SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(137th Edition)

MAY 2024

BY: CENTRAL LIBRARY

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

PREFACE

This is the **One Hundred Thirty-Seventh Issue** of the Current Awareness Bulletin started by Delhi Technological University, Central Library. The bulletin aims to compile, preserve, and disseminate information published by the faculty, students, and alumni for mutual benefit. It 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 **May, 2023**. The arrangement of the contents is alphabetical. The full text of the article which is either subscribed by the university or available in the web is provided in this bulletin.

Central Library

CONTENTS

1. A deep face spoof detection framework using multi-level ELBPs and stacked LSTMs, ***3.Chhavi Dhimanm***, ***6.Aashania Antil***, Arnav Anand and ***8.Soham Gakhar***, Electronics and CSE, DTU

2. A logistic regression model to facilitate setting of organic waste composting policy for sustainable waste management, ***6.Majed Ibrahim Al-Sari*** and ***3.A. K. Haritash***, Environmental, DTU

3. A Passive Floating C-MOS Memristor with Analog Applications, ***6.Rahul Kumar Gupta***, ***3.Mahipal Singh Choudhry*** and Varun Saxena, Electronics, DTU

4. A Survey on Artificial Intelligence-based Cyber Security in IoT Networks, ***8.Akshit Tyagi***, ***8.Aniket Singh***, ***8.Amit Yadav*** and ***3.Pawan Singh Mehra***, CSE, DTU

5. A Survey on Blockchain for Rental Lease Management, ***8.Anushka Jain***, ***8.Aparna Gangwar***, ***8.Chandni Kumari*** and ***3.Pawan Singh Mehra***, CSE, DTU

6. A Survey on Blockchain in Financial Institutions, ***8.Anshu Tiwari***, ***8.Aryan Kanjolia***, ***8.Bhishek Kumar*** and ***3.Pawan Singh Mehra***, CSE, DTU

7. A Survey on Skin Lesion Detection and Classification using Machine Learning, ***6. Rashmi Yadav*** and ***3.Dr. Aruna Bhat***, CSE, DTU

8. Adoption of Metaverse in Healthcare Sector: An Empirical Analysis of its Enablers, Vipulesh Shardeo, Bishal Dey Sarkar, Umar Bashir Mir and ***8.1.Prerna Kaushik***, Physics, DTU

9. An Analytical Comparison of Deep Learning Frameworks for 2D Image-Based Hatha-Yoga Pose Identification, ***7.Debashree Debalaxmi***, ***3.Virender Ranga*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

10. An empirical examination of the efect of domestic monetary policy on external commercial borrowings to India, ***3.Virender Kumar*** and Poonam, USME, DTU

11. An Improved Coot Optimization Algorithm for Optimal Power Flow in Small-Scale Power System, ***6.Udit Mittal***, ***3.Uma Nangia*** and ***3.Narender Kumar Jain***, Electrical, DTU

12. Analysis of Area-Controlerror in Automatic Generation Control, ***8.Jitesh Wadhwa***, ***8.Hayatunisha*** and ***3.Narendra Kumar***, Electrical, DTU

13. Artificial neural network-based adaptive control for nonlinear dynamical systems, ***6.Kartik Saini***, ***3.Narendra Kumar***, ***3.Bharat Bhushan*** and Rajesh Kumar, Electrical, DTU

14. Assessment of air quality benefits of vegetation in an urban-industrial region of India by integrating air monitoring with i-Tree Eco model, ***6.Mallika Vashist***, ***3.Thangamani Vijaya Kumar*** and ***3.Santosh Kumar Singh***, Environmental and Civil, DTU

15. Assessment of Solar-Biomass Hybrid Power System for Decarbonizing and Sustainable Energy Transition for Academic Building, ***6.Neelam Baghel***, ***3.Manjunath K.*** and ***3.Anil Kumar***, Mechanical and Environmental, DTU

16. Attribute-Based Encryption Scheme for Secure and Efficient Access in Blockchain, ***6.Anita Thakur***, ***3.Virender Ranga*** and ***3.Ritu Agarwal***, IT, DTU

17. Characterization of Luminescent Li3Ba2Gd3Mo8O32:Tb3+ Phosphors Prepared via Sol–Gel Method, Vijay Singh, B. R. V. Rao, Aadil Ahmad Bhat, Ch. B. Annapurna Devi, ***3.A. S. Rao*** and Ji Bong Joo, Applied Physics, DTU

