in radii, CN stands for the coordination number, R_h (CN) refers to the radius of the host cation and R_d (CN) denotes the radius of the doped ion. The difference between substituted and doped ions cannot be higher than 30% for the R_r value [42]. The difference between the Tb^{3+} ion (CN = 8, R_d = 1.04 Å) and the La^{3+} ion (CN = 8,

R_h = 1.16 Å) in the $NaLa(MoO_4)_2$ host, namely, R_r value, was found to be 10.34%. This indicates that Tb^{3+} ions can be substituted at the La^{3+} site in the $NaLa(MoO_4)_2$ matrix. Similarly, the XRD patterns of $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$ co-doped phosphors are shown in Figure S1. It is important to maintain the purity of the samples as the impurity, even in trace amounts, can significantly influence the optical and structural properties of the material. Impurities can introduce unwanted energy levels within the band-gap, creating non-radiative recombination centres that compete with the desired emission pathways. This competition often leads to quenching effects, reducing the overall photoluminescence intensity and affecting the material's efficiency. In our prepared Tb^{3+}/Eu^{3+} co-doped $NaLa(MoO_4)_2$ phosphors, impurities may disturb the energy transfer mechanism between Tb^{3+} and Eu^{3+} ions by altering the crystal field around the activator ions. This alteration can reduce colour purity and hinder tunable colour emission. Therefore, it is important to maintain the material purity of samples for optimal properties. To check the sample purity, all the diffraction peaks observed in the XRD patterns were matched with the JCPDS File No. 85–1751. No impurity diffraction peaks were observed, revealing that the impact of co-doped Eu^{3+} and Tb^{3+} ions had minimal impact on the crystal structure of the host. Thus, this indicates that all the above-prepared phosphors have a pure tetragonal phase.

The Rietveld refinement is used to determine the structural parameters of $NaLa_{1-x}Tb_x(MoO_4)_2$. In Fig. 2 are the observed, calculated and the difference XRD patterns of (a) $NaLa_{0.99}Tb_{0.01}(MoO_4)_2$, (b) $NaLa_{0.95}Tb_{0.05}(MoO_4)_2$, (c) $NaLa_{0.91}Tb_{0.09}(MoO_4)_2$, (d) $NaLa_{0.89}Tb_{0.11}(MoO_4)_2$, (e) $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ and (f) $NaLa_{0.83}Tb_{0.17}(MoO_4)_2$ molybdates. The outcomes of the refinement demonstrate that $NaLa_{1-x}Tb_x(MoO_4)_2$ crystallized into a tetragonal scheelite structure. The Rietveld-refined structure parameters for both undoped and Tb^{3+}-doped phosphors are provided in Table S1. It is evident from Table S1 that the lattice parameters (a and c) and unit cell volume gradually decrease as the Tb^{3+} content increases up to 17 mol%. This can be attributed to the smaller 8-coordinated Tb^{3+} ion ($r_{Tb}{}^{3+}$ = 1.04 Å) substituted for larger 8-coordinated La^{3+} ion ($r_{La}{}^{3+}$ = 1.160 Å) in the scheelite-type tetragonal $NaLa(MoO_4)_2$ system. The doping of smaller Tb^{3+} ions in larger La^{3+} ions sites lead to denser crystal structure. The denser crystal structure improves structural stability, which is advantageous for high-temperature applications, as it can help reduce lattice vibrations and maintain the integrity of the material under thermal stress [35, 43].

The unit cell volume and cell parameters (a and c) are presented in Fig. 3a with respect to the Tb^{3+} concentration. Further, the VESTA package was used to model the crystal structure of $NaLa_{0.93}Tb_{0.07}(MoO_4)_2$ phosphor using refined cell parameters and atomic coordinates, as shown in Fig. 3b [44]. It is observed that each Mo^{6+} was coordinated by four O^{2-} in a MoO_4 tetrahedral sites, which are connected to the (Na/La)O_8 dodecahedron by their shared oxygen atoms and stabilize the overall structure. Additionally, it is observed that eight O^{2-} anion coordinating with Na^+/La^{3+}, making MoO_4 dodecahedron (Na/La)O_8. In the $NaLa(MoO_4)_2$ system, substituted Tb^{3+} ions were also occupied by La^{3+} ions; these Tb^{3+} ions are furthermore bound to eight oxygen atoms [45].

3.2 XPS analysis

An XPS analysis was conducted to identify the oxidation state of the constituent elements in the $NaLa_{0.93}Tb_{0.07}(MoO_4)_2$ phosphor. The binding energy scale was calibrated in reference to C 1 s core level fixed at 284.6 eV in to the XPS data of all elements and survey scan. The survey scan of the $NaLa_{0.93}Tb_{0.07}(MoO_4)_2$ phosphor is shown in Fig. 4a. The National Institute of Standard Technology (NIST) XPS database assigns each La, Na, Mo, O and Tb element peak in the survey scan [46]. The survey scan ensures the existence of all relevant elements and validates the absence of impurities other than carbon. The XPS binding energy spectra of Na (1 s), La (3d), Tb (3d), Mo (3d) and O (1 s) are shown in Figs. 4b–f. In Fig. 4b, the Na 1 s peak is observed at an energy level of 1071.2 eV [47]. In Fig. 4c, it can be seen that the La ions are positioned at around 835.04 eV and 851.72 eV, corresponding to the $3d_{3/2}$ and $3d_{5/2}$ energy levels, respectively [47–49]. The XPS spectra validate the presence of lanthanum in the + 3-oxidation state. Figure 4d presents the two peaks centred at 1277.1 eV and 1242.3 eV attributed to the $3d_{3/2}$ and $3d_{5/2}$ core levels of Tb^{3+}, respectively. The core-level XPS spectra of Tb^{3+} validate its + 3-oxidation state [16]. In Fig. 4e is the Mo 3d doublet, showing peaks at approximately 235.6 and 232.5 eV [36, 50]. The observed peaks of Mo are from the $3d_{3/2}$ and $3d_{5/2}$ components, validating the + 6-oxidation state of Mo [26, 51]. The O 1 s peak, centred at ~ 530.4 eV, is shown in Fig. 4g. [26]

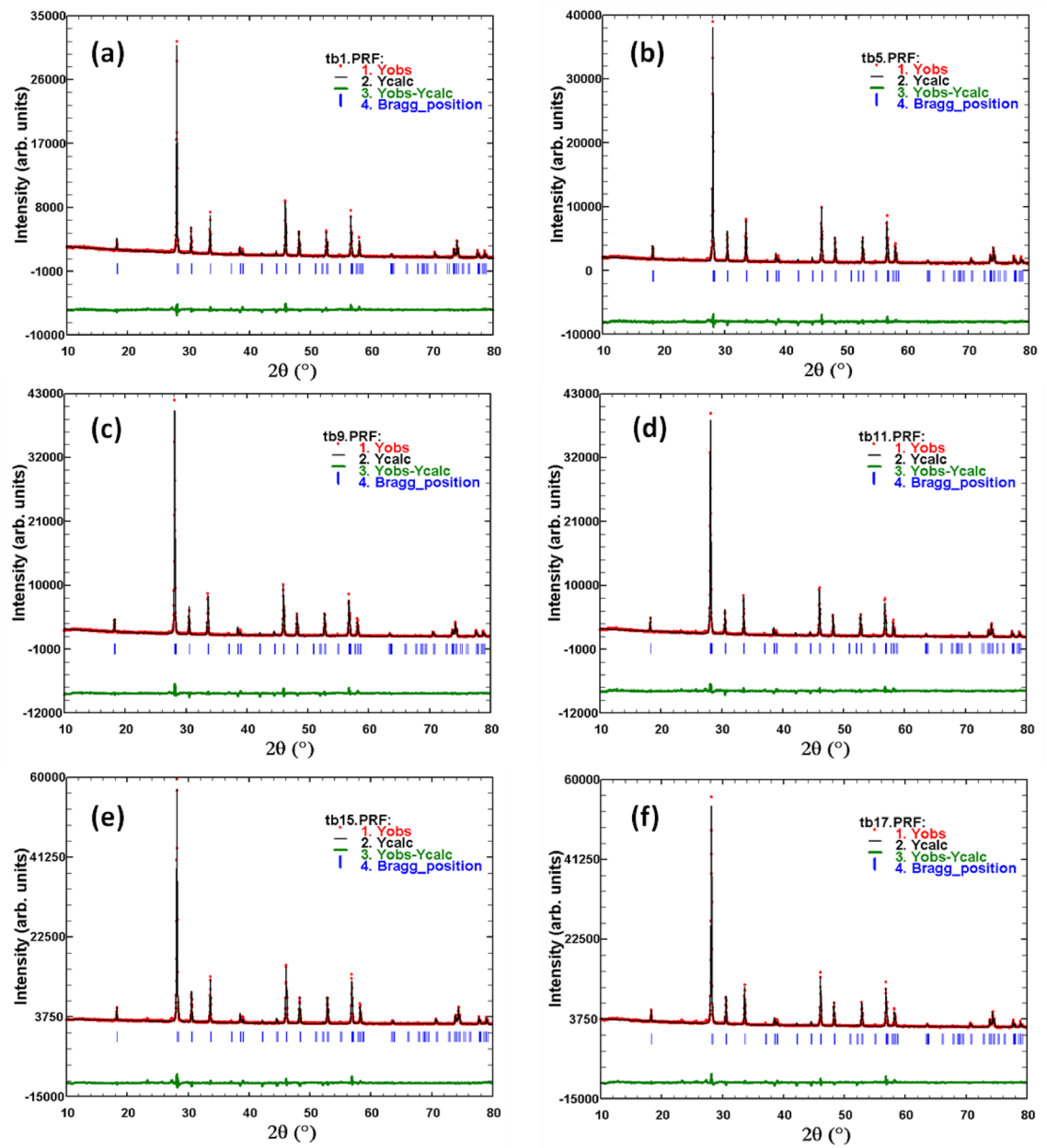


Fig. 2 Rietveld-refined XRD patterns of **a** $NaLa_{0.99}Tb_{0.01}(MoO_4)_2$, **b** $NaLa_{0.95}Tb_{0.05}(MoO_4)_2$, **c** $NaLa_{0.91}Tb_{0.09}(MoO_4)_2$, **d** $NaLa_{0.89}Tb_{0.11}(MoO_4)_2$, **e** $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ and **f** $NaLa_{0.83}Tb_{0.17}(MoO_4)_2$ phosphors

3.3 FTIR analysis

The FTIR spectra of the compounds $NaLa_{1-x}Tb_x(MoO_4)_2$ ($0.00 \le x \le 0.17$) are shown in Fig. 5. The spectra were collected over the wave number range of 350–4000 cm^{-1}. The peak at 401 cm^{-1} is due to the anti-symmetric bending of bonds in the $[MoO_4]_2$ cluster present in the $NaLa_{1-x}Tb_x(MoO_4)_2$ compound. The 401 cm^{-1} peak is attributed to the bending vibration of the Mo–O bonds within the MoO_4^{2-} tetrahedra. The 698 cm^{-1} and 760 cm^{-1} peaks correspond to the symmetric and asymmetric stretching vibrations of the O–Mo–O bonds in the MoO_4^{2-} units. These stretching modes are characteristic of the MoO_4^{2-} tetrahedron structure and indicate the presence of well-defined tetrahedral coordination [43]. The 905 cm^{-1} peak is typically associated with the anti-symmetric stretching mode of the Mo–O

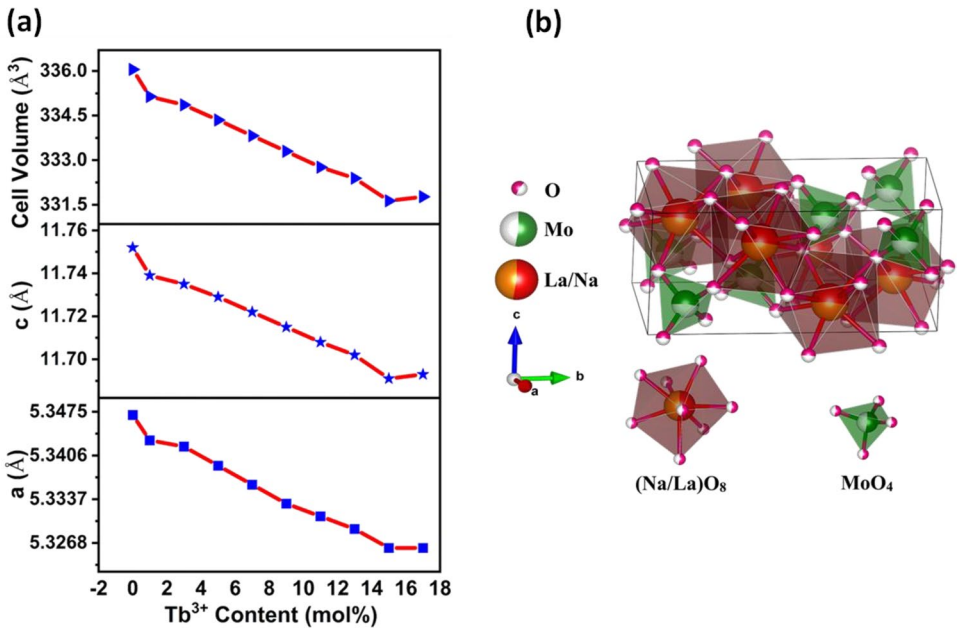


Fig. 3 **a** The unit cell volume and lattice parameters as a function of Tb^{3+} content in $NaLa(MoO_4)_2$ phosphors and **b** Crystal structure of $NaLa_{0.93}Tb_{0.07}(MoO_4)_2$ phosphor with MoO_4 tetrahedron and $(Na/La)O_8$ dodecahedron

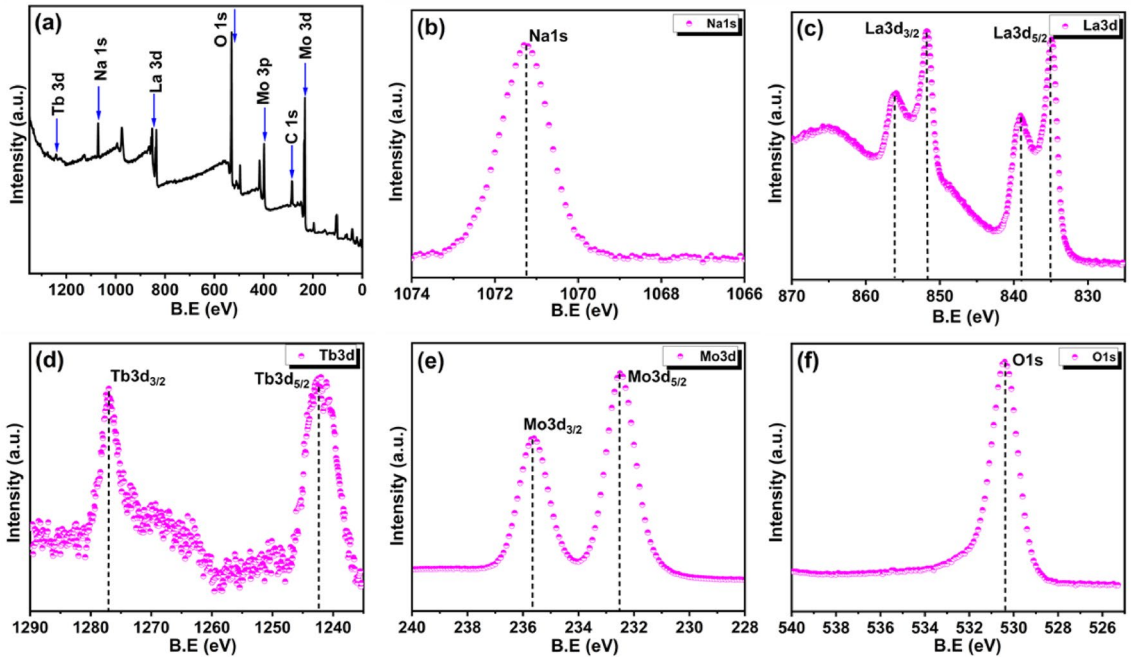


Fig. 4 High-resolution XPS spectra of $NaLa_{0.93}Tb_{0.07}(MoO_4)_2$: **a** XPS scan Survey, **b** Na 1 s spectra, **c** La 3d spectra, **d** Tb 3d spectra, **e** Mo 3d spectra, **f** O 1 s spectra

bonds. It signifies strong Mo–O bonding interactions within the tetrahedral framework [52]. The absence of extra bands further supports that every compound was in the pure phase. The stability of the observed vibrational bands at 401 cm^{-1}, 698 cm^{-1}, 760 cm^{-1} and 905 cm^{-1} in the FTIR spectra, even after doping with Tb^{3+}, indicates that the fundamental MoO_4^{2-} structure remains intact. This stability suggests that the incorporation of Tb^{3+} ions does not significantly disturb the local bonding environment of the MoO_4^{2-} tetrahedra.

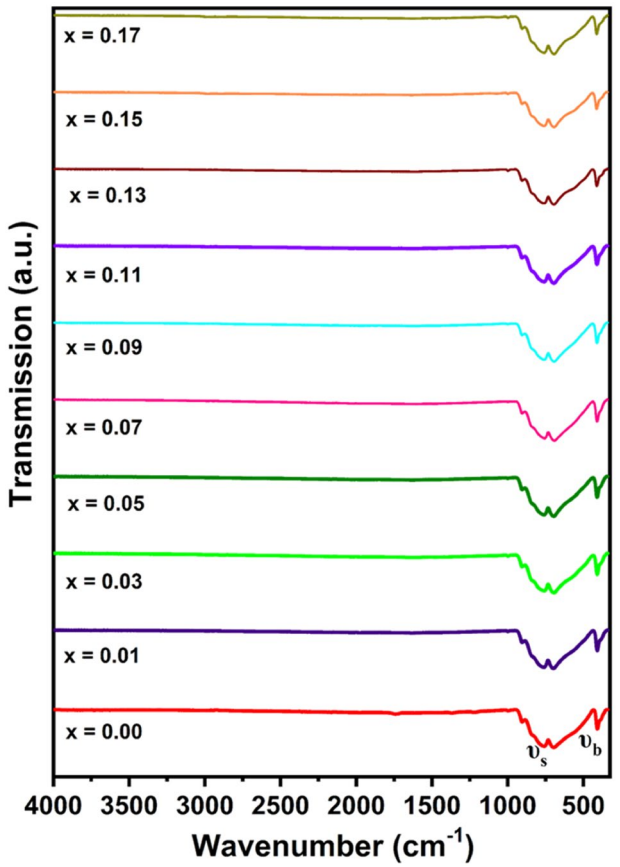


Fig. 5 FTIR spectra of $NaLa_{1-x}Tb_x(MoO_4)_2$ $(0.00 \le x \le 0.17)$

3.4 UV–Visible analysis

The DRS spectra and Kubelka–Munk plot of $NaLa_{1-x}Tb_x(MoO_4)_2$ are presented in Fig. 6a, b. The reflectance spectra exhibit charge transfer band around 298 nm for all the samples. From the DRS of the host and $NaLa_{1-x}Tb_x(MoO_4)_2$, the optical bandgap is calculated by using the Kubelka–Munk equation and Tauc equation, which is shown as [53]

$$F(R_\infty) = \frac{(1-R_\infty)^2}{2R_\infty} = \frac{K}{S}. \quad (2)$$

The Kubelka and Munk theory establishes the mathematical relationship between the absorption coefficient (K), scattering coefficient (s) and diffusion reflectance (R_∞).

The function $F(R_\infty)$ is commonly referred to as the Kubelka–Munk function. Here, R_∞ represents the ratio of the diffuse reflectance of the test sample to that of an ideal non-absorbing and scattering coefficient, denoted as $\frac{R_{sample}}{R_{reference}}$. Barium sulphate ($BaSO_4$) was employed as a reference sample. The relationship between the optical bandgap (E_g) and the linear coefficient (α) is established by the Tauc equation [54]:

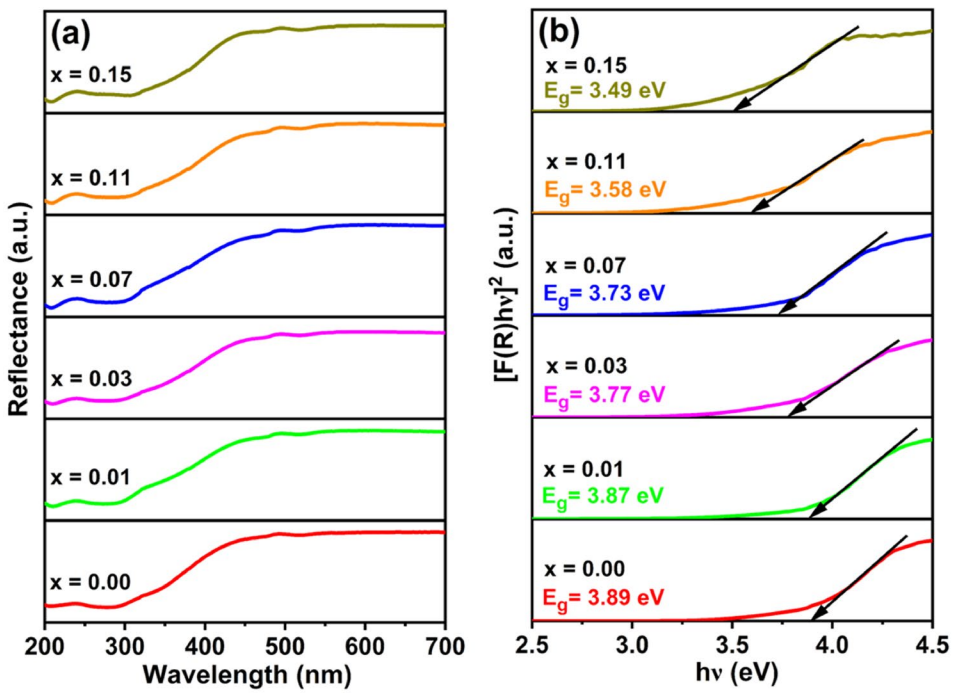


Fig. 6 **a** Diffuse reflectance spectra of $NaLa_{1-x}Tb_x(MoO_4)_2$ and **b** Kubelka–Munk plots of $NaLa_{1-x}Tb_x(MoO_4)_2$

$$\alpha h\vartheta = C_0\left(h\vartheta - E_g\right)^n. \quad (3)$$

In the given context, $h\vartheta$ represents the energy of a photon, whereas A is a constant. The values of n, which are 1/2, 2, 3/2 and 3, correspond to different types of transitions [55]. Specifically, when $n = 1/2$, it signifies a direct allowed transition; when $n = 2$, it indicates an indirect allowed transition; $n = 3/2$ denotes a direct forbidden transition, and $n = 3$ represents an indirect forbidden transition. The direct allowed transition was taken into account in this investigation. The bandgap energy of the of $NaLa_{1-x}Tb_x(MoO_4)_2$ compounds was determined using the Tauc plot method. This involved extrapolating the linear portion of the graph to obtain a value of $[F(R_\infty)h\vartheta]^2$ equal to zero. As it is seen in Fig. 6b, the optical bandgap of the material decreases gradually, as the Tb^{3+} content increases. This phenomenon can be linked to the formation of defect energy levels associated with Tb^{3+} ions within the bandgap of the host material. Introducing defect levels facilitates the transfer of electrons from negatively charged oxygen ions (O^{2-}) to these levels, hence causing a red shift in the bandgap [51].

3.5 Photoluminescence studies

3.5.1 *Photoluminescence properties of $NaLa(MoO_4)_2$:Tb^{3+} phosphors*

The PL study provides an analysis of the luminescent properties exhibited by $NaLa_{1-x}Tb_x(MoO_4)_2$ phosphors under ambient conditions. Figure 7a shows the typical (PLE) excitation spectrum of

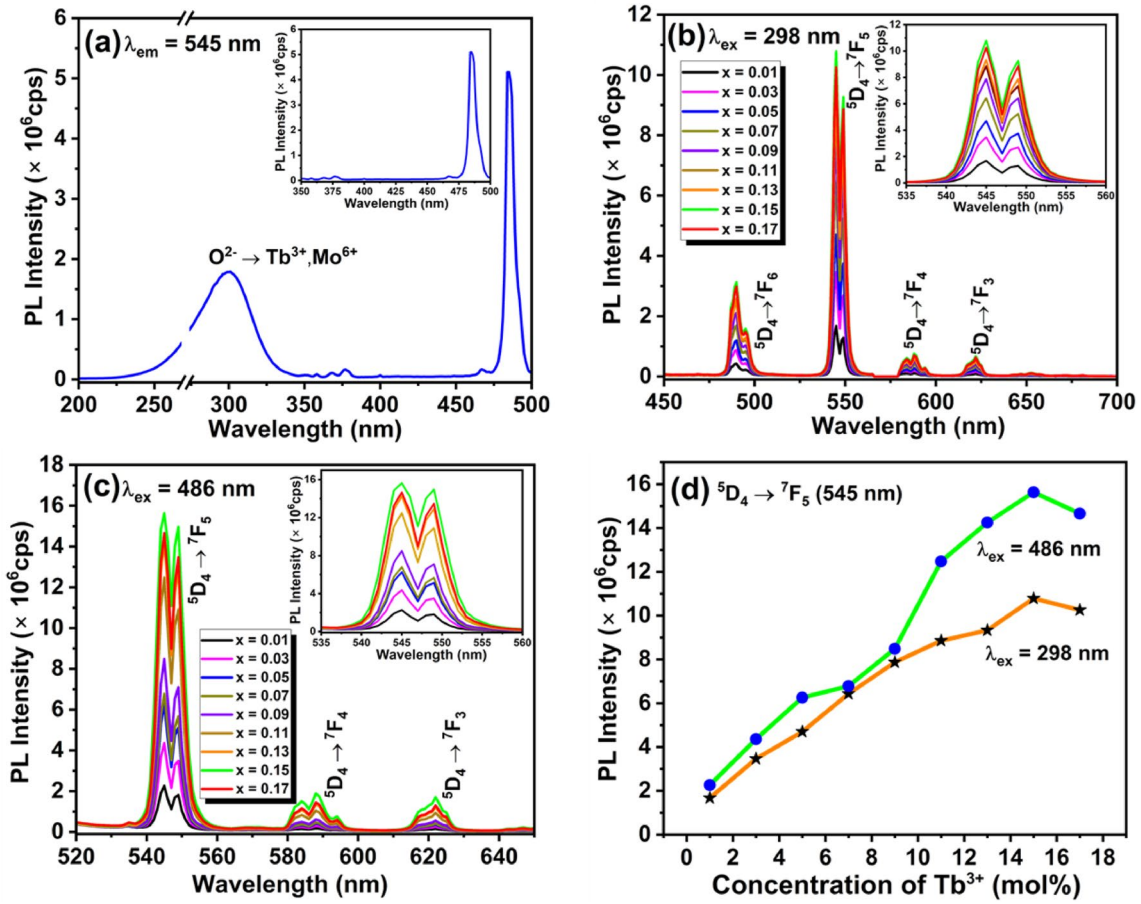


Fig. 7 **a** Excitation spectra of $NaLa_{0.93}Tb_{0.07}(MoO_4)_2$ phosphor monitored at 545 nm emission wavelength. The emission spectrum of $NaLa_{1-x}Tb_x(MoO_4)_2$ phosphors monitored at **b** $\lambda_{ex} = 298$ nm and **c** $\lambda_{ex} = 486$ nm. Inset shows the zoomed region of 545 nm ($^5D_4 \rightarrow {}^7F_5$) emission peak, **d** Variation of the emission intensity with the percentage of Tb^{3+} doping in $NaLa(MoO_4)_2$ phosphors monitored at $\lambda_{ex} = 486$ nm and $\lambda_{ex} = 298$ nm

$NaLa_{0.93}Tb_{0.07}(MoO_4)_2$. The PL emission spectra of the synthesized $NaLa_{1-x}Tb_x(MoO_4)_2$ phosphors are shown in Figs. 7b, c. The PL excitation spectrum monitored at 545 nm showcases two distinct parts. The first component is a broad range covering 200 to 350 nm, with the highest intensity observed at 298 nm. This portion is attributed to the charge transfer band (CBT), wherein electrons from the 2p level of O_2 orbitals are excited to the vacant orbitals of the central Mo^{6+} within the MoO_4^{2-} group. The observation of a prominent peak corresponding to the MoO_4^{2-} group in the excitation spectrum of Tb^{3+} ions suggests the occurrence of energy transfer from the MoO_4^{2-} group to the Tb^{3+} ions in $NaLa(MoO_4)_2$ phosphors. Another spectrum component covers from 350 to 500 nm and exhibits a distinct, narrow peak centred at 486 nm. This peak arises from the $4f^8 \rightarrow 4f^7$ 5d spin-forbidden transition of Tb^{3+} ions, specifically attributed to the $^7F_6 \rightarrow {}^5D_4$ transition of these ions, Fig. 7a [16]. The emission spectra of $NaLa_{1-x}Tb_x(MoO_4)_2$ phosphors with different concentrations of Tb^{3+} ions were recorded under excitation at 298 and 486 nm. The phosphors exhibit a prominent emission peak at 545 nm, corresponding to the $^5D_4 \rightarrow {}^7F_5$ transition of Tb^{3+}. Three additional peaks are detected at 488 nm ($^5D_4 \rightarrow {}^7F_6$), 590 nm ($^5D_4 \rightarrow {}^7F_4$) and 622 nm ($^5D_4 \rightarrow {}^7F_3$) [56, 57]. The profiles and positions of the photoluminescence (PL) spectra exhibit similarity, with no discernible changes observed, except for variations in the emission intensity when excited by different wavelengths. Figure 7d is the emission intensity of green light at 545 nm as a function of Tb^{3+} concentration. The observation reveals that the intensity of photoluminescence (PL) gradually increases as the concentration of Tb^{3+} ions increases. This increase continues until it reaches an optimal value at $x = 0.15$. Beyond this concentration, the PL intensity decreases with further increase in Tb^{3+} concentration. This occurrence is ascribed to the concentration quenching effect due to electric multipolar interaction.

3.5.2 *Photoluminescence properties of $NaLa(MoO_4)_2$:0.07Tb^{3+}/yEu^{3+} phosphors*

To investigate colour-tunable properties of $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$, phosphors were synthesized at different Eu^{3+} ions. The PLE and PL spectra of $NaLa_{0.93}Tb_{0.07}(MoO_4)_2$ are shown in Fig. 8. The excitation spectra consist of a broadband over 200–300 nm with band maximum at 298 nm, which corresponds to charge transfer from O^{2-} ions to Tb^{3+}, Mo^{6+} ions and some sharp peaks located at 486 nm in the range of 350–500 nm due to $^7F_6 \rightarrow {}^5D_4$ transition of the Tb^{3+} ions. In addition, the emission spectrum shows several peaks at 488, 545, 589, 619 and 651 nm, and they are related to $^5D_4 \rightarrow {}^7F_6$, $^5D_4 \rightarrow {}^7F_5$, $^5D_4 \rightarrow {}^7F_4$, $^5D_4 \rightarrow {}^7F_3$ and $^5D_4 \rightarrow {}^7F_2$ transitions, respectively [56]. Similarly, Fig. 8b shows $NaLa_{0.93}Eu_{0.07}(MoO_4)_2$ monitored at 615 nm and excited at 298 nm. The excitation spectra, centred at 298 nm, consist of CTB. This is caused by the charge transfer from O^{2-} ions to Eu^{3+}, Mo^{6+} ions in the range of 200 to 350 nm. Additionally, there are a series of strong peaks in the higher wavelength region of 350–500 nm, which can be attributed to the 4f–4f transitions of Eu^{3+}. Further, in a longer wavelength region, some sharp peaks were allocated at approximately 359, 379, 391, 414 and 463 nm, corresponding to the transition from the ground state of $^7F_0 \rightarrow {}^5D_4$, $^5G_2$, $^5L_6$, $^5D_3$ and $^5D_2$ to the excited levels of Eu^{3+} ions. Moreover, emission spectrum excited at 298 nm consists of a group of some weak emissions at about 536, 592, 655 and 700 nm assigned to $^5D_0 \rightarrow {}^7F_J$ (J = 0, 1, 3 and 4) and one strong red emission at about 615 nm is due to $^5D_1 \rightarrow {}^7F_2$ is due to intra-4f shell transition of Eu^{3+} ions [58–60]. Among these, two sharp emission lines peaking at 615 nm ($^5D_0 \rightarrow {}^7F_2$) and 591 nm ($^5D_0 \rightarrow {}^7F_1$) transitions were attributed to the characteristic emission of Eu^{3+} ions.

Figure 8c shows the PLE and PL spectra of $NaLa_{0.86}Tb_{0.07}Eu_{0.07}(MoO_4)_2$ monitored at 575 and 615 nm and excited at 298 nm. It can be seen that the excitation spectra consist of two components. The first component is the charge transfer band (CTB) of the host, which is exhibited in the shorter wavelength range of 250–350 nm. The second component is a sequence of strong peaks, which are owing to the 4f–4f configuration of Tb^{3+} and Eu^{3+}, and they are observed in the higher wavelength region of 350–500 nm. The wide spectral band ranging from 250 to 350 nm, monitored at 545 nm and 615 nm emission of Tb^{3+} and Eu^{3+} ions, respectively, is caused by the transfer of charge from O^{2-} ions to Tb^{3+}/Eu^{3+} and Mo^{6+} ions. Furthermore, in the Tb^{3+}/Eu^{3+} co-doped $NaLa(MoO_4)_2$ host, energy migration between the host and both Tb^{3+} and Eu^{3+} ions may occur due to the existence of the host excitation band during monitoring at 545 nm of Tb^{3+} ions and at 615 nm of Eu^{3+} ions. In addition, a peak at 486 nm is detected when the emission of Eu^{3+} monitored, which is caused by the Tb^{3+} transition from

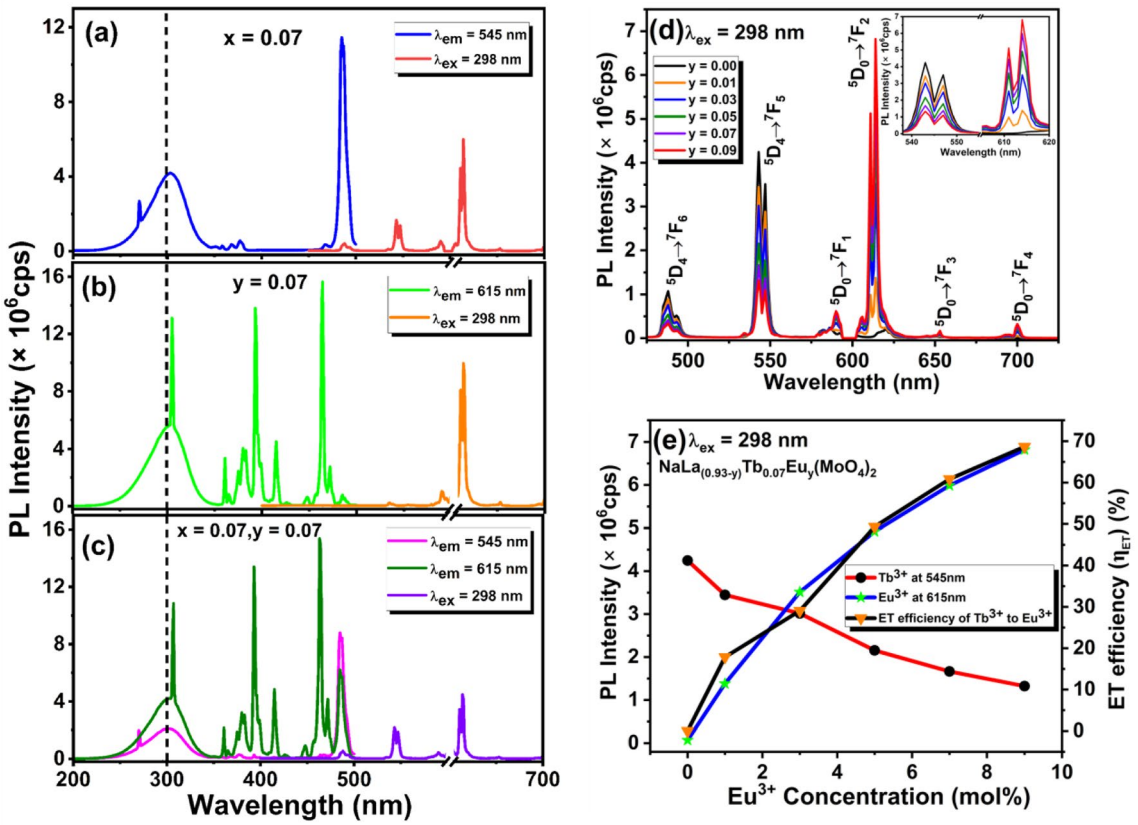


Fig. 8 PLE and PL spectra of **a** $NaLa_{0.93}Tb_{0.07}(MoO_4)_2$ monitored at 545 nm and excited at 298 nm, **b** $NaLa_{0.93}Eu_{0.07}(MoO_4)_2$ monitored at 615 nm and excited at 298 nm, **c** $NaLa_{0.86}Tb_{0.07}Eu_{0.07}(MoO_4)_2$ monitored at 575 nm and 615 nm and excited at 298 nm, **d** Emission spectra of $NaLa_{0.93\text{-}y}Tb_{0.07}Eu_y(MoO_4)_2$ phosphor monitored at 298 nm excitation wavelength and **e** PL intensity of the 545 nm and 615 nm peaks and ET efficiency (η_{ET}) variation with increasing Eu^{3+} concentration

$^7F_6 \rightarrow {}^5D_4$. This indicates that Tb^{3+} ions are transferring energy to Eu^{3+} ions in the $NaLa(MoO_4)_2$ host.

Further, the emission spectra of $NaLa_{0.93\text{-}y}Tb_{0.07}Eu_y(MoO_4)_2$ under excitation at 298 nm are presented in Fig. 8d. The emission spectra of each sample exhibit the same profile with varying intensities and contain the unique peaks of Tb^{3+} and Eu^{3+} ions resulting from the 4f–4f transitions. The observed spectra displayed emission peaks at specific wavelengths, namely 488 nm, 545 nm and 590 nm. These peaks were attributed to the transitions $^5D_4 \rightarrow {}^7F_6$, $^5D_4 \rightarrow {}^7F_5$ and $^5D_4 \rightarrow {}^7F_4$ of the Tb^{3+} ion. Additionally, emission peaks were observed at 590, 615, 654 and 701 nm. These peaks were attributed to the transitions $^5D_0 \rightarrow {}^7F_1$, $^5D_0 \rightarrow {}^7F_2$, $^5D_0 \rightarrow {}^7F_3$ and $^5D_0 \rightarrow {}^7F_4$ of the Eu^{3+} ion, respectively [58]. In Fig. 8e, it is shown that as the Eu^{3+} concentration rises, the luminescence intensity of Eu^{3+} ions progressively increases, whereas the luminescence intensity of Tb^{3+} ions gradually decreases. This phenomenon indicates that the energy emitted by Eu^{3+} is transmitted through energy transfer by Tb^{3+}. By controlling the concentration of Eu^{3+} ions, it is possible to achieve the desired emission colour and tune the photoluminescence emission within the green–red region.

The energy transfer efficiency (η_{ET}) in $NaLa_{(0.93\text{-}y)}Tb_{0.07}Eu_y(MoO_4)_2$ can be determined by the following equation [61]:

$$\eta_{ET} = 1 - \frac{I_s}{I_{so}}, \tag{4}$$

where I_s and I_{so} are the emission intensities with and without sensitizer. As we can see in Fig. 8e, η_{ET} increases as a function of Eu^{3+} concentration. It also increased gradually and reached the maximum value of 68.65%. This finding supports the presence of an effective ET route for Tb^{3+} to Eu^{3+} in the $NaLa(MoO_4)_2$ host. Because of this property, it is a promising option

for developing white light-emitting diodes (WLEDs) and display devices.

3.5.3 PL decay analysis and energy transfer dynamics

Figure 9a shows the PL decay curves of $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$ phosphors monitored at 298 nm excitation and 545 nm emission wavelengths. The decay curve for Tb^{3+} emission diverges gradually from a single exponential function and was accurately fitted by the following monoexponential equation [62]:

$$I(t) = I_0 + A\exp\left(-\frac{t}{\tau}\right). \quad (5)$$

I (t) signifies the emission intensity at a specific time t in the above equation. A is constant, whereas τ is the lifetime associated with the exponential decay component.

The parameters used to fit the fluorescence decay curves and decay lifetime (τ) values are presented in Table 1. It is observed that the lifetime of the Tb^{3+} energy level decreases monotonously with the increase in Eu^{3+} concentration. The variation of decay time with increasing Eu^{3+} concentration is represented in Fig. 9b. The decreasing lifetime of the Tb^{3+} energy level is a consequence of the possible energy transfer from the Tb^{3+} ion to the Eu^{3+} energy level.