18. Comparative study of luminescence in alkali-metal-based yttrium fluoride nanophosphor for biophotonic applications, , Simran Arora and ***3.A. Srinivasa Rao***, Applied Physics, DTU

19. Cosmic evolution in $f(Q, T)$ gravity: Exploring a higher-order time-dependent function of deceleration parameter with observational constraints, ***7.Madhur Khurana***, ***7.Himanshu Chaudhary***, Saadia Mumtaz, S.K.J. Pacif and G. Mustafa, Applied Physics and Applied Mathematics, DTU

20. Cross-efficiency aggregation by ordered visibility graph averaging: method, and application in portfolio selection, ***6.Reenu Kumari***, Abha Aggarwal and ***3.Anjana Gupta***, Applied Mathematics, DTU

21. Cyclic Heap Permutation Based Reconfiguration of Solar PV Array for Optimal Shade Dispersion, ***8.Viraj Jain***, ***8.Vikas Kumar***, ***8.Shlok Garg*** and ***3.Vinod Kumar Yadav***, Electrical, DTU

22. Design and simulation of solid-core octagonal photonic crystal fiber for terahertz wave propagation, ***7.Gurmeet Singh***, ***7.Shubham Sharma*** and ***3.Ajeet Kumar***, Applied Physics, DTU

23. DESIGNING OF A TWO STAGE TO ORBIT LAUNCH VEHICLE, ***8.Priyanshu Nailwal***, ***6.Prakash N. Budhaulia***, ***6.Rahul Kumar*** and ***3.Raj K. Singh***, Mechanical, DTU

24. Detection and Classification of Pulsed Doppler Radar Signals Using Deep Learning, ***8.Suhani Grover***, ***8.Sujal Gupta*** and ***3.Anup Mandpura***, Applied Physics and Electrical, DTU

25. Development of Ti3C2Tx-based novel immunosensor for cancer biomarker detection, ***6.Sweety*** and ***3.Devendra Kumar***, Applied Chemistry, DTU

26. Diurnal rail noise measurement, analysis, and evaluation of associated health impacts on residents living in the proximity of rail track area, Kranti Kumar, Arun Bhartia, ***3.Rajeev Kumar Mishra***, ***6.Ravi Pratap Singh Jadon*** and Jitendra Kumar, Environmental, DTU

27. EEG-Based Emotion Recognition Using Optimized Deep-Learning Techniques, ***Amit Kumar Dwivedi***, ***3.Om Prakash Verma*** and ***3.Sachin Taran***, Electronics, DTU

28. Emotion Recognition in Speech Using MFCC and Energy Based Ratio Features, ***6.Ravi*** and ***3.Sachin Taran***, Electronics, DTU

29. End-to-end handwritten Ge'ez multiple numerals recognition using deep learning, ***3.Ruchika Malhotra*** and ***7.Maru Tesfaye Addis***, Software and CSE, DTU

30. Energetic, exergetic, exergoeconomic, and enviroeconomic analysis of solar dish-integrated water desalination system with various nanofluids, Ravi Kant, ***3.Anil Kumar*** and Mukul Kumar Dubey, Mechanical and Environment, DTU

31. ERCU-Net: segmentation of road potholes using enhanced residual convolutional block based on U-Net for ADAS, ***6.Ruchi Tripathi***, ***3.S. Indu*** and ***3.Rohit Kumar***, Electronics, DTU

32. Experimental Investigation on Microstructure, Mechanical Behavior and Tribological analysis of Al2017/SiC/TiO_2 Composed by double stir casting route, ***8.Shlok Agarwal***, ***8.Kaustubh Kambiri*** and ***3.Roop Lal***, Mechanical, DTU

33. Exploring interacting bulk viscous model with decaying vacuum density, ***6.Vinita Khatri***, ***3.C. P. Singh*** and Milan Srivastava, Applied Mathematics, DTU

34. Fabrication and assessment of carboxymethyl guar gum-based sustainable flms for packaging application, ***6.Subodh K. Juikar*** and ***3.Sudhir G. Warkar***, Applied Chemistry, DTU

35. FDT Dr2T: a unified Dense Radiology Report Generation Transformer framework for X-ray images, ***6.Dhruv Sharma***, ***3.Chhavi Dhiman*** and ***3.Dinesh Kumar***, Electronics, DTU

36. Feature Aggregation with Latent Generative Replay for Federated Continual Learning of Socially Appropriate Robot Behaviours, Nikhil Churamani, ***8.Saksham Checker***, Hao-Tien Lewis Chiang and Hatice Gunes, CSE, DTU

37. Filter based saturated controller design for a class of nonlinear singularly perturbed systems, ***6.DIPAK PRASAD***, ***3.SUDARSHAN K VALLURU*** and MADAN MOHAN RAYGURU, Electrical, DTU

38. Flavonoids and their Therapeutic Applications in Neurological Disorders, ***8.Kanjam Manocha*** and ***3.Navneeta Bharadvaja***, Biotechnology, DTU

39. Fructose-Based Non-Isocyanate Polyurethane/Poly (Sodium Acrylate) Hydrogels: Design, Synthesis and Environmental Applications, ***6.Pooja Singh*** and ***3.Raminder Kaur***, Applied Chemistry, DTU

40. Generative Artificial Intelligence: A Systematic Review and Applications, Sandeep Singh Sengar, Affan Bin Hasan, ***3.Sanjay Kumar*** and Fiona Carroll, CSE, DTU

41. Graph-based Semi-supervised Learning for 3D Pose Estimation, ***8.Ravi Kant Sharma***, ***8.Raghav Kumar Jha*** and ***3.Shweta Meena***, Software Engineering, DTU

42. Gratitude research: Review and future agenda using bibliometric analysis of the studies published in the last 20years, ***3.Naval Garg***, USME, DTU

43. Green Emission in Thermally Stable Er3+ Doped Lead-Free Perovskite Phosphor for Solid-State Lighting and Optical Thermometry Applications, Ishant Kumar, Avinash Kumar, Sandeep Kumar, ***6.Vikas Sangwan***, and Arvind K. Gathania, Applied Physics, DTU

44. HumourHindiNet: Humour detection in Hindi web series using word embedding and convolutional neural network, AKSHI KUMAR, ***8.ABHISHEK MALLIK*** and ***3.SANJAY KUMAR***, CSE, DTU

45. Hypervisor Based Intrusion Detection Using Enhanced Radial Basis Neural Network on Cloud Environment, ***6.Naga Seshu Kumar Anumukonda***, ***3.Rajesh Kumar Yadav*** and ***3.Raghava N S***, CSE and ECE, DTU

46. Image Encryption Using Novel Chaotic System, ***3.Kriti Suneja***, ***8.Bhushan Kumar***, ***8.Kshitiz Jha*** and ***8.Lakshya Singh Chupha***, Electronics, DTU

47. Implementation of Secure Communication System using Various Chaotic Systems, ***8.S Suchit***, ***8.Piyush Goenka*** and ***3.Deva Nand***, Electronics, DTU

48. Implicit and Explicit Sarcasm Detection via Target of Offensive Language and Cognition of Logistic Tsallis Entropy Distribution in Social Networks, ***6.Himani Pokhriya*** and ***3.Goonjan Jain***, Applied Mathematics, DTU

49. Improved Performance of Oppositely Doped Core-Shell Junctionless Nanowire FET Using Gate Material Engineering, ***6.Bharti*** and ***3.Poornima Mittal***, Electronics, DTU

50. Information Visualization and Documentation for themes of Visual Culture, Nanki Nath and ***3.Ranganath M Singari***, Design, DTU

51. Machine Learning Approaches in Spatial Data Mining, ***7.1.Shalini Bhaskar Bajaj***, Ashima Narang and Priyanka Vashisth, CSE, DTU

52. Modelling student issues with MOOCs using TISM-P linkages, ***3.Shikha N. Khera*** and ***6.Himanshu Pawar***, DSM, DTU

53. Modified Six State Cryptographic Protocol with Entangled Ancilla Component States, ***6.Rashi Jain*** and ***3.Satyabrata Adhikari***, Applied Mathematics, DTU

54. Multi-armed bandit based online model selection for concept-drift adaptation, Jobin Wilson, Santanu Chaudhury and ***7.1.Brejesh Lall***, Electronics, DTU