In order to comprehend the understanding of the energy transfer process, the rate of energy transfer (W_{ET}) from Tb^{3+} to Eu^{3+} can be calculated by the following equation [63]:

Table 1 Fitting parameters of the monoexponential of decay lifetime A, lifetime τ and energy transfer efficiency (η_{ET}%) of the $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$ phosphors

Samples	A	τ(ms)	η_{ET}%
$y=0.00$	3.31898	20.54	0
$y=0.01$	0.00898	11.73	6.73
$y=0.03$	0.02987	10.08	20.63
$y=0.05$	2.76474	6.57	31.46
$y=0.07$	0.078	5.26	40.65
$y=0.09$	2.31822	4.62	48.20

$$W_{ET} = \frac{1}{\tau} - \frac{1}{\tau_0}. \quad (6)$$

Figure 9 (b) shows the increasing energy transfer rate from Tb^{3+} to Eu^{3+} with increasing Eu^{3+} concentration. Additionally, the estimation of energy transfer efficiency (η_{ET}) from the sensitizer ion (Tb^{3+}) to the activator ion (Eu^{3+}) can be implemented using the lifetimes of Tb^{3+} ion and the equation proposed by Paulose [64]:

$$\eta_{ET} = 1 - \frac{\tau}{\tau_0}. \quad (7)$$

The lifetimes of sensitizer Tb^{3+} ions, with and without Eu^{3+} ions, are represented by τ and τ_0, respectively. The variation of the (η_{ET}) value with respect to the concentration of Eu^{3+} is shown in Fig. 9b. The values of η_{ET}% are presented in Table 1. Based on the curve given in Fig. 9b, it is evident that both the energy transfer efficiency η_{ET} and the energy transfer rate W_{ET} exhibit a consistent upward trend with increasing doping

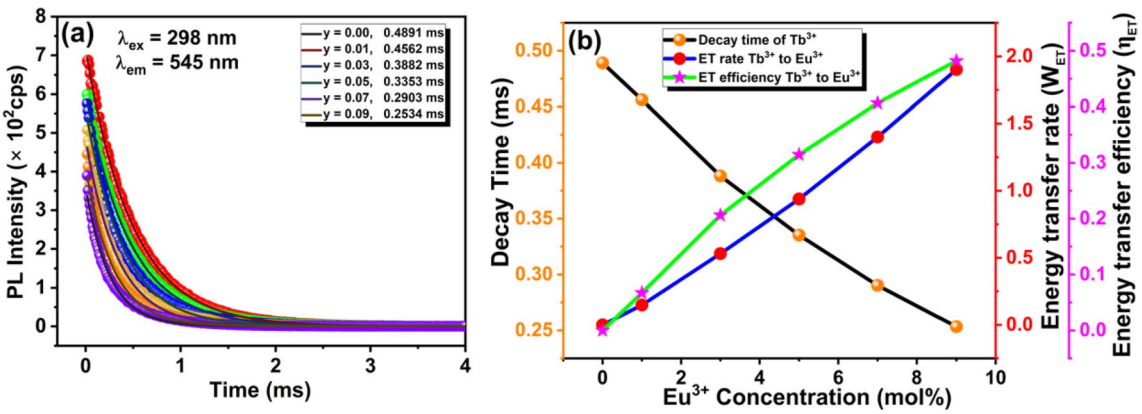


Fig. 9 **a** PL decay curves of $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$ phosphors monitored at 298 nm excitation and 545 nm emission wavelength. **b** Variation in decay time, energy transfer rate (W_{ET}) and energy transfer efficiency (η_{ET}) with increasing Eu^{3+} concentration

concentration of Eu^{3+} ions. This trend serves as direct proof supporting the occurrence of energy transfer between Tb^{3+} and Eu^{3+} ions in the $NaLa(MoO_4)_2$ host.

Furthermore, the radiative energy transfer dynamics in $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$ phosphors can be understood by the energy level diagram shown in Fig. 10. The PLE spectrum of the $NaLa_{0.86}Tb_{0.07}Eu_{0.07}(MoO_4)_2$ phosphors has shown the occurrence of broadband centred at 298 nm Fig. 8c. The broadband is the consequence of the direct overlap of $[MoO_4]^{2-}$ LMCT band and $O^{2-} \rightarrow Tb^{3+}/Eu^{3+}$ CTB. As depicted in Fig. 10, electrons absorbing 298 nm UV photons are excited to the $^1T_1$ and $^1T_2$ excited energy levels of the Mo^{+6} ions. Afterwards, the electrons relax to the excited levels of the Tb^{3+} and Eu^{3+} ions, resulting in the characteristic emission peaks of these rare-earth ions. However, it is important to point that, when the concentration of Eu^{3+} is increased, the intensity of the Eu^{3+} peaks in Fig. 8e also increases, while the intensity of the Tb^{3+} peaks diminishes. This results in an increasing trend for the 615 nm (Eu^{3+}) peak and a decreasing trend for the 545 nm (Tb^{3+}) peak. The reason behind this could be the increase in the possible energy transfer from Tb^{3+} to Eu^{3+} energy levels when the concentration of the Eu^{3+} is increased for the fixed concentration of Tb^{3+} in $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$ phosphor. This claim is supported by the variation in the energy transfer efficiency discussed from the results of the PL decay lifetime.

Concentration quenching of emission peaks is caused by non-radiative energy transfer among the RE ions at increasing doping concentrations. One of three processes—exchange interaction, radiative reabsorption or electric multipolar interaction—can result in a non-radiative energy transfer. A significant overlap between the emission and excitation spectra is necessary for radiative reabsorption to occur. Radiation reabsorption is disregarded in this case because of the narrow spectrum overlap between the Eu^{3+} excitation and the Tb^{3+} emission. For exchange interactions,

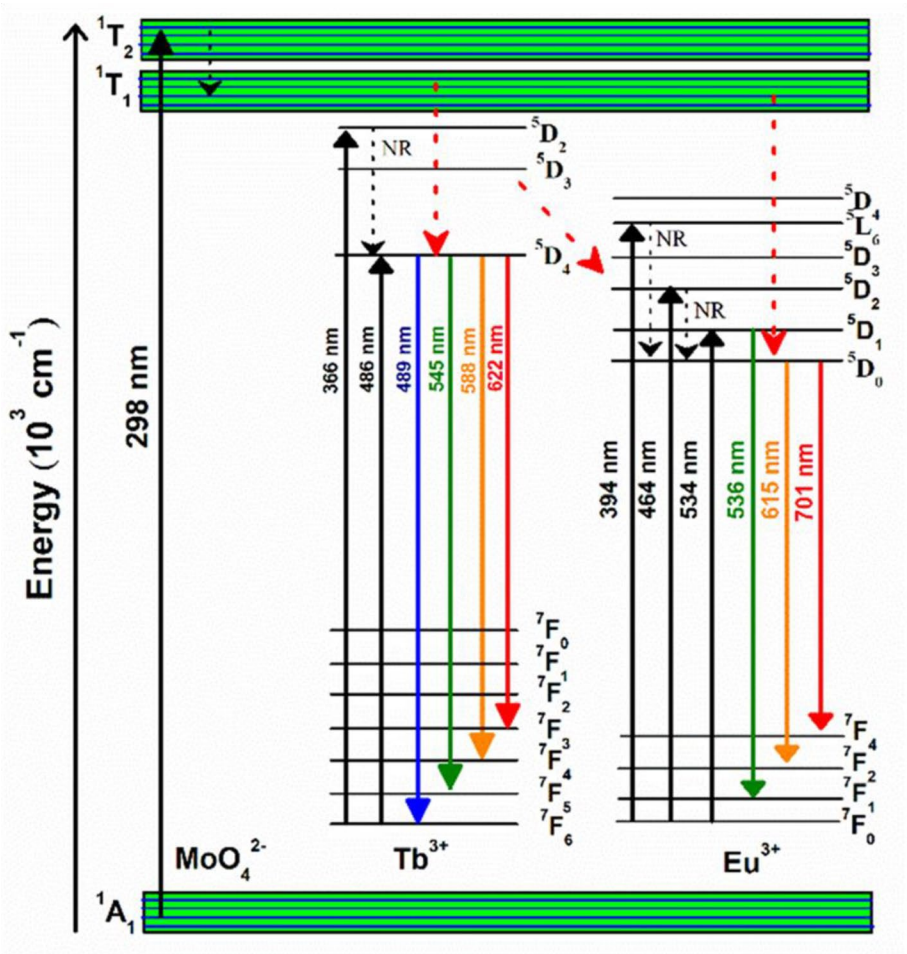


Fig. 10 Energy level diagram of the energy transfer process in $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$ phosphors

there must be a significant direct or indirect overlap between the donor and acceptor wave functions. The Van Uitert theory states that for wavefunction overlapping, there must be a critical separation distance (R_c) less than 5 Å between the emitting ions [65]. The R_c value for Tb^{3+}/Eu^{3+} co-doped $NaLa(MoO_4)_2$ system is determined by the following formula [66]:

$$R_c = 2\left[\frac{3V}{4\pi x_c N}\right]^{1/3}. \quad (8)$$

In this case, the variables V and N represent the unit cell volume and the number of cation sites within the unit cell that can be doped with Tb^{3+} and Eu^{3+} ions, respectively. x_c represents the total concentration of sensitizer (Tb^{3+}) and activator (Eu^{3+}) when the energy transfer efficiency is 50%. Here, the variables $V = 333.14$ $Å^3$ and $x_c = 0.1$ when the Eu^{3+} concentration was 0.03 and $N = 2$. Therefore, the critical distance (R_c) was determined to be 14.7 Å. Given that the calculated Rc value is greater than 5 Å, it is likely that the electric multipolar interaction plays a major role in the energy transfer between the sensitizer (Tb^{3+}) and activator (Eu^{3+}) within this host material.

The confirmation of the specific type of electric multipolar interaction can be obtained through Dexter formula and Reisfeld approximation, as expressed by the following equation [67]:

$$\frac{I_{s0}}{I_s} \propto C^{n}/3, \quad (9)$$

where I_s and I_{s0} are the PL emission intensity of Tb^{3+} ion ($^5D_4 \rightarrow {}^7F_5$) in the sample with and without doping of Eu^{3+} ions, respectively. C represents the sum of Tb^{3+} and Eu^{3+} ions doping concentrations, while $n = 6$, 8 and 10 correspond to the dipole–dipole, dipole-quadrupole and quadrupole–quadrupole interaction, respectively. Figure 11a–c shows the plot of I_{so}/I_s as a function of $C^{n/3}$ with $n = 6$, 8 and 10 for $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$

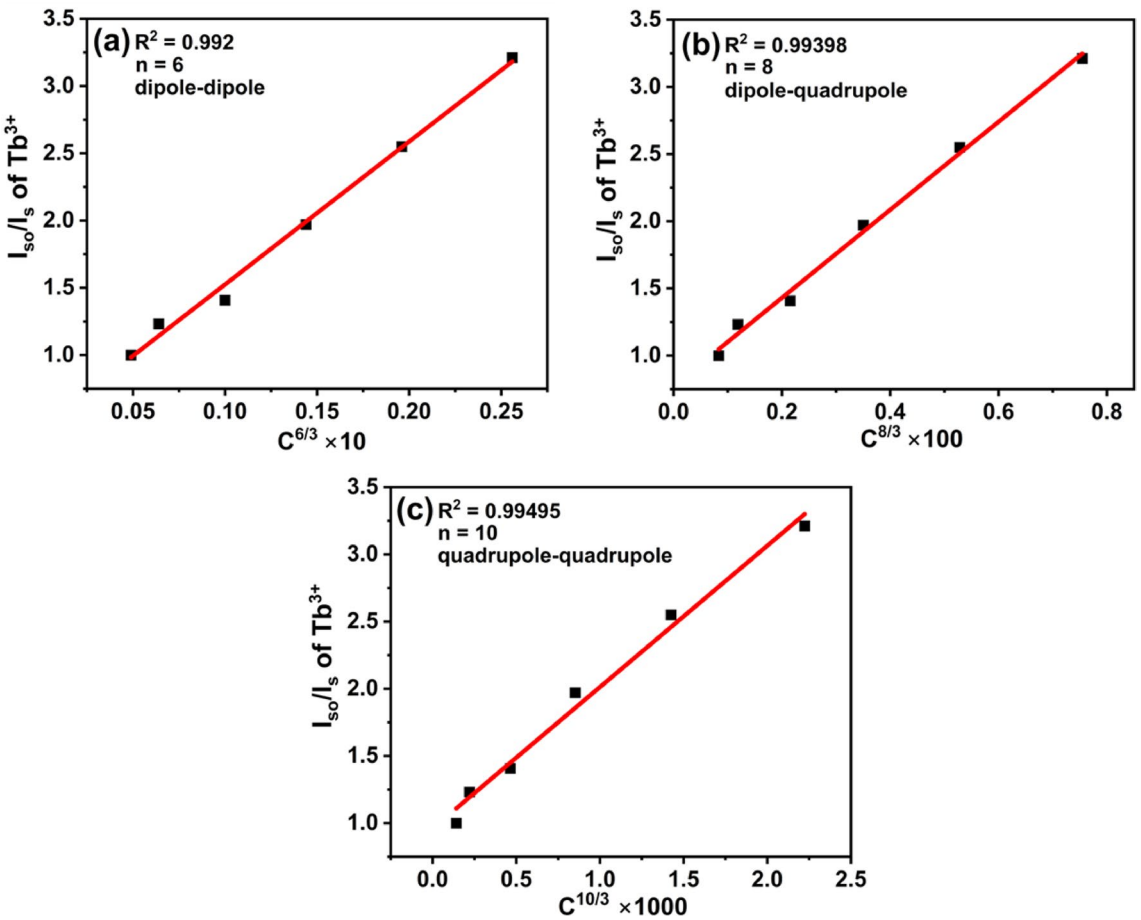


Fig. 11 The plots of I_{so}/I_s of Tb^{3+} as a function of **a** $C^{6/3}$, **b** $C^{8/3}$ and **c** $C.^{10/3}$

phosphors, respectively. The best linear fitting was reached at $n = 10$, indicating that the energy transfer from Tb^{3+} to Eu^{3+} can be attributed to the quadrupole–quadrupole interaction.

3.6 Thermal stability analysis

Knowledge of the thermal stability of phosphors is essential for their application in the lighting and display field. Previous studies on the Tb^{3+}/Eu^{3+} co-doped $NaLa(MoO_4)_2$ phosphors do not give information about the thermal stability of the phosphor [37, 39, 68]. The thermal stability of the Tb^{3+} doped and Tb^{3+}/Eu^{3+} co-doped $NaLa(MoO_4)_2$ phosphors is performed after analysing the TDPL spectra of both samples. The TDPL spectra of $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ and $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$ phosphors were monitored in the temperature range of 298 to 448 K and under 298 nm excitation, as shown in Figs. 12a, 13a. The peak position and form are found to be unchanged as the temperature rises. Figure 12b shows the normalized intensity variation of two major peaks of Tb^{3+} ion with rising temperature. The emission intensity of both peaks decreases with a temperature rise; however, it is evident that the peak intensities do not decrease in the same ratio. The normalized integrated intensity variation with rising temperature is presented in Fig. 12c. At 423 K, the intensity of emission intensity of $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ is approximately 61% that of 298 K. For $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$, the normalized intensity variation of 615 nm and 545 nm peaks is shown in Fig. 13b. At 423 K, the emission intensity of $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$ is about 66% that of 298 K.

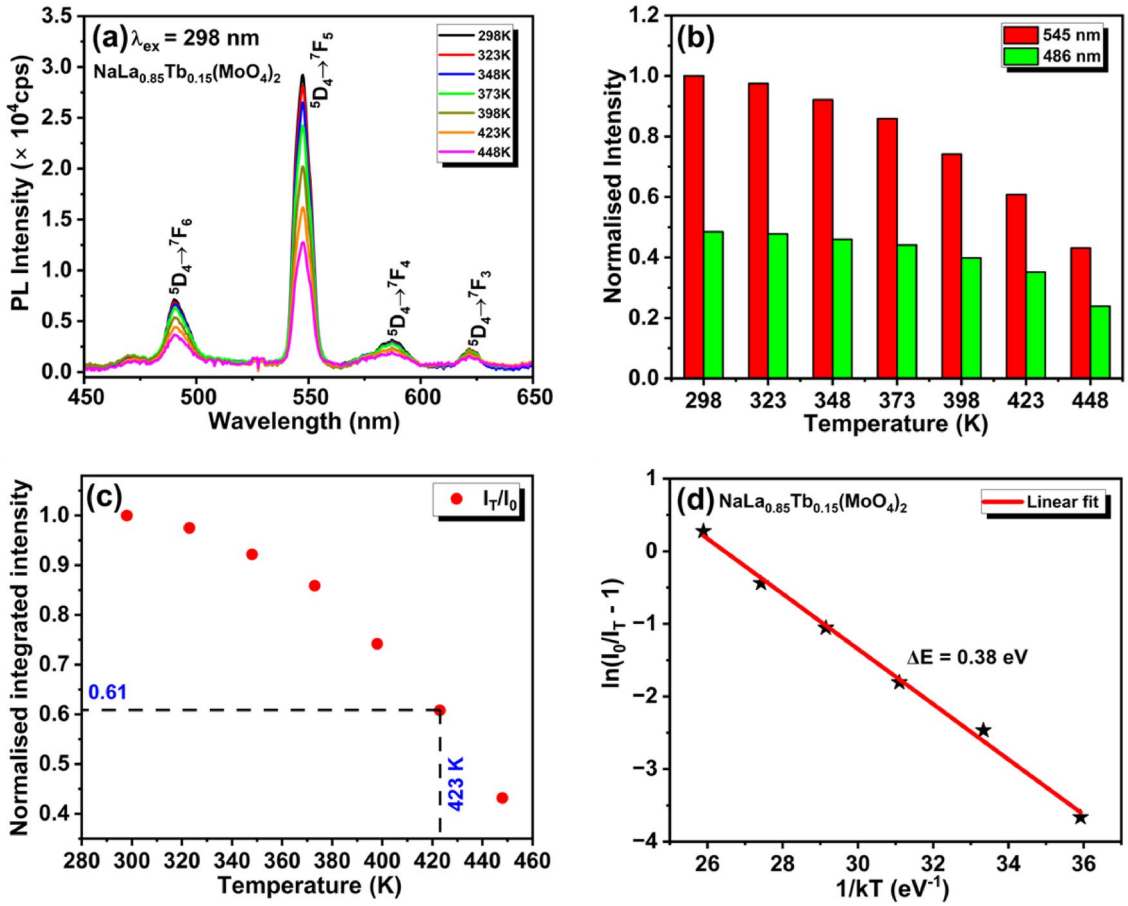


Fig. 12 **a** TDPL spectra of $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ phosphor excited at 298 nm, **b** Bar diagram representing the normalized intensity variation as a function of temperature for two major peaks of Tb^{3+}. **c** Normalized integral intensity variation as a function of temperature. **b** $\ln\left(\frac{I_0}{I_T} - 1\right)$ *vs* 1/kT plot for determination of activation energy

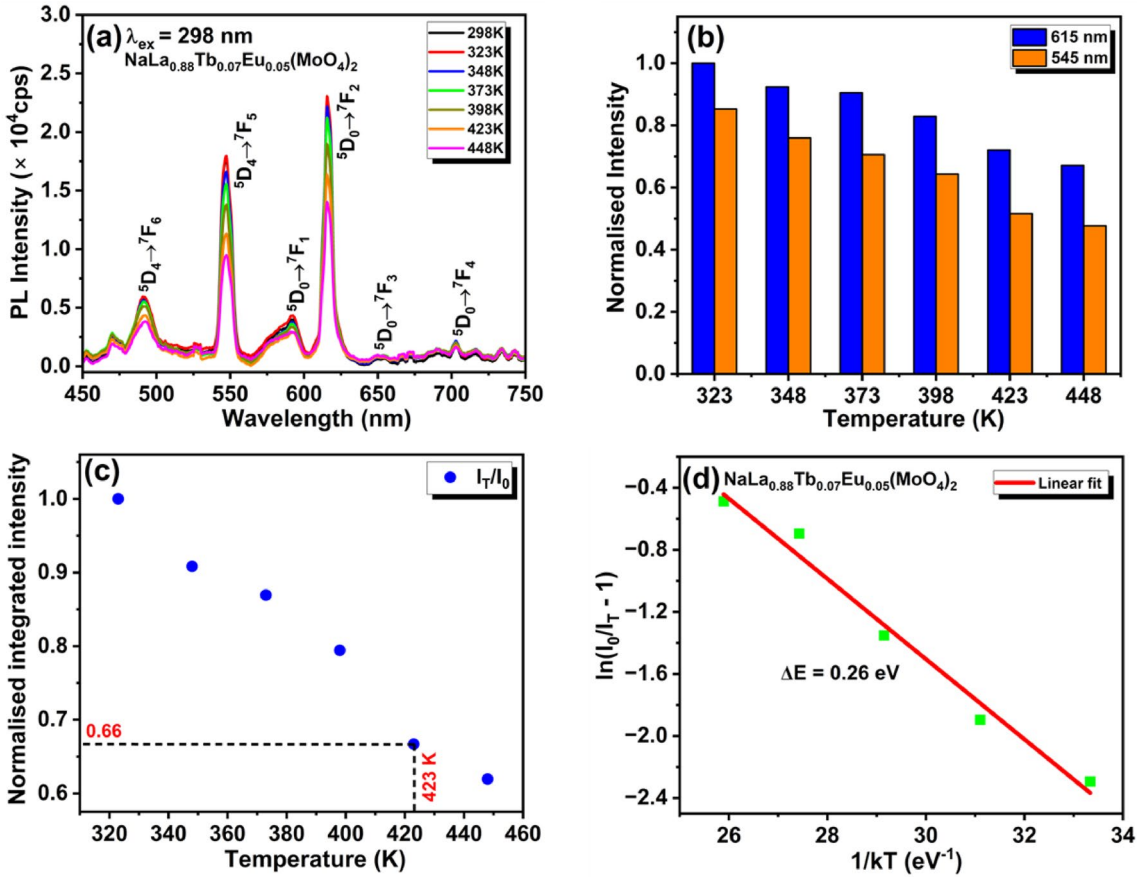


Fig. 13 **a** TDPL spectra of $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$ phosphor excited at 298 nm, **b** Bar diagram representing the normalized intensity variation as a function of temperature for two major peaks of Eu^{3+} and Tb^{3+}. **c** Normalized integral intensity variation as a function of temperature. **b** $\ln\left(\frac{I_0}{I_T}-1\right)$ *vs* 1/kT plot for determination of activation energy

The activation energy (ΔE) is another important parameter to evaluate the thermal stability of the phosphor obtained from its TDPL analysis. The ΔE value indicates the amount of thermal energy required to initiate non-radiative relaxation processes that compete with the radiative emission of the phosphor. The following Arrhenius equation can be employed to estimate the activation energy

$$I_T = \frac{I_0}{I + Ae^{(-\Delta E/kT)}}, \tag{10}$$

where A is the constant frequency factor, k represents the Boltzmann constant (8.629×10^{-5} eV/K) and the emission intensities at temperature 298 K and T are represented by I_0 and I_T, respectively. The linearly fitted $\ln\left(\frac{I_0}{I_T}-1\right)$ *vs* $1/kT$ plot for $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ is depicted in Fig. 12d. The ΔE is represented by the slope of the fitting line, which is 0.38 eV for $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ phosphor. The high value of ΔE is indicative that the phosphor has good thermal stability. The obtained ΔE of about 0.38 eV is relatively high, indicating the good thermal stability of the $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ phosphor. High ΔE also suggests a reasonable resistance to thermal quenching, allowing the material to retain its luminescent properties at elevated temperatures, which is beneficial for applications in high-temperature environments. The ΔE for $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$ phosphor is also evaluated. The slope of the linearly fitted line shown in Fig. 13d represents the ΔE of the $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$ phosphor, which is about 0.26 eV. The good thermal stability of the prepared phosphors is attributed to many factors. Firstly, the presence of strong Mo–O covalent bonds within the MoO_4^{2-} tetrahedra makes the structure more resilient

to high temperatures, helping to reduce thermal quenching. Secondly, it is observed that with doping the lattice parameters and the overall unit cell volume decrease, making the crystal structure denser. The denser crystal structure improves structural stability, which is advantageous for high-temperature applications, as it can help reduce lattice vibrations and maintain the integrity of the material under thermal stress.

3.7 Temperature sensing

The $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$ phosphor was studied for the temperature-sensing application as well. The variation in the intensity of 545 nm (Tb^{3+}) and 615 nm (Eu^{3+}) peaks with rising temperature is not similar. Therefore, dual-mode luminescence thermometry is evaluated by analysing I_{615}/I_{545} intensity ratio. The following equation can be employed to determine the FIR of the two peaks:

$$FIR = \frac{I_1}{I_2} = A + C\exp\left(-\frac{\Delta E}{kT}\right), \tag{11}$$

where I_1 represents the integrated peak intensity at 615 nm, I_2 denotes the integrated peak intensity at 545 nm, k is the Boltzmann constant, ΔE represents the effective energy gap of Tb^{3+} and Eu^{3+}, the proportional and offset parameters are denoted by A and C, respectively. The variation of FIR *vs* T is shown in the inset of Fig. 14a. The ln(FIR) variation with 1/T is presented in Fig. 14b. The slope of ln(FIR) *vs* 1/T plot gives the value of ΔE/k, which is equal to 157.45 eV.

The thermal sensitivity can be determined by calculating the absolute and relative sensitivity values by following two equations:

$$S_A = \frac{d(FIR)}{dT} = \frac{\Delta E}{KT^2} \times FIR \tag{12}$$

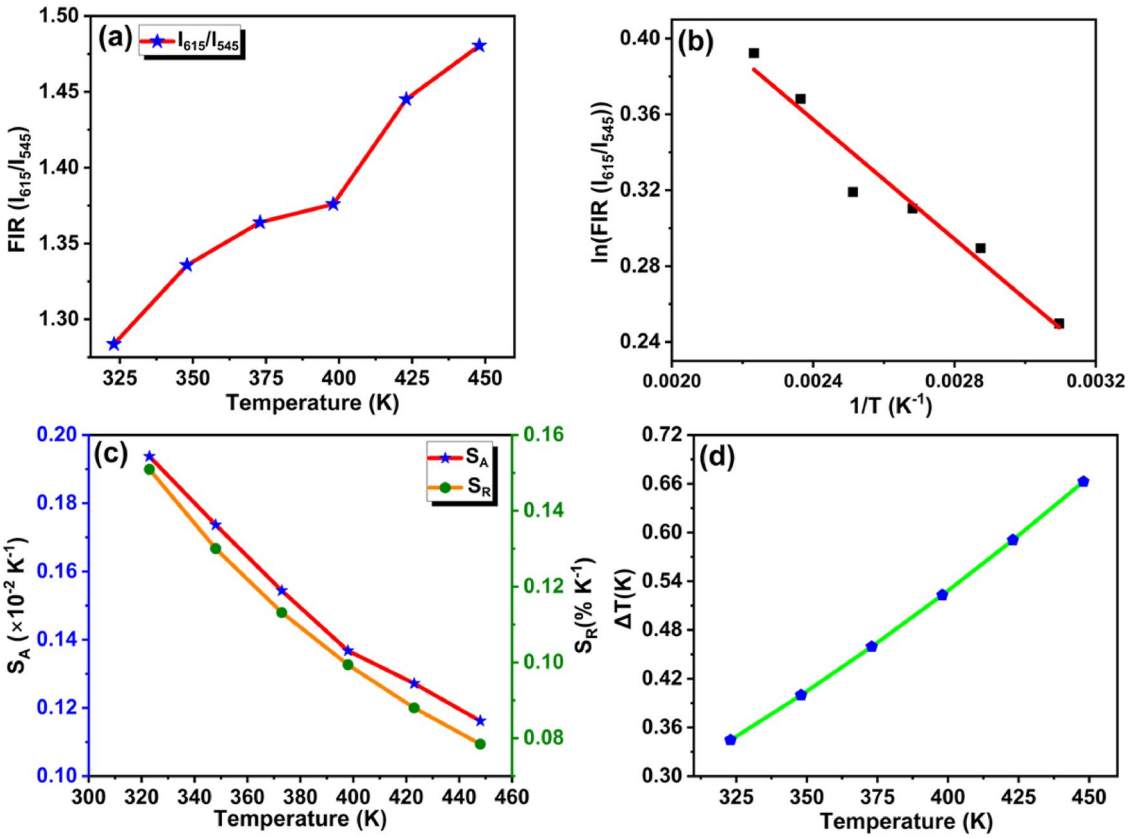


Fig. 14 **a** FIR level as a function of temperature, **b** ln(FIR) as a function of 1/T, **c** absolute sensitivity (S_A), and relative sensitivity (S_R) *vs* temperature, and **d** Temperature uncertainty (ΔT) variation with temperature for I_{615}/I_{545} intensity ratio, obtained after many experiments

$$S_R = \frac{S_A}{FIR} = \frac{1}{FIR} \times \frac{d(FIR)}{dT} = \frac{\Delta E}{KT^2}, \quad (13)$$

where S_A and S_R represent the absolute and relative sensitivities. The dependence of S_A and S_R on temperature for I_{615}/I_{545} is shown in Fig. 14c. The S_A and S_R at 323 K are about 0.19×10^{-2} K^{-1} and 0.15% K^{-1}, respectively. The obtained S_A and S_R values are relatively good for the temperature-sensing application.

Another important parameter that controls the material applicability for the temperature-sensing application is temperature uncertainty. The temperature uncertainty (ΔT) can be calculated from the following expression:

$$\Delta T = \frac{1}{S_R} \frac{\Delta FIR}{FIR}, \quad (14)$$

where standard deviation in the *FIR* value is denoted by ΔFIR. From Fig. 14d, it is deduced that the $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$ phosphor has a temperature measurement uncertainty of ~ 0.7 K, which is also superior to several temperature-sensing materials. Consequently, the relative sensitivity and temperature uncertainty measurement validate the usefulness of the phosphor as an excellent contactless temperature-sensing medium.

3.8 Photometric characterization

The Commission Internationale de I'Eclairage (CIE) chromaticity diagram with the points related to the PL emission spectra of $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$ phosphors under the excitation at 298 nm is displayed in Fig. 15a and corresponding point coordinates are

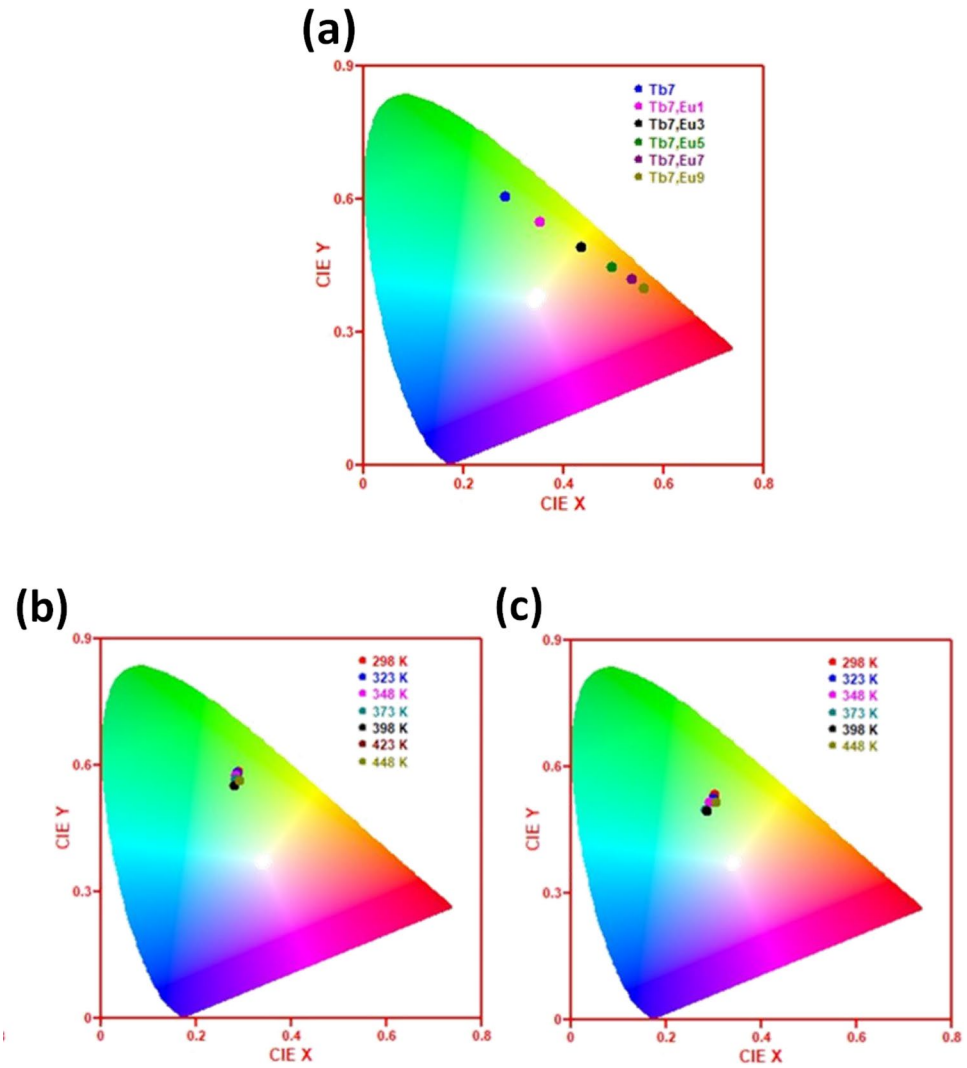


Fig. 15 **a** CIE chromaticity diagram of $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$ phosphor excited at $\lambda_{ex} = 298$ nm. CIE diagram for **b** $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ and **c** $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$ phosphor as a function of temperature

provided in Table S2 [41]. As the Eu^{3+} concentration increases, the emission colour shifts from green (0.2815, 0.6044) to red (0.5613, 0.4003), as seen in the image. Hence, due to the varied concentration of Eu^{3+} ions, the phosphor shows a range of colourful hues. The variation in colour coordinates of $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ and $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$ phosphors with rising temperature is shown in Fig. 15b, c. It is observed that, with a temperature rise, the colour coordinates of the $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ and $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$ phosphors remain nearly unchanged. This result is significant since a steady emission colour at operating temperatures is preferred for lighting applications.

The CCT value of the phosphor can be evaluated from CIE colour coordinates (x, y) using McCamy's following relation [69]:

$$CCT = -437n^{(3)} + 3601n^2 - 6861n + 5514.31 \quad (15)$$

and

$$n = \frac{(x - x_e)}{(y - y_e)}, \quad (16)$$

where ($x_e = 0.3320$, $y_e = 0.1858$) are the epicentre colour coordinates. The computed CCT values for $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$ phosphor under the excitation at 298 nm were found to be in the 1705 to 6395 K range and are listed in Table S3. Typical CCT values less than 5000 K are classified as warm white light sources for domestic applications. Therefore, these materials are appropriate for manufacturing colour display devices and solid-state lighting applications. At the same time, the CCT values for $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ and $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$ phosphors with rising temperature lie above 5000 K and they are listed in Table S5 and S6, respectively. Typically, a CCT value above 5000 K is considered a cold light, which can be used in commercial applications.

In addition, the $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ phosphor PVA composite film was fabricated and is presented in Fig. 16a, b. As seen in Fig. 16a, the film is white in daylight, but when exposed to UV light ($\lambda_{ex} = 254$ nm), it emits a green colour with the flexible behaviours displayed in Fig. 16b. This finding indicated that the $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ phosphor PVA film is an appropriate option for the flexible display device application.

4 Conclusion

In summary, the $NaLa(MoO_4)_2$-based phosphors co-doped by Tb^{3+} and Eu^{3+} ions were synthesized and their crystal structure were verified by the Rietveld refinement analysis. The photoluminescence (PL) emission of the $NaLa_{1-x}Tb_x(MoO_4)_2$ samples was monitored at two distinct excitation wavelengths of 298 and 486 nm. The optimal doping concentration of

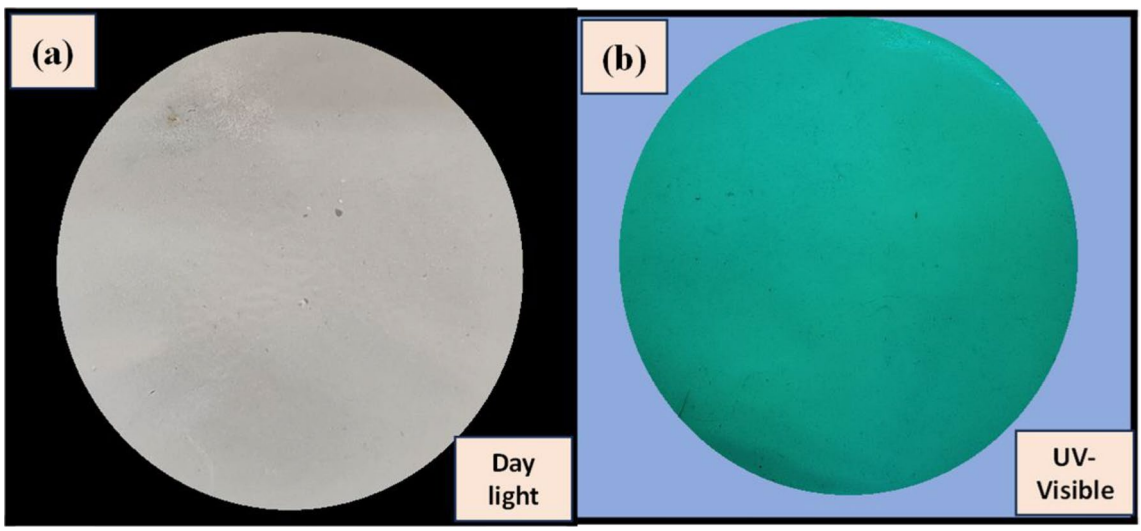


Fig. 16 PVA composite film of the $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ phosphor **a** under daylight and **b** UV light ($\lambda_{ex} = 254$ nm)

Tb^{3+} in the PL is 15 mol%. Furthermore, it is observed that the photoluminescence emission intensity of the $NaLa_{0.93-y}Tb_{0.07}Eu_y(MoO_4)_2$ samples can be adjusted within the green to red range. This phenomenon indicates the presence of energy transfer from the Tb^{3+} to Eu^{3+} ion in the $NaLa(MoO_4)_2$ host matrix, and it can be attributed to quadrupole–quadrupole interaction. The determination of the critical distance was found to be 14.7 Å. Moreover, energy transfer efficiency was determined using the PL intensity technique and the lifetime method, and the values are 68.65 and 48.20%, respectively. Furthermore, this study investigates the changes in photoluminescence (PL) intensity as a function of temperature for two phosphor materials: $NaLa_{0.85}Tb_{0.15}(MoO_4)_2$ and $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$. Based on the FIR mode, the S_A and S_R at 323 K are about 0.19×10^{-2} K^{-1} and 0.15% K^{-1}, respectively, for $NaLa_{0.88}Tb_{0.07}Eu_{0.05}(MoO_4)_2$ phosphor. These values indicate that the phosphor demonstrates favourable characteristics for its potential utilization in temperature-sensing applications. The chromaticity diagram developed by the International Commission on Illumination (CIE) has shown the utility of green to red-emitting phosphors in white light-emitting diode (WLED) applications.

Acknowledgements

The authors acknowledge the Chemical Science Division, Indian Institute of Science, Bangalore, for extending the instrumental facility.

Author contributions

Sonali Tomar: Synthesis, characterization, conceptualization, data analysis and paper writing; Vaibhav Chauhan: data analysis, reviewing and paper writing; M. Jayasimhadri: Characterization; C. Shivakumara: Supervision, Conceptualization, Data analysis, reviewing and editing.

Funding

The authors have not disclosed any funding.

Data availability

The datasets generated during and analysed during the current study are not publicly available but are available from the corresponding author on reasonable request.

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.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10854-024-14109-8.