55. MULTIBAND SUNFLOWER FRACTAL ANTENNA FOR WIRELESS ENERGY HARVESTING, ***6.Neeraj Sharma***, ***3.Asok De*** and ***3.Anukul Pandey***, Electronics, DTU

56. Multilingual Depression Detection in Online Social Media Across Eight Indian Languages, ***7.Viraj Rajderkar*** and ***3.Dr. Aruna Bhat***, CSE, DTU

57. Non-Alpha-Num: a novel architecture for generating adversarial examples for bypassing NLP-based clickbait detection mechanisms, ***6.Ashish Bajaj*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

58. Nonlinear dynamical system approximation and adaptive control based on hybrid-feed-forward recurrent neural network: Simulation and stability analysis, ***3.R. Shobana***, Rajesh Kumar and ***3.Bhavnesh Jaint***, Electrical, DTU

59. ON ESTIMATION OF HANKEL DETERMINANTS FOR CERTAIN CLASS OF STARLIKE FUNCTIONS, ***3.S. SIVAPRASAD KUMAR*** AND ***6.NEHA VERMA***, MATHEMATICS, DTU

60. Online-Learning-Based Multi-RIS-Aided Wireless Systems, ***6.Ishaan Sharma***, ***3.Rohit Kumar*** and Sumit J. Darak, Electronics, DTU

61. On-To-Methodology: Ontology Development Methodology, ***7.Magendra Singh*** and ***3.Dr. Daya Gupta***, CSE, DTU

62. Parametric Bernstein operators based on contagion distribution, ***7.Kanita*** and ***3.Naokant Deo***, Mathematics, DTU

63. Performance evaluation of absorption cooling system for air conditioning-based novel trigeneration system using solar energy, ***6.Shobhit Mishra*** and ***3.Raj Kumar Singh***, Mechanical, DTU

64. Positional Analysis and Comparison of Different Charge Generation Layer based Blue OLEDs, ***6.Sugandha Yadav***, ***3.Poornima Mittal*** and Shubham Negi, Electronics, DTU

65. Potassium tert-butoxide promoted a direct one-pot synthesis of nitriles from aldehydes at room temperature, ***8.SAURAV KUMAR***, ***6.RITIKA KUBBA***, NITYANANDA AGASTI, ANITHA SELVARAJ and ***3.ANIL KUMAR***, Applied Chemistry, DTU

66. Power quality enhancement in single phase two level/five level converters using adaptive-RBFNN algorithm, ***6.Sudhanshu Mittal***, ***3.Alka Singh*** and ***3.Prakash Chittora***, Electrical, DTU

67. Prediction of shear stress distribution in compound channel with smooth converging floodplains, ***6.Vijay Kaushik*** and ***3.Munendra Kumar***, Civil, DTU

68. Prediction of the flow resistance in non-prismatic compound channels, ***6.Vijay Kaushik***, Bandita Naik, ***3.Munendra Kumar*** and ***3.Vijay K. Minocha***, Civil, DTU

69. Remote Sensing and Machine Learning Fusion: A Robust Framework for Land Use and Land Cover Change Detection, Kriti Tripathi and ***6.Abhishek Verma***, IT, DTU

70. Rethinking Latency-Aware DNN Design with GPU Tail Effect Analysis, Fuxun Yu, Zirui Xu, Longfei Shangguan, Di Wang, Dimitrios Stamoulis, ***8.1.Rishi Madhok***, Nikolaos Karianakis, Ang Li, ChenChen Liu, Yiran Chen and Xiang Chen, CSE, DTU

71. Retraction Note: CanarDeep: a hybrid deep neural model with mixed fusion for rumour detection in social data streams, Deepak Kumar Jain, Akshi Kumar and ***7.Akshat Shrivastava***, CSE, DTU

72. R-Peak Detection Using Wavelet Scattering Transform for Pre-Term Infant ECG Dataset, ***6.Shikha Singhal*** and ***3.Manjeet Kumar***, Electronics, DTU

73. RUHate-MM: Identification of Hate Speech and Targets using Multimodal Data from Russia-Ukraine Crisis, Surendrabikram Thapa, Farhan Ahmad Jafri, ***8.Kritesh Rauniyar***, Mehwish Nasim and Usman Naseem, CSE, DTU