References

1. N.C. George, A.J. Pell, G. Dantelle, K. Page, A. Llobet, M. Balasubramanian, G. Pintacuda, B.F. Chmelka, R. Seshadri, Chem. Mater. **25**, 3979 (2013)
2. A.G. Bispo-Jr, L.F. Saraiva, S.A.M. Lima, A.M. Pires, M.R. Davolos, J. Lumin. **237**, 118167 (2021)
3. Z. Xia, Y. Zhang, M.S. Molokeev, V.V. Atuchin, J. Phys. Chem. C **117**, 20847 (2013)
4. H. Ji, L. Wang, M.S. Molokeev, N. Hirosaki, R. Xie, Z. Huang, Z. Xia, O.M. Ten Kate, L. Liu, V.V. Atuchin, J. Mater. Chem. C Mater. **4**, 6855 (2016)
5. Y. Wang, H. Wu, W. Hu, J.F. Stoddart, Y. Wang, H. Wu, J.F. Stoddart, W. Hu, Adv. Mater. **34**, 2105405 (2022)
6. L. Gu, H. Wu, H. Ma, W. Ye, W. Jia, H. Wang, H. Chen, N. Zhang, D. Wang, C. Qian, Z. An, W. Huang, Y. Zhao, Nat. Commun. (2020). https://doi.org/10.1038/s41467-020-14792-1
7. X. Ma, P. Feng, Y. Wang, S. Ding, S. Tian, Y. Wang, J. Mater. Chem. C Mater. **10**, 1105 (2022)
8. Y. Zhao, M. Yu, C. Liu, S. Li, Z. Li, F. Jiang, L. Chen, M. Hong, J. Mater. Chem. C Mater. **9**, 2890 (2021)
9. H. Ji, Z. Huang, Z. Xia, M.S. Molokeev, V.V. Atuchin, S. Huang, Inorg. Chem. **53**, 11119 (2014)
10. M. Tan, B. Del Rosal, Y. Zhang, E. Martín Rodríguez, J. Hu, Z. Zhou, R. Fan, D.H. Ortgies, N. Fernández, I. Chaves-Coira, Á. Núñez, D. Jaque, G. Chen, Nanoscale **10**, 17771 (2018)

11. E. Carrasco, B. Del Rosal, F. Sanz-Rodríguez, Á.J. De La Fuente, P.H. Gonzalez, U. Rocha, K.U. Kumar, C. Jacinto, J.G. Solé, D. Jaque, Adv. Funct. Mater. **25**, 615 (2015)
12. S.K. Karunakaran, G.M. Arumugam, W. Yang, S. Ge, S.N. Khan, X. Lin, G. Yang, ACS Sustain. Chem. Eng. **9**, 1035 (2021)
13. Y. Kobayashi, E.H. Sekiya, K. Saito, R. Nishimura, K. Ichii, T. Araki, J. Lightwave Technol. **36**, 2723 (2018)
14. Q.Q. Zhu, Y. Meng, H. Zhang, S. Li, L. Wang, R.J. Xie, ACS Appl. Electron. Mater. **2**, 2644 (2020)
15. C. Feldmann, T. Jüstel, C.R. Ronda, P.J. Schmidt, Adv. Funct. Mater. **13**, 511 (2003)
16. P. Dixit, V. Chauhan, K. Pandey, P.C. Pandey, Mater. Chem. Phys. **305**, 127913 (2023)
17. P. Dixit, V. Chauhan, P. Kumar, P.C. Pandey, J. Lumin. (2020). https://doi.org/10.1016/j.jlumin.2020.117240
18. H. Ji, Z. Huang, Z. Xia, M.S. Molokeev, X. Jiang, Z. Lin, V.V. Atuchin, Dalton Trans. **44**, 7679 (2015)
19. H. Kim, J. Kim, S. Lim, K. Park, J. Nanosci. Nanotechnol. **16**, 1827 (2016)
20. P.K. Pandey, P. Dixit, V. Chauhan, P.C. Pandey, J. Alloys Compd. **952**, 169911 (2023)
21. L.S. Archana, D.N. Rajendran, J. Cyriac, J. Lumin. **243**, 118679 (2022)
22. Y. Gao, Y. Cheng, F. Huang, H. Lin, J. Xu, Y. Wang, J. Alloys Compd. **735**, 1546 (2018)
23. D. Verma, R.P. Patel, M.L. Verma, Mater. Sci.- Poland **37**, 55 (2019)
24. D. Song, C. Guo, J. Zhao, H. Suo, X. Zhao, X. Zhou, G. Liu, Ceram. Int. **42**, 12988 (2016)
25. K. Panigrahi, S. Saha, S. Sain, R. Chatterjee, A. Das, U.K. Ghorai, N. Sankar Das, K.K. Chattopadhyay, Dalton Trans. **47**, 12228 (2018)
26. V. Chauhan, P.K. Pandey, P. Dixit, P. Deshmukh, S. Satapathy, P.C. Pandey, J. Lumin. **248**, 118994 (2022)
27. B.P. Singh, A.K. Parchur, R.S. Ningthoujam, A.A. Ansari, P. Singh, S.B. Rai, Dalton Trans. **43**, 4770 (2014)
28. M. Prasad, V.K. Rai, Mater. Res. Bull. **160**, 112116 (2023)
29. H.S. Kim, K.I. Machida, T. Horikawa, H. Hanzawa, J. Alloys Compd. **633**, 97 (2015)
30. R.J. Xie, N. Hirosaki, Sci. Technol. Adv. Mater. **8**, 588 (2007)
31. V.V. Atuchin, N.F. Beisel, E.N. Galashov, E.M. Mandrik, M.S. Molokeev, A.P. Yelisseyev, A.A. Yusuf, Z. Xia, ACS Appl. Mater. Interfaces **7**, 26235 (2015)
32. J.L. Leaño, S.Y. Lin, A. Lazarowska, S. Mahlik, M. Grinberg, C. Liang, W. Zhou, M.S. Molokeev, V.V. Atuchin, Y.T. Tsai, C.C. Lin, H.S. Sheu, R.S. Liu, Chem. Mater. **28**, 6822 (2016)
33. G. Li, C.C. Lin, W.T. Chen, M.S. Molokeev, V.V. Atuchin, C.Y. Chiang, W. Zhou, C.W. Wang, W.H. Li, H.S. Sheu, T.S. Chan, C. Ma, R.S. Liu, Chem. Mater. **26**, 2991 (2014)
34. P. Dixit, P.K. Pandey, V. Chauhan, P. Deshmukh, S. Satapathy, P.C. Pandey, Methods Appl Fluoresc **10**, 044003 (2022)
35. Sonali, V. Chauhan, P.C. Pandey, C. Shivakumara, ACS Appl. Opt. Mater. **2**, 41 (2024)
36. V.V. Atuchin, A.S. Aleksandrovsky, O.D. Chimitova, C.P. Diao, T.A. Gavrilova, V.G. Kesler, M.S. Molokeev, A.S. Krylov, B.G. Bazarov, J.G. Bazarova, Z. Lin, Dalton Trans. **44**, 1805 (2015)
37. L. Peng, Q. Meng, W. Sun, Ceram. Int. **45**, 20656 (2019)
38. G. Li, L. Li, M. Li, W. Bao, Y. Song, S. Gan, H. Zou, X. Xu, J. Alloys Compd. **550**, 1–8 (2012)
39. Y. Zhai, J. Li, X. Li, Y. Dong, Y. Wang, S. Song, J. Solgel. Sci. Technol. **2**, 3 (2015)
40. S.B. Stevens, C.A. Morrison, Phys. Rev. B **43**, 7386 (1991)
41. G. Sanjay, Sonali, G.P. Abhilash, S. Joshi, C. Shivakumara, Ceram. Int. **49**, 18219 (2023)
42. R. Gopal, J. Manam, Ceram. Int. **48**, 30724 (2022)
43. S. Tomar, N.K. Mishra, V. Chauhan, K. Kumar, C. Shivakumara, J. Phys. Chem. C (2024). https://doi.org/10.1021/acs.jpcc.4c01353
44. Sonali, C. Shivakumara, Methods Appl. Fluoresc. **11**, 024001 (2023)
45. L. Jiang, Z. Wang, H. Chen, Y. Chen, P. Chen, Z. Xu, J. Alloys Compd. **734**, 179 (2018)
46. C. D. Wagner, *NIST X-Ray Photoelectron Spectroscopy (XPS) Database* (National Bureau of Standards, Gaithersburg, MD, 1990).
47. R. Krishnan, J Thirumalai, I. B. Shameem Banu, and R Chandramohan, (n.d.)
48. C.V. Ramana, R.S. Vemuri, V.V. Kaichev, V.A. Kochubey, A.A. Saraev, V.V. Atuchin, ACS Appl. Mater. Interfaces **3**, 4370 (2011)
49. V.V. Atuchin, T.A. Gavrilova, J.C. Grivel, V.G. Kesler, J. Phys. D Appl. Phys. (2009). https://doi.org/10.1088/0022-3727/42/3/035305
50. C.V. Ramana, V.V. Atuchin, L.D. Pokrovsky, U. Becker, C.M. Julien, J. Vacuum Sci. Technol. A: Vacuum Surf. Films **25**, 1166 (2007)
51. V. Chauhan, P. Dixit, P.K. Pandey, S. Chaturvedi, P.C. Pandey, J. Phys. Chem. C **127**, 19159 (2023)
52. V. Chauhan, P. Dixit, P.C. Pandey, J. Rare Earths **39**, 1336 (2021)
53. R. Gopal, A. Kumar, J. Manam, Mater. Chem. Phys. **272**, 124960 (2021)

54. D.L. Wood, B.L. Aboxatoxies, M. Hill, T. Jersey, J. Tauc, Phys. Rev. B (1972). https://doi.org/10.1103/PhysRevB.5.3144
55. X. Ding, G. Zhu, W. Geng, Q. Wang, Y. Wang, Inorg. Chem. **55**, 154 (2016)
56. Y. Hua, S. Khaja Hussain, J.S. Yu, New J. Chem. **43**, 10645 (2019)
57. V.V. Atuchin, A.S. Aleksandrovsky, O.D. Chimitova, A.S. Krylov, M.S. Molokeev, B.G. Bazarov, J.G. Bazarova, Z. Xia, Opt Mater (Amst) **36**, 1631 (2014)
58. N. Jain, B.P. Singh, R.K. Singh, J. Singh, R.A. Singh, J. Lumin. **188**, 504 (2017)
59. V.V. Atuchin, A.S. Aleksandrovsky, B.G. Bazarov, J.G. Bazarova, O.D. Chimitova, Y.G. Denisenko, T.A. Gavrilova, A.S. Krylov, E.A. Maximovskiy, M.S. Molokeev, A.S. Oreshonkov, A.M. Pugachev, N.V. Surovtsev, J. Alloys Compd. **785**, 692 (2019)
60. V.V. Atuchin, A.S. Aleksandrovsky, O.D. Chimitova, T.A. Gavrilova, A.S. Krylov, M.S. Molokeev, A.S. Oreshonkov, B.G. Bazarov, J.G. Bazarova, J. Phys. Chem. C **118**, 15404 (2014)
61. Y. Feng, J. Huang, C. Li, G. Hu, J. Liu, X. Yu, J. Alloys Compd. **706**, 478 (2017)
62. P. Halappa, H.M. Rajashekar, C. Shivakumara, J. Alloys Compd. **785**, 169 (2019)
63. W. Liu, J. Zhang, Z. Hao, G. Xiang, L.L. Zhang, X. Zhang, G. Pan, Y. Luo, H. Zhao, H. Wu, J. Mater. Sci. Mater. Electron. **28**, 8017 (2017)
64. P.I. Paulose, G. Jose, V. Thomas, N.V. Unnikrishnan, M.K.R. Warrier, J. Phys. Chem. Solids **64**, 841 (2003)
65. L.G. Van Uitert, J. Electrochem. Soc. **114**, 1048 (1967)
66. G. Blasse, Phys. Lett. **28**, 444 (1968)
67. D.L. Dexter, J. Chem. Phys. **21**, 836 (1953)
68. G. Li, L. Li, M. Li, W. Bao, Y. Song, S. Gan, H. Zou, X. Xu, J. Alloys Compd. **550**, 1 (2013)
69. C.S. McCamy, Color. Res. Appl. **17**, 142 (1992)

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

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

Air Quality, Atmosphere & Health
https://doi.org/10.1007/s11869-024-01689-x

Traffic-induced nanoparticle emissions and associated respiratory risk analysis using measurements conducted in a roadside environment

Kanagaraj Rajagopal[1] · **S. Ramachandran**[2] · **Rajeev Kumar Mishra**[1]

Received: 1 May 2024 / Accepted: 10 December 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract
Air quality is a major concern for healthy living as particles present in air are directly associated with human health, especially in urban regions that are more vulnerable to different types of atmospheric pollutants. In this study, for the first time, the concentration of atmospheric nanoparticles in an urban roadside environment are analysed, and the deposition of these nanoparticles in human respiratory tract are quantified using Multiple-path particle dosimetry model for different age groups over Delhi. The measured seasonal mean nanoparticle concentration ranges from 24,000 to 41,000 particles cm^{-3}. The seasonal sequence of deposition of nanoparticles is winter > monsoon > summer > autumn > spring. The deposition of nanoparticle in adults is 30 to 40% higher than in children and infants, and further, the deposition is higher in the alveolar region than in the bronchiole and trachea regions. About 90% of the particles get deposited in the alveolar regions, 6 to 8% in the bronchiole region, and 2% in the trachea region. The estimated deposition of nanoparticles for an individual working 8 h a day in the near road conditions is 338 µg/year in Delhi. The deposition increases almost linearly as a function of time, and is 3 times higher (1016 µg/year) for a person residing near the road throughout the day (24 h). The deposition fraction of particles ranges between 0.05 and 0.10 µg/day in the alveolar region, < 0.05 µg/day in the bronchiole region, and lies between 0.02 and 0.04 µg/day in the trachea. The nanoparticles deposited in the respiratory system can lead to the development of various diseases such as asthma, chronic obstructive pulmonary disease, and can lead to carcinogenicity. The new insights from the study gained on the deposition of nanoparticles in humans residing near roadside conditions are crucial to estimate the human health risk potential, and to formulate mitigation measures for exposure reduction which can result in a better and sustainable future.

Keywords Air quality · Exposure · Health impacts · Urban aerosols · Nanoparticles · Sustainability

Introduction

Urban regions experience a complex mixture of nanoparticle and aerosol emissions from a variety of sources. The major emission sources include activities such as construction, industrial, waste burning, biomass burning, traffic, and road dust (Kulkarni et al. 2020; Mishra, 2023; Mohan and Mishra 2022; Sengupta et al. 2022). All these activities emit different kinds of aerosols which result in degrading air quality (Kanawade et al. 2020; Shivani, Gadi et al. 2019). Air quality is a major parameter for determining the quality and standard of living (Kumar and Mishra 2018; Nagpure et al. 2014; Oliveira et al. 2019; Rizwan et al. 2013). Particulate matter and gaseous pollutants are among the primary air pollutants in urban regions (Sharma et al. 2024). This classification is based on the physical properties of the pollutants. Recent advancements in air pollutant measurement techniques paved the way for measuring and analyzing particles such as nano pollutants in the atmosphere, which have a higher impact on human health (Ezz et al. 2015; Gupta et al. 2020; Mosonik et al. 2019). Nanoparticles in the atmosphere are generated through various natural and anthropogenic processes. In urban regions, nanoparticle emissions occur mainly from anthropogenic sources, especially vehicular ones (Buonanno et al. 2014; Joshi et al. 2016). The vehicular sources emit different size nanoparticles based

✉ Rajeev Kumar Mishra
rajeevkumarmishra@dtu.ac.in

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

[2] Space and Atmospheric Sciences Division, Physical Research Laboratory, Ahmedabad 380009, India

Published online: 24 December 2024

on the type of fuel used, combustion technique, and from break and tyre wear emissions (Gani et al. 2021; Kwon et al. 2020). The nanoparticles in the atmosphere are measured using the electrical mobility technique in terms of number concentrations (Rajagopal et al. 2023). In urban regions, the concentration of the nanoparticles varies significantly based on the sources (Rajagopal et al. 2024a, b).

The physicochemical characteristics of particulate matter (PM) including nanoparticles which impact human health (carcinogenicity, inflammation, lung and respiratory diseases) are their (a) physical: size - which decides the deposition of pollutants in different regions of the respiratory system and (b) chemical: composition – the species type (black carbon, organics, sulfate, nitrates, dust, sea salt, smoke etc.) which is based on their emission sources. Further, their health impacts can vary depending on the environmental setting, meteorological conditions (relative humidity, winds, planetary boundary layer height, rainfall, and solar radiation), and atmospheric dynamics (horizontal and vertical transport), and as a function of season. Unfavorable weather conditions (such as low wind speed, higher relative humidity, lower planetary boundary layer/mixing height, and less/no rainfall) may further worsen the pollution scenario and associated health effects, in contrast, favorable weather conditions (such as high wind speed, low relative humidity, higher planetary boundary layer/mixing height, and rainfall) will favor an early dispersion of pollutants. Thus, the dispersion of nanoparticles in urban space depends on windspeed, ventilation, mixing height and street canyon effect. The nanoparticles emitted from automobile exhaust are smaller in size but their concentrations are high which then undergo coagulation and condensation once emitted into the atmosphere (Mohan et al. 2024). The secondary process (condensation and coagulation) increases the size of the particles and reduces the number concentration.

Traffic-related air pollution and associated health impacts are gaining global attention from government authorities, researchers, and industrial sectors. The nanoparticles which are smaller in size and have higher surface area can pose serious health issues (Mohan et al. 2024). Several studies investigated the impact of particulate matter on human health and found that air pollution is responsible for asthma, lung cancer, chronic obstructive pulmonary diseases, and cardiovascular diseases (Oliveira et al. 2017; Zhu et al. 2021), however, the scientific studies on deposition potential and fraction of nano size pollutants are limited. Due to their smaller size, the particles can escape the defence mechanism in the human body and penetrate deep into the lungs and even through the bloodstream (Chalupa et al. 2004; Rajagopal et al. 2024a). The nanoparticles are also responsible for neurodegenerative diseases like Alzheimer's (Ma et al. 2022). Invitro studies show that air pollutants after exposure can cause significant biological responses such as inflammation, oxidative stress, and genotoxicity (Khan et al. 2022). Therefore, it is vital to analyse the acute/chronic health effects due to the atmospheric nanoparticles exposure. It can be done by estimating the deposited particle concentration in the human lungs, which can be correlated with health effects. Annually the short-term exposure to $PM_{2.5}$ is attributed to about 0.25 million deaths globally. The present study estimates the concentration of the nanoparticles deposited as a function of age and associated breathing frequency from infants to adults in roadside conditions. Multiple Path Particle Dosimetry (MPPD) model, developed by Applied Research Associates (ARA), is a computational model for estimating the deposition of nanoparticles in the human respiratory tract (Manojkumar et al. 2019). The model uses size-segregated data to predict the deposition of the nanoparticles in the human respiratory tract for ideal lung geometries (Vu et al. 2017).

In this study, for the first time, the concentration of atmospheric nanoparticles (10 to 1000 nm) in an urban roadside environment over Delhi, a megalopolis, are analysed, and the deposition of these nanoparticles in human respiratory tract are quantified using Multiple-path particle dosimetry model for different age groups over Delhi. Atmospheric nanoparticles in the above size range were measured over Delhi for a year covering all the five seasons (winter, spring, summer, monsoon and autumn). The novelty of the study is that it estimates the nanoparticle deposition in different seasons for multiple age groups in roadside conditions, a first-of-its-kind analysis in the Indian scenario. The study provides new and comprehensive insights into potential health risks from nanoparticle deposition in the human respiratory system, and can be termed crucial as it provides quantitative and valuable inputs for initiating mitigation measures aimed towards improving air quality and human health.

Study area

The nanoparticle measurements were performed at a major pedestrian road (< 1 m) at the northwest part of Delhi. Delhi is a major urban city in India where air quality is a major concern. The measurement site is a typical representative of roadside microenvironment and commercial, educational and residential areas with different types of roads surrounding it. The study region experiences continuous emissions throughout the year due to transportation sources. The population of the Rohini district, where the study area is located, is ~0.9 million, and undergoes diverse climatic conditions, varying from extreme cold winter to hot summer. The road adjacent to the study location is a state highway on which

two and three-wheelers, motorized and non-motorized rickshaws, heavy vehicles, personal and commercial cars, cranes, trucks, and e-vehicles ply on regular basis. The geography of the urban area and the monitoring location are shown in Fig. 1.

Instrumentation

Nanoparticle measurements

The measurements of nanoparticles were done using a GRIMM Aerosol Technik GmbH & Co. K. G made SMPS + C (scanning mobility particle sizer (SMPS) with butanol-based Condensation unit) (S. No 54031801). The instrument has the capacity to measure nanoparticles from 10 to 1000 nm with different size distribution channels. The atmospheric air is sampled at a flow rate of 0.3 LPM and passed through the neutralizer and Universal-DMA for detection. The DMA, which works on the principle of electrical mobility, is attached with a condensation unit. The charged nanoparticles are passed through a N-butyl alcohol saturation unit. After passing through the condensation unit, charged particles in the sampled air size increase by 10 folds for better visualization. The particles are then sent for magnification in a laser beam chamber that has a photodiode. The detected concentration of nanoparticles is then displayed at a time interval of 7 min for each cycle. The instrument can measure a maximum concentration of about 10^{10} particles per liter, and the concentrations are expressed in particles/cm^3.

Multiple path particle dosimetry model

The Multiple Path Particle Dosimetry model (MPPD) is a globally used computational model for dosimetry analysis due to exposure in human lungs (Asgharian et al. 2022). The model is sophisticated and well-validated, which can give more accurate results. The model predicts the nanoparticles deposited in different regions of the lungs using the lung geometries and assumptions (Deng et al. 2019). The model can predict the deposition of mono (particles of single size bin of particular size), and poly dispersed (particles in different size bins within a particular size range) nanoparticles in the size range of 10 to 10^6 nm (Asgharian et al. 2001). The lung geometry in the model is differentiated into three major regions as head/throat, tracheobronchial, and alveolar/pulmonary. The input parameters of the model include the concentration, density and size of the particles, which are used in estimating the concentration of nanoparticles deposited in the lungs for estimating the deposition dosage. In the present study, we have used the recent version of MPPD model (version 3.04) which can handle up to 4 overlapping lognormal distributions at a time. The model

Fig. 1 Geographical location of the study area and the monitoring station

provides options for single- and multiple-path air flows and resembles the most realistic human airway tract structure for deposition analysis compared to other computational models, making the model more suitable for use (Manojkumar et al. 2019; Miller et al. 2016). The model mimics the biological functions of the lungs, such as structure, gas exchange ratio and role of defence mechanism.

Methodology

The nanoparticles measured in the 10 to 1000 nm size range using SMPS for a year covering all the seasons of winter, spring, summer, monsoon, and autumn in the roadside microenvironment are utilised for predicting the deposition of nanoparticles. The size distribution analysis is carried out using the GRIMM Nano software, which follows the ISO 15,900 recommendations for particle mobility spectrometers. The deposition dose of the inhaled particle for the specific size range is calculated using Eq. 1 (also referred to as the dose rate), where Dp is the diameter of the particle, and DF is the deposition fraction. The step-wise execution of the model working procedure and the schematic of the model are given in Fig. 2. The basic assumptions in the model as per model recommendations which serve as inputs are given in Fig. 2.

$$Dose\ Rate = \int_{Dp1}^{Dp2} V_E.DF(D_P).n_N^0(D_P).fdlog(D_P) \quad (1)$$

Limitations of the study and statistical analysis

This study uses computational techniques to simulate the deposition of nanoparticles in human lungs and does not represent the concentration measured through any clinical studies. The deposition that occurred in the human respiratory tract may vary from predictions/estimations due to the morphological changes of nanoparticles in the respiratory tract, which is more humid in nature, and as a result hygroscopic growth of the particle inhaled may occur, which are not accounted for in the study. Further, the chemical characteristics of nanoparticles could not be measured and investigated, which could be termed as one of the limitations of the work. The Kruskal-Wallis test, a non-parametric statistical test, was employed to determine the differences, if any, in the seasonal variability of nanoparticle concentrations.

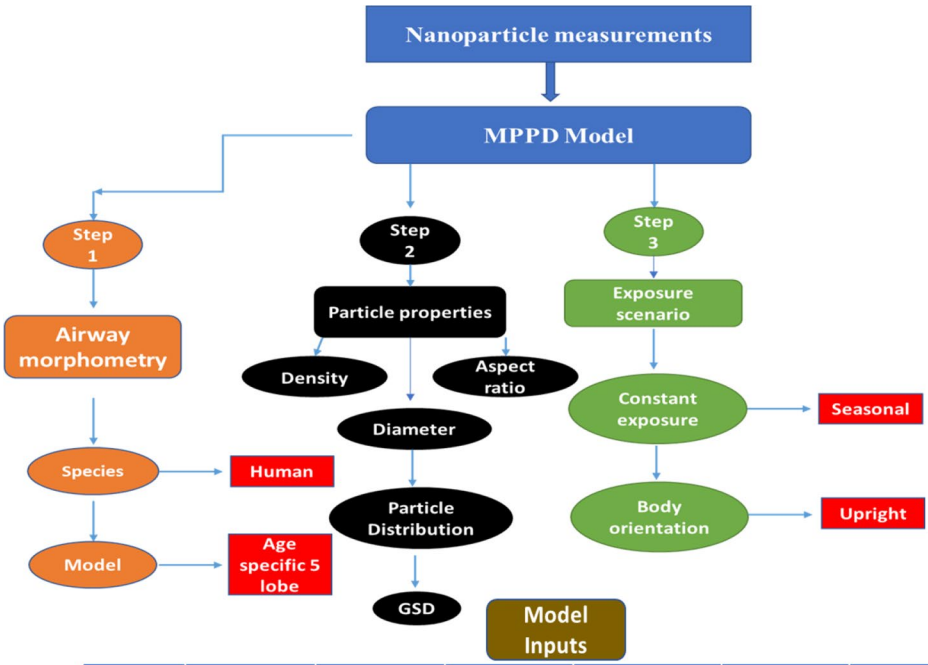


Fig. 2 A schematic of the MPPD model along with working procedure and model inputs. *URT-Upper respiratory tract volume, FRC- Functional residual capacity, TV- Tidal volume, BF-Breathing frequency. Constant values used in the model are Density = 1.0 g/cm^3, Aspect Ratio = 1.0, and Geometric Standard Deviation (GSD) = 1

Model value	Infant(3 months)	Toddler 28 months	Children 8 years	Adolescence 14 years	Adult 18 years	Adult 21 years
URT	2.45	7.92	21.03	30.63	37.38	42.27
FRC	17.97	30.81	501.32	987.56	1159.38	2123.75
TV	30.44	100.1	278.2	388.1	446.7	477.2
BF	39	26	17	16	15	14

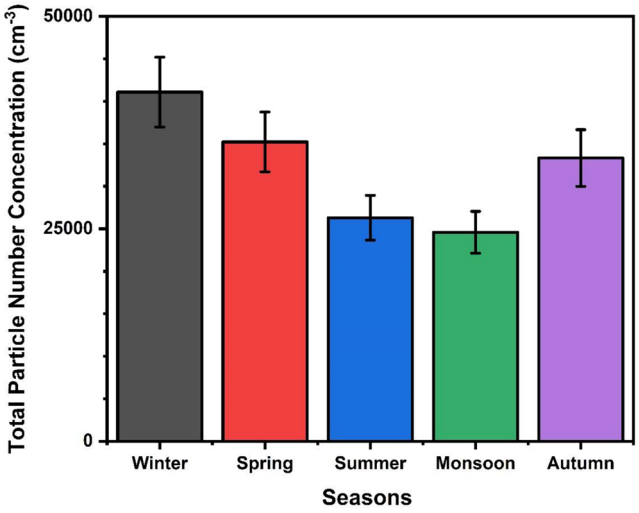


Fig. 3 Seasonal mean total particle concentration of nanoparticles (10 to 1000 nm) over Delhi. Vertical bars correspond to ±1σ (standard deviation from the mean)

Further, the Mann-Whitney U test, a non-parametric statistical test of the null hypothesis, was conducted to analyse the variation of independent seasonal groups. The test revealed statistically significant differences in nanoparticle concentration between different seasons with a p-value < 0.05. Both these tests were performed using the IBM SPSS software.

Results and discussion

Seasonal variation of nanoparticles

The particle number concentration (PNC) of atmospheric nanoparticles in the 10 to 1000 nm size range exhibits significant seasonal variations over Delhi. The total PNC is maximum in winter (4.1×10^4 cm^{-3}) and lowest in monsoon (2.4×10^4 cm^{-3}) (Fig. 3). The seasonal order of total PNC from high to low goes as winter, spring, autumn, summer, and monsoon. In the study the year is divided into 5 seasons that includes the monsoon season; during the monsoon season 77 days of data were available which are used in the study. The seasonal classification in the study follows the classification prescribed by the India Meteorological Department (IMD). The monsoon season is marked by copious amount of rainfall (~730 mm in 2021 monsoon over North-west Delhi where the measurements were conducted) (https://hydro.imd.gov.in/hydrometweb/(S(vo3wtanhjge1rk5500ogip55))/PRODUCTS/Publications/Rainfall%20Statistics%20of%20India%20%202021/Rainfall%20Statistics%20of%20India%202021.pdf), and due to wet scavenging the particle concentration decreases significantly over the study region (Fig. 3), thus, in this study the monsoon season is considered separately. The monsoon rainfall over Delhi, a megalopolis, exhibits significant variations across different zones (districts). The rainfall varied from ~500 mm in West and Northeast Delhi to > 1100 mm over North and Central Delhi. The seasonal changes in nanoparticle concentrations in Delhi occur due to the variation in the intensity of the emission sources and local/regional meteorological factors including rainfall.

The concentration of particles observed during the winter is nearly twice (~2 times) higher than the lowest measured concentration in monsoon. The particle concentration in the study region is quite high and lies in the range of 10^4 cm^{-3} throughout the year (Table 1). Studies show that the penetration depth of nanoparticles is higher than that of fine and coarse particles in the human respiratory system. The nanoparticles cannot be removed through the preventive mechanism of the respiratory system. Higher-sized particles, such as fine and coarse-range particles, are found to be deposited more in the head and tracheobronchial regions of the human respiratory system.

Seasonal particle number and mass distribution

The particle number size distribution (PNSD) obtained in different seasons is analyzed to understand better the transformation of particles from one size regime to the other. In winter particle concentration peaks in the N_{latk} mode (50–100 nm) (Fig. 4a). Smaller particles in the nucleation (10–30 nm) and Aitken modes (30–100 nm) have higher concentrations, whereas the concentration in the accumulation mode (100–1000 nm) is lower during spring (Fig. 4). A linear shift of particles from a smaller to a larger size, i.e., from N_{nuc} to N_{acc} is observed during autumn. The concentration increases from 1×10^4 cm^{-3} (N_{nuc} 10–30 nm) to 3×10^4 cm^{-3} (N_{acc} 100–1000 nm), which occurs gradually. During summer, the concentration decreases from smaller to larger size. The decline in concentration from 2.5×10^4 cm^{-3} to 2×10^4 cm^{-3} is not significant as that of the increase seen in autumn (Fig. 4). Due to the prevalent meteorology in winter,

Table 1 Seasonal characteristics of total particle number concentration (10 to 1000 nm) (N_{total}) over Delhi

Season	Parameter (cm^{-3})	Minimum	Maximum	Mean ± S. D	Median
Winter	N_{total}	2691	147,182	41,089 ± 23,669	35,134
Spring	N_{total}	8694	139,019	35,226 ± 17,909	30,413
Summer	N_{total}	1752	130,433	26,307 ± 13,719	23,702
Monsoon	N_{total}	4337	112,587	24,590 ± 13,256	21,459
Autumn	N_{total}	3774	111,316	33,329 ± 17,883	29,893

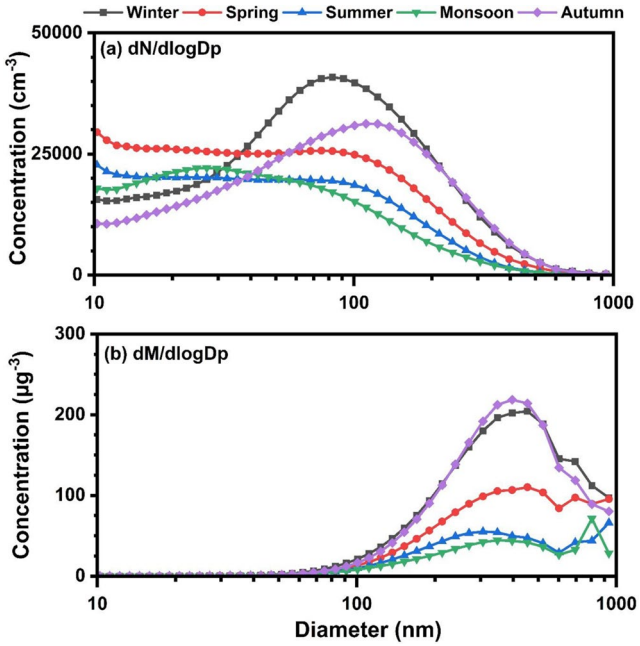


Fig. 4 (**a**) Number and (**b**) mass size distribution of particles measured over Delhi covering the 10 to 1000 nm size range and classified as a function of size into nucleation (10–30 nm), small Aitken (30–50 nm), large Aitken (50–100 nm) and accumulation (100–1000 nm) modes in different seasons (winter, spring, summer, monsoon, and autumn)

apart from the direct primary emissions, other secondary processes, such as condensation and coagulation also play a role in shifting the particle concentration to higher size from smaller size, thus, accounting for the presence of larger size particles in higher concentration during winter (Yadav et al. 2021). A comparative analysis of particle number and mass size distributions (Fig. 4) clearly shows that the smaller size contributes significantly less to particle mass. The particles in the size range of N_{acc} (100 to 1000 nm) contribute a higher proportion to mass concentration.

Quantification of nanoparticle deposition in the human respiratory system

The deposition concentrations estimated by the MPPD model are 1.31×10^{-3} µg/hr during winter with deposition rates of 0.29 µg/hr, followed by spring (1.13×10^{-3} µg/hr, 0.25 µg/hr), summer (8.8×10^{-4} µg/hr, 0.01 µg/hr), monsoon (1.1×10^{-3} µg/hr, 0.02 µg/hr) and autumn (7.88×10^{-4} µg/hr, 0.01 µg/hr), respectively (Fig. 5). The deposition concentration estimates are based on the concentration of particles and exposure time. The length of the seasons varies from 91 days (summer) to 70 to 80 days (winter, monsoon and autumn) and 45 days (spring). An adult who works around 8 h a day near roadside conditions experiences a deposition concentration of around 338 µg/year

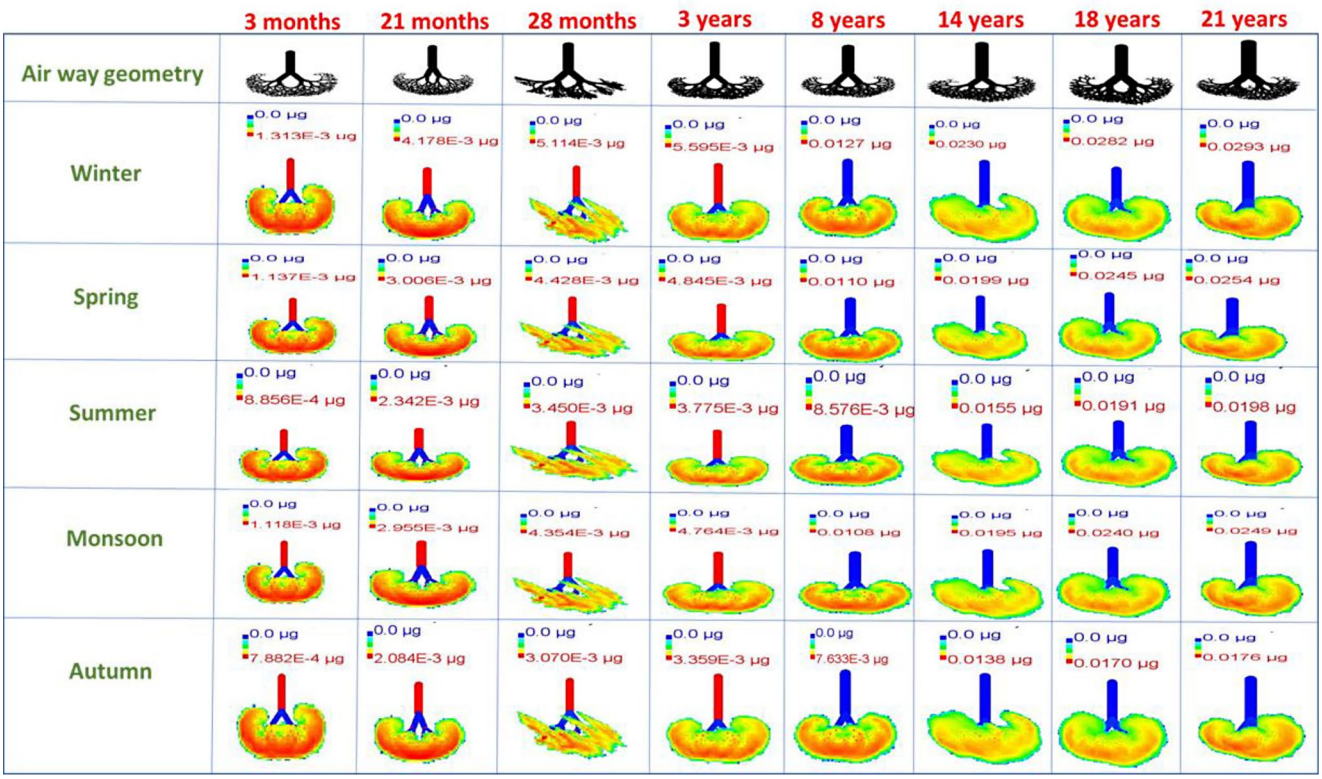


Fig. 5 Estimated deposition of nanoparticles during different seasons (winter, spring, summer, monsoon, autumn) in the study region of Delhi for age groups from 3 months to 21 years using MPPD model

of nanoparticles. Similarly, the adult resident who resides/spends 24 h a day near the road experiences a deposition of nanoparticles > 1000 µg/year. The exposure concentration of particles is 25 times higher than the national annual standards for $PM_{2.5}$ (40 µg/year) and 16 times higher than the PM_{10} (60 µg/year). When compared to the WHO 2021 standard, the exposure concentration is ~200 times higher than the $PM_{2.5}$ (5 µg/year) and 67 times higher than the PM_{10} (15 µg/year) standards, emphasizing the vulnerability of citizens working or living in the vicinity of the road in the study region. It may be noted that the estimated particle deposition only corresponds to nanoparticles and not fine and coarse particles. The exposure estimation shows that the deposition of nanoparticles is found in the head airway tract for children below three years, however, for adults, the deposition is found only in the inner regions of the lungs. The variation in the deposition arises due to the differences in the development of respiratory tract among infants, children and adults. The study region is surrounded by a mixture of educational and residential areas where the population (infants, children and adults) vulnerable to these exposures vary and leads to various health hazards (Balakrishnan et al. 2019; Pandey et al. 2021; Yadav et al. 2022).

Quantification of total mass deposited during different seasons

The estimated mass of nanoparticles deposited in the respiratory system in a day in the road environment varies between 0.02 and 0.40 µg (Fig. 6). In almost all seasons, the bronchiole deposition is less than 0.05 µg/day for different age groups. The alveolar deposition for infants and

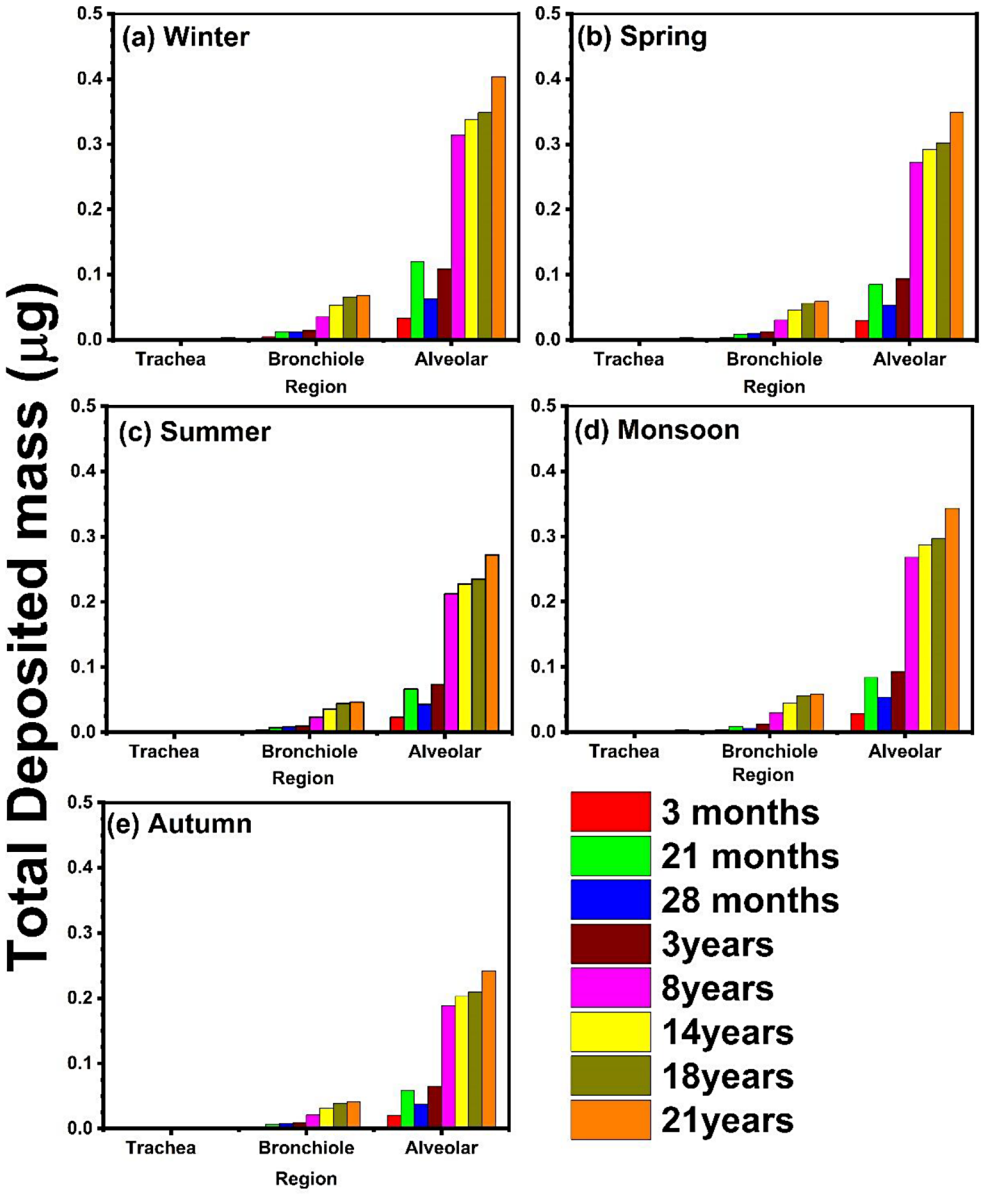


Fig. 6 Estimated total deposited mass in different regions of respiratory system (trachea, bronchiole, and alveolar) during (**a**) winter, (**b**) spring, (**c**) summer, (**d**) monsoon and (**e**) autumn in the study region of Delhi for different age groups ranging from from 3 months to 21 years using MPPD model

kids ranges from 0.05 μg/day to 0.10 μg/day, but for the age groups from 8 years to 21 years, the alveolar deposition is higher, and it varies from 0.30 μg/day to 0.40 μg/day (Fig. 6).