74. Scalable and Cost-Efcient PoA Consensus-Based Blockchain Solution for Vaccination Record Management, ***6.Neetu Sharma*** and ***3.Rajesh Rohilla***, Electronics, DTU

75. Semi-Empirical DFT Based Investigation of Electronic and Quantum Transport Properties of Novel GS-AGNR (N) FET, ***6.Anshul*** and ***3.Rishu Chaujar***, Applied Physics, DTU

76. Skin Cancer Detection using Knowledge Distillation, ***6.Rashmi Yadav*** and ***3.Dr. Aruna Bhat***, CSE, DTU

77. Spectral studies of thermally stable Dy3+/Sm3+ co-doped Li2Ba5W3O15 phosphors for warm white LEDs, ***6.Anu***, ***6.Sheetal Kumari***, Nisha Deopa and ***3.A. S. Rao***, Applied Physics, DTU

78. SSR-GAN: super resolution-based generative adversarial networks model for flood image enhancement, ***6.Vinay Dubey*** and ***3.Rahul Katarya***, CSE, DTU

79. Stock assessment using cumulative prospect theory in DEA cross-efficiency model: A case study of the Indian stock market, ***6.Reenu Kumari***, ***3.Anjana Gupta*** and Abha Aggarwal, Mathematics, DTU

80. Sustainable building materials: A comprehensive study on eco-friendly alternatives for construction, ***6.Yonatan Ayele Abera***, Civil, DTU

81. Synthesis, analysis and characterizations of Dy3+ ions-doped CaBi2Nb2O9 phosphors for optoelectronic device applications, ***6.Sheetal Kumari***, ***7.Pooja Rohilla***, Samarthya Diwakar, Rupesh A. Talewar, Ankur Shandilya, Yasha Tayal, K. Swapna, Aman Prasad and ***3.A. S. Rao***, Applied Physics, DTU

82. Techno-Economic, Environmental, and Policy Perspectives of Carbon Capture to Fuel Technologies, ***8.Aditya Singh***, ***8.Vishesh Saini***, ***8.Sambhav Jain*** and ***3.Anunay Gour***, Environmental, DTU

83. Textual Context guided Vision Transformer with Rotated Multi-Head Attention for Sentiment Analysis, ***3.Chhavi Dhiman*** and ***6.Gaurav Kumar***, Electronics, DTU

84. The estimation of parameters of generalized cosmic Chaplygin gas and viscous modified Chaplygin gas and Accretions around Black Hole in the background of Einstein-Aether gravity, Puja Mukherjee, Ujjal Debnath, ***7.Himanshu Chaudhary*** and G. Mustafa, Mathematics, DTU

85. Toward Multimodal Complaint Severity Detection From Social Media, Apoorva Singh, Prince Jha, Souryadip Das, ***8.1.Raghav Jain*** and Sriparna Saha, Software Engineering, DTU

86. Towards Energy-Efficiency: Integrating DRL and ze-RIS for Task Offloading in UAV-MEC Environments, MUHAMMAD NAQQASH TARIQ, JINGYU WANG, SAIFULLAH MEMON, ***7.1.MOHAMMAD SIRAJ***, MAJID ALTAMIMI and MUHAMMAD AYZED MIRZA, CSE, DTU

87. Trihalomethanes monitoring and their seasonal variation in urban municipal water supply system in North India, M. A. Mazhar, S. Ahmed, ***3.G. Singh***, A. Husain and Rahisuddin, Environmental, DTU

88. Ultrafast catalytic reduction of organic pollutants using ternary zinc–copper–nickel layered double hydroxide, ***6.Jigyasa Pathak*** and ***3.Poonam Singh***, Applied Chemistry, DTU

89. Understanding the internationalization of Indian state-owned enterprises, ***3.Aakanksha Kaushik***, USME, DTU

90. User Experience Design And Its Impact On Human Psychology, ***6.Shahid Ahmad***, ***3.Ranganath M Singari*** and S.L. Bhandarkar, Design, DTU

91. Visual Servoing using FPGA-based Hardware accelerated Deep Learning Solution, ***8.Prabhav Saxena***, ***8.Arjun Gupta***, ***8.Shivansh Chadha***, ***8.Suhani Grover***, ***8.Sujal Gupta*** and ***3.Dr. Anup Kumar Mandpura***, Mechanical, Electronics, Physics and Electrical, DTU