The estimated alveolar deposition for the 8 to 21 years age group is 0.32–0.40 μg/day in winter, 0.25–0.32 μg/day in spring, 0.20–0.27 μg/day in summer, 0.25–0.33 μg/day in monsoon and 0.20–0.25 μg/day in autumn. In winter, the deposition concentration is 2 times higher than in autumn for adults. The average nanoparticle deposition in the alveolar region of a particular individual, irrespective of the age groups residing near the road in different seasons is 26.6 μg/day in winter, 14.6 μg/day in spring, 22 μg/day in summer, 23.1 μg/day in monsoon and 17.1 μg/day in autumn. The seasonal order of deposition is winter > monsoon > summer > autumn > spring. The seasons with more number of days in them account for more exposure periods and vice versa. It is estimated that in a year atleast about 100 μg of nanoparticles are deposited in the alveolar regions alone in adults in the study region. The age wise order of deposition of nanoparticles is adult (18 years and above) > adolescent (14 years) > child (3 to 8 years) > toddler > (21, 28 months) > infant (3 months). Thus, it is clear that in urban areas where most residents are located near the road, the vulnerability to nanoparticles is higher compared to other environmental regimes (rural, semi-urban) irrespective of the exposure time or specified age group.

Quantification of total particle numbers deposited in different seasons

For adults, the daily deposition of nanoparticles in the alveolar regions ranges between 5 and 8×10^5 in different seasons (Fig. 7). The deposition of nanoparticles is estimated to be less in infants (3 months), gradually increases in children, and becomes higher in adults. Less than 2% of particles are deposited in the trachea region and 6–8% in the bronchiole region, whereas the remaining 90% get deposited in the alveolar regions (Fig. 7). Due to its physical properties, most particles get deposited in the alveolar regions. The nanoparticles having lesser mass efficiently penetrate the head and TB regions, so more deposition is found in the alveolar regions (Viitanen et al. 2017). The estimated total nanoparticle deposition in different regions of the human respiratory tract varies based on respiratory parameters such as tidal volume, functional breathing capacity, and breathing frequency. The age group of 21 years and above experiences more nanoparticle deposition than other age groups (Fig. 7). Nanoparticle deposition in the human respiratory tract can cause various health issues such as asthma, leukemia, prostate, and olfactory disorders (Manigrasso and Avino 2012), thus, necessary mitigation measures need to be taken to reduce nanoparticle exposure, especially in urban regions.

Deposition fraction of particles in different regions in the human respiratory tract

Deposition fraction indicates the fraction of inhaled nanoparticles deposited in the respiratory system. The factors that influence the deposition include the mechanism and the airway structure which are based on the mass median diameter of the aerodynamic particles. The analysis shows that the deposition fraction in trachea region is minimal (< 0.01). In the bronchiole region, the deposition fraction is almost the same (~ 0.03) in all the seasons, suggesting that the deposition in the bronchiole region is less influenced by the seasonal variation in concentration and meteorology. The deposition fraction in the alveolar regions varies based on age and seasonal concentration.

The deposition fraction varies between 0.02 and 0.04 in the alveolar regions for different seasons. In winter, the fraction exceeds 0.04 for infants, suggesting that infants are more vulnerable during winter. The deposition in the alveolar region is 4–5 times greater than the deposition in the bronchiole region (Fig. 8). The difference in deposition in alveolar regions for adults is around 30 to 40% higher compared to infants and children. A previous $PM_{2.5}$ concentration-based deposition study also reported that the total deposition of particulate matter is seen more in children compared to other age groups (Manojkumar and Srimuruganandam 2022) consistent with the present study. Several other studies showed that the particles with higher diameter deposit in the head region, and those with lower diameter deposit in the alveolar regions, which are corroborated by the present study.

Conclusions

For the first time, this study estimates the deposition of nanoparticles (size ranging from 10 to 1000 nm) in the human respiratory system using a Multiple Path Particle Dosimetry (MPPD) over the urban megalopolis Delhi. The seasonal mean concentration of total nanoparticle concentration ranges between 24, 000 to 41, 000 particles cm^{-3} over Delhi. The concentration in winter is the highest among all the seasons. The winter concentration is 2-times higher than the autumn concentration. The deposition of nanoparticles ranges between 0.29 - 1.31×10^{-3} μg/hr in winter, followed by spring ($0.25–1.13 \times 10^{-3}$ μg/hr), summer ($0.01 \times 10^{-3} - 1.13 \times 10^{-4}$ μg/hr), monsoon ($0.02–1.1 \times 10^{-3}$ μg/hr) and autumn ($0.01–7.88 \times 10^{-4}$ μg/hr). The analysis reveals that an adult working 8 h a day near the road environment in

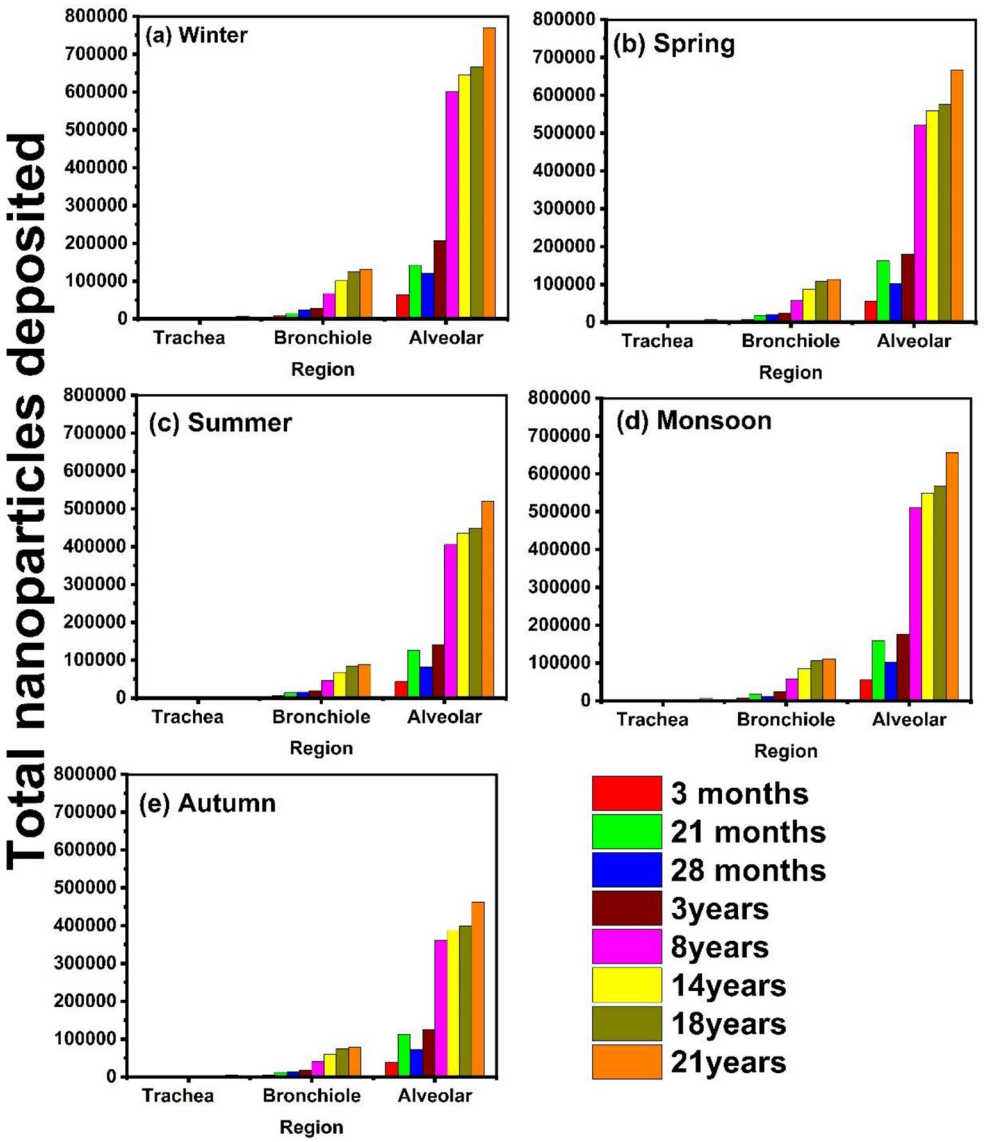


Fig. 7 Total nanoparticle deposition estimated using MPPD model for different age groups in different regions of respiratory system, namely, trachea, bronchiole, and alveolar in (**a**) winter, (**b**) spring, (**c**) summer, (**d**) monsoon and (**e**) autumn) over Delhi

the study region experiences a nanoparticle deposition of 338 µg/year. Further, on a 24-h scale (day), the residents residing near the roadside experience a 3-times higher deposition (> 1000 µg/year of nanoparticles) in the study region, clearly indicating a linearity between the deposition of nanoparticles and the hours of exposure. The daily deposited mass concentration ranges between 0.02 and 0.04 µg in the trachea region, followed by bronchiole < 0.05 µg; in alveolar regions, the amount of particle mass deposited varies between 0.05 and 0.10 µg/day. The rate of deposition in the alveolar region is higher for the age group of 8 to 21 years due to the increased breathing frequency. In terms of deposition rate in different regions of the human respiratory tract, 90% of particles get deposited in the alveolar regions, 6–8% in the bronchiole region, and 2% in the trachea region, highlighting that the deposition of nanoparticles is dominant in the alveolar regions. The deposition fraction of particles ranges between 0.02 and 0.04 in the alveolar region which is 4–5 times higher than in the bronchiole region. Compared to children and infants, adults experience 30 to 40% higher nanoparticle deposition in the respiratory system. It may be noted that the model estimates of nanoparticle deposition may vary in real time on a day to day basis as the deposition depends on lung parameters and humidity. The new quantitative insights gained on the seasonal variation in the deposition of nanoparticles in humans residing near roadside conditions are crucial to estimate the human health risk potential, especially in winter when the nanoparticles and their deposition rates are significantly higher. These findings, with implications for improved air quality and public health, are vital for formulating mitigation measures on a seasonal scale for exposure reduction to different age groups, hitherto unavailable, can lead to a better and sustainable future.

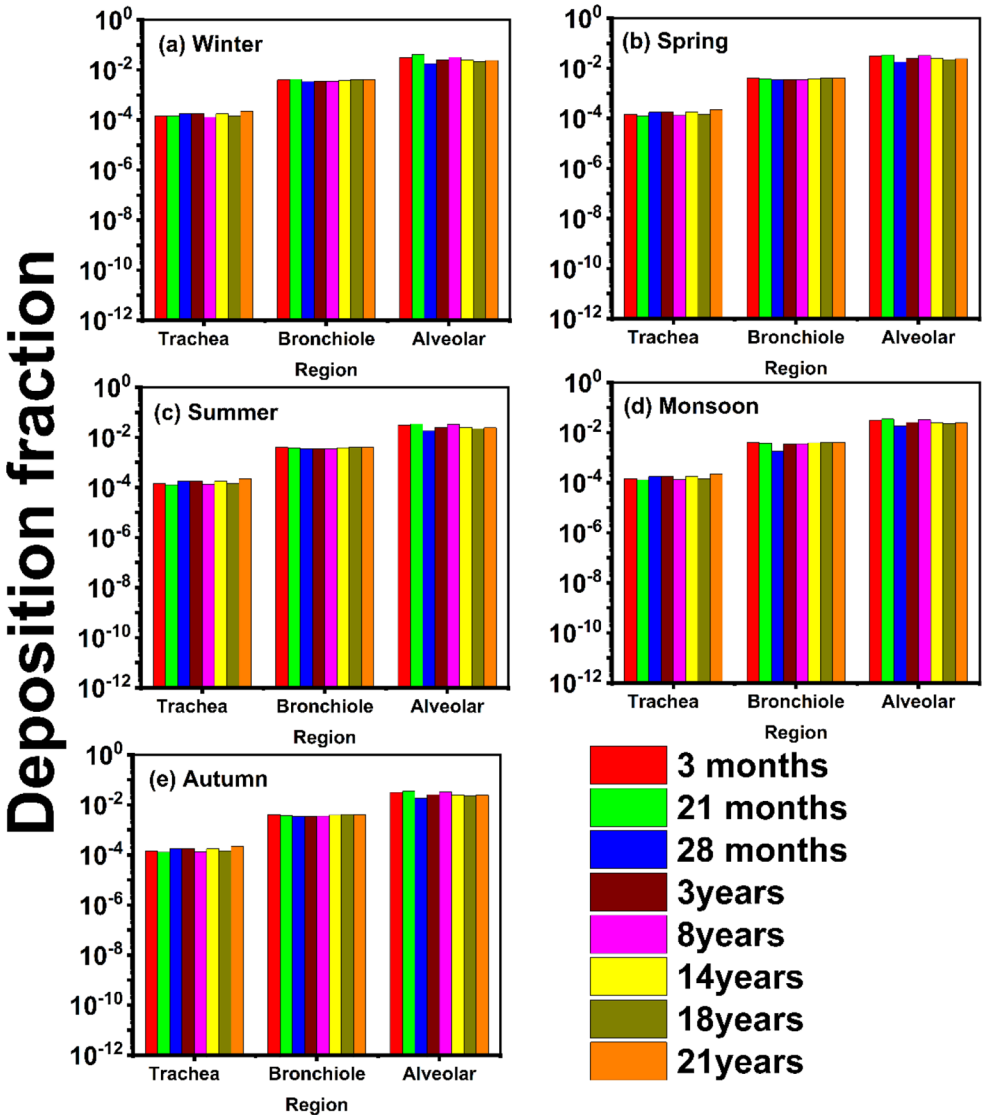


Fig. 8 Deposition fraction of nanoparticles estimated using MPPD model for different age groups in the trachea, bronchiole, and alveolar regions of respiratory system during (**a**) winter, (**b**) spring, (**c**) summer, (**d**) monsoon and (**e**) autumn in the urban roadside study region of Delhi

Acknowledgements The measurements were conducted in the Advance Air and Acoustics Research Laboratory, Department of Environmental Engineering, Delhi Technological University. KR thanks Dr. Manoj Kumar, Centre for Research on Energy and Clean Air, for his help in executing the model simulations.

Author contributions K. R conducted Data collection, analysis, interpretation, writing original draft, S.R contributed in Conceptualization, methodology, interpretation, writing, review, and editing and R.K. M contributed in Conceptualization, methodology, supervision, writing, review, and editing. All the authors read and approved the final manuscript.

Funding Not applicable.

Data availability All the data used in the manuscript can be made available on request.

Declarations

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication The work has not been published anywhere previously in any form, and it is not being considered for publication elsewhere. All the authors approve the publication.

Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Asgharian B, Hofmann W, Bergmann R (2001) Particle deposition in a multiple-path model of the human lung. Aero Sci Tech 34:332–339. https://doi.org/10.1080/02786820119122

Asgharian B, Price O, Creel A, Chesnutt J, Schroeter J, Fallica J, Erives G, Rasheed N, Chemerynski S (2022) Simulation modeling of air and droplet temperatures in the human respiratory tract for inhaled tobacco products. J Aero Sci 166:106050. https://doi.org/10.1016/j.jaerosci.2022.106050

Balakrishnan K, Dey S, Gupta T, Dhaliwal RS, Brauer M, Cohen AJ, Stanaway JD, Beig G, Joshi TK, Aggarwal AN, Sabde Y, Sadhu H, Frostad J, Causey K, Godwin W, Shukla DK, Kumar GA, Varghese CM, Muraleedharan P, Dandona L (2019) The impact of air pollution on deaths, disease burden, and life expectancy across the states of India: the Global Burden of Disease Study 2017. Lanc Plan Health 3:e26–e39. https://doi.org/10.1016/S2542-5196(18)30261-4

Buonanno G, Stabile L, Morawska L (2014) Personal exposure to ultrafine particles: The influence of time-activity patterns. Sci Total Environ 468–469:903–907. https://doi.org/10.1016/j.scitotenv.2013.09.016

Chalupa DC, Morrow PE, Oberdörster G, Utell MJ, Frampton MW (2004) Ultrafine particle deposition in subjects with asthma. Enviro Health Persp 112:879–882. https://doi.org/10.1289/ehp.6851

Deng Q, Deng L, Miao Y, Guo X, Li Y (2019) Particle deposition in the human lung: Health implications of particulate matter from different sources. Environ Res 169:237–245. https://doi.org/10.1016/j.envres.2018.11.014

Ezz WN, Mazaheri M, Robinson P, Johnson GR, Clifford S, He C, Morawska L, Marks GB (2015) Ultrafine particles from traffic emissions and children's health (UPTECH) in Brisbane, Queensland (Australia): Study design and implementation. Int J Environ Res Publ Health 12:1687–1702. https://doi.org/10.3390/ijerph120201687

Gani S, Chambliss SE, Messier KP, Lunden MM, Apte JS (2021) Spatiotemporal profiles of ultrafine particles differ from other traffic-related air pollutants: Lessons from long-term measurements at fixed sites and mobile monitoring. Environ Sci Atmos 1:558–568. https://doi.org/10.1039/d1ea00058f

Gupta P, Satsangi M, Satsangi GP, Jangid A, Liu Y, Pani SK, Kumar R (2020) Exposure to respirable and fine dust particle over North-Central India: chemical characterization, source interpretation, and health risk analysis. Environ Geochem Health 42:2081–2099. https://doi.org/10.1007/s10653-019-00461-w

Joshi M, Khan A, Anand S, Sapra BK (2016) Size evolution of ultrafine particles: Differential signatures of normal and episodic events. Environ Poll 208:354–360. https://doi.org/10.1016/j.envpol.2015.10.001

Kanawade VP, Srivastava AK, Ram K, Asmi E, Vakkari V, Soni VK, Varaprasad V, Sarangi C (2020) What caused severe air pollution episode of November 2016 in New Delhi? Atmos. Environ 222:117125. https://doi.org/10.1016/j.atmosenv.2019.117125

Khan S, Gurjar BR, Sahu V (2022) Deposition modeling of ambient particulate matter in the human respiratory tract. Atmos Poll Res 13:101565. https://doi.org/10.1016/j.apr.2022.101565

Kulkarni SH, Ghude SD, Jena C, Karumuri RK, Sinha B, Sinha V, Kumar R, Soni VK, Khare M (2020) How Much Does Large-Scale Crop Residue Burning Affect the Air Quality in Delhi? Environ Sci Tech 54:4790–4799. https://doi.org/10.1021/acs.est.0c00329

Kumar A, Mishra RK (2018) Human health risk assessment of major air pollutants at transport corridors of Delhi, India. J Tran Health 10:132–143. https://doi.org/10.1016/j.jth.2018.05.013

Kwon HS, Ryu MH, Carlsten C (2020) Ultrafine particles: unique physicochemical properties relevant to health and disease. Exp Mol Med 52:318–328. https://doi.org/10.1038/s12276-020-0405-1

Ma L, Zhang Y, Lin Z, Zhou Y, Yan C, Zhang Y, Zhou W, Ma W, Hua C, Li X, Deng C, Qi Y, Dada L, Li H, Bianchi F, Petaj T, Kangasluoma J, Jiang J, Liu S, Hussein Y, Ma YL (2022) Deposition potential of 0.003–10 µm ambient particles in the humidified. Eco Environ Saf 243:114023. https://doi.org/10.1016/j.ecoenv.2022.114023

Manigrasso M, Avino P (2012) Fast evolution of urban ultrafine particles: Implications for deposition doses in the human respiratory system. Atmos Environ 51:116–123. https://doi.org/10.1016/j.atmosenv.2012.01.039

Manojkumar N, Srimuruganandam B (2022) Age-specific and seasonal deposition of outdoor and indoor particulate matter in human respiratory tract. Atmos Poll Res 13:101298. https://doi.org/10.1016/j.apr.2021.101298

Manojkumar N, Srimuruganandam B, Nagendra SMS (2019) Application of multiple-path particle dosimetry model for quantifying age specified deposition of particulate matter in human airway. Eco Environ Saf 168:241–248. https://doi.org/10.1016/j.ecoenv.2018.10.091

Miller FJ, Asgharian B, Schroeter JD, Price O (2016) Improvements and additions to the Multiple Path Particle Dosimetry model. J Aero Sci 99:14–26. https://doi.org/10.1016/j.jaerosci.2016.01.018

Mohan V, Mishra RK (2022) A picture of Delhi' s regional air quality during diminished anthropogenic activities in the COVID 19 era. Arab J Geosci 2:1–14. https://doi.org/10.1007/s12517-022-10567-8

Mohan V, Soni VK, Mishra RK (2024) Analysing the impact of day-night road traffic variation on ultrafine particle number size distribution and concentration at an urban site in the megacity Delhi. Atmos Poll Res 15:102065. https://doi.org/10.1016/j.apr.2024.102065

Mosonik BC, Kibet JK, Ngari SM (2019) Simulating the Health Impact of Particulate Emissions from Transport Fuels Using Multipath Particle Deposition Model (MPPD). Opn J Mod Sim 7:115–124. https://doi.org/10.4236/ojmsi.2019.72006

Nagpure AS, Gurjar BR, Martel JC (2014) Human health risks in national capital territory of Delhi due to air pollution. Atmos Poll Res 5:371–380. https://doi.org/10.5094/APR.2014.043

Oliveira MLS, Navarro OG, Crissien TJ, Tutikian BF, da Boit K, Teixeira EC, Cabello JJ, Agudelo-Castañeda DM, Silva LFO (2017) Coal emissions adverse human health effects associated with ultrafine/nano-particles role and resultant engineering controls. Environ Res 158:450–455. https://doi.org/10.1016/j.envres.2017.07.002

Oliveira MLS, Izquierdo M, Querol X, Lieberman RN, Saikia BK, Silva LFO (2019) Nanoparticles from construction wastes: A problem to health and the environment. J Clean Prod 219:236–243. https://doi.org/10.1016/j.jclepro.2019.02.096

Pandey A, Brauer M, Cropper ML, Balakrishnan K, Mathur P, Dey S, Turkgulu B, Kumar GA, Khare M, Beig G, Gupta T, Krishnankutty RP, Causey K, Cohen AJ, Bhargava S, Aggarwal AN, Agrawal A, Awasthi S, Bennitt F, Dandona L (2021) Health and economic impact of air pollution in the states of India: the Global Burden of Disease Study 2019. Lan Plan Health 5:e25–e38. https://doi.org/10.1016/S2542-5196(20)30298-9

Rajagopal K, Ramachandran S, Mishra RK (2023) Roadside measurements of nanoparticles and their dynamics in relation to traffic sources in Delhi: Impact of restrictions and pollution events. Urban Clim 51:101625. https://doi.org/10.1016/j.uclim.2023.101625

Rajagopal K, Mohan V, Mishra RK (2024a) Are Delhi residents exposed to lesser particle number concentration due to the firework ban in the city? 24, 01532. Air Qual Atmos Health. https://doi.org/10.1007/s11869-024-01532-3

Rajagopal K, Ramachandran S, Mishra RK (2024b) Size resolved particle contribution to vehicle induced ultrafine particle number concentration in a metropolitan curbside region. Atmos Environ 337:120773. https://doi.org/10.1016/j.atmosenv.2024.120773

Rizwan SA, Nongkynrih B, Gupta SK (2013) Air pollution in Delhi: Its Magnitude and Effects on Health. Ind J Com Med 38:4–8. https://doi.org/10.4103/0970-0218.106617

Sengupta A, Govardhan G, Debnath S, Yadav P, Kulkarni SH, Parde AN, Lonkar P, Dhangar N, Gunwani P, Wagh S, Nivdange S, Jena C, Kumar R, Ghude SD (2022) Probing into the wintertime meteorology and particulate matter (PM2.5 and PM10) forecast

over Delhi. Atmos Poll Res 13:101426. https://doi.org/10.1016/j.apr.2022.101426

Sharma M, Khare M, Mishra RK (2024) Air quality changes in Delhi due to open waste burning: an accidental fire in Bhalswa landfill. Intl J Environ Sci Tech 21:655–664. https://doi.org/10.1007/s13762-023-04921-w

Shivani, Gadi R, Saxena M, Sharma SK, Mandal TK (2019) Short-term degradation of air quality during major firework events in Delhi, India. Met Atmos Phys 131:753–764. https://doi.org/10.1007/s00703-018-0602-9

Viitanen AK, Uuksulainen S, Koivisto AJ, Hämeri K, Kauppinen T (2017) Workplace measurements of ultrafine particles-A literature review. Anls Work Exp Health 61:749–758. https://doi.org/10.1093/annweh/wxx049

Vu TV, Ondracek J, Zdímal V, Schwarz J, Delgado-Saborit JM, Harrison RM (2017) Physical properties and lung deposition of particles emitted from five major indoor sources. Air Qual Atmos Health 10:1–14. https://doi.org/10.1007/s11869-016-0424-1

Yadav SK, Kompalli SK, Gurjar BR, Mishra RK (2021) Aerosol number concentrations and new particle formation events over a polluted megacity during the COVID-19 lockdown. Atmos Environ 259:118526. https://doi.org/10.1016/j.atmosenv.2021.118526

Yadav SK, Mishra RK, Gurjar BR (2022) Ultrafine particle number concentration and its size distribution during Diwali festival in megacity Delhi, India: Are 'green crackers safe? J Environ Mgmt 317:115459. https://doi.org/10.1016/j.jenvman.2022.115459

Zhu L, Ranasinghe D, Chamecki M, Brown MJ, Paulson SE (2021) Clean air in cities: Impact of the layout of buildings in urban areas on pedestrian exposure to ultrafine particles from traffic. Atmos Environ 252:118267. https://doi.org/10.1016/j.atmosenv.2021.118267

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

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

Supporting Information

Ultra-sensitive NO_2 gas sensor at room temperature based on glycerol-cross-linked PEDOT: PSS-MoS_2 nanocomposite

Priyanka Dutta,[1] **Anuj Sharma**,[1,2] **Videsh Kumar**,[1,3] **Govind Gupta**[1, 2*]

[1]CSIR National Physical Laboratory, Dr. K. S. Krishnan Road, New Delhi, 110012, India.

[2]Academy of Scientific and Innovative Research, CSIR-HRDC Campus, Ghaziabad, Uttar Pradesh-201002, India.

[3]Delhi Technological University, Rohini, New Delhi 110042, India.

Priyanka Dutta: dutpriya2600@gmail.com

Anuj Sharma: anujsharma.npl@gmail.com

Videsh Kumar: videshk93@gmail.com

*Govind Gupta: govindnpl@gmail.com

govind@nplindia.org

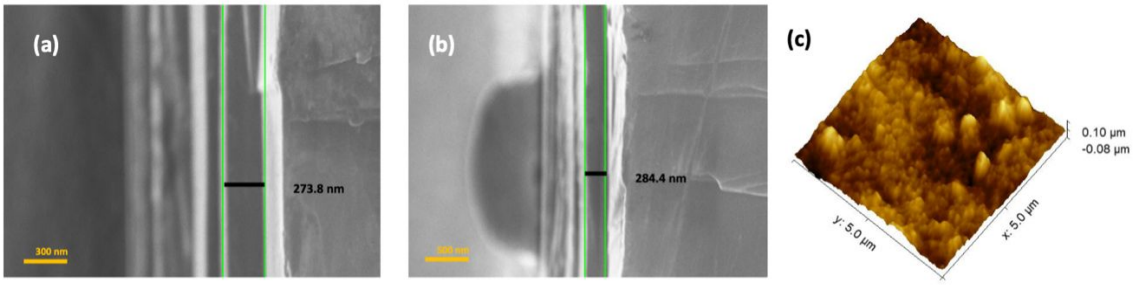


Figure S1: *Cross-sectional FESEM image of the polymer nanocomposite (a) PDGM1 and (b) PDGM2; (c) AFM image of PDGM3 deposited over a silicon substrate etched with 1 % HF solution.*

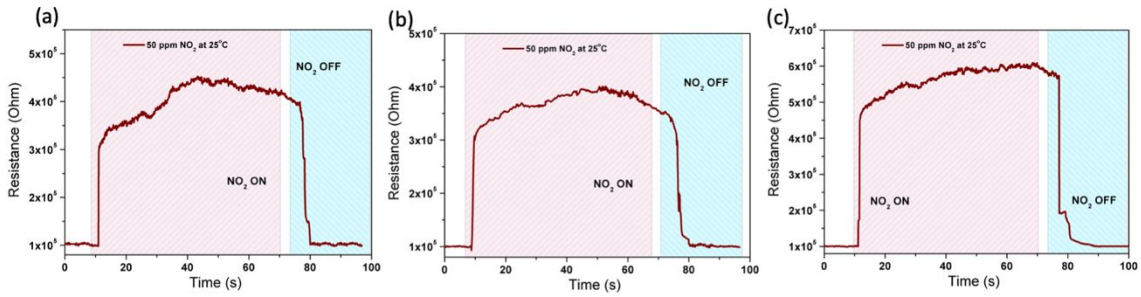


Figure S2: *Gas sensing response of (a) PDGM 1, (b) PDGM 2, and (c) PDGM 3 in the presence of 50 ppm NO_2 at RT.*

Table S1: *Gas sensing properties of the fabricated sensors and pure PEDOT: PSS and pure pure MoS_2 nanosheet.*

Sl. No.	**Sensor name**	**Sensor type**	**Gas detected**	**Operating Temperature**	**Sensor Response % with 50 ppm NO_2**	**Response time (s) with 50 ppm NO_2**	**Recovery time (s) with 50 ppm NO_2**
1.	PEDOT: PSS	p-type	NO_2	25 °C	110	24	16.5
2.	MoS_2 nanosheet	n-type	NO_2	25 °C	45	58	48
3.	PDGM 1	n-type	NO_2	25 °C	280	21.5	14
4.	PDGM 2	n-type	NO_2	25 °C	330	19	10.2
5.	PDGM 3	n-type	NO_2	25 °C	500.3	14.5	9

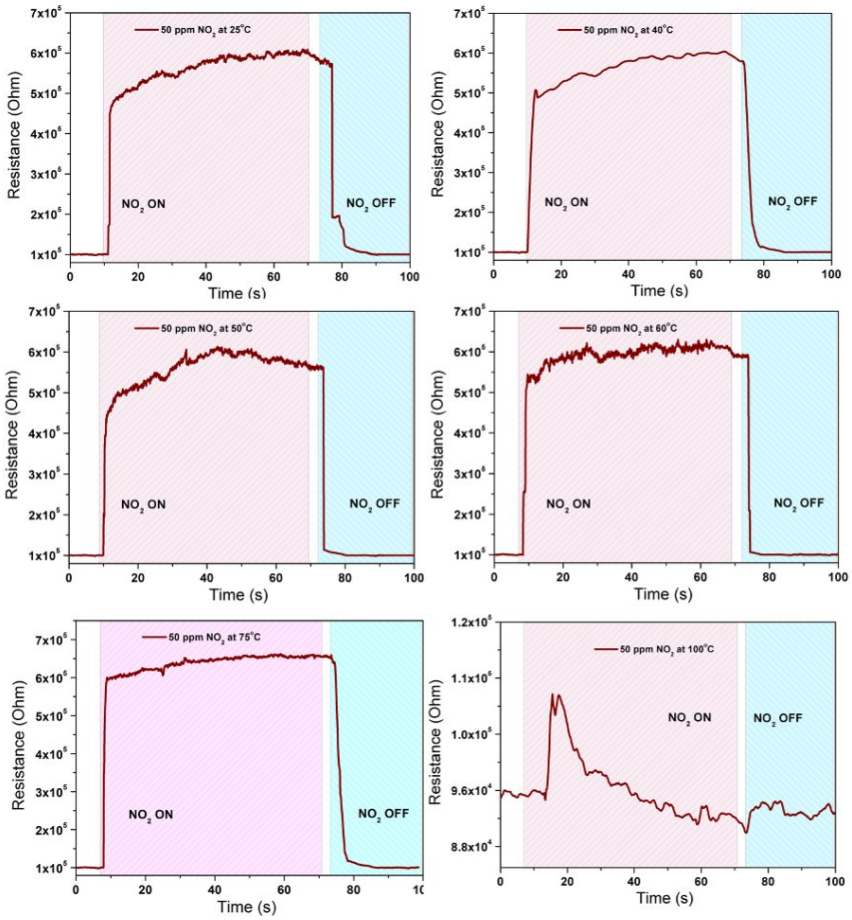


Figure S3: *Gas sensing response of PDGM 3 in the presence of 50 ppm NO_2 at different temperatures (25°C to 100°C).*

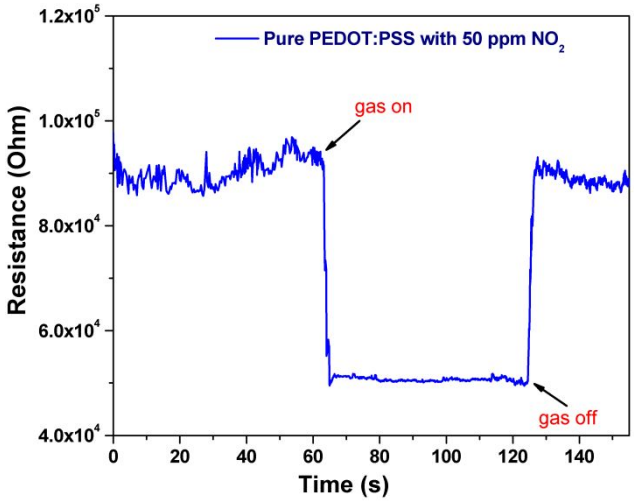


Figure S4: *Gas sensing response of PEDOT: PSS in the presence of 50 ppm NO_2 at RT.*

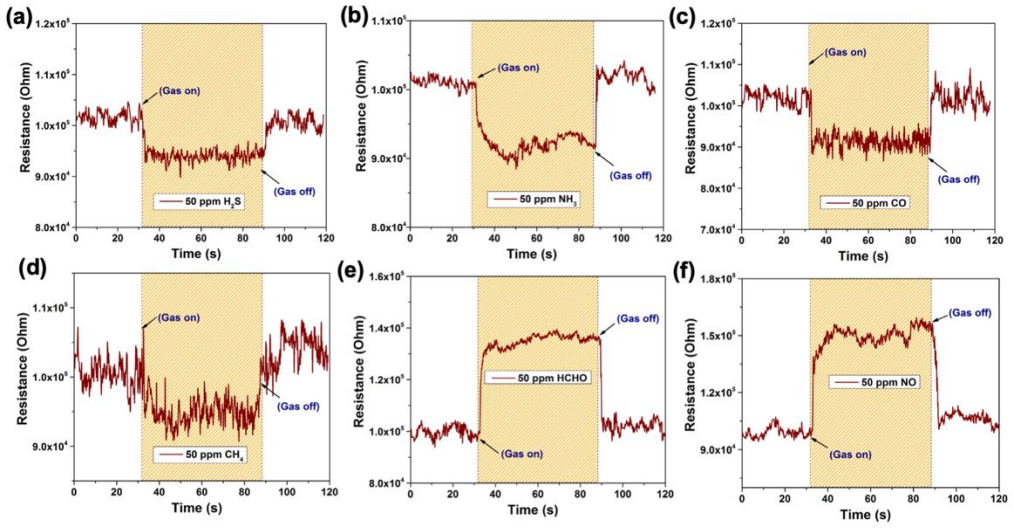


Figure S5(a-f): *Real-time gas sensing plots of PDGM 3 in the presence of 50 ppm of different oxidizing and reducing gases (50 ppm) at RT.*

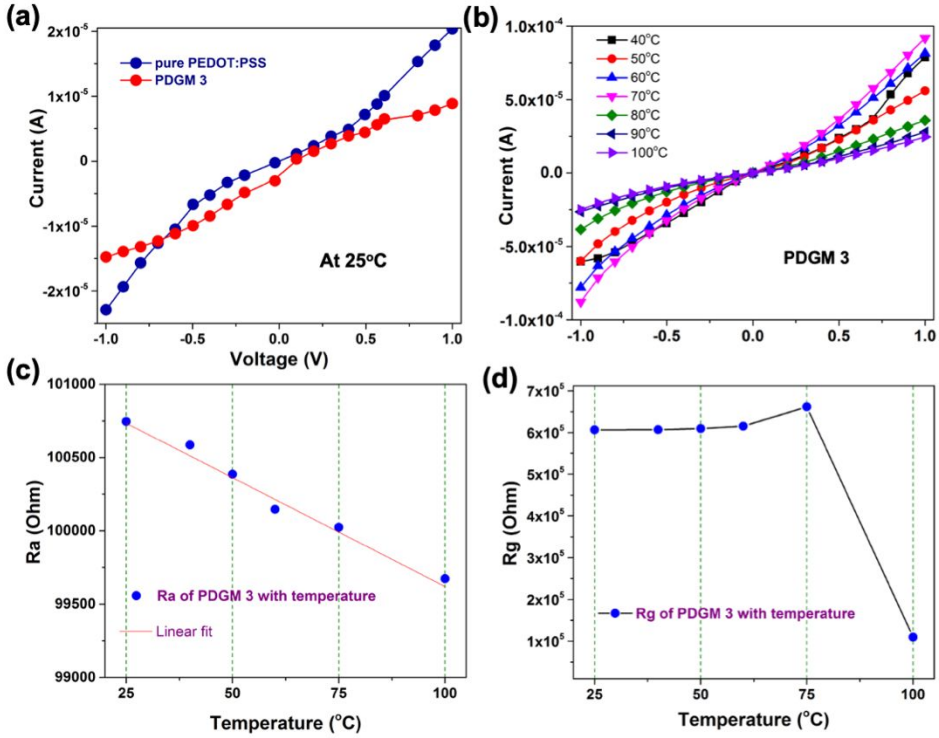


Figure S6: *IV plot of (a) PEDOT: PSS and PDGM 3 at RT, (b) PDGM 3 at different temperatures from 40°C to 100°C, (c) Ra vs temperature plot of PDGM 3, and (d) Rg vs temperature plot of PDGM 3.*

UnMA-CapSumT: Unified and Multi-Head Attention-driven Caption Summarization Transformer

[1]Dhruv Sharma, [2]Chhavi Dhiman, [3]Dinesh Kumar
[1,2,3]Deapartment of Electronics and Communication Engineering,
Delhi Technological University, Delhi, India
[1]*dhruv.0906@yahoo.in*, [2]*chhavi.dhiman@dtu.ac.in*, [3]*dineshkumar@dtu.ac.in*

***Abstract*— Image captioning is the generation of natural language descriptions of images which have increased immense popularity in the recent past. With this different deep-learning techniques are devised for the development of factual and stylized image captioning models. Previous models focused more on the generation of factual and stylized captions separately providing more than one caption for a single image. The descriptions generated from these suffer from out-of-vocabulary and repetition issues. To the best of our knowledge, no such work exists that provided a description that integrates different captioning methods to describe the contents of an image with factual and stylized (romantic and humorous) elements. To overcome these limitations, this paper presents a novel Unified Attention and Multi-Head Attention-driven Caption Summarization Transformer (*UnMA*-CapSumT) based Captioning Framework. It utilizes both factual captions and stylized captions generated by the Modified Adaptive Attention-based factual image captioning model (MAA-FIC) and Style Factored Bi-LSTM with attention (SF-Bi-ALSTM) driven stylized image captioning model respectively. SF-Bi-ALSTM-based stylized IC model generates two prominent styles of expression- {romance, and humor}. The proposed summarizer $UnMHA-ST$ combines both factual and stylized descriptions of an input image to generate styled rich coherent summarized captions. The proposed $UnMHA-ST$ transformer learns and summarizes different linguistic styles efficiently by incorporating proposed word embedding fastText with Attention Word Embedding (fTA-WE) and pointer-generator network with coverage mechanism concept to solve the out-of-vocabulary issues and repetition problem. Extensive experiments are conducted on Flickr8K and a subset of FlickrStyle10K with supporting ablation studies to prove the efficiency and efficacy of the proposed framework. The code of the proposed work is available at: https://github.com/dhruvsharma09/UnMA-CapSumT.git**

***Index Terms*— computer vision, factual image captioning, natural language processing, stylized image captioning, text summarization, word embedding,**

1. INTRODUCTION

Image captioning [1] [2] identifies objects within an image, comprehending their interactions to describe the contents of an image in the form of descriptive sentences. Caption Generation is a very challenging task as it generates descriptive and procedural texts that are more effective for image comprehension. Hence, it is employed for numerous applications like human-robot interaction, aid to the blind [3] [4], visual question answering [5], etc. Most studies [6] [7] [8] on image captioning primarily focus on the description of the factual content of an image while ignoring the underlying emotion or style of the same image. On the other hand, very few studies [9] [10] learned the knowledge of factual content and its linguistic style for the generation of style-based descriptions of images. This on the other hand facilitates many real-world applications, like chatbots, or enlightening users in photo captioning for different social media platforms, storytelling, etc. as it greatly enriches the expressibility of the caption, making it more attractive. Further, style-based caption generation not only describes the contents of an image but also analyzes the intrinsic style-based elements in the captioning [11] [12]. Also, this requires paired image-sentiment [11] caption data and word-level supervision to emphasize the sentiment words, therefore making style-based captioning very expensive.

With the advancements in deep-learning technology, different models for factual [7] [8] and stylized image captioning [11] [12] are developed. These works generate separate sentences for factual and stylized (romantic and humorous) descriptions of an image. Therefore, these methods sometimes may give erroneous descriptions by either describing the factual content in dull language or may lead to the

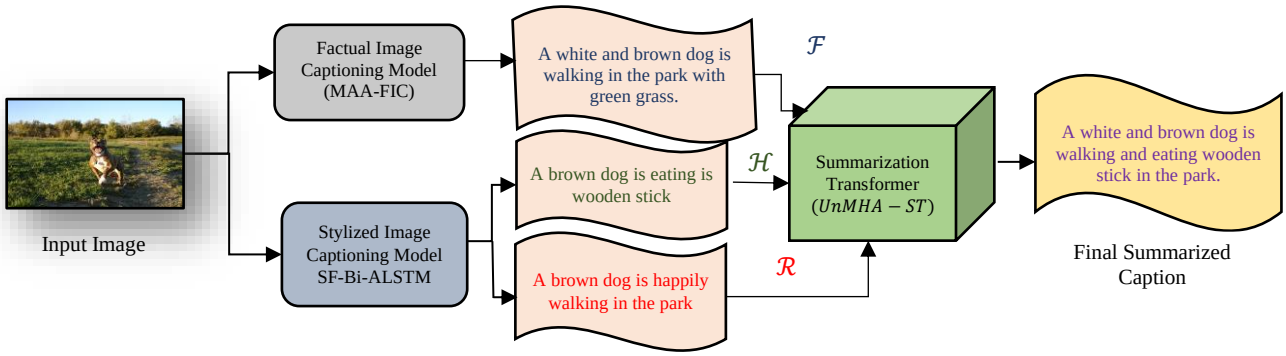


Fig. 1: Block Diagram Representation of the proposed $UnMA$ −CapSum Transformer based Captioning Framework

incorrect representation of the style. This may be due to poor learning knowledge of factual content and its associated linguistic styles. To the best of our knowledge, there are no such works in literature that provide single-sentence summarized descriptions incorporating factual, romantic, and humorous content.

To address the abovementioned issues and to improve the learning of knowledge of factual content and its associated linguistic styles this paper extended the task of image captioning to abstractive text summarization. The task is to automatically summarize the source article into an accurate short version that reflects its central content comprehensively. Though the summaries generated also suffers from major issues of rare words that arises due to repetition of words or due to out-of-vocabulary issues [13] [14]. In this direction, this paper presents a novel Unified Attention (Un-A) and Multi-Head Attention-driven Caption Summarization Transformer ($UnMA$-CapSumT) based Captioning Framework. An overview of the proposed framework is depicted in Fig. 1, the framework is twofold: (i) it integrates MAA-FIC-based factual image captioning and SF-Bi-ALSTM based stylized image captioning techniques for the generation of factual, stylized – {romantic, and humorous} description of an image, and (ii) the proposed summarizer $UnMHA-ST$ combines both factual and stylized descriptions of input image to generate styled rich coherent summarized captions.

The main contributions of this work are:

i. This paper proposed a Unified and Multi-head Attention-based Caption Summarization Transformer $UnMHA-ST$ to capture the intra- and intermodal interactions of multimodal information and generates refined attended representations. Also, it utilizes the concept of a pointer generator network and coverage mechanism to solve the problems of rare words which arise due to out-of-vocabulary (OOV) and repetition issues.
ii. The proposed framework integrates Modified Adaptive Attention-based (MAA-FIC) and SF-Bi-ALSTM-based factual and stylized image captioning models for the generation of factual and stylized descriptions of images which are utilized by $UnMHA-ST$ for generation of styled rich coherent summarized captions.
iii. The paper also presents an efficient Attention enabled fastText word embedding, fTA-WE, that integrates the attention mechanism into the Continuous Bag of words (CBOW) model of fastText that efficiently learns vector representation of words and enhances the performance of the proposed $UnMHA-ST$.
iv. Extensive experiments on Flickr8K [15], and a subset of FlickrStyle10K [9] demonstrate the efficiency and efficacy of the proposed $UnMA$-CapSumT framework.

The organization of the rest of the paper is as follows: Section 2 presents the different state-of-the-art for factual image captioning, stylized image captioning, and text summarization. Section 3 discusses the proposed methodology. Section 4 presents the Dataset Used, Implementation details with the results obtained. Finally, Conclusions are drawn in Section 5.

2. RELATED WORKS

This section presents different techniques from past for factual image captioning, stylized image captioning, and text summarization.

2.1 Factual Image Captioning

Traditional image captioning methods [16] [17] [18] [19] fail for many practical applications as they are unable to generate variable-length descriptions and the descriptions generated are less natural. Analysis of complex, high-dimensional, and noise-contaminated data is extremely difficult. Therefore, it is essential to create new algorithms that can categorize, summarize, extract, and transform data into understandable forms. Hence, deep-learning-based models have demonstrated exceptional achievements in the past few years. To overcome these limitations deep learning methods for image captioning were encouraged. These methods [1] [20] rely on large-scale datasets which efficiently and effectively generate variable-length descriptions of images.

Deep-learning-based models have demonstrated exceptional achievements in the past few years. These methods [20] [21] proposed a generative model that combined recent advances in CV and machine translation which is trained to maximize the likelihood of the target description sentence for a given training image. In addition to this, [20] directly models the probability distribution of generating a word given previous words and an image. To generate more natural and descriptive long-length sentences, guided-LSTM or g-LSTM [22] is proposed. g-LSTM adds semantic information extracted from the image as extra input to each unit of the LSTM block. For the generation of captions for static as well as dynamic images, Donhaue et al. [23] stacked multiple LSTMs to jointly train and learn temporal dynamics and convolutional perceptual representations. Wang et al. [24]generated semantically rich descriptions of images using deeper Bi-directional LSTM framework. This framework utilized past and future context information to learn long-term visual language interactions.

The above-discussed methods provide captions for the entire scene rather than taking into account the spatial elements of an image that are important for captioning. To overcome these limitations soft-attention [1], spatial and channel-wise attention [6], adaptive [7] [25] [26], multi-head and self-attention [27], and Semantic attention [28] based captioning of images came into existence. The attention-based methods [20] [21] have successfully surpassed the performance of traditional deep-learning models by highlighting salient parts of the image and by providing higher-order interactions between objects and the scene. However, to concentrate more on the visual salient regions Xiao et al. [29] proposed Dual-LSTM and adaptive attention mechanism-based model for image captioning and visual question answering. The proposed model flexibly makes a trade-off between the visual semantic region and textual content to achieve a significant performance of prior state-of- the-art approaches. Yan et al. [30] proposed a task-adaptive attention module that introduced diversity regularization to enhance the expression ability of the attention mechanism. Further, Xiao et al. [31] presented a new variant of LSTM, named ALSTM (Attention-LSTM) for image captioning for generation of captions by paying more attention to the most relevant context words.

2.2 Stylized Image Captioning

Stylized or style-based image captioning is the description of images with different styles or sentiments. Stylized image captioning when compared with factual image captioning provides a better understanding of images. Ideally, a stylized image captioning model should fulfill two conditions, (i) provides the correct description of images, and (ii) generates correct stylized words or phrases in the correct positions of descriptions. To fulfill these conditions, a stylized image captioning model utilizes non-parallel [9] [12] [32] [33] or parallel [10] [11] [34] [35] stylized image captioning data. Mathews et al. [34] described an image using positive and negative sentiments. The model proposed utilized [34] word-level regularization for producing emotional image captions. Another work [32] developed a novel semantic term representation to disentangle content and style in descriptions. Gan et al. [9] proposed a novel factored LSTM which automatically distills the style factors in the monolingual text corpus and generated romantic and humorous captions for images. To efficiently describe the visual contents of an image using linguistic styles very few works [10] [35] focused on splitting the stylized sentence into (i) style-related part that reflects the linguistic style and (ii) a content-related part that contains the visual content. To generate controllable stylized descriptions of images Chen et al. [11] suggested adaptive learning with attention. Also, it provided more stylish image descriptions relevant to the content of the images. MSCap [12] is the first attempt toward the generation of image descriptions in multiple styles with a single model. Zhao et al. [33] proposed a novel model with a memory module

that memorizes the knowledge of linguistic style and distills the content-relevant style knowledge with an attention mechanism for generating stylized descriptions. Further, Li et al. [36] proposed a sentimental image captioning to reflect the inherited sentiments of an image. The methods discussed above focus on the generation of image descriptions with linguistic styles in a single decoding process. In contrast, the proposed *UnMA* CapSumT generates a single summarized image description with factual and stylized elements.

2.3 Text Summarization

Researchers in the past have developed many traditional techniques or methods for summarization of text, such as feature score [37], semantic information [38], template [39], classification [40], and dynamic programming [41]. These methods generated incomplete, redundant summaries for the input text data. Also, these methods are easily affected by abnormal data which leads to the generation of uncorrelated text at the output. To overcome the generation of incomplete, redundant, and uncorrelated text at the output deep-learning-based automatic abstractive summarization methods are encouraged. Rush et al. [42] proposed a fully data-driven attention-based approach by combining probabilistic models and generation algorithms to produce accurate summaries. Nallapati et al. [13] [14] described summarization tasks using attentional encoder-decoder RNN and provided state-of-the-art results. The abstractive summaries generated by [42] [13] [14] suffer from OOV and low-frequency issues. To overcome these limitations text summarization with a pointer-generator network and coverage mechanism [43] was proposed. This solves the issues related to OOV and also avoids the generation of repetitive and incoherent phrases in summarized text. Hsu et al. [44] introduced a novel inconsistency loss function for abstractive and extractive text summarization with a pointer-generator network and coverage mechanism. Inconsistency loss penalized the inconsistency between two levels of attention and efficiently solves the issues related to OOV and repetitive phrases in the generated summary.
Li et al. [45] proposed a multi-head attention-driven abstractive text summarization network that learned relevant information in different representation subspaces. Yao et al. [46] proposed dual-encoder-based abstractive text summarization in which the primary encoder conducts coarse encoding and the second one models the importance of words. In recent past transformer-based architectures are now being encouraged for summarization tasks. In this direction, Deaton et al. [47] proposed a transformer-based approach for abstractive summarization tasks. The model [47] incorporated basic transformer-based architecture [48] with a pointer-generator network and coverage mechanism. Further, this paper presented multi-head attention and unified attention-driven $UnMHA - ST$ transformer model with a pointer-generator network and coverage mechanism. this provides improvements in the performance of the summarization task by reducing redundancy and factual errors.

3. PROPOSED WORK

This section discusses the proposed model, which is divided into two phases i.e., image captioning, and text summarization. In the proposed model, as depicted in Fig. 2, two image captioning techniques are integrated to generate factual, romantic, and humorous captions of an image. Further, these captions are fed to the transformer-based text summarization model to provide the final summarized caption which contains all three (factual, romantic, and humorous) elements in a single caption.

3.1 Modified Adaptive Attention-based Factual Image Captioning Model (MAA-FIC)

This section describes in detail the proposed MAA-FIC factual image captioning model. The input to the model is an image $\mathcal{I}$ and the output is the descriptive factual sentence $\mathcal{F}$ with $\mathcal{k}$ encoded words: $\mathcal{F} = \{w_1, w_2, \ldots . w_{\mathcal{k}}\}$.

3.1.1 Feature Extraction

For object detection, the proposed MAA-FIC model incorporates Faster R-CNN for the detection of objects. After the detection of objects, each detected object is mapped to a feature vector. For each image $\mathcal{I}$, n-objects are detected, given by, $\{\Phi_1, \Phi_2, \ldots, \Phi_n\}$; $\Phi_i \in \mathbb{D}^d$ where, $\mathbb{D}^d$ is the $\mathbb{D} -$

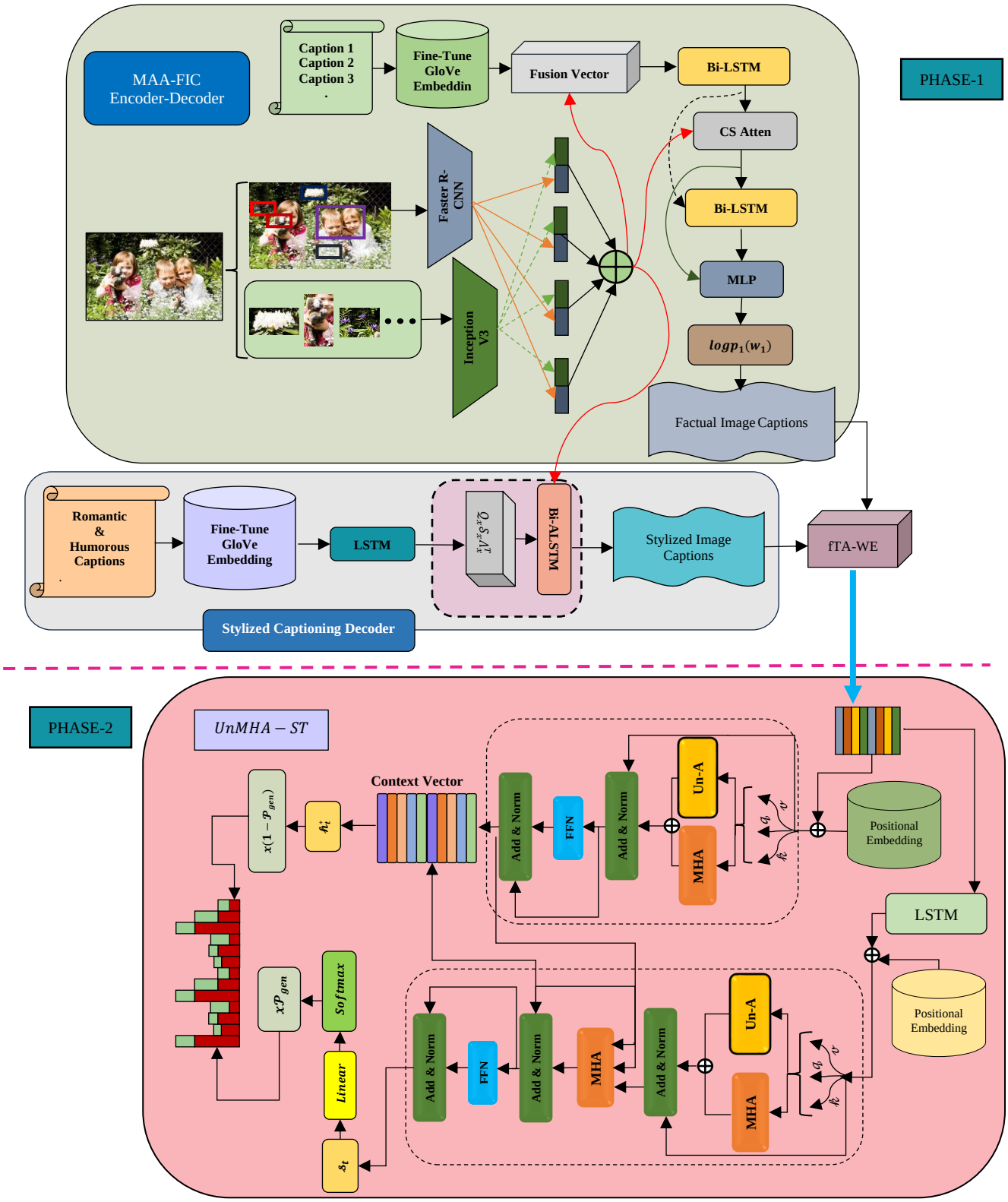


Fig. 2: Proposed Methodology: Phase-1: MAA-FIC module and Stylized Image Captioning Module, Phase 2: $UnMHA-ST$: Unified and Multi-head Attention based Summarization Transformer

dimensional vector of top n-boxes as the region of objects.

The localization of objects is performed to extract spatial relationships between objects. The n-objects detected using Faster R-CNN [49] for each image are fed to Inception-V3 [50] and output the feature vector that represents the spatial location of each object, represented as $\{\theta_1, \theta_2, \ldots, \theta_n\}$; $\theta_i \in \mathbb{D}^{\mathbb{T}}$,

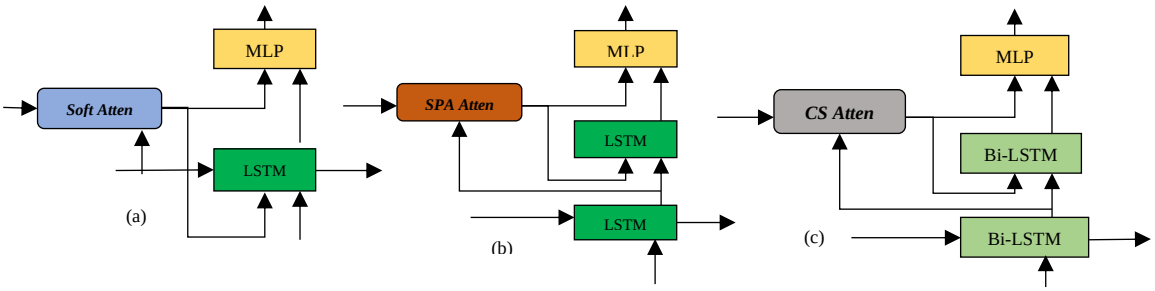


Fig. 3: (a) Soft-attention Mechanism, (b) Adaptive Attention Mechanism, and (c) Proposed Modified Adaptive Attention

where, $\mathbb{D}^{\mathbb{T}}$ is the $t - dimensional$ vector of the spatial location of each object. The features extracted from the Faster R-CNN module and Inception-V3 module are concatenated. Therefore, each concatenated feature vector consists of a detected object feature vector (Φ_i) and the feature vector obtained from the localization of objects (θ_i). Mathematically, the concatenated image feature vector is represented as:

$$\mathbb{F}_i = [\Phi_i; \theta_i], \mathbb{F}_i \in d^D, D = d + \mathbb{T} \quad (1)$$

3.1.2 Text Generation Network

This section introduces Modified Adaptive based Attention (MAA) and Bi-LSTM-based factual language decoder module. The MAA utilizes both channel and spatial attention mechanism with bidirectional sequence learning. From the definition of attention mechanism, it utilizes the most relevant part of the input in a flexible manner. This technique combines the weighted combination of all the encoded input vectors, with the most relevant vectors being attributed with the highest weights. For the model with attention mechanism:

$$a_t = \mathbb{f}(\mathbb{F}, \mathbb{h}_t) \quad (2)$$

where $\mathbb{f}$ is the attention mechanism and $\mathbb{F} = [\mathbb{F}_1, \mathbb{F}_2 \ldots, \mathbb{F}_{\mathbb{k}}]$ represents the concatenated image features and $\mathbb{h}_t$ is the hidden state of the RNN at time t. Channel-wise attention distribution for $\mathbb{k}$ regions is defined as:

$$\tilde{\mathcal{C}} = \tilde{\mathcal{C}}_{\mathcal{W}} = \mathcal{W}_{hc}^T \tanh\left((\mathcal{W}_{c\mathbb{F}} \otimes \mathbb{F} + \beta_c) \oplus \left(\mathcal{W}_{\mathbb{f}c}\mathbb{h}_t\right)\mathcal{I}^T\right) \quad (3)$$

$$\psi = softmax(\tilde{\mathcal{C}}) \quad (4)$$

Further, spatial attention weights of the $\mathbb{k}$ regions mathematically expressed as:

$$\tilde{\mathcal{S}} = \tilde{\mathcal{S}}_{\mathcal{W}} = \mathcal{W}_{hs}^T \tanh\left(\left(\mathcal{W}_{s\mathbb{F}}\mathbb{F} + \beta_s\right) \oplus \left(\mathcal{W}_{\mathbb{f}s}\mathbb{h}_t\right)\mathcal{I}^T\right) \quad (5)$$

$$\varphi = softmax(\tilde{\mathcal{S}}) \quad (6)$$

where, $\beta_s$and $\beta_c$are bias terms. After attaining the channel and spatial attention weights, channel-spatial weights are thus combined to obtain a modulated feature map. This provides the modified attentional distributions for the $\mathbb{k}$ regions as:

$$\psi = \phi_c(h_t, \mathbb{F}) \quad (7)$$

$$\varphi = \phi_s(h_t, \mathbb{f}(\mathbb{F}, \psi)) \quad (8)$$

$$\widetilde{\mathcal{X}} = \mathbb{f}(\mathbb{F}, \psi, \varphi) \quad (9)$$

Therefore, the final modified attention weights are defined mathematically as:

$$a_t = \sum_{i=1}^{\mathbb{k}} (\mathcal{X}_{it}\mathbb{F}_{it}) \quad (10)$$

Based on the channel and spatial attention mechanism, modified adaptive attention is proposed with visual sentinel and Bi-LSTM layers. Fig. 3 presents the basic architecture of MAA. The visual sentinel helps the model to focus more on wither image information or on the language rules. The memory unit in Bi-LSTM stores the information of both the previous partial visual information and the language rules. Therefore, visual sentinel $\mathbb{s}_t$ formula based on the Bi-LSTM memory unit is:

$$m_t = \theta(\mathcal{W}_x\mathcal{X}_t + \mathcal{W}_h h_{t-1}) \quad (11)$$

$$\mathbb{s}_t = m_t \odot \tanh(\tilde{h}_t) \quad (12)$$

Where, $\mathcal{X}_t$ is the input of the Bi-LSTM, m_t is the sentinel gate, is the output of the last time node h_{t-1}), and $\tilde{h}_t$ is the memory cell. Also, the modified adaptive attention (MAA) mechanism is obtained following the sentinel gate $\mathbb{s}_t$

$$\hat{a}_t = \delta_t s_t + (1 - \delta_t) a_t \quad (13)$$

$\delta_t \in [0,1]$ and is the new sentinel gate at time t. The value of δ_t decides whether the model focuses on image information or language rules. If the value of δ_t is equal to zero, it denotes focus on image information whereas the value of 1 indicates that the focus is on the language rules. The channel-spatial attention distribution of k regions obtained by splicing an element and is given by:

$$\widehat{\mathcal{X}} = softmax([\widetilde{\mathcal{X}}; \mathcal{W}_h^T \tanh(\mathcal{W}_s s_t + \mathcal{W}_m h_t))]) \quad (14)$$

To generate the natural language description of an input image, fine-tuned GloVe text embeddings are employed and are represented as $\mathbb{T}$. The input to the text generation module is the weighted fused vector obtained after the fusion of image features $\mathbb{F}$ and text features $\mathbb{T}$.

$$\mathcal{Q}_t = \{\mathbb{F}_t \mathbb{T}_t\} \quad (15)$$

The fused feature vector $\mathcal{Q}_t$ is incorporated for the generation of accurate description of an image. Further, the resulting output generates a word on the next time node using the MLP function that incorporates backpropagation through a time algorithm to update the parameters of the LSTM network. The objective function defined for the optimization of the proposed model is given by:

$$\Delta = arg \max_{\vartheta} \sum_{r,y} \sum_{t=0}^{N} \log p(\mathcal{F}_t | r, \Delta, \mathcal{F}_1, \mathcal{F}_2 \ldots \mathcal{F}_{t-1}) \quad (16)$$

ϑ is the learnable parameter with m feature maps whose weight is $r \in \{r_1, r_2 \ldots r_t\}$. The cross-entropy loss is incorporated to minimize the loss function and to maximize the probability of each correct word appearing. The cross entropy-loss function is given by:

$$\mathcal{L}(\Delta) = -\sum_{t=1}^{N} \log (p_t(w_t^* | \mathbb{G}_{1:t-1}^*)) \quad (17)$$

here *N* represents total words in a sentence; Δ denotes all the parameters in the model; $\mathbb{G}_{1:t-1}^*$defines the ground truth.

3.2 SF-Bi-ALSTM Based Stylized Image Captioning

This section introduces the SF-Bi-ALSTM module, which serves as the building block for the generation of style-based descriptions of images. The input to the proposed style-based caption generation model is image $\mathcal{I}$ and the output is the descriptive romantic and humorous sentences $\mathcal{R}$ and $\mathcal{H}$ with n and m encoded words: $\mathcal{R} = \{w_1, w_2, \ldots . w_n\}$, $\mathcal{H} = \{w_1, w_2, \ldots . w_m\}$. The proposed stylized captioning model is based on encoder-decoder-based architecture. For the encoder, this model employs the same strategy as employed by the MAA-FIC model as discussed in *Section A* whereas the SF-Bi-ALSTM model is utilized for the generation of style-based captions.

3.2.1 SF-Bi-ALSTM Module

This section presents a new and modified variant of LSTM and Factored LSTM namely, Modified Style Factored Bi-LSTM. Traditional LSTM captures the long-term dependencies among words in sentences but fails to capture the styles in sentences. Also, Factored LSTM incorporated style factors into the visual caption generation model but fails to generate more attractive and coherent style-based descriptions. To overcome these limitations SF-Bi-ALSTM Module is designed that produces more meaningful stylized descriptions, combining LSTM with attention layers from both directions. Further, the proposed Bi-LSTM learns to refine the input vector from network hidden states and sequential context information.

Consider, $\mathcal{M}_x \in \mathbb{R}^{m \times n}$, then $\mathcal{Q}_x \in \mathbb{R}^{m \times r}$, $\mathcal{S}_x \in \mathbb{R}^{r \times r}$, and $\Lambda_x \in \mathbb{R}^{r \times n}$. $\mathcal{M}_x$ can be represented in the form of three matrices:

$$\mathcal{M}_x = \mathcal{Q}_x \mathcal{S}_x \Lambda_x^T \quad (18)$$

Further, the style-specific matrix $\{\mathcal{S}_x\}$ can be represented as:

$\mathcal{S}_x = \frac{1}{N} |\mathcal{S}\mathcal{S}^T|$ (19) The memory gates and cells in the traditional Bi-LSTM are

modified with respect to $\mathcal{M}_x$ and attention mechanism:

$$\tilde{x}_t = sigmoid(\mathcal{Q}_{x_t} \mathcal{S}_{x_t} \Lambda_{x_t} x_t + W_h \tilde{h}_{t-1}) \odot x_t \quad (20)$$

$$\tilde{\imath}_t = sigmoid(\mathcal{Q}_{ix} \mathcal{S}_{ix} \Lambda_{ix} \tilde{x}_t + W_{ih} \tilde{h}_{t-1} + \tilde{b}_i) \quad (21)$$

$$\tilde{f}_t = sigmoid(\mathcal{Q}_{fx} \mathcal{S}_{fx} \Lambda_{fx} \tilde{x}_t + W_{fh} \tilde{h}_{t-1} + \tilde{b}_f) \quad (22)$$

$$\tilde{o}_t = sigmoid(\mathcal{Q}_{ox} \mathcal{S}_{ox} \Lambda_{ox} \tilde{x}_t + W_{oh} \tilde{h}_{t-1} + \tilde{b}_o) \quad (23)$$

$$\tilde{g}_t = sigmoid(\mathcal{Q}_{gx}\mathcal{S}_{gx}\Lambda_{gx}\tilde{x}_t + W_{gh}\tilde{h}_{t-1} + \tilde{b}_g) \quad (24)$$
$$\tilde{c}_t = \tilde{f}_t \odot c_{t-1} + \tilde{x}_t \odot \tilde{g}_t \quad (25)$$
$$h_t = \tilde{o}_t \odot tanh\tilde{c}_t \quad (26)$$
$$\mathcal{p}_{t+1} = (\mathbb{C}h_t) \quad (27)$$

where, x_t is the input and an input update gate is also defined as $\tilde{x}_t$ according to input and the hidden state h_{t-1}. Also, Λ_{ix}, $\Lambda_{fx}, \Lambda_{ox}$, and Λ_{gx} are the input weight matrices, and W_{ix}, W_{fx}, W_{ox}, and W_{gx} are the weight matrices which are applied to recurrently update the matrices of hidden states. Also $\{\mathcal{M}\}, \{\mathcal{Q}\}$, and $\{\Lambda\}$ are the matrices shared by two styles of romantic $\mathcal{S}_{\mathcal{R}}$ and humorous $\mathcal{S}_{\mathcal{H}}$ respectively. The style factored Bi-ALSTM model is implemented as:

$$h_t^f = \tilde{o}_t^f \odot tanh\tilde{c}_t^f \quad (28)$$
$$h_t^b = \tilde{o}_t^b \odot tanh\tilde{c}_t^b \quad (29)$$
$$y_t = W_{hy}^f h_t^f + W_{hy}^b h_t^b + b_y \quad (30)$$

3.2.2 Style-Based Language Generation

To generate the style-based descriptions of images the decoder of the proposed model leverages multi-task learning for sequence tagging [51]. This helps to learn to disentangle the style factors from the text corpus. The input to the SFA-Bi-LSTM module is the image feature matrix $\mathcal{I}_{\mathcal{f}}$ extracted from the image encoder and the text feature embedding matrix $\mathcal{T}_{\mathcal{f}}$ for romantic and humorous text corpus. The matrix $\mathcal{T}_{\mathcal{f}}$ is obtained via fine-tuned GloVe embedding. The extracted features are utilized by the SFA-Bi-ALSTM module according to equations 20-30. The output generated from the proposed Bi-LSTM is soft-attended and generates either romantic or humorous descriptions of images. Also, at time step t, negative log-likelihood loss is incorporated to minimize the loss function and to maximize the probability of each word appearing for both romantic and humorous styles.

$$\mathcal{L}(\Theta) = -\sum_{i=1}^{n} s_{ij} log\hat{s}_{\Theta,ij} + (1 - s_{ij})\log(1 - \hat{s}_{\Theta,ij}) \quad (31)$$

C. Text Summarization Module

The image descriptions generated from the MAA-FIC model and SF-Bi-ALSTM based stylized captioning model are collected and summarized to generate a caption that contains factual, romantic, and humorous content in one caption. This section discusses about the proposed novel transformer-based summarization model, UnMHA-ST with *fTA-Word-Embeddings*, to generate a summarized caption for an image.

3.3 fTA-Word-Embedding (fTA-WE)

To improve the performance and interpretability of the word embedding models, this work proposes a fTA (fastText with Attention) word embedding as depicted in Fig. 4. The proposed embedding module utilizes fastText [52] word embeddings with soft-attention [53] to learn vector representation of words such that the words which are semantically related and close to each other. Also, fTA-WE leverage sub-words to enrich the vector space for rare and unseen words. The context vector c_v for vocabulary $\mathcal{V} = [v_1, \dots, v_{\mathcal{N}}]^T \in \mathbb{R}^{\mathcal{N}\times\mathcal{D}}$ of size $\mathcal{N}$ and word vector size $\mathcal{D}$ for fastText is represented by:

$$c_v = \sum_{i\in[-b,b]-\{0\}} u_{e_j} \quad (32)$$

where, $\mathcal{U} = [u_1, \dots, u_{\mathcal{N}}] \in \mathbb{R}^{\mathcal{N}\times\mathcal{D}}$ represents the sub-words and is used in the calculations of context vectors to efficiently learn a representation for each word. Taking the word $'white'$ as an example the sub words formed for $n = 3$ $grams$ are $< wh, whi, hit, \dots >$. Further, e_j and b are the index of each word and the size of the context window. If w_0 is the masked word index with vector v_{w_0}, the probability of w_0 to occur in the context of $\{w_{-b}, \dots, w_{-1}, w_1, \dots, w_b\}$

$$\mathcal{p}(w_0|W_{[-b,b]-\{0\}}) \propto expv_{w_0}^T c_v \quad (33)$$

The embedding of each word is represented by the embeddings of all sub-words which can be expressed mathematically as:

$$\tilde{u}_w = \sum_{e_j\in\mathbb{S}_w} u_{e_j} \quad (34)$$

The attention mechanism of the fTA-WE embeddings, utilizes key matrix $\mathcal{K}$, value V, and query matrix Q the attention weight and the context vector with attention, defined as:

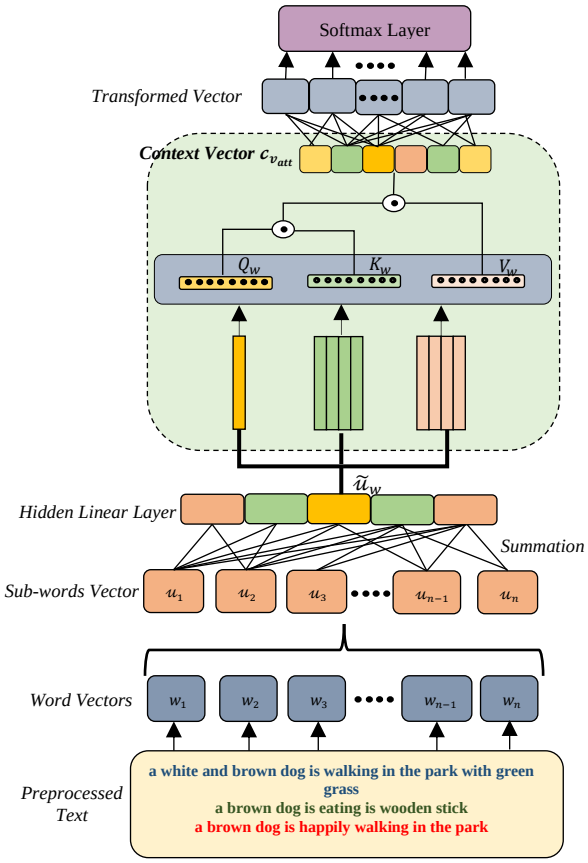


Fig. 4: Proposed fTA-WE

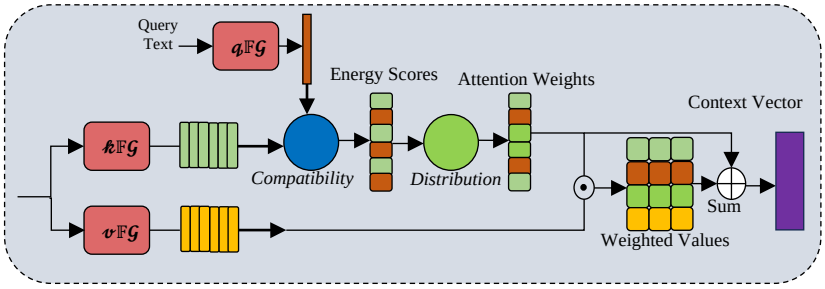


Fig. 5: Proposed Unified Attention Module

$$a_{w_i} = \exp\left(v_{w_0}^T c_v\right) \quad (35)$$

$$c_{v_{att}} = \sum_{i\in[-b,b]-\{0\}} a_{w_i}\left(\sum_{e_j\in\mathbb{S}_w} \tilde{u}_{e_j}\right) \quad (36)$$

The probability of the masked word is given by:

$$p(w_0|W_{[-b,b]-\{0\}}) \propto \exp\left(\tilde{u}_{w_0}^T c_{v_{att}}\right) \quad (37)$$

The attended vector is linearly transformed to obtain the transformed vector representation of the input text. Furthermore, the obtained transformed vector representation is used as an input to the text summarization transformer to generate the summarized caption for the images.

3.4 $UnMHA-ST$ Text Summarization Transformer

This section provides a detailed description of the abstractive transformer-based text summarization model. The summarization model takes input from two captioning mechanisms discussed in section 3.1 and 3.2 respectively. It generates a summarized caption of an image that reflects the factual, romantic, and humorous contents in a single caption. The proposed $UnMHA-ST$ as depicted in Fig. 2, utilizes a unified attention mechanism in addition to multi-head attention. This helps in concurrently capturing

the intra-modal and intermodal interactions of multimodal information which further helps in the generation of their related attended representations. With a unified attention module, the proposed $UnMHA-ST$ also leverages a pointer-generator network and coverage mechanism to solve the problems of rare words which arise due to out-of-vocabulary (OOV) and repetition issues. The input to the transformer network is the text embedding that is generated from the fTA-WE model. These embeddings are added with the positional embedding and are further split into key $\mathcal{k}$, values $\mathcal{v}$, and query $\mathcal{q}$ respectively. To jointly attend information from different representation subspaces at different positions, MHA is employed which is mathematically represented as:

$$Attention(\mathcal{k},\mathcal{q},\mathcal{v}) = softmax(\frac{\mathcal{q}\mathcal{k}^T}{\sqrt{d_k}})\mathcal{v} \quad (38)$$

$$MHA(\mathcal{k},\mathcal{q},\mathcal{v}) = Concat(\mathcal{h}_1,\mathcal{h}_2,\dots,\mathcal{h}_i)W^0 \quad (39)$$

$$\text{where, } \mathcal{h}_j = Attention(\mathcal{k}W_j^{\mathcal{k}},\mathcal{q}W_j^{\mathcal{q}},\mathcal{v}W_j^{\mathcal{v}}) \quad (40)$$

The unified attention (Un-A) module is depicted in Fig. 5 utilizes an iterative alternating attention mechanism [54] that allows a fine-grained exploration of both the query and the document. Un-A do not collapse the query into a single vector instead it evaluated keys and query by using a compatibility function $e = \mathbb{F}(\mathcal{q},\mathcal{k})$. The function $\mathbb{F}$ is the energy function [51]. The energy scores thus obtained are transferred into the attention weights using a distribution function $\mathcal{G}$, $\mathcal{E} = \mathcal{G}(e)$. Therefore, the vector obtained from the compatibility and distribution function is combined with the transformer values v, which are merged or added to represent the context vector in a more compact form.

$$\mathcal{Z}_i = \mathcal{E}_i\mathcal{v}_i \quad (41)$$

$$C = \sum_{i=1}^{d_k}\mathcal{Z}_i \quad (42)$$

The vectors obtained from both the attention mechanisms are further added and passed through the subsequent section of the encoder layer which is FFN.

$$\Omega_t = f(MHA(\mathcal{k},\mathcal{q},\mathcal{v}),C) \quad (43)$$

Also, the decoder structure follows the traditional transformer structure [48] with a unified attention mechanism. The embeddings input to the decoder network are shifted right version of the embeddings at the encoder network. These embeddings are made to pass through the multi-head attention network and the unified attention network. The dumped results are then fed through a series of FC layers and finally a linear activation function followed by a softmax layer, generating the output probabilities. In addition to this, the whole model incorporates a pointer generator network [43] and a coverage mechanism [43]. The context vector is obtained from the encoder output and the output state of the hidden layer inside the decoder and predicts new words for the dictionary. The pointer network's attention provides information about the most significant words at any given time, which is helpful for prediction. The current moment's attention distribution and the sum of weights for the encoder's hidden layer are represented as:

$$b^t = softmax(\Omega_t) \quad (44)$$

$$n_t = \sum_i b^t \mathrm{h}_i \quad (45)$$

Further, vocabulary probability distribution and the probability of distribution of words is mathematically expressed as:

$$\mathcal{P}_{vocab} = softmax(\mathcal{t}'(\mathcal{t}[\Omega_t,n_t] + \mathcal{u}) + \mathcal{u}') \quad (46)$$

Where, $\mathcal{t},\mathcal{t}',\mathcal{u},\mathcal{u}'$ are the learnable parameters. Also, the probability distribution of words is:

$$\mathcal{P}(\mathcal{W}) = \mathcal{P}_{vocab}(\mathcal{W}) \quad (47)$$

Hence, the model through a pointer network must either directly copy words from the source or create new terms to address the out-of-vocabulary issues. Therefore, the pointer-generator network is employed with probability to copy a word from the source text or to generate the word:

$$\mathcal{P}_{gen} = sigmoid\left(w_h^T h_t^* + w_n^T n_t + w_x^T x_t + \mathcal{r}_{ptr}\right) \quad (48)$$

Where, w_n^T, w_h^T, w_x^T, and $\mathcal{r}_{ptr}$ are learnable parameters and $\mathcal{P}_{gen}\epsilon[0,1]$.

Further, $\mathcal{P}_{gen}$ acts as a switch that either generates a word from the vocabulary using sampling from $\mathcal{P}_{vocab}$, or copies a word from the input sequence by sampling from the attention distribution. Also, there exists an extended vocabulary for each document which union of the vocabulary, and all words appearing in the source document.

$$\mathcal{P}(\mathcal{W}) = \mathcal{P}_{gen}\mathcal{P}_{vocab}(\mathcal{W}) + (1-\mathcal{P}_{gen})\sum_{j:w_j} a_{ij} \quad (49)$$

Note, if w is an OOV word, then $\mathcal{P}_{vocab}(\mathcal{W}) = 0$. Also, if w does not appear in the source document,

then $\sum_{j:w_j} a_{ij} = 0$.

For time-step i, negative log-likelihood loss is evaluated for the target word w_i and is given by:

$$\mathcal{L}_i = -log\mathcal{P}(\mathcal{W}_i) \quad (50)$$

$$\mathcal{L} = \frac{1}{T}\sum_{i=0}^{T} \mathcal{L}_i \quad (51)$$

where T is the target sequence length. To overcome the issues related to the repetition of words, a coverage mechanism is employed to direct the attention version to non-repeating words. The attention distributions generated using all previous prediction steps are added to create the coverage vector χ_i^v.

$$\chi_t = \sum_{t'=0}^{t-1} \Omega_{t'} \quad (52)$$

Therefore, equation (43) is modified by considering the effect of the coverage mechanism. Hence, this makes it easier for the attention mechanism to stop attending to the same places repeatedly and stop producing repetitive text as a result.

$$\Omega_{t'} = f(MHA(\mathcal{k}, \mathcal{q}, \mathcal{v}), C, \chi_t) \quad (53)$$

The coverage loss is also defined to penalize repeatedly attending to the same locations.

$$\mathcal{L}_c = \lambda \sum_i \min(\chi_t, \Omega_{t'}) \quad (54)$$

where λ is the balancing parameter. Hence, the total loss for the proposed $UnMHA - ST$ is the combination of equations (53) and (54) respectively,

$$\mathcal{L}_{total} = \mathcal{L} + \mathcal{L}_c \quad (55)$$

4. RESULTS AND DISCUSSIONS

To validate the effectiveness of the proposed MAA-FIC experiments are conducted on the Flickr8K dataset. This dataset contains around 8K images paired with five descriptions for each image. To generate a style-based description of an image FlickrStyle10K dataset is utilized. This dataset contains textual annotations for romantic and humorous styles respectively. For this dataset, only 7K annotations are publicly available. For the task of summarization, the text data obtained from the two captioning methods (discussed in Section III) is collected and utilized for the generation of summarized description of an image reflecting factual, romantic, and humorous elements. The data used for summarization contains around 7000 paragraphs with three sentences (one each for factual, romantic, and humorous caption).

4.1. Implementation Details

For the factual caption generation model, Adam [53] optimizer is utilized for the minimization of cross-entropy loss with a learning rate of $2e-5$. Further, the batch size is set as 64 and the proposed model is trained for 70 epochs. Also, to extract. the text features, fine-tuned GloVe embeddings are utilized with embedding size as 300. To evaluate the performance of the proposed factual image captioning model, BLEU@N [55] and METEOR [56] scores are evaluated.

To generate romantic and humorous captions, ADAM [57] optimizer is utilized at the encoder end with a learning rate of $2e-5$. For language decoder embedding size is set as 300 and the dimensions of the hidden layer of the proposed SF-Bi-ALSTM module is set as 512. The whole model is trained for about 60 epochs with a batch size of 96. To evaluate the performance of generated stylized captions, style transfer accuracy and perplexity of the proposed model are evaluated. Also, the relevancy of the captions generated is evaluated in terms of BLEU@N [55] and METEOR [56].