1. *Chancellor*

2. *Pro Vice Chancellor* ***2.1.*** *Ex Pro Vice Chancellor*

3.*Faculty* ***3.1.*** *Ex Faculty*

4.*Teaching-cum-Research Fellow* ***4.1.****Asst. Librarian*

5.*Alumni*

6.*Research Scholar* ***6.1.*** *Ex Research Scholar*

7.*PG Scholar* ***7.1.*** *Ex PG Scholar*

8.*Undergraduate Student* ***8.1.*** *Ex Undergraduate Student*

Signal, Image and Video Processing
https://doi.org/10.1007/s11760-024-03169-2

ORIGINAL PAPER

A deep face spoof detection framework using multi-level ELBPs and stacked LSTMs

Chhavi Dhiman[1] · Aashania Antil[1] · Arnav Anand[3] · Soham Gakhar[2]

Received: 18 October 2023 / Revised: 13 March 2024 / Accepted: 20 March 2024
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

Abstract
Facial recognition technology has emerged as the most important element in many interactive AI systems due to its ease and accuracy on par with that of humans. Nevertheless, its dependable deployment is constrained by vulnerability to presentation attacks. Therefore, for facial recognition technology to be used safely in unsupervised situations, automatic detection of presentation attacks is crucial. In this context, we propose a novel deep face spoof detection framework, which employs multi-level Elliptical Local Binary Pattern (ELBP) and stacked LSTMs. The ELBP, a variant of Local Binary Patterns (LBPs), is utilized in three levels for three color spaces—RGB, HSV, and RGB + HSV—to acquire discriminating features. We evaluate our framework through extensive experiments on two publicly available and challenging datasets—CASIA-FASD, CASIA-SURF, and OULU-NPU. The experimental results demonstrate that our framework achieves better performance in terms of APCER, NPCER, ACER, EER, ROCs, and confusion matrix.

Keywords Face anti-spoofing · Texture descriptor · Multi-level features · Color space · LSTMs

1 Introduction

Face recognition is one of the most used biometrics these days. The simplicity of usage and non-intrusive data collection is primarily responsible for facial recognition's widespread adoption. Even many recent studies have reported its equivalency on the human level [1]. Therefore, face recognition for access control is becoming more popular, but its reliable implementation is constrained by its susceptibility to presentation attacks (PA), also known as spoofing. Face recognition systems could be tricked by simply showing printed pictures or videos to the biometric sensor. Printed photographs, replayed videos, and 3D masks are examples of presentation attacks [2, 3]. Thus, face anti-spoofing has become a need of the hour for creating a secure and impregnable biometric system.

In recent years, numerous techniques have been proposed to protect face authentication systems against spoofing attacks. These techniques have shown better performance against spoofing attacks. However, when exposed to real-world scenarios such as illumination variation and camera equipment, their performances dropped significantly. Most spoof attacks are performed either by displaying 2D prints or presenting video displays. Discriminative characteristics that may help identify spoofing have been identified by researchers through examination of the textures and quality of attack attempts, such as a facial texture spoofing detection approach based on the co-occurrence of adjacent local binary patterns (CoALBP) proposed by Gragnaniello et al. [4] Similarly, YCbCr and HSV color spaces were used by Boulkenafet et al. [5] to produce a uniform LBP for genuine vs. false categorization. In addition to textural details, temporal information [6, 7] contained in face liveness is also equally beneficial in capturing spoofing, and experts are leveraging this temporal information for face liveness detection. Several

✉ Chhavi Dhiman
chhavi.dhiman@dtu.ac.in

Aashania Antil
aashiantil40@gmail.com

Arnav Anand
anandarnav2016@gmail.com

Soham Gakhar
sohumgakhar_2k19co381@dtu.ac.in

[1] Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India

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

[3] Wobot Intelligence, Delhi, India

Published online: 28 April 2024

works have connected recurrent neural networks [8], such as Long Short-Term Memory Networks (LSTM) [9], after spatial embeddings to capture temporal cues between sequential frames. Although binary cross-entropy supervised methods can arrive at promising performances, they fail to mine spoof patterns and may be stuck in poor generalization.