For the summarization module, a single-layer $UnMHA - ST$ transformer with a pointer generator network and coverage mechanism is incorporated. The model is trained for 100 epochs with a dropout rate of 0.1 and a batch size of 4. Further, the pointer-generator network leverages 300-dimensional GloVe embeddings with a vocabulary size of 400K. To evaluate the performance of the proposed model ROUGE-1, ROUGE-2, and ROUGE-L scores are evaluated.

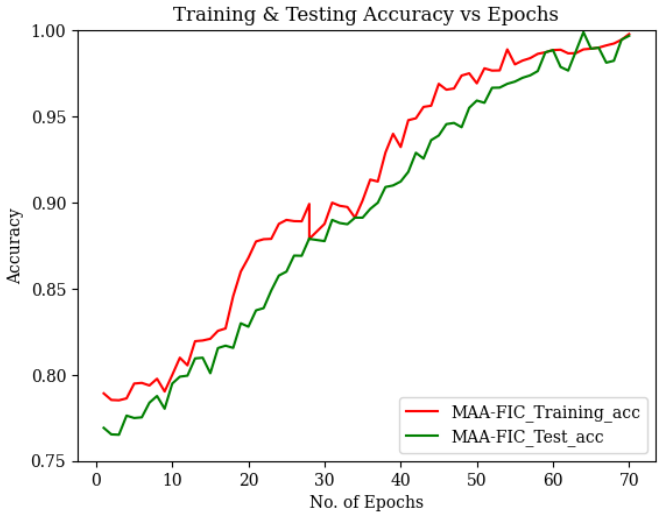
4.2 Experimental Results

This section presents the results for (i) Modified Adaptive Attention-based Factual Image Captioning (MAA-FIC) Model, (ii) Style-based caption generation model, and (iii) Caption Summarization model.

4.2.1 MAA- FIC Model

Table I reports the comparison of the proposed MAA-FIC model with other state-of-the-art on the Flickr8K dataset. The comparison of the proposed model is made in terms of $BLEU@N$ and $METEOR$

(a)



(b)

Fig. 6: (a) Accuracy and (b) Loss curves for the proposed MAA-FIC

TABLE I
RESULTS ON FLICKR8K DATASET

Model	B-1	B-2	B-3	B-4	M
LSTM [21]	66.0	42.0	27.0	18.0	-
g-LSTM [58]	64.7	45.9	31.8	21.6	20.60
Log Bilinear [59]	65.6	42.4	27.7	17.7	17.31
SCA-CNN [6]	68.2	49.6	35.9	25.8	-
Hard Attention [1]	67.0	45.7	31.4	21.3	20.30
m-RNN [20]	48.2	35.7	26.9	20.8	-
AA + Bi-LSTM [26]	70.2	49.1	35.9	26.6	-
Deep-Bi-LSTM [24]	65.5	46.8	32.0	21.5	-
SDCD [60]	67.2	45.1	30.5	21.5	-
JRAN [61]	67.7	47.4	33.2	22.7	20.9
MAA-FIC	**74.8**	**52.3**	**38.7**	**26.8**	**22.6**

scores. MAA-FIC utilized fusion of Inception-V3 and Faster R-CNN at the encoder and provided provides state-of-the-art results when compared with ImageNet-CNN [21], VGGNet-CNN [20] based feature extraction method. Also, Jia et al. [58] utilized semantic information from the image as an extra input to each unit of the LSTM block. Very few works [1] [6] [26] focused on attention mechanism-based architectures and the proposed MAA-FIC utilized modified adaptive attention (MAA). MAA combines channel and spatial attention mechanisms to focus more on most of the important image information as well as the position of the image regions. Furthermore, the MAA provided a significant increase in the BLEU@N and METEOR scores when compared with the [59] [24]. Fig. 6 (a) and (b) depict the accuracy and loss curves for the proposed MAA-FIC. The curves make it evident that with the increase in the number of epochs, the accuracy of the proposed model increases while the losses decrease. Fig.7 represents the qualitative results obtained for the proposed MAA-FIC model. The factual captions obtained for Flickr8K produces syntactically and semantically correct descriptions by focusing more on the salient object regions. Also, by leveraging the modified adaptive attention the proposed model provides a better correlation between the objects. However, there are a few instances where the proposed MAA-FIC model could not detect the correct object. For example, in Fig. 7(b) the MAA-FIC model predicted "*mobile"* instead of "*camera".*

4.2.3 SF-Bi-ALSTM Based Stylized Image Captioning

Table II reports the results using both the romantic references and the humorous references for the proposed SF-Bi-ALSTM-based stylized image captioning module with other state-of-the-art. The comparison is carried out with respect to $BLEU-1$, $BLEU-3$, and $METEOR$ (M) scores with style accuracy (cls) and perplexity (ppl). The results reported in Table II make it evident that (1) given a specified style, the proposed model is tailored to that style outperforming the baseline approaches across different automatic evaluation metrics; (2) the relative performance variation shows that the proposed model can successfully simulate the style factors in caption generation. Further, StyleNet [11] and SF-LSTM [9] incorporate factored LSTM and style-factual LSTM decoders for the generation of stylized captions. These methods proposed end-to-end learning. framework for the generation of factual and style-based descriptions. On the other hand, the proposed model defines a modified style factored Bi-LSTM with captions. Further, the works [10] [12] implemented the task of multi-style-based caption learning using unpaired data but these methods when compared with the proposed stylized captioning model generate more attractive and coherent style-based descriptions.

TABLE II
RESULTS ON FLICKRSTYLE10K DATASET

Style	Model	B-1	B-3	M	cls	ppl
Romantic	StyleNet [9]	13.3	1.5	4.5	57.1	6.9
	MSCap [12]	17.0	2.0	5.4	91.3	-
	SF-LSTM [11]	27.8	8.2	11.2	-	-
	SAN [10]	28.3	8.7	11.5	90.9	9.1
	Detach and attach [62]	24.3	-	-	82.4	-
	Wu et al. [63]	25.4	5.7	9.2	-	-
	Proposed	**29.8**	**9.2**	**11.9**	**92.0**	**9.3**
Humorous	StyleNet [9]	13.4	0.9	4.3	42.5	7.3
	MSCap [12]	16.3	1.9	5.3	88.7	-
	SF-LSTM [11]	27.4	8.5	11.0	-	-
	SAN [10]	27.6	8.1	11.2	87.8	8.4
	Detach and attach [62]	23.0	-	-	89.2	-
	Wu et al. [63]	27.2	5.9	9.0	-	-
	Proposed	**29.2**	**9.0**	**11.8**	**89.8**	**8.7**

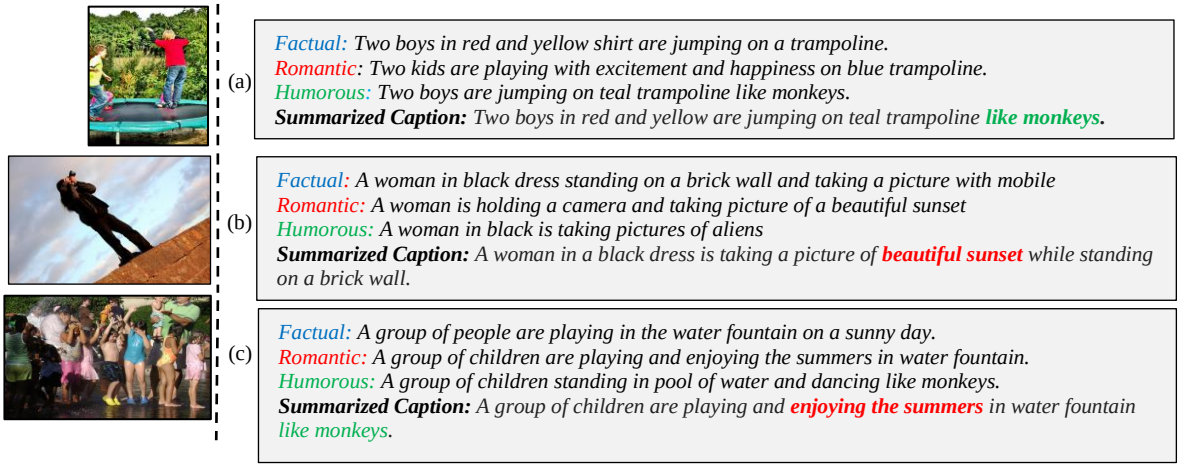


Fig. 7: Qualitative results obtained for the proposed $UnMA$ –CapSumT (a) UnMA-CapSumT Factual + Humorous (b) UnMA-CapSumT Factual + Romantic (c) UnMA-CapSumT Factual + Romantic + Humorous

The qualitative results presented in Fig.7 for the generation of romantic and humorous proves that the generated descriptions describe the content of the image well. The generated captions express the content in a romantic *(A little girl enjoying the joys of childhood with her brother in background)* or humorous *(A group of children standing in pool of water and dancing like monkeys)*. More intriguingly, the descriptions generated are not only romantic or humorous but they also suit the visual content of the image coherently, making the caption visually appealing and relevant. Also, qualitative results were obtained to verify that the sentences so generated provide a strong correlation with the human evaluation.

4.2.4 UnMHA-ST Text Summarization Transformer

The summarized captions generated from the proposed UnMHA-ST is presented in Fig. 7. From Fig. 7 it is evident that there exist some cases as shown in Fig. 7(a), according to the context of the image the summarized captions depict only the factual and humorous elements and for a few images as shown in Fig. 7(b) romantic style has dominated with the factual element. Further, there are cases (Fig. 7(c)) in which the proposed summarization framework can successfully include romantic and humorous styles with factual elements.

TABLE III
ABLATION RESULTS OBTAINED FOR SUMMARIZED CAPTIONS

Model	Transformer	Embedding	R-1	R-2	R-L
Model-1	Baseline [48]	Doc2Vec	20.18	4.33	13.71
Model-2	Baseline [48]	GloVe	21.3	4.43	14.19
Model-3	**Baseline** [48]	**fastText**	**23.46**	**4.65**	**14.77**
Model-4	$UnMHA-ST$	fTA-WE	26.72	5.71	15.24
Model-5	$UnMHA-ST$ + Pointer Generator Network	fTA-WE	28.32	6.43	17.72
Model-6	**$UnMHA-ST$ + Pointer Generator Network + Coverage Mechanism**	**fTA-WE**	**30.11**	**6.73**	**18.22**

Table III reports the ROUGE-1, ROUGE-2, and ROUGE-L scores for the baseline transformer model [48] with Doc2Vec, GloVe, and fastText word embeddings. From the results reported it is evident that the fastText word embeddings are superior as they can derive word vectors for unknown words or out of vocabulary words. Further, the baseline transformer model is replaced with the proposed $UnMHA-ST$ and fTA- WE word embedding which results in significant improvement of R-1, R-2, and R-L scores. To overcome the OOV issues the proposed $UnMHA-ST$ transformer is utilized with pointer-generator and to further enhance the performance of the proposed framework and also to avoid the repetition issues $UnMHA-ST$ is equipped with coverage mechanism. This provides significant improvements by generating fluent summarized captions. Also, box plot-based model performance

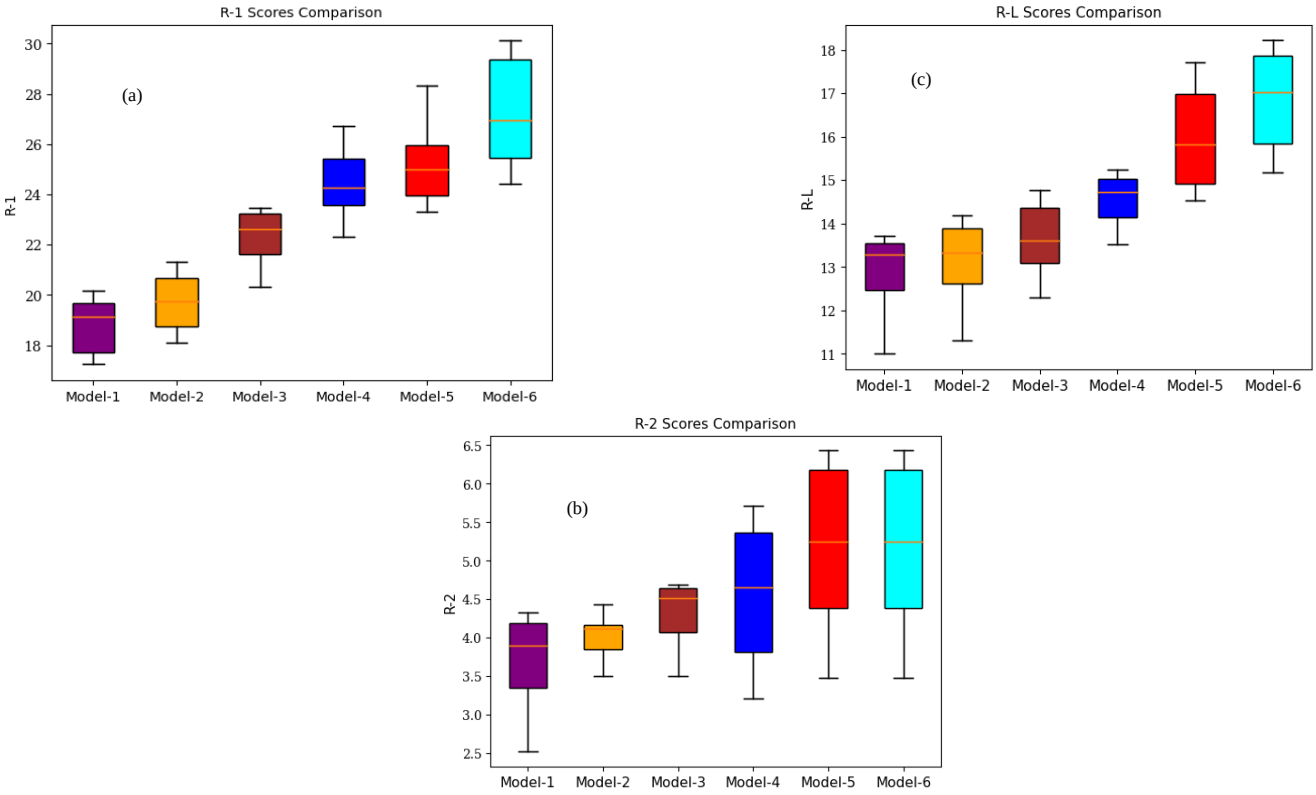


Fig. 8: Comparison for Quantitate Results Obtained for the proposed Framework (a) R-1, (b) R-2, and (c) R-L

GT: A group of people are playing in the water fountain on a sunny day. A group of children are playing and enjoying the summers in water fountain. A group of children standing in pool of water and dancing like monkeys.
Model-1: ***A children are playing with water in the the summers in water pool***
Model-2: Children are playing in water and dancing in water enjoying in water in the the summers
Model-3: Children are dancing in water like monkeys enjoying in water fountain in the the summers
Model-4: ***A group of people are playing in water fountain and dancing in pool of the summers in water fountain.***
Model-5: ***A group of people are enjoying and enjoying in water fountain like monkeys.***
Model-6: ***A group of children are playing and enjoying the summers in water fountain like monkeys.***

GT: A women in blue clothes sits alone against a brick wall and looking outside a window. Woman sits against brick wall in coffee shop having sweet coffee waiting for her lover. A woman sits in a café looking at a lizard on the brick red wall
Model-1: ***A women sitting against red wall alone in café drinking coffee in café waiting for lover in coffee shop.***
Model-2: A women sitting near red wall in cafeteria wait for lizards at red wall and waiting for lover in cafeteria.
Model-3: A women in blue is looking at lizards on red wall and sitting against red wall in cafeteria.
Model-4: ***A women in blue sitting in café against red wall waiting for lover in café while looking at lizards in wall.***
Model-5: ***A women sitting in cafeteria against red wall waiting for her coffee lover and looking at lizards on red wall in cafeteria.***
Model-6: ***A women in blue clothes sitting alone against red wall in café having a sweet coffee and waiting for her lover while looking at lizards***

Fig.9: Observed Qualitative Results for the proposed *UnMA*-CapSumT *(Text in Red represents the repeated words that appeared in the summarized captions)*: Model -6 (proposed model) provides a summarized caption for an image without repetition of words and out-of-vocabulary issues

comparison in terms of R-1, R-2, and R-L is shown in Fig. 8.

Further, Fig. 9 presents the qualitative ablated summarized captions for Model 1 to Model-6. Model-1, the baseline transformer model with Doc2Vec word embedding generated a poor-quality summarized caption with redundant and repeated information. Further, fastText embedding with the baseline transformer model (Model-3) provided a slight improvement in the generated summarized captions. With the use of the proposed $UnMHA - ST$ and fTA-WE in Model-4 provided an improvement in results but still there exists repetition of words like “in water fountain” in first case and “café” and “wall” in second case. In Model-5, the pointer-generator network adaptively points to the contextual words with appropriate styles hence resulting in generation of romantic and humorous elements in the summarized captions. For the second case, Model-5 generates a word “cafeteria” that is synonym to the word “café” or “coffee shop”. Further, Model-6 generated a summarized caption that provided a strong

correlation with the human evaluation by generating syntactically and semantically correct descriptions by highlighting both factual and stylized elements. Also, the summarized caption generated is free from OOV and repetition problems with the use of a pointer-generator network and coverage mechanism.

5. CONCLUSION

This paper presents a novel caption summarization framework, $UnMA-$CapSumT to generate summarized captions highlighting the factual, romantic, and humorous elements in a single caption. The proposed framework is divided into two stages: (i) generation of factual, romantic, and humorous descriptions of images using the MAA-FIC model and SF-Bi-ALSTM-based stylized image captioning model, (ii) using the descriptions generated in (i) to produce a summarized single line caption for the images by incorporating multi-head attention and unified attention driven $UnMHA-ST$ transformer. The proposed $UnMHA-ST$ transformer utilizes a pointer-generator network and coverage mechanism to avoid the issues related to OOV and repetition problems. The summarized caption generated provides an improvement in learning knowledge of factual content and its associated linguistic styles. Also, the observed qualitative results make it evident that the proposed framework provided a strong correlation with human evaluation by generating semantically and syntactically correct descriptions. Further, the improvements in the performance and interpretability of the proposed framework are enhanced with the utilization of an efficient word embedding fTA-WE. To validate the effectiveness of the proposed framework extensive experiments are conducted on Flickr8K and FlickrStyle10K. Experimental results prove that the proposed MAA-FIC, SF-Bi-ALSTM-based image captioning model, and $UnMHA-ST$ summarization transformer model provided state-of-the-art results and generates visually and grammatically correct factual, romantic, humorous, and summarized captions for a given image.

REFERENCES

[1] P. Anderson, X. He, C. Buehler, D. Teney, M. Johnson, S. Gould and L. Zhang, "Bottom-up and top-down attention for image captioning and visual question answering," in *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018.

[2] G. Li, L. Zhu, P. Liu and Y. Yang, "Entangled transformer for image captioning," in *IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, 2019.

[3] S. Wu, J. Wieland, O. Farivar and J. Schiller, "Automatic alt-text: Computer-generated image descriptions for blind users on a social network service," in *ACM Conference on Computer-Supported Cooperative Work and Social Computing*, Portland, Oregon, 2017.

[4] D. Gurari, Q. Li, A. Stangl, A. Guo, C. Lin, K. Grauman, J. Luo and J. Bigham, "Vizwiz grand challenge: Answering visual questions from blind people," in *e IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[5] S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. Zitnick and D. Parikh, "VQA: Visual question answering," in *IEEE International Conference on Computer Vision*, 2015.

[6] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu and T.-S. Chua, "SCA-CNN: Spatial and Channel-wise Attention in Convolutional Networks for Image Captioning," in *arXiv:1611.05594v2*, 2017.

[7] Z. Deng, Z. Jiang, R. Lan, W. Huang and X. Luo, "Image captioning using DenseNet network and adaptive attention," *Signal Processing: Image Communication*, vol. 85, no. 12, July 2020.

[8] Y. Wang, N. Xu, A.-A. Liu, W. Li and Y. Zhang, "High-Order Interaction Learning for Image Captioning," *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, 2021.

[9] C. Gan, Z. Gan, X. He and J. Gao, "Stylenet: Generating attractive visual captions with styles," in *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Hawaii, 2017.

[10] G. Li, Y. Zhai, Z. Lin and Y. Zhang, "Similar Scenes arouse Similar Emotions: Parallel Data Augmentation for Stylized Image Captioning," in *ACM Multimedia Conference*, Virtual Event, China, 2021.

[11] T. Chen, Z. Zhang, Q. You, C. Fang, Z. Wang, H. Jin and J. Luo, "“Factual” or “Emotional”: Stylized Image Captioning with Adaptive Learning and Attention," in *European Conference on Computer Vision*, Munich, Germany, 2018.

[12] L. Guo, J. Liu, P. Yao, J. Li and H. Lu, "MSap: Multi-Style Image Captioning With Unpaired Stylized Text," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[13] R. Nallapatti, F. Zhai and B. Zhou, "SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[14] R. Nallapatti, B. Zhou, C. N. dos santos, C. Gulcehre and B. Xiang, "Abstractive text summarization using sequence-tosequence rnns and beyond," in *arXiv:1602.06023*, 2016.

[15] P. Young, A. Lai, M. Hodosh, and J. Hockenmaier, "From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions," *Transactions of the Association for Computational Linguistics*, vol. 2, pp. 67-78, 2014.

[16] Y. Yang, C. L. Teo, H. Daume and Y. Aloimono, "Corpus-guided sentence generation of natural images," in *Empirical Methods in Natural Language Processing*, Edinburgh United Kingdom, 2011.

[17] M. Mitchell, J. Dodge, A. Goyal, K. Yamaguchi, K. Stratos, X. Han, A. Mensch, A. Berg, T. Berg and H. Daume, "Generating image descriptions from computer vision detections," in *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, Avignon , 2012.

[18] A. Farhadi, M. Hejrati, M. Sadeghi, P. Young, C. Rashtchian, J. Hockenmaier and D. Forsyth, "Every picture tells a story: Generating sentences from images," in *Proceedings of the European Conference on Computer Vision*, Crete, Greece , 2010.

[19] M. Hodosh, P. Young and J. Hockenmaier, "Framing image description as a ranking task: Data, models and evaluation metrics," *Journal of Artificial Intelligence Research*, vol. 47, pp. 853-899, August 2013.

[20] J. Mao, W. Xu, Y. Yang, J. Wang, Z. Huang and A. Yuille, "Deep captioning with multimodal recurrent neural networks (m-rnn)," in *arXiv:1412.6632*, 2015.

[21] O. Vinyals, A. Toshev, S. Bengio and D. Erhan, "Show and Tell: A Neural Image Caption Generator," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* , Boston, USA, 2015.

[22] X. Jia, E. Gavves, B. Fernando and T. Tuytelaars, "Guiding the long-short term memory model for image caption generation," in *IEEE International Conference on Computer Vision*, Santiago, Chile, 2015.

[23] J. Donahue, L. Hendricks, M. Rohrbach, . S. Venugopalan, . S. Guadarrama, K. Saenko and . T. Darrell, " Long-Term Recurrent Convolutional Networks for Visual Recognition and Description," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 677-691, 2015.

[24] C. Wang, H. Yang and C. Meinel, "Image Captioning with Deep Bidirectional LSTMs," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 14, no. 2s, pp. 1-20, 2018.

[25] J. Lu, C. Xiong, D. Parikh and R. Socher, "Knowing When to Look: Adaptive Attention via A Visual Sentinel for Image Captioning," in *arXiv:1612.01887v2*, 2017.

[26] D. Sharma, C. Dhiman and D. Kumar, "Automated Image Caption Generation Framework using Adaptive Attention and Bi-LSTM," in *IEEE Delhi Section Conference (DELCON)*, Delhi, 2022.

[27] Z. Zhang, Q. Wu, Y. Wang and F. Chen, "Exploring region relationships implicitly: Image captioning with visual relationship attention," *Image and Vision Computing*, vol. 109, May 2021.

[28] Q. You, H. Jin, Z. Wang, C. Fang and J. Luo, "Image captioning with semantic attention," in *IEEE Conference on Computer Vision and Pattern Recognition*, Caesars Palace, 2016.

[29] F. Xiao, X. Gong, Y. Zhang, Y. Shen, J. Li and X. Gao, "DAA: Dual LSTMs with adaptive attention for image captioning," *Neurocomputing*, vol. 364, pp. 322-329, 2019.

[30] C. Yan, Y. Hao, L. Li, J. Yin, A. Liu, Z. Mao, Z. Chen and X. Gao, "Task-Adaptive Attention for Image Captioning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 1, pp. 43-51, 2021.

[31] F. Xiao, W. Xue, Y. Shen and X. Gao, "A New Attention-Based LSTM for Image Captioning," *Neural Process Letters*, vol. 54, p. 3157–3171, Feburary 2022.

[32] A. Mathews, L. Xie and X. He, "Semstyle: Learning to generate stylised image captions using unaligned text," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[33] W. Zhao, X. Wu and X. Zhang, "MemCap: Memorizing Style Knowledge for Image Captioning," in *AAAI Conference on Artificial Intelligence*, 2020.

[34] A. P. Mathews, L. Xie and X. He, "SentiCap: Generating Image Descriptions with Sentiments," in *AAAI'16: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, Phoenix, Arizona, 2016.

[35] K. Shuster, S. Humeau, H. Hu, A. Bordes and J. Weston, "Engaging Image Captioning via Personality," in *arXiv:1810.10665*, 2018.

[36] T. Li, Y. Hu and X. Wu, "Image Captioning with Inherent Sentiment," in *IEEE International Conference on Multimedia and Expo (ICME)*, Shenzhen, China, 2021.

[37] H. Oliveira, R. Ferreira, R. Lima, R. D. Lins, F. Freitas, M. Riss and S. J. Simske, "Assessing shallow sentence scoring techniques and combinations for single and multi-document summarization," *Expert Systems with Applications*, vol. 65, pp. 68-86, 2016.

[38] M. Kar, S. Nunes and C. Ribeiro, "Summarization of changes in dynamic text collections using latent dirichlet allocation model," *Information Processing & Management*, vol. 51, no. 6, pp. 809-833, 2015.

[39] P.-E. Genes and G. Lapalme, "Fully abstractive approach to guided summarization," in *50th Annual Meeting of the Association for Computational Linguistics*, 2012.

[40] A. Louis, "A bayesian method to incorporate background knowledge during automatic text summarization," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, Maryland, 2014.

[41] S. Banarjee, P. Mitra and K. Sugiyama, "Multi-document abstractive summarization using ilp based multi-sentence compression," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[42] A. M. Rush, S. Chopra and J. Weston, "A Neural Attention Model for Abstractive Sentence Summarization," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, 2015.

[43] A. See, P. J. Liu and C. D. Manning, "Get To The Point: Summarization with Pointer-Generator Networks," in *arXiv preprint arXiv:1704.04368*, 2017.

[44] W. -T. Hsu, C.-K. Lin, M.-Y. Lee, K. Min, J. Tang and M. Sun, "A Unified Model for Extractive and Abstractive Summarization using Inconsistency Loss," *56th Annual Meeting of the Association for Computational Linguistics*, p. 132–141, 2018.

[45] J. Li, C. Zhang, X. Chen, Y. Cao, P. Liao and P. Zhang, "Abstractive Text Summarization with Multi-Head Attention," in *International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, 2019.

[46] K. Yao, L. Zhang, D. Du, T. Luo, L. Tao and Y. Wu, "Dual Encoding for Abstractive Text Summarization," *IEEE Transactions on Cybernetics*, vol. 50, no. 3, pp. 985 - 996, 2020.

[47] J. Deaton, A. Jacobs, K. Kenealy and A. See, "Transformers and Pointer-Generator Networks for Abstractive Summarization," pp. 1-9, 2019.

[48] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and . I. Polosukhin, "Attention is all you need," in *NIPS*, 2017.

[49] S. Ren, K. He, G. Ross and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 1137-1149, 2017.

[50] C. Szegedy, V. Vanhoucke, S. Loffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016.

[51] M.-T. Luong, Q. V. Le, I. Sutskever, O. Vinyals and L. KAiser, "Multi-task Sequence to Sequence Learning," in *arXiv:1511.06114v4*, 2015.

[52] P. Bojanowski, E. Grave, A. Joulin and T. Mikolov, "Enriching Word Vectors with Subword Information," in *arXiv:1607.04606*, 2017.

[53] D. Bahdanau, . K. Cho and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," in *arXiv:1409.0473*, 2014.

[54] M. Daniluk, T. Rocktäschel, J. Welbl and S. Riedel, "Frustratingly Short Attention Spans in Neural Language Modeling," in *arXiv:1702.04521*, 2017.

[55] K. Papineni, S. Roukos, T. Ward and W.-J. Zhu, ""Bleu: a method for automatic evaluation of machine," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 2002.

[56] S. Banerjee and A. Lavie, "Meteor: An automatic Metric for MT Evaluation with Improved Correlation with Human Judgments," in *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, 2005.

[57] D. P. Kingma and J. Ba, "Adam: A method for stochastic," in *arXiv:1412.6980*, 2014.

[58] X. Jia, E. Gavves, B. Fernando and T. Tuytelaars, "Guiding the long-short term memory model for image caption Generation," in *arXiv:1509.04942*, 2015.

[59] R. Kiros, R. Salakhutdinov and R. Zemel, "Multimodal Neural Language Models," in *31st International Conference on Machine Learning*, 2014.

[60] S. Ding, S. Qu, Y. Xi and S. Wan, "Stimulus-driven and concept-driven analysis for image caption generation," *Neurocomputing*, vol. 398, pp. 520-530, 2020.

[61] C. Wang and X. Gu, "Learning joint relationship attention network for image captioning," *Expert Systems with Applications,* vol. 211, p. 118474, 2023.

[62] Y. Tan, Z. Lin, P. Fu, L. Wang, Y. Cao and W. Wang, "Detach and Attach: Stylized Image Captioning without Paired Stylized Dataset," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022.

[63] X. Wu, W. Zhao and J. Luo, "Learning Cooperative Neural Modules for Stylized Image Captioning," *International Journal of Computer Vision,* vol. 130, p. 2305–2320, 2022.

Check for updates

Conflict Resolution Quarterly

WILEY

RESEARCH ARTICLE

Validation of Nonviolent Work Behavior (NVWB) Scale and Its Relation With Workplace Conflict

Suhans Bansal[1] | Sarika Kumari[2] | Naval Garg[3]

[1]School of Business, University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India | [2]Dept. of Telecommunication, CGCA, New Delhi, India | [3]Department of University School of Management and Entrepreneurship, Delhi Technological University, New Delhi, India

Correspondence: Suhans Bansal (suhansbansal@gmail.com)

Received: 12 June 2024 | **Revised:** 19 September 2024 | **Accepted:** 30 October 2024

Funding: The authors received no specific funding for this work.

Keywords: nonviolent work behavior | psychometric properties | workplace conflict

ABSTRACT

The current study assessed the reliability and validity of the nonviolent work behavior (NVWB) scale with the help of three studies. Study 1 examined the factor structure of the scale using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Study 2 explored the internal consistency reliability, convergent, divergent, and predictive validity using different statistical estimates. It also examined the measurement invariance of the scale. Similar to the original scale, the EFA and CFA also yielded four factors: "NVT-Stress," "NVT-Ability," "NVT-Constructive," and "NVT-Ignore." The results also revealed adequate Cronbach's alpha, composite reliability, and McDonald's omega, which established the reliability of the scale. The positive correlation between NVWB and attitude to nonviolence scores confirmed the scale's convergent validity. It was also revealed that four dimensions of NVWB significantly predict relationship and task conflict. The validation of the scale was based on three independent studies conducted on three different samples.

1 | Introduction

1.1 | Violence at the Workplace

The rising cases of workplace violence have drawn the attention of managers, researchers, and policymakers worldwide. The instances of violence in the workplace are inevitable as they are linked with human interactions, which are influenced by various individual and organizational factors like personality, viewpoints, value systems, and leadership style (Leiß and Rausch 2022). Other such reasons for the increase in workplace violence are workplace conflict and the tendency of employees to respond in a similar way in which they have been attacked, resulting in a conflict spiral (Arnetz et al. 2018; Biswas and Chakraborty 2023). This spiral leads to an exponential growth of hatred and mistrust among the employees, resulting in an increase in aggression and violence across the organization (Sarkar and Garg 2020). With time, organizational culture starts tolerating and accommodating workplace violence, wherein employees' mild and moderate aggressions are often ignored (Julia, Rajkumar, and John 2022; Sakka and Ahammad 2020). Therefore, researchers like Bodhi, Singh, and Joshi (2021), Deng, Hao, and Yang (2022), and Sarkar and Garg (2020) have stressed that employees and employers must incorporate various mechanisms to respond to conflict and violence in the workplace. One such solution is practicing nonviolence in the workplace.

1.2 | Nonviolence

Social scientists have suggested using the tools of Positive Psychology 2.0 (PP2.0), such as gratitude and nonviolence, among others, to modify human behavior (Chakraborty and Biswas 2019; Bansal, Garg, and Singh 2022; De Kerpel and

© 2024 Wiley Periodicals LLC.

Van Kerckhove 2023). Nonviolence is "more than nonappearance of violence or non-harming. Rather, it is a condition of mental ablution and positive acts through body, mind, and speech" (Sarkar and Palo 2019, p. 65). It has also been defined as efforts to counter and limit human anger, hatred, and hostility (Asthana 1990). It includes self-introspection that helps in leveraging one's internal resources of compassion, love, and respect (Walz and Ritchie 2000). The notion of nonviolence has been studied extensively in social, political, and spiritual contexts. Unfortunately, there have been very few efforts to ingrain the tenants of nonviolence in the organizational contexts, even though exploring nonviolence at the workplace is essentially pivotal for two critical reasons. First, restrictive measures like disciplinary actions and zero violence tolerance policy have their own negative repercussions. Excessive restrictions are perceived as an infringement of personal freedom. It also creates an atmosphere of mistrust, fear, and suspicion among employees, which leads to poor employee morale and motivation (Lee 2021; Deng, Hao, and Yang 2022; Badar, Aboramadan, and Plimmer 2023). Second, contrary to violence, nonviolence is a positive phenomenon that breeds love, compassion, forgiveness, trust, and faith among employees (Mayton, Diessner, and Granby 1996; Mayton 2009; Bhawuk 2011), thereby generating resources that can be used to peacefully tackle negative events in the workplace. Positive psychologists have enumerated the benefits associated with these human virtues. Compassion promotes workplace spirituality and organizational citizenship behavior among educators (Mahipalan 2022). Thus, workplace nonviolence seems doubly beneficial for an organization. On the one hand, it counters workplace aggression and violence, and on the other hand, it is often accompanied by the above-mentioned desirable human values.

1.3 | Nonviolent Work Behavior

Only a handful of researchers, including Bhalerao and Kumar (2015), Chakraborty and Biswas 2020, Sarkar and Palo (2019), and Lata and Garg (2023), have emphasized the importance of nonviolence in the workplace. One such seminal work is by Bhalerao and Kumar (2015). The authors defined nonviolent behavior as a "propensity of an individual to constructively respond to provocation such as violation of norms of respect and justice at workplace or any provocation in a way so as to not cause harm to the opponent in thoughts, words, and deeds" (Bhalerao and Kumar 2015, p. 41). According to their definition, nonviolent work behaviours (NVWBs) are one's attitudinal and behavioral inclinations to respond positively and constructively when one faces challenges and violence in the workplace. The underlying philosophy is to respond or react nonviolently to others' incivility, aggression, and violence. Even when situations include violation of the norms of justice and respect, NVWB promotes one to not cause any harm to others.

Further, Bhalerao and Kumar (2015) have suggested four dimensions of NVWBs. The first dimension is NVT-Ignore, wherein the victims of workplace violence ignore the attackers' violent/aggressive acts. The employees do not feel conflict/violence is worthy of responding to. In the second dimension, NVT-Stress, the employee sees violence as an outcome of attackers' stress (Verma and Garg 2024). The employee gives the benefit of the doubt to the aggressor and diffuses the spiral of violence (Bansal and Garg 2024). The third dimension is NVT-Ability, wherein violence is seen as a result of the inability of the aggressor to cope with his/her present situation. The victim gives a compassionate and considerate view of workplace violence and diffuses the spiral of violence. The last dimension is the NVT-Constructive. In this dimension, the employee brings out a change in self and responds pro-socially to the violence, thereby diffusing its spiral. Instead of a revengeful response, employees resort to NVWB to break the shackles of the cycle of violence by ignoring, understanding, or reacting constructively to others' aggression (Bhalerao and Kumar 2015).

1.4 | Why Does the NVWB Scale Need Revalidation?

Although the development of the NVWB scale was a significant contribution, the scale's validation was marred with several shortcomings. Firstly, the scale was developed on a sample of 158 respondents only, and the sample was highly skewed in favor of males ($n = 112$) and young professionals in the age category of 20–29 years ($n = 123$). Secondly, EFA was applied using varimax rotation, which is an orthogonal rotation that is based on the statistical assumptions that no relationship exists among extracted factors, and they are not correlated (Watkins 2018). However, researchers suggest using oblique rotation while examining psychological constructs, like nonviolence, as it allows factors to correlate (Reio and Shuck 2014; Treiblmaier and Filzmoser 2010). Watkins (2018, p. 233) highlights that "everything measured in the social sciences is correlated to some degree," therefore, "oblique rotations will often be a more realistic representation of the data, will provide a solution that allows for easier interpretation, and will give the researcher additional information not available in orthogonal rotations" (Fabrigar and Wegener 2012, p. 78). Thirdly, the factorial structure extracted by EFA was not confirmed with the help of confirmatory factor analysis (CFA). It is imperative to check the fitness of the EFA model using CFA and on a separate sample (Koenig and Zaben 2021; Boateng et al. 2018) in order to confirm the extracted factor structure. Fourthly, internal consistency reliability was assessed using Cronbach's alpha (CA) values only. Given the limitations of CA (Boateng et al. 2018), it is suggested that other stringent indicators like composite reliability (CR) and McDonald's omega (MO) be used. Lastly, the original study was limited to the examination of construct validity only, whereas the predictive validity was not evaluated.

1.5 | Contributions

Given the limitations of the validation process of the scale, the present study revisits the psychometric properties of the scale using stringent and rigorous statistical analysis. For instance, this study uses a larger sample size to explore and validate the factor structure. Second, the current study adopts a rigorous scale validation process as suggested by Koenig and Zaben (2021) and American Educational Research Association (2014). Third, the factor extraction is based on oblique rotation rather than orthogonal rotation, thereby satisfying the

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21456 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

statistical assumption that factors are correlated (according to the suggestions of Watkins 2018; Reio and Shuck 2014; American Educational Research Association 2014; Treiblmaier and Filzmoser 2010). Fourth, in accordance with the suggestions of Koenig and Zaben (2021) and Boateng et al. (2018), this study uses two independent samples to explore the EFA and CFA. Fifth, the current research uses more stringent reliability measures like CR and MO to assess the reliability of the scale. Last, the present study also explores predictive validity and test–retest reliability of the scale.

Further, this study is structured in the following sections. The first section discusses the theoretical framework of the study. The second section elaborates upon the three studies that assess the psychometric properties of the NVWB scale. The last sections provide the discussion, implications, and limitations along with the future scope of the study.

1.6 | Theoretical Framework

This study is based on two theories, which are the concept of *Ahinsa* and the *Broaden and Built Theory* (Fredrickson 2004). The Concept of Ahinsa advocates for practicing nonviolence even in the event of one being attacked by others, and thereby nonviolence practicing individual can diffuse the spiral of violence (Bhalerao and Kumar 2015). Simply put, this concept suggests that when an individual practices nonviolence, they keep themselves calm and assume that conflicting and violent individuals may have their own disabilities due to which they are being in conflict or are violent. Such disabilities may include not being able to handle work-related stress or problems at home which might be hampering that person to focus on work related tasks (Sarkar and Garg 2021). Hence, the person practicing nonviolence may choose to ignore such individuals or situations and try to improve their abilities that might be a root cause of conflict among other measures to control conflicting situations at the workplace (Sarkar and Garg 2020). Moreover, broaden-and-build theory of positive emotions states that positive emotions like gratitude "broaden people's momentary thought-action repertoires and build their enduring psychological and social resources" (Fredrickson 2004, p. 147). Therefore, it was hypothesized that NVWB scale would have a negative relationship with the relationship and task relationship scale and would be negative predictors of workplace conflict. Also, based on the concept and theory, it was hypothesized that nonviolent work scale behaviorwould have a positive correlation with attitude to nonviolence (ANV) scale. Therefore, the ANV scale has been used as a convergent indicator for NVWB scale as both the scales are intended to measure scores of nonviolence.

2 | Common Methodology

The present study is based on three independent studies. The first study explored and validated the factorial structure of the scale using EFA and CFA. The second study examined the reliability and validity and evaluated the measurement invariance (MI) of the scale. Lastly, the third study evaluated test–retest reliability of the NVBW scale. A detailed methodology was provided at the start of each study.

3 | Study 1: Factorial Structure

3.1 | Methods

The sample of this study was divided into two subsamples, EFA and CFA, which were to be conducted on two different samples. EFA "is a variable reduction technique that identifies the number of latent constructs present within a set of variables, assuming that the underlying latent construct/s cannot be measured directly" (Koenig and Zaben 2021, p. 6) and CFA "allows for the systematic comparison of an alternative a priori factor structure [identified by EFA] based on systematic assessment procedures and estimates the relationship between latent construct, which have been corrected for measurement errors" (Boateng et al. 2018, p. 11). EFA was conducted using IBM SPSS v21, and CFA was conducted on AMOS v4. The sample of 643 Indian working professionals was split into two independent subsamples using the sample split option in the IBM SPSS v21. The sample was divided based on odd and even serial numbers. The first subsample had participants from even serial number with $n_1 = 321$, and the second subsample had participants from odd serial number with $n_2 = 322$. The Ssubsample 1 was subjected to EFA, and the Subsample 2 was subjected to CFA.

3.1.1 | Data Collection Procedure

We conducted a series of workshops on violence in the workplace and its consequences. Topics of the workshop included and were not limited to an introduction to violence at the workplace and traditional and modern forms of workplace violence among others. Also, we created groups related to the workshop on Whatsapp and Telegram group. A structured form created on Google forms was circulated over these groups prior to the start of the workshop. Moreover, to receive the maximum participation, we approached the human resource managers of different organizations in Indian industries. To convince these managers to send their employees to the workshops, we told them that we would be conducting series of studies as well, whose findings and implications will be shared with them as well. Average time taken by the participants to complete the self-administered questionnaire was 25 min.

3.1.2 | Participants

There were 811 participants in the first set of workshops conducted in the month of October. The form was circulated to all of them, and we received data from 643 participants. There were 325 male and 318 female participants. There were 155, 171, 186, and 131 employees belonging to "below 30 years," "31–40 years," "41–50 years," and "above 51 years" age categories. Of these 643 participants, 329 were unmarried, 238 married, 26 widowed, and 50 were separated from their partners (including divorced). There were 185 participants who had work experience of 2–5 years, 169 had 6–15 years of work experience, 153 had 16–25 years of work experience, and 136 had more than 25 years of work experience. Table 1 provides an account of respondents' demographics.

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21456 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.1.3 | Instrument

A structured questionnaire in the form of google form was shared with the participants, which was divided into three sections. The first section provided a short description of the study and ethical statements and captured the consent of the participants. The second section captured the demographic information, and the third section measured the NVWB of the participants using a scale developed by Bhalerao and Kumar (2015). A few of the statements are "I inquire if the subordinate is upset due to some reason, if he/she threatens me" and "I excuse verbal insults made by coworkers." This scale is a 32-item and five-point Likert scale rated between 1 (*never*) and 5 (*usually*). The scores of the scale were interpreted using average scores, that is, participants with less than 3 as average scores were interpreted to show lower NVWBs and those with scores above 3 were interpreted to show higher NVWBs (Bhalerao and Kumar 2015).

Moreover, the form was programmed to end the survey if any participant opted not to participate in the study after reading the ethical and privacy statements. The original scale in English was used in this study for two reasons. First, English language is introduced to Indian students at very early stage of their schooling (Garg and Mehak 2021). Thus, they become well-acquainted with the language till the time they join the labor market. Second, Indian industry has lingual diverse workforce, and English is preferred as a primary language for communication in Indian companies especially at supervisory and above levels (Tsimpli et al. 2020). Thus, English version of the scale was proposed for the study.

3.1.4 | Ethical Concerns

In the form, short information on the research was provided to understudies, and participants' informed consent was obtained. Their anonymity was promised, while strict academic use of data was assured. They were also promised that their data would not be shared with any other party, which reduced the hesitancy of the respondents.

3.2 | Results

Kaiser–Meyer–Olkin (KMO) test of sampling adequacy and Bartlett's test of sphericity were applied to the first subsample to check the suitability of the data for EFA (Shkeer and Awang 2019). The value of KMO was above the recommended value of 0.70 (KMO = 0.817), and Bartlett's test of sphericity was significant at 0.05 level of significance, which asserted the appropriateness of data for EFA. Principal axis factoring with promax-oblique-rotation was selected in IBM SPSS v21 to conduct EFA. A quad-dimensional factor solution with an Eigenvalue of 13.802 that explains 60.78% of the total variance was obtained from the test. All items of the scale were retained as they had factor loadings above the threshold value of 0.5, as suggested by Maskey, Fei, and Nguyen (2018) and Hair et al. (2015). Similar to the original scale, they were named as NVT_Stress, NVT_Ability, NVT_Constructive, and NVT_Ignore. Each dimension has eight statements. The results of EFA are in congruence with the original study by Bhalerao and Kumar (2015). Table 2 reports the factor structure along with factor loadings, variance explained, commonalities (h^2), and eigenvalues of the NVWB scale.

TABLE 1 | Descriptive statistics.

		Study 1				
Variables	**Category**	**Subsample 1**	**Subsample 2**	**Total of Subsamples 1 and 2**	**Study 2**	**Study 3**
Gender	Male	167	158	325	324	95
	Female	155	163	318	267	93
Age	Below 30 years	64	91	155	247	38
	31–40 years	89	82	171	138	57
	41–50 years	99	87	186	107	48
	Above 51 years	70	61	131	99	45
Marital status	Unmarried	178	151	329	288	90
	Married	114	124	238	228	82
	Widowed	11	15	26	27	5
	Divorced	19	31	50	48	11
Work experience	Below 5 years	88	97	185	178	37
	6–15 years	90	79	469	214	62
	16–25 years	79	74	153	111	48
	Above 25 years	65	71	136	88	41

Source: Primary data.

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21456 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 | Factor loadings, communalities, and eigenvalues.

S. no.	Factors	Statements	Loadings	h^2	Eigenvalues
1	NVT Stress	"I gently inquire if a subordinate is under some sort of pressure, if he/she gives me cold look."	0.821*	0.674	4.424
2		"I quietly probe if something is disturbing my subordinate if he/she is impolite."	0.768*	0.590	
3		"I inquire if coworker is under any anxiety, when he/she takes full credit for my work."	0.708*	0.501	
4		"I query if my coworker is under stress if he/she interrupts me continuously."	0.717*	0.514	
5		"I make queries if there are budgetary constraints when resources are refused to me."	0.703*	0.494	
6		"I try to find out if a colleague is under nervous tension, when he/she insults me."	0.752*	0.566	
7		"I inquire if the subordinate is upset due to some reason if he/she threatens me."	0.827*	0.684	
8		"I query if coworker is under some pressure if he/she criticizes me."	0.633*	0.401	
9	NVT Ability	"I feel a coworker lacks communication skills if he/she interrupts me continuously when I speak."	0.666*	0.444	3.650
10		"I feel colleagues lack etiquettes when they verbally insult me."	0.689*	0.475	
11		"I feel threats of coworkers may be due to inability to deal with issues."	0.700*	0.490	
12		"I feel a coworker is critical when he/she does not understand my work."	0.808*	0.653	
13		"I feel resources are refused at workplace due to inability to manage budget."	0.632*	0.399	
14		"I feel covert negative glances of colleagues are due to lack of social etiquettes."	0.650*	0.423	
15		"I feel rudeness of a colleague is due to inability to express."	0.611*	0.373	
16		"I feel a coworker may not be realizing when he/she takes full credit for my work."	0.627*	0.393	
17	NVT Constructive	"I put in extra effort to improve my work quality when a coworker criticizes me in an unjust manner."	0.584*	0.341	3.020
18		"I seek a private confrontation when a coworker insults me publicly at workplace."	0.501*	0.251	
19		"I re-evaluate if the goals were too aggressive when a subordinate fails to achieve targets."	0.592*	0.350	
20		"I try to improve my feedback when a person fails to perform as per expectation."	0.689*	0.475	
21		"When resources are refused at work, I try to improve the quality of the proposal requesting resources."	0.578*	0.334	
22		"I try a different approach to explain my rationale when a coworker snubs /brushes-off my views."	0.591*	0.349	
23		"I review if I have been disciplined enough when I see my subordinate's indiscipline."	0.605*	0.366	
24		"I improve performance monitoring when a subordinate falls short of his /her responsibility."	0.744*	0.554	

(Continues)

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21456 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 | (Continued)

S. no.	Factors	Statements	Loadings	h^2	Eigenvalues
25	NVT Ignore	"I excuse verbal insults made by coworkers."	0.587*	0.345	2.709
26		"I ignore threats of coworkers."	0.582*	0.339	
27		"I ignore unjust criticism of my work."	0.617*	0.381	
28		"I ignore cold looks of coworkers."	0.516*	0.266	
29		"I ignore impoliteness of coworkers."	0.636*	0.404	
30		"I excuse when a coworker takes full credit for my work."	0.559*	0.312	
31		"I ignore continuous interruptions of a coworker while I speak."	0.506*	0.256	
32		"I ignore it if I am refused resources at workplace."	0.637*	0.406	
Number of statements					32
Eigenvalues					13.802
Explained variance					60.781

*Significant as factor loadings > 0.5.
Source: Primary data.

As suggested by previous researchers, like Koenig and Zaben (2021) and American Educational Research Associati on (2014), the extracted factor structure from EFA should be to CFA to examine the model fitness. Accordingly, the CFA was conducted on the second subsample using IBM AMOS. Koenig and Zaben (2021), Boateng et al. (2018), and Xia and Yang (2018) have suggested threshold values of these model fit indices as RMSEA < 0.08, GFI ≥ 0.95, CFI ≥ 0.97, TLI > 0.90, RFI ≥ 0.90, and SRMR ≤ 0.05. The quad-factor model showed satisfactory values of RMSEA = 0.032, TLI = 0.957, SRMR = 0.030, CFI = 0.985, GFI = 0.961, RFI = 0.938, $\chi^2 = 460.233$ at degree of freedom (df) = 458 and $p = 0.002$ ($p < 0.05$). These indices suggested the ad equate fitness of the model. The results of EFA and CFA thus confirmed the factorial structure of NVWB scale (Figure 1).

4 | Study 2: Reliability and Validity Testing

4.1 | Methods

This study examines the reliability of NVWB scale using CA, CR, and MO values. The internal convergent validity is evaluated using the average variance explained (AVE) values, and int ernal divergent validity is investigated with the help of Fornell and Larcker's (1981) testing method. Further, convergent and divergent validity is also ensured by calculating correlation bet ween NVBW, ANV, and opposite-sex self-concept scores. The p redictive validity of the scale is further examined through regr essing NVWB over workplace conflict. Lastly, this study also tests the MI using gender invariance test (GIT). MI "concerns the extent to which the psychometric properties of the observed indicators are transportable (generalizable) across groups or o ver time" (Boateng et al. 2018, p. 149). Such properties include t he factor structure of the scale (Boateng et al. 2018).

4.1.1 | Data Collection Procedure

As mentioned in Section 3, we conducted the second set of series of workshop on the workplace violence in the month of November. Again, Whatsapp and Telegram groups were created, where a structured Google form was shared.Meth od as mentioned in the Section 3 was used to ensure maximu m participation. The only difference was that this time, we approached the human resource managers of those organizations that were not approached in the first study. Participants took an average time of 33 min to complete the self-admini stered questionnaire.

4.1.2 | Participants

There were 658 participants this time, and of those, 591 respo nded to the questionnaire. There were 324 male and 267 female participants in the sample. There were 247, 138, 107, and 99 e mployees belonging to "below 30 years," "31–40 years," "41–50 years," and "above 51 years" age categories. Of these 591 parti cipants, 288 were unmarried, 228 married, 27 widowed, and 48 were separated from their partners (including divorced). There were 178, 214, 111, and 88 participants with work experience of 2–5 years, 6–15 years, 16–25 years, and more than 25 years. Table 1 reports demographic description of the sample.

4.1.3 | Instruments

The structured questionnaire had six sections. The first section provided a short description of the study and ethical statements and captured the consent of the participants. The second secti on captured the demographic information of the respondents.

4.1.3.1 | NVWB Scale. The third scale measured NVWB u sing scale whose factor structure was validated in Section 3. The original scale is developed by Bhalerao and Kumar (2015). It is a 32-item and five-point Likert scale rated between 1 (*never*) and 5 (*usually*).

4.1.3.2 | ANV Scale. The fourth section assessed nonviolence attitudes using the ANV scale developed and validated by Craven, Seaton, and Yeung (2015). The s

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21456 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

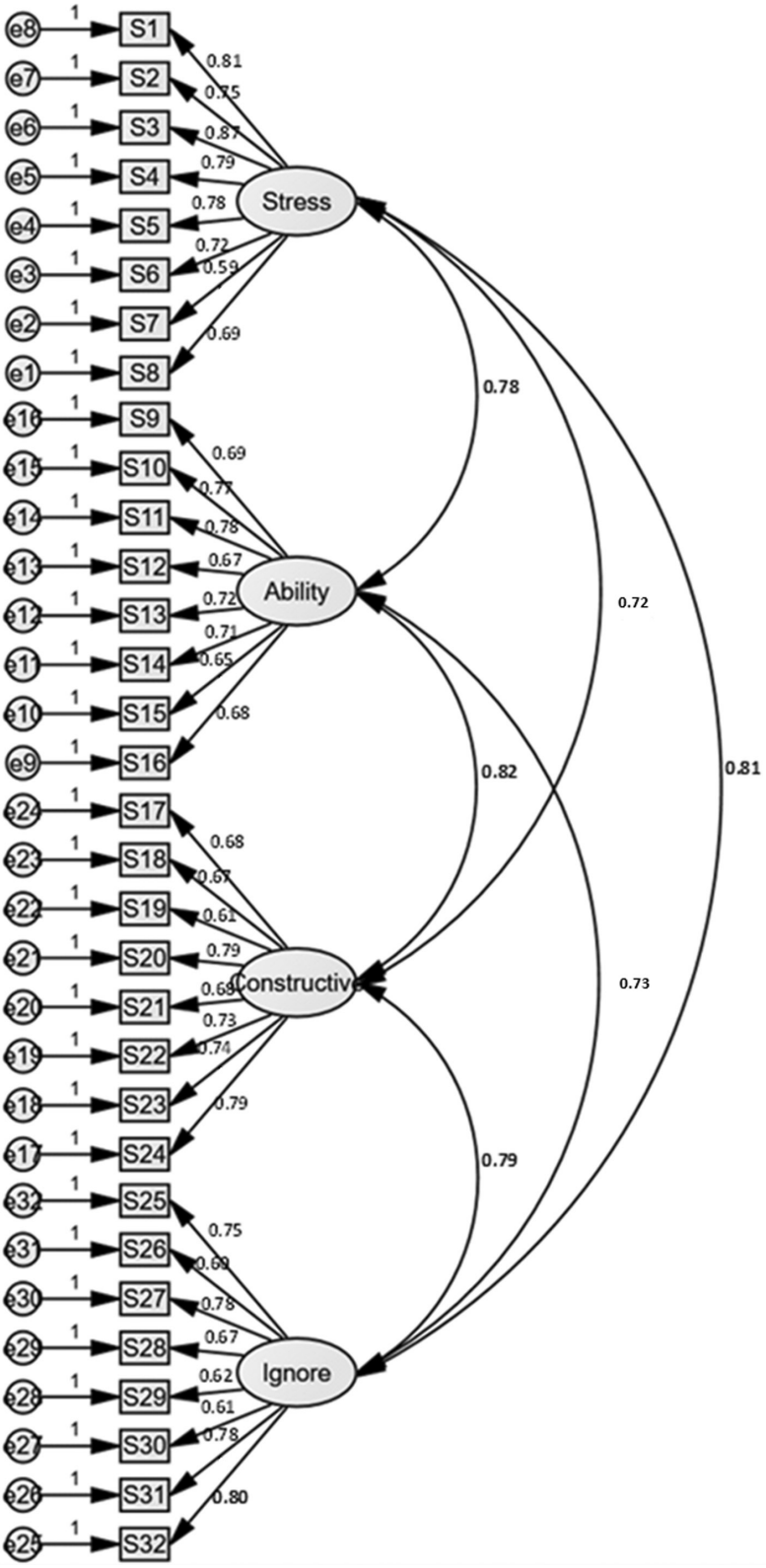


FIGURE 1 | Model.
Source: authors' own, "S" = statement.

cale measured two dimensions of nonviolence, cognitive and affective, using eight statements rated on a six-point sca le ranging from 1 (*strongly disagree*) to 6 (*strongly agree*) (Garg and Mehak 2021).

4.1.3.3 | **Opposite-Sex Colleague Relations.** The fifth section measured the opposite-sex colleague relations using adapted version of Marsh's (1993) scale. The four-itemed scale was concerned about employees' interactions with colleagues

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21456 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of the opposite sex. It is a five-point rating scale ranging betwee n 1 (*strongly disagree*) and 5 (*strongly agree*).

4.1.3.4 | Relationship and Task Conflict Scale. The last section measured the scores on workplace conflict using the Relationship and Task Conflict Scale (RTCS), which was captured using a scale developed by JehnJehn et al. (2008) and reworded by Meier et al. (2013). It has eight measures, which measure the two dimensions of workplace conflict: relati onship conflict and task conflict. The statements were rated on a seven-point scale ranging between 1 (*never*) to 7 (*always*) (Ga rg and Mehak 2021).

The scores of all scales were interpreted using average scores, that is, participants with average scores less than 3, 3.5, and 4 were interpreted to show lower expected behaviors/attitudes fo r the respective scales rated on five-point Likert scale, six-point Likert scale, and seven-point Likert scale, respectively, and vi ce-versa.

4.1.4 | Ethical Concerns

Similar to Section 3, ethical and technical issues like short descr iption of the study, an informed consent, anonymity, strict aca-demic use, and nonsharing of data were assured to the participa nts.

4.2 | Results

The reliability of the scale was assessed using CA, CR, and MO values, whose acceptable values are 0.70 and above (Cronbach and Shavelson 2004; Kyriazos et al. 2018). The values of CA for "NVT_Stress," "NVT_Ability," "NVT_Constructive," "NVT_ Ignore" (Bhalerao and Kumar 2015, p. 41), and overall NVWB were 0.91, 0.90, 0.87, 0.92, and 0.94, respectively. The values of CR were reported as 0.90, 0.88, 0.90, 0.89, and 0.95, for "NVT_ Stress," "NVT_Ability," "NVT_Constructive," and "NVT_Igno re," and the overall NVWB, respectively. Similarly, the MO val-ues are above the threshold limit of 0.70. These estimates estab-lished the reliability of the NVWB scale. Also, the AVE value s for four dimensions of NVWB were 0.58, 0.53, 0.53, and 0.52, which confirmed the convergent validity of the scale. The For nell and Larcker (1981) testing method asserted that the square root of the AVE values of the factors should be greater than the c orrelation coefficient between the factors to conclude the diver-gent validity of the scale. Table 3 demonstrates the results of the Fornell and Larcker (1981) testing method, which established s cale's divergent validity.

Further, convergent validity was also assessed by examining the relationship between NVWB scores and ANV scores. Th e ANV scores were calculated using the ANV score developed by Craven, Seaton, and Yeung (2015). The scale has two subs cales: cognitive (proactive understanding) and affective (do not endorse violence). The scale's items were suitably adjusted for w orkplace context. Since both the scales captured nonviolence at work, a positive association between two scores was hypothe-sized. Divergent validity suggests no or low correlation between two theoretically distinct independent constructs. Thus, no or low correlation was hypothesized between NVWB scores and opposite-sex self-concept scores. There is no linkage between one's inclination toward nonviolence and opposite-sex self -concept (Craven, Seaton, and Yeung 2015). Accordingly, the cr iterion validity of the NVWB scale was assessed by examining t he correlation coefficients of NWBS and ANV scores. Table 3 re-ports statistically significant positive correlations between four dimensions of NVWB and two dimensions of ANV scores. Also, a positive correlation was reported between the overall NVBW score and ANV score ($r = 0.532$, $p < 0.05$), which concluded the scale's convergent validity. A statistically insignificant correla-tion ($p < 0.05$) was reported between the NVWB score and op posite-sex self-concept, which established divergent validity of the scale. Further, Predictive validity is "the extent to which a measure predicts the answers to some other question or a re-sult to which it ought to be related with" (Boateng et al. 2018, p. 14). Thus, based on the concept of *Ahinsa*, it was hypothesized that NVWB should lead to lesser workplace (relationship and task) conflict. Table 4 represents the results of regression analy-sis, wherein four dimensions of NVWB were regressed over the overall relationship and RTCS to ascertain NVWB's predictive validity. The results revealed that all four dimensions of NVWB were statistically significant negative predictors of workplace (relationship and task) conflict. These four dimensions collec-tively explained 41.58% variance in the conflict, which confirm ed the predictive validity of NVWB scale.

GIT was conducted to assess the gender-based MI of the scale. A multi-group CFA was conducted on the quad-dimensional NWBS, and the grouping was done on the basis of gender. The fit indices of the configural invariance (Model 1) were CFI= 0.971 and RMSEA = 0.035. As these values were appropriate to their respective threshold values suggested by Putnick and Bornstein (2016), the configural invariance was proved across male and fe-male groups. The fit indices of the metric invariance (Model 2) also lay within their threshold limits. Once statistically signifi-cant model fit values of Model 2 were received, the metric invaria nce was assessed by comparing the CFI and RMSEA values of Model 1 and Model 2. The values of $\Delta CFI = 0.002$ and $\Delta RMSEA = 0.003$ were well within the threshold limits, which asserted the metric invariance across groups based on gender. Chen et al. (2018) mentioned that changes in RMSEA ($\Delta RMSEA$) and CFI (ΔCFI) should be fewer than 0.010 and 0.015, respectively. Further, the values of CFI, ΔCFI, RMSEA, and $\Delta RMSEA$ (refer t o Model 3 in Table 5) values were within the suggested limits and confirmed the metric invariance. Lastly, the Model 4 assessed the strict invariance of the scale. Its CFI, ΔCFI, RMSEA, and ΔRM SEA were also within the suggested threshold values. This con firmed the strict invariance of NWBS across gender cohorts. The results of the four models confirmed the MI of the scale, thereby concluding the suitability of the NVWB scale across both gender s. Table 5 reports the results of GIT.

5 | Study 3: Test–Retest Reliability

5.1 | Methods

This study assesses the correlation between the scores of the scale administered at one point of time (T1) with those ob-tained after a few days to months at Time period 2 (T2). Boate

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21456 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 | Correlation matrix and Fornell and Larcker testing system.

	Correlation matrix										
Scales	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
NVT Stress	1 ***(0.761)***										
NVT Ability	0.489*	1 ***(0.714)***									
NVT Constructive	0.508*	0.529*	1 ***(0.721)***								
NVT Ignore	0.556*	0.519*	0.538*	1 ***(0.714)***							
NWBS	0.525*	0.517*	0.561*	0.602*	1						
RTCS	−0.422*	−0.417*	−0.430*	−0.444*	−0.499*	1					
NVC	0.434*	0.417*	0.440*	0.461*	0.499*	−0.589*	1				
NVA	0.487*	0.477*	0.501*	0.499*	0.521*	−0.577*	0.601*	1			
ANVS	0.500*	0.503*	0.511*	0.505*	0.564*	−0.490*	0.612*	0.637*	1		
FARS	0.417*	0.423*	0.454*	0.449*	0.476*	−0.433*	0.587*	0.564*	0.618*	1	
OSSCS	0.120	0.108	0.099	0.117	0.138	0.289	0.311	0.289	0.344	0.199	1

Note: Value of AVE—square root of AVE in bold and italics, correlation significant at $p < 0.05$.
Abbreviations: ANVS, Attitude to Nonviolence Scale; FARS, Forgiveness, Reconciliation/Amend, Revenge Scale (Forgiveness); NVA, non-violence affective; NVC, non-violence cognitive; NWBS, Nonviolent Work Behavior Scale; OSSCS, Opposite-Sex Self Concept Scale; RTCS, relationship and task conflict (workplace conflict).
*Significant at $p < 0.05$.
Source: Primary data.

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21456 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ng et al. (2018) suggest that to confirm the test–retest reliability, the population should remain the same and a higher correlation indicates higher test–retest reliability. However, to further strengthen this test, we changed the order of the statement as well.

5.1.1 | Data Collection Procedure

Since test–retest reliability testing required repeated collection of data from the same respondents, working professionals enrolled in MBA-Executive courses in the universities situated in the Delhi-NCR region were approached in the third set of workshops. Each participant was given an exclusive ID, which helped in reapproaching them at Time 2. We collected the data through a Google form shared on the WhatsApp and Telegram groups created for this set of workshops. Participants took an average time of 22 min to complete the self-administered questionnaire.

5.1.2 | Participants

At time T1, 310 students responded to the NVWB scale and at time T2 (after 1 month), 188 respondents reverted. Also, the language of the items was slightly modified. Table 1 reported the demographic description of the sample.

5.1.3 | Instrument

A structured Google form divided into three sections was used to collect the data. The first section described the study, elaborated on the ethical practices, and obtained the informed consent of the participants. The second section collected demographic information, and the last section collected the scores on NVWBs. The third section used the NVWB scale validated in the previous two studies, which was originally developed by Bhalerao and Kumar (2015). The scores of the NVWB scale were interpreted using average scores, that is, participants with average scores less than 3 were interpreted to show lower NVWBs and vice versa.

5.1.4 | Ethical Concerns

Similar to Sections 3 and 4, ethical and technical issues like a short description of the study, informed consent, anonymity, strict academic use, and non-sharing of data were assured to the participants.

5.2 | Analysis of Data and Results

The correlation coefficients were calculated to check consistency over repeated scores. The correlation coefficient values for NVT_Stress, NVT_Ability, NVT_Constructive, NVT_Ignore, and overall NVWB were 0.661, 0.574, 0.637, 0.670, and 0.525. These values were statistically significant (at $p < 0.05$), which concluded the test–retest reliability of the scale.

6 | Discussion

In today's toxic workplace, where employees are engaged in retaliatory behavior, which augments workplace conflict, managers, and policymakers are resorting to various action plans, including conflict management schemes, policies to avoid conflicts, and punishments to resolve conflict and behavioral issues. However, previous studies and evidence from industrial disputes have suggested that such redressal mechanisms are both time-consuming and costly affairs (Tao et al. 2022; Kool and Agrawal 2020). Hence, social scientists have suggested incorporating behavioral modulation techniques and practicing nonviolence in the workplace to reduce the intricacies of adopting conflict resolution schemes (Sarkar and Palo 2019). A reliable and valid scale is the foremost requirement for the scientific

TABLE 4 | Regression analysis for predictive validity (workplace conflict).

Model	Variable	Unstandardized coefficient		Standardized coefficient		
		B	Std. error	Beta	*t*	Sig.
1	Constant	−1.207	0.601		−8.111	0.000*
	NVT stress	−0.396	0.217	−0.591	−10.654	0.000*
	NVT ability	−0.431	0.185	−0.626	−11.003	0.000*
	NVT constructive	−0.449	0.194	−0.644	−9.987	0.000*
	NVT ignore	−0.467	0.201	−0.662	−10.196	0.000*

Model summary

Model	R	*R* square	Adjusted *R* square	Change statistics		ANOVA	
				F changes	Sig. *F* changes	F	Sig.
1	0.618	0.381	0.379	28.031	0.000*	28.031	0.000*

*Significant at $p < 0.05$.
Source: Primary data.

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21456 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 5 | Measurement of invariance of NWBS across gender.

Measurement invariance	CFI	ΔCFI	RMSEA	ΔRMSEA
Configural invariance (Model 1)	0.971	—	0.035	—
Metric invariance (Model 2)	0.973	0.002	0.038	0.003
Scalar invariance (Model 3)	0.977	0.004	0.042	0.004
Strict invariance (Model 4)	0.982	0.005	0.048	0.006

Source: Primary data, Δ = difference between current model and previous one.

investigation of any phenomenon (Chakraborty and Biswas 2021; Sharma et al. 2022). A scale developed following a logical and scientific methodology allows accurate measurement of the variable, which leads to significant empirical studies required for the expansion of the knowledge base. A scale with adequate psychometric properties makes sure that any observed changes are due to actual variations in the measured construct and are not due to any measurement error, which supports the robustness of the results. The idea of integrating nonviolence has been around for quite some time now, but it has yet to get the desired impetus owing to an acute lack of reliable and valid scale.

Bhalerao and Kumar (2015) developed a comprehensive scale to measure NVWB that captures four inherent dimensions of nonviolence at work. The authors argued that behavioral modificati on is the key to nonviolence wherein employees do not resort to a reciprocal approach and diffuse the spiral of workplace violence. Although their contribution was considered significant, their scale validation methodology has several limitations, whi ch necessitates a logical and comprehensive study focusing on th e validation of the scale. This paper examined the reliability and validity of the NVWB scale using three independent studies. Th e first study validated the four-dimensional factorial structure of the scale with the help of EFA and CFA and was conducted on two independent samples. The EFA yielded a four-factor so lution, which is in line with the results obtained by Bhalerao and Kumar (2015). The model fit indices established the four -factor structure of the NVWB scale. The second study evalua ted and confirmed the scale's reliability and validity of the scale. Adequate CA, CR, and MO values established the scale's inte rnal consistency and reliability. Previous studies, including tho se by Bhalerao and Kumar (2015), Sarkar and Garg (2020), and Lata and Garg (2023), have also reported acceptable values of CA. The significant positive correlation between NVWB scores and ANV scores recommended the construct validity of the sca le. Bhalerao and Kumar (2015) also concluded the scale's con struct validity though they examined the relationship between NVWB and forgiveness scores. This study also investigated the predictive validity of the scale and concluded that four dimensio ns of NVWB are significant predictors of workplace conflict. Th e result supported the idea that nonviolent employees are less li kely to experience relationship- and task-related workplace conflicts. The second study also evaluated gender MI and concluded that the NVWB scale is equally reliable and valid for both male and female respondents. The third study investigated test–retest validity by collecting the data twice. The significant positive correlation between the two NVWB scores established the test–rete st reliability of the scale.

7 | Implications

This study offers several significant theoretical and practical implications. Firstly, the study provided sufficient statistical evidence of the reliability and validity of the scale, and thereby, it addresses an important research gap of inadequate validation of the NVWB scale. Given the importance of promoting nonviolence in the workplace, a reliable and valid scale is poised to encourage future researchers to examine different facets of the nonviolent behavior of employees. For instance, social scientists can explore how practicing nonviolent behavior can reduce/diffuse the spiral of workplace conflict that otherwise could have encouraged entirely quitting in employees. Secondly, the study also enriches the literature on nonviolence and peace. Nonviolence has been a topic of great interest among researchers in sociology, political science, and religious and spiritual domains. There is an emergent requi rement to put in efforts to institutionalize nonviolence in the context of professional organizations to counter workplace conflict. This scale will open new dimensions about nonviolence in the w orkplace and allow the parties of interest to understand the phenomenon more precisely. Moreover, industrial laws are formula ted to promote peace and harmony among employees, employees and employers, and the government. Therefore, having a valid m easure promotes the government's view on encouraging peaceful conflict resolution within the workplace. By understanding the d ynamics of nonviolent behaviors, managers can develop training programs that inculcate the essence of nonviolence in the work place and develop nonviolent behaviors within employees. For instance, HR managers can use this scale to conduct an impac t study. In other words, managers are encouraged to administer this scale prior to engaging employees in workshops and training regimes based on practising nonviolence in the workplace to analyze how employees demonstrate nonviolence at the workplace, and thereafter readministering the scale to check improvements in the scores of nonviolence among trainees. Also, managers can use this scale as one of the tools in the psychometric assessment of candidates prior to selecting them, and it can help managers recruit and select employees who can demonstrate nonviolence d uring conflicts in the workplace. Further, the scale will help the employees understand their focus during a conflicting situation and reorient themselves if they find themselves to be more aggressive during a conflict. Also, with the knowledge about their nonviolent nature, employees can reorient their strategies to app roach a conflict. Moreover, Sy and Choi (2013) explain that leader s' attitudes are contagious and are propagated by their followers. Hence, a valid instrument to measure NVWBs can be helpful in developing leadership development programs. Leaders who are t rained to practice and promote nonviolence at the workplace wil

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21456 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

l inspire their subordinates to propagate NVWBs in every aspect of their work and help build a positive organizational environme nt and culture. Lastly, the results of our study will help managers develop programs and schemes that promote nonviolence in the workplace. For instance, the scale may be used to promote and r eward employees. Employees with highly NVWBs may be rew arded suitably.

8 | Limitations

Though the study provides exciting findings, it has its limitations too. Firstly, the study has measured the MI solely based on the gender of the cohort. In the future, researchers may explore it based on the generational classifications of the population (gene ration X, Y, and Z) or other demographic variables. Secondly, the study examines the reliability of the scale for internal consistenc y. Future researchers may also explore the inter-rater reliability of the scale. Third, the current study has assumed a positive cor-relation between nonviolence and reduced conflict, only. Future researchers are encouraged to test factors, such as management support and job stress, among others, that may impact the prop osed relationship. Fourth, the current NVWB scale measures no nviolent behaviors of the employees in response to certain ag-gression and provocation of his/her colleague. It does not cap-ture employees' routine affection for NVWB, which may be an important aspect of fostering a positive work culture. While thi s scale effectively measures responses to certain form of aggres-sion or provocation, it tends to overlook the routine expression of nonviolent behavior, which may contribute to a consistent harmonious work environment. Hence, future research may dev elop a separate tool or modify the statements of the current tool to measure employees' proactive nonviolent behavior. Last, this study was conducted in the Indian context, which might limit the generalizability of the scale elsewhere. Hence, future researchers are urged to test the psychometric properties of the scale in dif-ferent cultural and national settings so that better generalizable results can be achieved, and the scale can also be generalized.

9 | Conclusion

In a nutshell, this study explored the reliability and validity of the scale. The statistically significant values confirm the reli ability and validate the scale in the Indian context. Hence, the sc ale shall be used without any apprehensions.

Acknowledgments

We would like to thank all the participants of this longitudinal and mul-tistudy research for taking out time from their busy schedules and beco me a part of our study. Also, we would like to appreciate their patience for being part of this study in all the time frames. Lastly, we would like to express our gratitude to reviewers of this manuscript as your valuable inputs have made this manuscript more impactful for the literature and research community.

Ethics Statement

Ethical considerations and guidelines of the 1964 Helsinki Declaration were followed.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Research data are not shared.

References

American Educational Research Association, American Psychological Association, National Council On Measurement In Education, & Joint Committee On Standards For Educational And Psychological Testing. 2014. *Standards for Educational and Psychological Testing*, 230. Was hington, DC: American Educational Research Association.

Arnetz, J. E., L. E. Hamblin, S. Sudan, and B. B. Arnetz. 2018. "Organiza tional Determinants of Workplace Violence Against Hospital Workers." *Journal of Occupational and Environmental Medicine* 60, no. 8: 693–699. https://doi.org/10.1097/jom.0000000000001345.

Asthana, H. S. 1990. "T.A.T. Responses of Some Violent Individuals." In *Recent Research In Psychology*, 39–53. https://doi.org/10.1007/978-1-4612-4458-5_3.

Badar, K., M. Aboramadan, and G. Plimmer. 2023. "Despotic vs Narcissistic Leadership: Differences in Their Relationship to Emotional Exhaustion and Turnover Intentions." *International Journal of Conflict Management* 34, no. 4: 818–837. https://doi.org/10.1108/ijcma-12-2022-0210.

Bansal, S., and N. Garg. 2024. "Quitting Silently: A Longitudinal Research on the Impact of Workplace Conflict and Nonviolent Work Behavior." *Conflict Resolution Quarterly*: 21444. https://doi.org/10.1002/crq.21444.

Bansal, S., N. Garg, and J. Singh. 2022. "Exploring the Psychometric P roperties of the Cyberbullying Attitude Scale (CBAS) and Its Relation With Teasing and Gratitude in Indian Collegiates." *International Journal of Educational Management* 37, no. 1: 225–239. https://doi.org/10.1108/ijem-05-2022-0198.

Bhalerao, H., and S. Kumar. 2015. "Nonviolence at Workplace—Scale Development and Validation." *Business Perspectives and Research* 3, no. 1: 36–51. https://doi.org/10.1177/2278533714551863.

Bhawuk, D. P. 2011. "Spirituality and Indian Psychology." In *International and Cultural Psychology/International and Cultural Psychology Series*. https://doi.org/10.1007/978-1-4419-8110-3.

Biswas, W., and D. Chakraborty. 2023. "'As You Sow so You Reap': A Multi-Case Study on Reshaping Workforce Well-Being Through Pos itive Sustainability at Construction Firms." *South Asian Journal of Business and Management Cases* 12, no. 3: 333–351. https://doi.org/10.1177/22779779231203349.

Boateng, G. O., T. B. Neilands, E. A. Frongillo, H. Melgar-Quiñónez, and S. L. Young. 2018. "Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer." *Frontiers in Public Health* 6: 149. https://doi.org/10.3389/fpubh.2018.00149.

Bodhi, R., T. Singh, and Y. Joshi. 2021. "Modelling the Role of Social Media Usage in Improving Users Well-Being: A Social Enhancement Model Perspective." *Benchmarking: An International Journal* 29, no. 8: 2450–2470. https://doi.org/10.1108/bij-03-2021-0152.

Chakraborty, D., and W. Biswas. 2019. "Evaluating the Impact of Human Resource Planning Programs in Addressing the Strategic Goal of the Firm." *Journal of Advances in Management Research* 16, no. 5: 659–682. https://doi.org/10.1108/jamr-01-2019-0007.

Chakraborty, D., and W. Biswas. 2020. "Articulating the Value of Human Resource Planning (HRP) Activities in Augmenting Organizational Performance Toward a Sustained Competitive Firm." *Journal of Asia Business Studies* 14, no. 1: 62–90. https://doi.org/10.1108/jabs-01-2019-0025.

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21456 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Chakraborty, D., and W. Biswas. 2021. "Enlivening Workplace Climate Through Strategic Human Resource Management Initiatives: Unleashing Its Efficacy." *Business Perspectives and Research* 9, no. 3: 427–445. https://doi.org/10.1177/2278533720983069.

Chen, C. C., G. B. Richardson, M. H. C. Lai, C. L. Dai, and D. G. Hays. 2018. "Development and Cross-Cultural Validity of a Brief Measure of Separation-Individuation." *Journal of Child and Family Studies* 27, no. 9: 2797–2810. https://doi.org/10.1007/s10826-018-1140-2.

Craven, R. G., M. Seaton, and A. S. Yeung. 2015. "Attitude to Non-Violence Scale: Validity and Practical Use." *Journal of Interpersonal Violence* 32, no. 13: 2018–2045. https://doi.org/10.1177/0886260515590785.

Cronbach, L. J., and R. J. Shavelson. 2004. "My Current Thoughts on Coefficient Alpha and Successor Procedures." *Educational and Psychological Measurement* 64, no. 3: 391–418. https://doi.org/10.1177/0013164404266386.

De Kerpel, L., and A. Van Kerckhove. 2023. "Advocating Beyond Call of Duty: A Review of Employee Advocacy and a Call for Research." *Management Review Quarterly*: 1–40. https://doi.org/10.1007/s11301-023-00381-1.

Deng, J., X. Hao, and T. Yang. 2022. "The Increase of Counterproductive Work Behaviour From Organizational and Individual Level Due to Workplace Conflict: A Sequential Moderated Mediation Model." *International Journal of Conflict Management* 34, no. 2: 213–233. https://doi.org/10.1108/ijcma-04-2022-0079.

Fabrigar, L. R., and D. T. Wegener. 2012. *Exploratory Factor Analysis*, 159. USA: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199734177.001.0001.

Fornell, C., and D. F. Larcker. 1981. "Evaluating Structural Equation Models With Unobservable Variables and Measurement Error." *Journal of Marketing Research* 18, no. 1: 39–50. https://doi.org/10.1177/002224378101800104.

Fredrickson, B. L. 2004. "The Broaden–And–Build Theory of Positive Emotions." *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 359, no. 1449: 1367–1377. https://doi.org/10.1098/rstb.2004.1512.

Garg, N., and N. Mehak. 2021. "Gratitude Resentment and Appreciation Scale (GRAT-16): Analyzing Psychometrics Properties in the Indian Context." *Current Psychology* 42, no. 11: 8771–8780. https://doi.org/10.1007/s12144-021-02169-6.

Hair, J. F., M. Wolfinbarger, A. Money, P. Samouel, and M. Page. 2015. *Essentials of Business Research Methods*. New York: Routledge eBooks. https://doi.org/10.4324/9781315704562.

Jehn, K. A., L. L. Greer, S. S. Levine, and G. Szulanski. 2008. "The Effects of Conflict Types, Dimensions, and Emergent States on Group Outcomes." *Group Decision and Negotiation* 17, no. 6: 465–495. https://doi.org/10.1007/s10726-008-9107-0.

Julia, G. J., E. Rajkumar, and R. John. 2022. "Prevalence of Violent Communication Within Indian Organisations–A Systematic Review and Meta-Analysis." *Heliyon* 8, no. 12: e12336. https://doi.org/10.1016/j.heliyon.2022.e12336.

Koenig, H. G., and F. A. Zaben. 2021. "Psychometric Validation and Translation of Religious and Spiritual Measures." *Journal of Religion and Health* 60, no. 5: 3467–3483. https://doi.org/10.1007/s10943-021-01373-9.

Kyriazos, T. A., A. Stalikas, K. Prassa, M. Galanakis, V. Yotsidi, and A. Lakioti. 2018. "Psychometric Evidence of the Brief Resilience Scale (BRS) and Modeling Distinctiveness of Resilience from Depression and Stress." *Psychology* 9, no. 7: 1828–1857. https://doi.org/10.4236/psych.2018.97107.

Kool, V. K., and R. Agrawal. 2020. *The Psychology of Nonviolence: Models and Their Validation*, 195–263. Germany: Springer eBooks. https://doi.org/10.1007/978-3-030-56865-8_7.

Lata, K., and N. Garg. 2023. "Predicting Non-violent Work Behaviour Among Employees Using Machine Learning Techniques." *International Journal of Conflict Management* 34, no. 5: 931–944. https://doi.org/10.1108/ijcma-04-2023-0074.

Lee, Y. 2021. "Bridging Employee Advocacy in Anonymous Social Media and Internal Corporate Social Responsibility (CSR)." *Management Decision* 59, no. 10: 2473–2495. https://doi.org/10.1108/md-01-2020-0101.