Motivated from above discussion, both textural and temporal data are leveraged in our work using ELBP and LSTMs. The work is performed in a two-step steps. Firstly, we propose the multi-level Elliptical variant of Local Binary Patterns (ELBPs) [10]-based feature representation for three color spaces RGB, HSV, and RGB + HSV. Secondly, defining a stacked LSTMs (Long Short-Term Memory) learning model with three LSTM layers to learn the wider cohesive association among pixels in different image patches.

To summarize, this paper comprises the following sections: Sect. 2 presents a brief survey of face spoof detection techniques. Section 3 outlines the proposed methodology for face anti-spoofing. Experimentation details along with a proper comparison are provided in Sect. 4. Lastly, Sect. 5 holds the conclusion.

2 Related Works

In recent years, many researchers have come up with robust and sophisticated face anti-spoofing detection approaches [11–14]. These approaches have greatly helped the security aspect of facial-based biometric systems by improving them immensely. They perform several analyses on spoofing attacks centered upon images, replay videos, and counterfeited masks. Therefore, we can categorize most of the literature on FAS as static and dynamic approaches.

2.1 Static approaches

Static approaches rely on texture-based information between real and fake facial images from the view of reflections on the surface and variation in materials. These approaches perform detection using a single frame without considering the temporal details. Therefore, these approaches are often known for good performance, minimal computation, and minimal cost. Moreover, they have a quicker response than dynamic systems. For instance, in [12], the experts used variations of LBPs (Local Binary Patterns) [15] and found their potential as textural descriptors on various attacks. Work like [11] demonstrates the usage of RI-LBPs (rotational invariant) and other similar textural descriptors as color texture features. Further, these textural details have been fused with extracted deep features using CNN. For involving the temporal features to express the facial dynamic structure for face anti-spoofing, Pereira et al. [13] introduced the LOP-TOP approach, which took three orthogonal planes into account that would intersect the center of a pixel in the XT, YT, and XY directions, where T stands for time axis. Following their probing outcomes, the multi-resolution LBP-TOP with SVM classifier got the best Half Total Error Rate (HTER) of 7.6% on the Replay Attack dataset. The contrast in the structural texture between 2 and 3D images is observed in [16] to discover the spoofing attacks centered upon the analysis of Fourier spectra, in which the reflections of light on the two surfaces cause a difference in their frequency distributions, whereas Maatta et al. [17] extract the texture of 2D images using the multi-scale local binary pattern (MLBP) that is later passed to the SVM classifier to classify spoofed and real faces. On the other hand, Tan et al. [18] use of a retina-based approach along with DoG filters for extracting the reflectance-based features from the facial images to detect spoofs. Another work [19] combined color texture details from the brightness as well as the chrominance/colored channels utilizing a color LBP descriptor. The work showed great results under the observed experimental outcomes.

2.2 Dynamic approaches

These approaches utilize the temporal details from the replay videos presented to the sensor like the motion of the head [20], lips, and other facial movements such as eye blinking. Other dynamic approaches include the user to follow some challenges to prove authenticity as a real user. For instance, Pan et al. [7] used a non-intrusive approach to track the blinking of the eyes as a measure for better face anti-spoofing. A randomized conditional field was employed to model various phases of eye blinking. Another work [6] combined the motion of the mouth and eye blinking 3D characteristics to detect spoofing attacks in facial images, whereas in [21] the proposal of a temporal stain metric calculated from the patterns of optical flow on the facial regions to perceive the precise facial features in the video footage to detect spoofs. In [22], a model is proposed to improve the macro- and microexpressions of the face for an advanced anti-spoofing detection. Their work resulted in the best execution and better generalizability with lesser complexity in accordance with their experimental analysis as well as results.

Besides the above approaches, some other works include deep neural network-based architecture [23–25] learn features spontaneously for face spoofing detection. CNN has generated very efficient and accurate results when it comes to computer vision applications such as object recognition [26], image classification [27], and semantic segmentation [28]. However, random learning of features often leads to poor performance.

Face liveness detection based on replay videos may be considered a video classification problem instead of image classification. Xu et al. [29] introduced the concept of LSTMs

in the FAS field. Later, LSTMs were combined with CNNs to improve motion cues for face presentation attack detection [30]. It is noted that learning temporal variations in CNN-based spatial features have strengthened the architecture's capability to detect the spoofs in videos. The above discussion motivated us to unify the characteristics of stacked LSTMs with novel multi-level ELBP textural details.