Leiß, T. V., and A. Rausch. 2022. "How Personality, Emotions and Situational Characteristics Affect Learning From Social Interactions in the Workplace." *Vocations and Learning* 16, no. 1: 73–97. https://doi.org/10.1007/s12186-022-09303-w.

Mahipalan, M. 2022. "The Interplay Between Spirituality, Compassion and Citizenship Behaviours in the Lives of Educators: A Study." *International Journal of Ethics and Systems* 39, no. 4: 804–819. https://doi.org/10.1108/ijoes-12-2021-0233.

Marsh, H. W. 1993. "The Multidimensional Structure of Academic Self-Concept: Invariance Over Gender and Age." *American Educational Research Journal* 30, no. 4: 841. https://doi.org/10.2307/1163206.

Maskey, R., J. Fei, and H. Nguyen. 2018. "Use of Exploratory Factor Analysis in Maritime Research." *Asian Journal of Shipping and Logistics* 34, no. 2: 91–111. https://doi.org/10.1016/j.ajsl.2018.06.006.

Mayton, D. 2009. *Nonviolence and Peace Psychology*. Springer eBooks. https://doi.org/10.1007/978-0-387-89348-8.

Mayton, D. M., R. Diessner, and C. D. Granby. 1996. "Nonviolence and Human Values: Empirical Support for Theoretical Relations." *Peace and Conflict Journal of Peace Psychology* 2, no. 3: 245–253. https://doi.org/10.1207/s15327949pac0203_5.

Meier, L. L., S. Gross, P. E. Spector, and N. K. Semmer. 2013. "Relationship and Task Conflict at Work: Interactive Short-Term Effects on Angry Mood and Somatic Complaints." *Journal of Occupational Health Psychology* 18, no. 2: 144–156. https://doi.org/10.1037/a0032090.

Putnick, D. L., and M. H. Bornstein. 2016. "Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research." *Developmental Review* 41: 71–90. https://doi.org/10.1016/j.dr.2016.06.004.

Reio, T. G., and B. Shuck. 2014. "Exploratory Factor Analysis." *Advances in Developing Human Resources* 17, no. 1: 12–25. https://doi.org/10.1177/1523422314559804.

Sakka, G., and M. F. Ahammad. 2020. "Unpacking the Relationship Between Employee Brand Ambassadorship and Employee Social Media Usage Through Employee Wellbeing in Workplace: A Theoretical Contribution." *Journal of Business Research* 119: 354–363. https://doi.org/10.1016/j.jbusres.2020.03.038.

Sarkar, A., and N. Garg. 2020. "'Peaceful Workplace' Only a Myth?" *International Journal of Conflict Management* 31, no. 5: 709–728. https://doi.org/10.1108/ijcma-11-2019-0217.

Sarkar, A., and N. Garg. 2021. "A Multilevel Model Exploring the Relationships Between Individual Spirituality, Spiritual Climate, and Nonviolent Work Behavior." *Conflict Resolution Quarterly* 38, no. 3: 209–229. https://doi.org/10.1002/crq.21298.

Sarkar, A., and S. Palo. 2019. "Nonviolence Behaviour in the Workplace: Myth or Reality?" In *Palgrave Studies in Indian management*, 65–81. Germany: Springer International Publishing. https://doi.org/10.1007/978-3-030-13984-1_5.

Sharma, B., S. S. Lee, and B. K. Johnson. 2022. "The Dark at the End of the Tunnel: Doomscrolling on Social Media Newsfeeds." *Technology, Mind, and Behavior* 3, no. 1. https://doi.org/10.1037/tmb0000059.

Shkeer, A. S., and Z. Awang. 2019. "Exploring the Items for Measuring the Marketing Information System Construct: An Exploratory Factor

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21456 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Analysis." *International Review of Management and Marketing* 9, no. 6: 87–97. https://doi.org/10.32479/irmm.8622.

Sy, T., and J. N. Choi. 2013. "Contagious Leaders and Followers: Expl oring Multi-Stage Mood Contagion in a Leader Activation and Member Propagation (LAMP) Model." *Organizational Behavior and Human Decision Processes* 122, no. 2: 127–140. https://doi.org/10.1016/j.obhdp.2013.06.003.

Tao, W., Y. Lee, J. Li, and M. He. 2022. "How Dialogic Vaccine Communication in the Workplace Facilitates Employee Advocacy for COVID-19 Vaccine Uptake." *Journal of Public Relations Research* 35, no. 1: 17–36. https://doi.org/10.1080/1062726x.2022.2150624.

Treiblmaier, H., and P. Filzmoser. 2010. "Exploratory Factor Analysis Revisited: How Robust Methods Support the Detection of Hidden Multivariate Data Structures in IS Research." *Information and Management* 47, no. 4: 197–207. https://doi.org/10.1016/j.im.2010.02.002.

Tsimpli, I. M., M. Vogelzang, A. Balasubramanian, et al. 2020. "Linguistic Diversity, Multilingualism, and Cognitive Skills: A Study of Disadvantaged Children in India." *Languages* 5, no. 1: 10. https://doi.org/10.3390/languages5010010.

Verma, S., and N. Garg. 2024. "Validation and Confirmation of the Equanimity Scale-16 in India and Its Relationship With Well-Being." *Mindfulness* 15, no. 3: 689–699. https://doi.org/10.1007/s12671-024-02318-5.

Walz, T., and H. Ritchie. 2000. "Gandhian Principles in Social Work Practice: Ethics Revisited." *Social Work* 45, no. 3: 213–222. https://doi.org/10.1093/sw/45.3.213.

Watkins, M. W. 2018. "Exploratory Factor Analysis: A Guide to Best Practice." *Journal of Black Psychology* 44, no. 3: 219–246. https://doi.org/10.1177/0095798418771807.

Xia, Y., and Y. Yang. 2018. "The Influence of Number of Categories and Threshold Values on Fit Indices in Structural Equation Modeling with Ordered Categorical Data." *Multivariate Behavioral Research* 53, no. 5: 731–755. https://doi.org/10.1080/00273171.2018.1480346.

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21456 by Delhi Technological University, Wiley Online Library on [05/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Optical and Quantum Electronics (2025) 57:57
https://doi.org/10.1007/s11082-024-07978-0

Visible emission studies of samarium doped $Ca_2Ga_2SiO_7$ phosphor for luminescent device applications

Anand Parasar[1] · Vikas Sangwan[2] · Sunil Kumar[1] · Sanjay Kumar[1] · Kaushal Jha[1] · Kusum Rawat[3]

Received: 28 September 2024 / Accepted: 3 December 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
A series of $Ca_2Ga_2SiO_7$ phosphors doped with variable amounts of samarium ions (Sm^{3+}) were synthesized by the solid-state reaction method. XRD analysis validates the formation of a pure phase, while SEM demonstrates homogenous particle structure with irregular morphology in the Sm^{3+} doped $Ca_2Ga_2SiO_7$ (CGS) phosphor. CGS phosphors exhibited emission centered at 563, 599, and 646 nm, which corresponds to the transitions $^4G_{5/2} \rightarrow {}^6H_j$ (J = 5/2, 7/2, and 9/2) of the Sm^{3+} ions under n-UV excitation of 401 nm. The interaction between Sm^{3+} ions was primarily dipole-dipole, as confirmed by Dexter's theory and the Inokuti-Hirayama model. The decay curve analysis exhibited that the optimized phosphor had a lifetime of 2.11 ms. The chromaticity coordinates for this optimized phosphor, calculated according to the Commission Internationale de l'Eclairage (CIE) standards, were (0.579, 0.413), indicating a pure orange emission. These results propose that this phosphor is a strong candidate for application in solid-state lighting.

Keywords Phosphor · Emission · XRD · Solid-state lighting

✉ Kaushal Jha
kaushaljha096@gmail.com

✉ Kusum Rawat
kusum04rawat@gmail.com

[1] Department of Electronic Science, Babasaheb Bhimrao Ambedkar Bihar University, Muzaffarpur, Bihar 842001, India

[2] Luminescent Materials Research Laboratory (LMRL), Department of Applied Physics, Delhi Technological University, Delhi 110042, India

[3] Electronic Materials & Devices Research Laboratory, Department of Electronics, Deen Dayal Upadhyaya (DDU) Gorakhpur University, Gorakhpur, Uttar Pradesh 273009, India

Published online: 19 December 2024

1 Introduction

The advancement of lighting technology over the past decade has provided a great opportunity to minimize hazardous emissions and save energy. Solid-state lighting (SSL), hailed as a cutting-edge fourth-generation lighting technology, outperforms traditional incandescent and fluorescent bulbs in various aspects including energy efficiency, effectiveness, cost-effectiveness, and the quality of color-balanced light it delivers. The white light produced by conventional InGaN blue LEDs coated with yellow phosphor (YAG: Ce) not only ensures high efficiency but also offers a lightweight, compact design, superior reliability, and a longer lifespan (McKittrick and Shea-Rohwer 2014; Binnemans 2009; Shang et al. 2014). However, the lack of a red component in YAG phosphors leads to certain limitations, including a high correlated color temperature (CCT) and a low color rendering index (CRI). To address this, the incorporation of red or orange light-emitting phosphors can improve both CCT and CRI values. An alternative method for generating white light is the combination of a tri-color (RGB) phosphor with n-UV excitation. However, different phosphors have different degradation times and also there is a large mismatch among the efficiencies of red (Y_2O_2S: Eu^{3+}), green (ZnS: Cu^{+}, Al^{3+}) and blue ($BaMgAl_{10}O_{17}$:Eu^{2+}) phosphors. Compared with red phosphors, green and blue phosphors are superior of efficiency, chemical durability, and thermal stability (Cho et al. 2017; Phillips et al. 2007). An alternative method for producing white light involves the use of dichromatic phosphor. This technique utilizes cyan and orange-emitting phosphors that can be excited by n-UV radiation. Among various rare-earth (RE) ions, Sm^{3+} ions, when incorporated into a suitable host lattice, can generate orange or orange-red emission when exposed to UV and blue light excitations, due to their characteristic 4f-4f transitions. Consequently, Sm^{3+} is considered an appropriate dopant for orange-red emission, which can enhance the efficiency of white LEDs (Ratnam et al. 2014; Maske et al. 2024).

The structure of the silicate host influences the emission properties of the doped phosphor. The rigidity and symmetry of the silicate lattice affect the distribution of dopant ions and their interaction with the host material, impacting the efficiency of the energy transfer and the resulting luminescent output. Silicates, characterized by their robust three-dimensional network of silicon and oxygen atoms, offer excellent chemical and thermal stability, making them an ideal host matrix for rare-earth ions. Samarium-doped silicate phosphors are particularly promising due to their ability to produce visible light, especially in the orangish-red region, when excited via ultraviolet (UV) or blue light. This feature is critical for applications like white light-emitting diodes (LEDs), where a balance of red, green, and blue emissions is needed to produce high-quality white emissions. Furthermore, the distinct orange-red emission from Sm^{3+} ions is advantageous for display technologies and other devices requiring accurate color representation (Ratnam et al. 2014; Lin et al. 2005; Jha and Jayasimhadri 2018). Besides their structural benefits, silicate hosts are cost-effective and abundant, making them an attractive option for large-scale phosphor production. These factors make samarium-doped silicate phosphors an appealing subject for research, particularly for developing new luminescent materials that meet the requirements of contemporary technologies. Also, several well-known applications for gallate-based oxide materials include lasers, multiple-color phosphors, photocatalysts, LEDs, and detectors. In the melilite family, calcium gallate ($Ca_2Ga_2SiO_7$) has space group P421 m (Behrh et al. 2016). Recently, Kusum Rawat et al. reported about the luminescent ability of Terbium (Tb^{3+}) activated $Ca_2Ga_2SiO_7$

as a green phosphor for application in the lighting field (Rawat et al. 2020). To our knowledge, no prior reports have been on Sm^{3+} doped $Ca_2Ga_2SiO_7$ phosphors.

In this study, we focus on investigating the visible emission properties of samarium-doped $Ca_2Ga_2SiO_7$ phosphors, with an emphasis on how the varying dopant concentrations influence luminescent performance. We conduct a comprehensive analysis of the structural, morphological, and emission properties of these phosphor. The study aimed to explore the potential use of these phosphors in optoelectronic applications, particularly in the area of solid-state lighting.

2 Experimental procedure and characterization techniques

CGS: Sm^{3+}, phosphor powders with a chemical composition of $Ca_{2(1-x)}Ga_2SiO_7$:xSm^{3+} (x = 0.0, 0.1, 0.25, 0.50, 1.0, 2.0 mol%), were synthesised through a high-temperature solid-state reaction procedure. The required stoichiometric ratios of the basic materials $CaCO_3$ (calcium carbonate), Ga_2O_3 (gallium oxide), SiO_2 (silicon dioxide), and Sm_2O_3 (samarium oxide) were obtained to prepare the desired material. The weighed chemicals were then kept in the agate mortar and homogenously mixed with acetone as a wetting agent. The mixture was then positioned in a microprocessor-controlled muffle furnace and heated at 1300 °C for 12 h. The mixed powder was cooled to room temperature. The phosphor was named as CGSS1, CGSS2, CGSS3, CGSS4 and CGSS5, corresponding to 0.1, 0.25, 0.50, 1.0 and 2.0 mol % Sm^{3+} dopant concentration, respectively.

X-ray diffraction (XRD) measurements were performed out with a Rigaku X-ray diffractometer using Cu-Kα radiation (δ = 1.5406 x) in the range of 15 to 70°. The X-ray tube was set to an acceleration of 30 kV and a current of 15 mA. The morphology of the phosphors was studied using a scanning electron microscope (SEM) (Carl Zeiss EVO 40). Emission and excitation spectra along with quantum efficiency were analysed using a fluorescence spectrophotometer (Jasco FP-8500, Japan). Decay measurements were performed using the Edinburgh FLS920 system, which uses xenon light as the excitation source for emission and decay studies.

3 Results and discussion

The XRD patterns for $Ca_{2(1-x)}Ga_2SiO_7$:xSm^{3+} (0.0, and 1.0 mol%) phosphor were compared to the standard $Ca_2Ga_2SiO_7$ card number 00-043-0634 in the International Crystallographic Diffraction Database (ICDD), as illustrated in Fig. 1.

The diffraction spectra of the undoped and 1.0 mol% Sm^{3+} doped $Ca_2Ga_2SiO_7$ show that the observed peaks are in good agreement with the sample ICDD data (Card No.: 00-043-0634), indicating that a tetrahedral structure which belong to the P421 m space group (Rawat et al. 2020). There is a minor change in the volume (V) of the unit cell between the undoped CGS sample and the CGS phosphor doped with 1.0 mol% of Sm^{3+} ions. To calculate the lattice parameters of a tetragonal lattice, Bragg's law ($n\lambda = 2dsin\theta$) is used to determine the interplanar spacing d. The relation for a tetragonal lattice is ($\frac{1}{d^2} = \frac{h^2+k^2}{a^2} + \frac{l^2}{c^2}$). Here, *a* and *c* are the lattice parameters, and *h*, *k*, *l* are the Miller indices. By measuring diffraction angles (θ) for various reflections and solving the equation for multiple peaks, *a*, *b* and *c* were

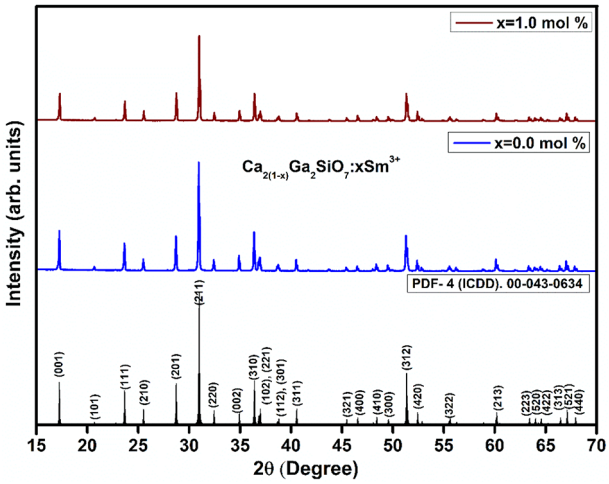


Fig. 1 XRD patterns of the ICDD data, undoped host lattiоce ($Ca_2Ga_2SiO_7$) and 1.0 mol % Sm^{3+} doped $Ca_2Ga_2SiO_7$

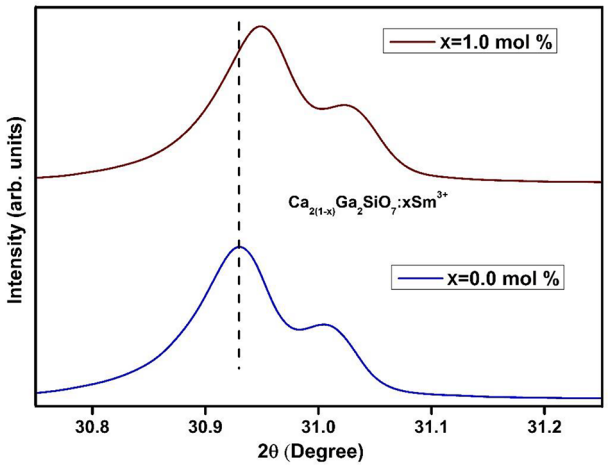


Fig. 2 XRD patterns of undoped host lattice, and 1.0 mol % Sm^{3+} doped $Ca_2Ga_2SiO_7$ for 2θ values between 30.7 to 31.3

calculated. The evaluated lattice parameters for undoped, and 1.0 mol % Sm^{3+} doped phosphor is (a = b = 7.8030 Å and c = 5.1398 Å), (a = b = 7.8032 Å and c = 5.1431 Å), respectively. The volume change happens from 312.9460 to 313.1630 Å³. Although this variation is very minimal, but this may due to the variation in the radii difference of the Sm^{3+} (1.08 Å) and Ca^{2+} (1.12 Å) (Vishwakarma and Jayasimhadri 2016a, b). We analyse the most intense band from the XRD pattern for the undoped CGS sample and 1.0 mol % doped CGS phosphor (CGSS4) and found that the diffraction peaks slightly shifted towards a higher 2θ angle for CGSS4 as depicted in Fig. 2. Now, this shift may arise due to the decrease in interatomic distance caused by replacement of Ca^{2+} ions with high concentration of Sm^{3+} ions, suggesting that Sm^{3+} ions were effectively incorporated into the host lattice at the Ca^{2+} sites. The percentage variation in the substituted (R_s) and dopant (R_d) ions ionic radii must be less than 30.0% for effective doping. This differential is estimated be calculated using the formula given below (Vishwakarma and Jayasimhadri 2016a, b).

$$D_r = \frac{R_s - R_d}{R_s} \quad (1)$$

The ionic radius of Sm^{3+} (Rd) is 1.08 Å and that of Ca^{2+} (Rs) is 1.12 Å in this instance; the computed value of D_r is determined to be 3.57%. To understand how samarium in +3 state replaces calcium is +2 state without making any charge variation in the crystal, the concept of charge compensation comes into play. The charge compensation is provided by a calcium vacancy ($3Ca^{2+} \rightarrow 2Sm^{3+}$ + calcium vacancy) (Liu et al. 2007). Therefore, it appears that Sm^{3+} ions efficiently replace the Ca^{2+} position in the $Ca_2Ga_2SiO_7$ host lattice without altering the structure of the lattice. The well-known Debye-Scherrer equation, which is provided by the relation below, is applied to calculate the average crystallite size (Sahu and Mula 2019).

$$D = \frac{k\lambda}{\beta\, COS\theta} \quad (2)$$

Where, "shape factor is represented by K (0.94), the diffraction angle is θ, the full width at half maximum (FWHM) is λ (1.5406 Å), and the wavelength of the X-ray is λ". The average crystallite size that was computed showed variation between 50 and 54 nm in relation to the Sm^{3+} concentration.

The SEM micrograph of 1.0 mol % Sm^{3+} doped $Ca_2Ga_2SiO_7$ phosphor is represented in Fig. 3. It is detected that the CGS phosphor exhibit micro size and uniformly distributed particles. It is to be noted that the grain size is much larger in comparison to the average crystallite size (XRD analysis), which is the typical nature of the oxide sample. The photoluminescence excitation (PLE) spectrum for Sm^{3+} doped $Ca_2Ga_2SiO_7$ phosphor is illustrated in Fig. 4. The excitation spectra for the CGS phosphor (1.0 mol% Sm^{3+}) was obtained by setting the emission at 599 nm. Multiple peaks in this spectrum were found at 342, 359, 372, 401, 413, 436, and 473 nm, which are indicative of transitions from the ground state $^4H_{9/2}$, $^4D_{3/2}$, $^6P_{7/2}$, $^4F_{7/2}$, $^6P_{5/2}$, $^4G_{9/2}$, and $^4M_{15/2}$, respectively (Lin et al. 2005; Ramachari et al. 2013; Yu et al. 2014; Ankoji and Rudramadevi 2019). Since the strength of the excita-

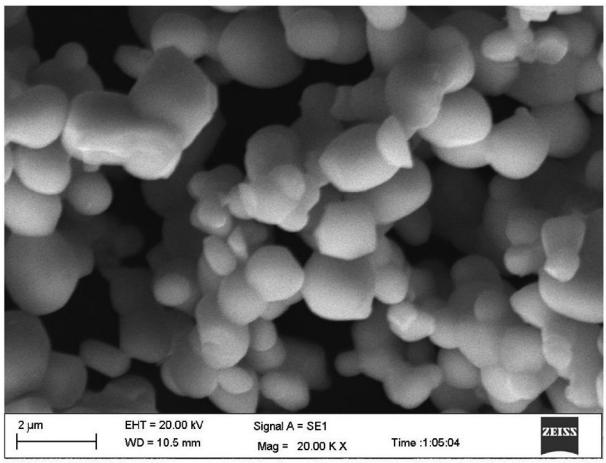


Fig. 3 SEM of 1.0 mol % Sm^{3+} doped $Ca_2Ga_2SiO_7$ phosphor

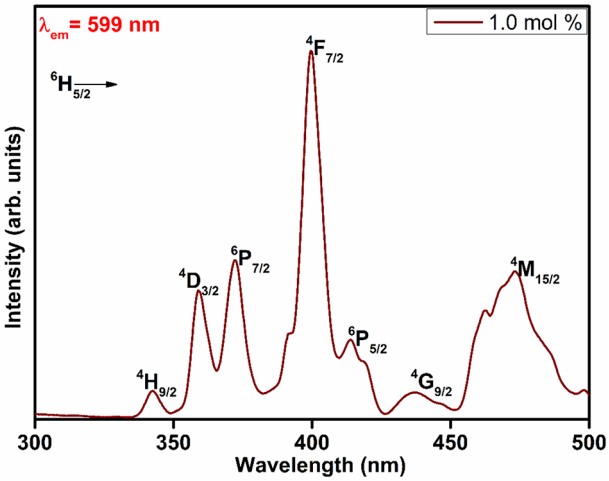


Fig. 4 Photoluminescence excitation (PLE) spectrum of 1.0 mol % Sm^{3+} doped $Ca_2Ga_2SiO_7$ Phosphor

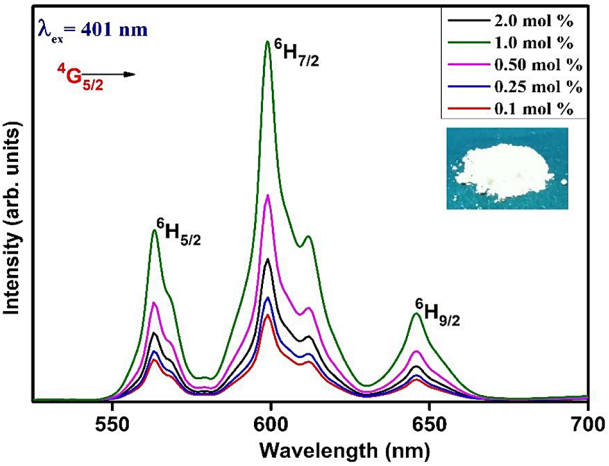


Fig. 5 Photoluminecence spectra of $Ca_2Ga_2SiO_7$ Phosphor at various concentration of Sm^{3+} under 401 nm excitation (Inset represents the physical image of phosphor powder)

tion band at 401 nm was the most significant, the emission spectra were recorded at this particular excitation.

The emission spectra represented in Fig. 5 and the image of the phosphor (powdered form) is presented in the inset. The spectra displayed peaks at 563, 599, and 646 nm, which are attributed to the $^4G_{5/2} \rightarrow {}^6H_{5/2}$, $^4G_{5/2} \rightarrow {}^6H_{7/2}$, and $^4G_{5/2} \rightarrow {}^6H_{9/2}$ electronic transition, respectively. The $^4G_{5/2} \rightarrow {}^6H_{5/2}$ is magnetic dipole (MD) electronic transition, $^4G_{5/2} \rightarrow {}^6H_{9/2}$ is forced electric dipole (ED) transition, and $^4G_{5/2} \rightarrow {}^6H_{7/2}$ belong to both MD and forced ED transition (Maske et al. 2024; Jha and Jayasimhadri 2018; Wantana et al. 2017; Sk et al. 2013; Nandanwar et al. 2024). To access the nature of the inversion symmetry around the dopant (Sm^{3+}), the ratio of the induced electric and magnetic transition is calculated. By comparing the ratio of the induced electric transition to the magnetic field. In the current scenario, the intensity of emission of the electric dipole (ED) transition is lower than that of magnetic dipole (MD) transition, which indicate that the inversion symmetry is retained (Maske et al. 2023, 2024). The emission spectra clearly indicate that the maximum intensity

is achieved at 1.0 mol% of Sm^{3+} ions. Therefore, the optimal doping concentration of Sm^{3+} in $Ca_2Ga_2SiO_7$ is 1.0 mol%. The quenching of emission takes place due to the exchange of energy among neighbouring Sm^{3+} ions, which occurs as the concentration of Sm^{3+} ions enhance which lead to decrement of the distance among nearby Sm^{3+}ions.

It was also observed that at different Sm^{3+} concentrations, the emission bands position and shape stayed constant. The reason for this is that the outermost $5s^25p^6$ electrons electric field acts as a shield for the 4f electrons. Applying Dexter's Theory to the emission spectra data enables one to analyze the nature of the interactions among Sm^{3+} ions. The relation of the Dexter theory is represented below (Dexter 1953; Dexter and Schulman 1954; Kaur et al. 2017): -

$$\frac{I}{x} = k[1 + \beta\ (x)_3^Q]^{-1} \tag{3}$$

here, the emission intensity is represented by I, the dopant's molar concentration (Sm^{3+}) by x, constants k and β, and a parameter Q about the kind of molar interaction. Exchange interactions, dipolar-dipolar interactions, dipolar-quadrupolar interactions, and quadrupolar-quadrupolar interactions are represented by Q values of 3, 6, 8, or 10.

In the above equation an assumption is that $\beta\ (x)_3^Q >> 1$, which makes the equation takes the form:

$$log\frac{I}{x} = K' - \frac{Qlog\,(x)}{3} \tag{4}$$

Where $K' = logk - log\beta$

Based on the above equation the graph of $log\frac{I}{x}$ versus $\log(x)$ is represented in Fig. 6 . The slope of the graph gives the value of -Q/3. In the present case, the value of -Q/3 is equal to -1.79. The evaluated value of the Q was found to 5.37, which is close to 6 indicating that the interaction among the Sm^{3+} ions is prominently due to dipolar-dipolar interaction. For practical applications of phosphor materials, quantum efficiency is a crucial factor to consider. The optimized phosphor in this study demonstrated an internal quantum efficiency of

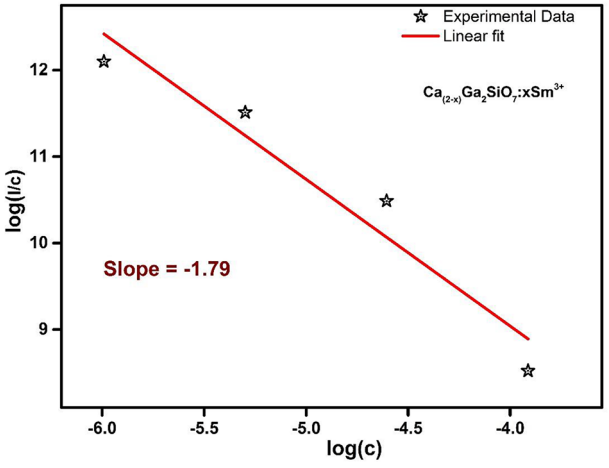


Fig. 6 Dexter plot representing graph of log(I/c) verses log(c) under 401 nm excitation

62.44% and an external quantum efficiency of 25.38%. This CIE value is in close proximity with the Amber LED NSPAR 70BS developed by the Nichia corporation having (0.570, 0.420) Chromaticity coordinates and is depicted in Fig. 7. Vishwakarma et al. reported the chromaticity values (0.557, 0.440) for Sm^{3+} doped $BaNb_2O_6$. Kaur et al. has reported the chromaticity values as (0.574,0.424) for Sm^{3+} doped calcium aluminozincate phosphor (Vishwakarma and Jayasimhadri 2016a, b; Nandanwar et al. 2024).

The comparison with the LED developed by Nichia corporation and as reported by other researchers indicate that the emission of CGS phosphor is very close to the industry standard. The chromaticity values for the CGSS3 phosphor lies on the periphery of the CIE 1931 plot, which indicates the emission to be in pure orange region. The value of the CCT was evaluated with the help of the relation given below by McCamy (McCamy 1992):

$$CCT = -449\, n^3 + 3525\, n^2 - 6823.3\, n + 5520.33 \quad (5)$$

where $n = (x - x_e) / (y - y_e)$, $x_e = 0.332$, and $y_e = 0.186$ is the coordinate value of the epicentre and n represents the inverse slope line. The measured CCT value is 1760 K, indicating that the phosphor is effective for producing warm white LEDs.

The lifetime decay patterns for the Sm^{3+} doped $Ca_2Ga_2SiO_7$ phosphor, stimulated at 401 nm, are shown in Fig. 8. The decay patterns were analysed by measuring the lifetime of the $^4G_{5/2}$ excited state at 599 nm emission. The decay curves were fitted using various exponential equations to identify the lifetime values; the bi-exponential model yielded the best match, as shown below (Choi et al. 2001; Yang et al. 2012; Umamaheswari et al. 2012).

$$I = I_0 + A_1 exp\left(-\frac{t}{\tau_1}\right) + A_2\, exp\left(-\frac{t}{\tau_2}\right) \quad (6)$$

Where "I and I_0 is the value of the emission strengths at time t and 0", "τ_1 and τ_2" illustrates the two components of the lifetime, corresponding to the fast and slow lifetimes, respectively; A_1 and A_2 are the parameter for fitting constant and t is the time.

In Fig. 9 (a), the bi-exponential fitting of the lifetime decay curve for the optimized phosphor is shown. It should be emphasized, nevertheless, that a bi-exponential fit was made for

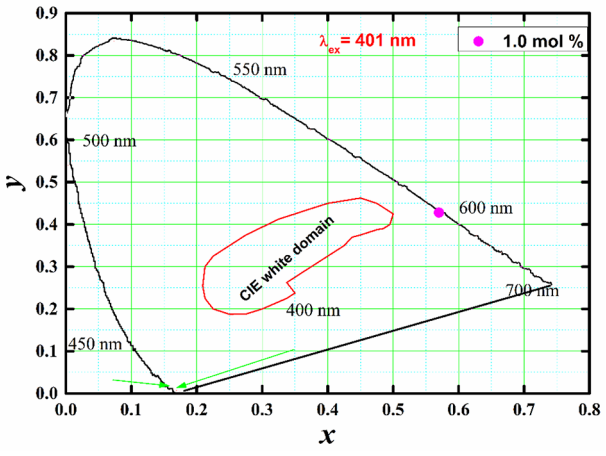


Fig. 7 CIE 1931 plot for the optimized phosphor (1.0 mol % of Sm^{3+} doped $Ca_2Ga_2SiO_7$) under 401 nm

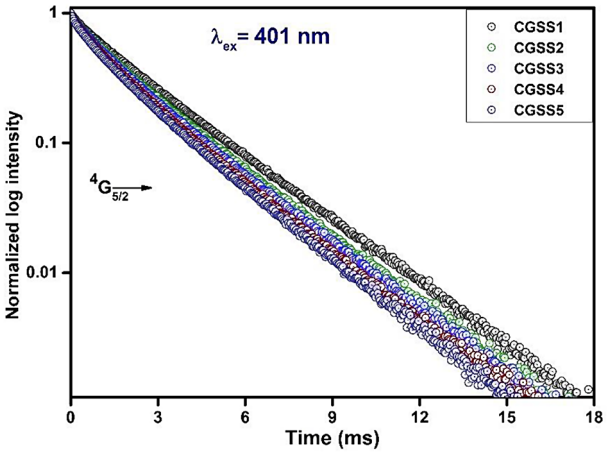


Fig. 8 Lifetime decay curves for $^4G_{5/2}$ level under 401 nm exciation for CGSS1, CGSS2, CGSS3, CGSS4, and CGSS5 samples

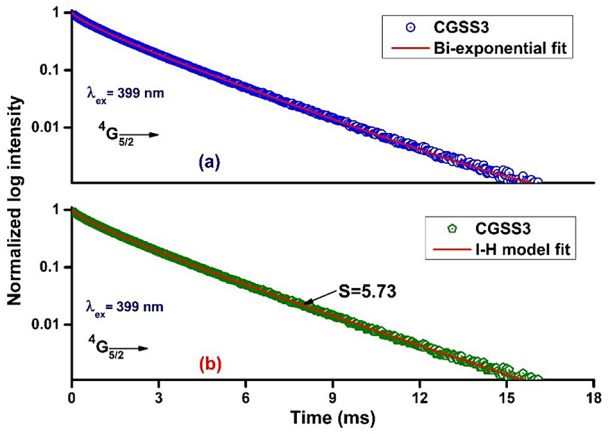


Fig. 9 Lifetime decay curves of the optimized sample (CGSS3) showing (**a**) bi-exponential fit (**b**) I-H model fitting

all of the samples (CGSS1, CGSS2, CGSS3, CGSS4, and CGSS5), and that the associated lifespan values were 2.51, 2.29, 2.22, 2.11, and 2.05 ms, respectively determined using the equation mentioned below(Jha and Jayasimhadri 2016).

$$\tau_{avg} = \frac{A_1\tau_1^2 + A_2\tau_2^2}{A_1\tau_1 + A_2\tau_2} \tag{7}$$

It was observed that the lifetime value lowered with the enhancement in the Sm^{3+} concentration, indicating the presence of the energy transfer between the dopant (Sm^{3+}) ions. The decay curve of the optimised phosphor was subjected to the Inokuti-Hirayama (I-H) model in order to inspect the nature of these interactions, as illustrated in Fig. 9 (b). The equation given by the I-H model describes the luminescence decay curve as follows. The luminescence decay curve is described by the equation given by the I-H model as described below (Rasool et al. 2013; Inokuti and Hirayama 1965). :

$$I_t = I_0 \exp\{\frac{-t}{\tau_o} - Q\left(\frac{t}{\tau_o}\right)^{\frac{3}{s}}\} \quad (8)$$

where I_t is the emission intensity at time t, t is time after excitation, τ_o the intrinsic decay time of donor in the absence of the acceptors and Q is the parameter related to energy transfer. The value of S depicts the interaction nature, which takes the values of 3, 6, 8 and 10 for exchange, dipolar-dipolar, dipolar-quadrupolar, quadrupolar-quadrupolar interactions, respectively. The dipolar-dipolar nature of interaction was determined to be the most prevalent in this instance, with a value of S of 5.73, which is quite close to 6. It is verified that the interaction falls into the dipole-dipole category based on the I-H model's result, which is consistent with the Dexter theory covered in the section above.

4 Conclusion

In summary, a series of Sm^{3+} doped $Ca_2Ga_2SiO_7$ phosphors were synthesized using the high-temperature solid-state reaction method. XRD analysis confirmed that the $Ca_2Ga_2SiO_7$ sample was a pure phase with P421 m space group. The excitation spectrum shows several peaks with the most intense being located at 401 nm, which falls in the blue region. The emission spectra showed peaks positioned at 563, 599, and 646 nm corresponding to $^4G_{5/2}\rightarrow{}^6H_{5/2}$, $^4G_{5/2}\rightarrow 6H_{7/2}$, $^4G_{5/2}\rightarrow{}^6H_{9/2}$, respectively. The ideal concentration of Sm^{3+} ions in the $Ca_2Ga_2SiO_7$ host lattice was determined to be 1.0 mol%. The results from Dexter's theory and the Inokuti-Hirayama (I-H) model, based on the PL spectra and decay curves, respectively, verified that the interaction between Sm^{3+} ions was dipolar-dipolar. The optimized sample's lifetime value was 2.11 ms, and the PL decay curves showed that typical lifetime values fell as the Sm^{3+} concentration increased. The CIE chromaticity coordinates indicate that the emission lies in the pure orange region on the CIE plot. According to all of the aforementioned findings, the phosphor can find applicability in lighting and photonic devices.

Acknowledgements The author, Dr. Kusum Rawat acknowledges DST-SERB research project no. SPG/2021/000574 for providing fund and support.

Author contributions Writing of original manuscript/draft was done by Anand Parasar and Vikas Sangwan. The conceptualization and Methodology part was performed by Sunil Kumar and Sanjay Kumar. The supervision, editing of the manuscript, and validation were done by Kaushal Jha and Kusum Rawat. The revision was primarily done by Kaushal Jha and Kusum Rawat.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

Ankoji, P., Rudramadevi, B.H.: Opt. Mater. (Amst). **95**, 109249 (2019)
Behrh, G.K., Gautier, R., Latouche, C., Jobic, S., Serier-Brault, H.: Inorg. Chem. **55**, 9144 (2016)

Binnemans, K.: Chem. Rev. **109**, 4283 (2009)
Cho, J., Park, J.H., Kim, J.K., Schubert, E.F.: Laser Photon Rev. **11**, 1600147 (2017)
Choi, Y.Y., Sohn, K.-S., Park, H.D., Choi, S.Y.: J. Mater. Res. **16**, 881 (2001)
Dexter, D.L.: J. Chem. Phys. **21**, 836 (1953)
Dexter, D.L., Schulman, J.H.: J. Chem. Phys. **22**, 1063 (1954)
Inokuti, M., Hirayama, F.: J. Chem. Phys. **43**, 1978 (1965)
Jha, K., Jayasimhadri, M.: J. Alloys Compd. **688**, 833 (2016)
Jha, K., Jayasimhadri, M.: J. Lumin. **194**, 102 (2018)
Kaur, S., Jayasimhadri, M., Rao, A.S.: J. Alloys Compd. **697**, 367 (2017)
Lin, H., Yang, D., Liu, G., Ma, T., Zhai, B., An, Q., Yu, J., Wang, X., Liu, X., Yue-Bun, E., Pun: J. Lumin. **113**, 121 (2005)
Liu, J., Lian, H., Shi, C.: Opt. Mater. (Amst). **29**, 1591 (2007)
Maske, R.T., Yerpude, A.N., Wandhare, R.S., Dhoble, S.J.: Luminescence. **38**, 1814 (2023)
Maske, R.T., Yerpude, A.N., Wandhare, R.S., Dhoble, S.J.: J. Mater. Sci.: Mater. Electron. **35**, 1 (2024)
McCamy, C.S.: Color. Res. Appl. **17**, 142 (1992)
McKittrick, J., Shea-Rohwer, L.E.: J. Am. Ceram. Soc. **97**, 1327 (2014)
Nandanwar, C.M., Yerojwar, R.M., Kokode, N.S., Yerpude, A.N., Dhoble, S.J.: Indian J. Phys. **98**, 1083 (2024)
Phillips, J.M., Coltrin, M.E., Crawford, M.H., Fischer, A.J., Krames, M.R., Mueller-Mach, R., Mueller, G.O., Ohno, Y., Rohwer, L.E.S., Simmons, J.A., Tsao, J.Y.: Laser Photon Rev. **1**, 307 (2007)
Ramachari, D., Rama Moorthy, L., Jayasankar, C.K.: Mater. Res. Bull. **48**, 3607 (2013)
Rasool, S.N., Moorthy, L.R., Jayasankar, C.K.: Opt. Commun. **311**, 156 (2013)
Ratnam, B.V., Jayasimhadri, M., Jang, K.: Spectrochim Acta Mol. Biomol. Spectrosc. **132**, 563 (2014)
Rawat, K., Vishwakarma, A.K., Jha, K.: Mater. Res. Bull. **124**, 110750 (2020)
Sahu, M.K., Mula, J.: J. Am. Ceram. Soc. **102**, 6087 (2019)
Shang, M., Li, C., Lin, J.: Chem. Soc. Rev. **43**, 1372 (2014)
Sk, N., Rasool, L., Rama Moorthy, Jayasankar, C.K.: Opt. Commun. **311**, 156 (2013)
Umamaheswari, D., Jamalaiah, B.C., Sasikala, T., Chengaiah, T., Kim, I., Moorthy, L.R.: J. Lumin. **132**, 1166 (2012)
Vishwakarma, A.K., Jayasimhadri, M.: J. Lumin. **176**, 112 (2016a)
Vishwakarma, A.K., Jayasimhadri, M.: J. Alloys Compd. **683**, 379 (2016b)
Wantana, N., Kaewjaeng, S., Kothan, S., Kim, H.J., Kaewkhao, J.: J. Lumin. **181**, 382 (2017)
Yang, L., Xu, X., Hao, L., Yang, X., Agathopoulos, S.: J. Lumin. **132**, 1540 (2012)
Yu, R., Mi Noh, H., Kee Moon, B., Chun Choi, B., Hyun Jeong, J., Sueb Lee, H., Jang, K.: S Soo Yi J. Lumin. **152**, 133 (2014)

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

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