3 Proposed methodology

The section briefly discusses the proposed methodology based on a multi-level deep face spoof detection framework. The proposed framework includes three steps: data preprocessing, feature formation, and feature learning.

3.1 Data preprocessing

In the preprocessing phase, the frames are organized and labeled according to the corresponding dataset methodology. The frames are then resized to [299 × 299] pixels and fed into our proposed framework. The viola-jones algorithm [31] is utilized for face detection and ROI cropping. Furthermore, the input data are randomly flipped and rotated. Additionally, data augmentation techniques such as color jitter, random flipping, and rotation are applied.

3.2 Feature formation

Feature formation is performed in a twofold process. Firstly, the region of interest (ROI) of the face is analyzed in two distinct color spaces, namely RGB and HSV. RGB color space alone is not sufficient to provide all the necessary details in an image. As per expert opinion [32, 33], the complementary channels in the color model contain more fraudulent information, such as print defects, noise, and moiré effects. Moreover, luminance plays a vital role in highlighting imperfections on any surface. By incorporating HSV in conjunction with RGB, the color information can be separated from the intensity information, which is not possible in RGB alone. This separation can be beneficial in detecting spoofing attacks involving printed images or video displays as they may have different color properties compared to a live face. Additionally, the skin color information can be extracted using HSV, which is significant in face anti-spoofing systems. By utilizing both RGB and HSV color spaces in face anti-spoofing, a broader range of visual features can be captured, thereby improving the system's ability to detect fake faces.

Hence, in this paper, the HSV color space has been employed alongside RGB to better incorporate illumination from input frames. Additionally, a combination of both RGB and HSV color spaces is used to extract both color and hue-saturation values. This approach leads to better classification results compared to using each color space individually. As shown in Fig. 1, each RGB and HSV input face ROI is given as input to the Multi-Level Feature Extraction (MLFE) Module. It processes the input image in three levels L_i, $i \in [0, 2]$. At each level L_i elliptical LBP (ELBP) features are extracted. ELBP features are extended version of LBPs [34], where, for each central pixel (x_c, y_c) in the input image with a gray level value g_c, the value of the LBP is computed by contrasting the value of g_c with the gray level values of pixels at R distance within its surrounding in P neighborhood pixels that are in a circular path either clockwise or anti-clockwise direction. If the central pixel value (x_c, y_c) has a bigger gray level value than the neighboring pixel P_i, the neighbor pixel is given a value of 0 else 1 giving a P-bit binary number and for a neighborhood of eight pixels. LBP in total generates 256 different binary patterns. The $\mathrm{LBP}^{\mathbf{P,R}}$ is calculated as follows:

$$\mathrm{LBP}^{\mathrm{P,R}}(x_c, y_c) = \sum_{i=1}^{P} s(g_i^{\mathrm{P,R}} - g_c)^{i-1} \tag{1}$$

where $s(x)$ is defined as

$$s(x) = \begin{cases} 1, \, if x \geq 0 \\ 0, \, if x < 0 \end{cases} \tag{2}$$

A circular path neighborhood of pixels can be advantageous for a texture descriptor, but in certain applications, important texture details that could differentiate objects may be overlooked due to the circular configuration's presence. An ellipse, on the other hand, is more helpful in obtaining features from different directions [35]. So, for calculating the Elliptical Local Binary Patterns (ELBPs) [36] for each central pixel (x_c, y_c), we consider the neighboring pixels P that lie on an ellipse (with radii R_1 and R_2) estimated as:

$$\mathrm{ELBP}^{P,R1,R2}(x_c, y_c) = \sum_{i=1}^{P} s(g_i^{P,R1,R2} - g_c)2^{i-1} \tag{3}$$

where the ith neighboring pixel coordinates of (x_c, y_c), are calculated as:

$$anglestep = 2 \times \pi / P \tag{4}$$

$$x_i = x_c + R1 \times \cos(i-1) \times \mathrm{anglestep} \tag{5}$$

$$y_i = y_c - R2 \times \sin(i-1) \times \mathrm{anglestep} \tag{6}$$

MLFE module divides a face ROI in 4^i parts at each ith level. It results in formation of a histogram h_i at each level with dimension $[1 \times 256 \times 4^i]$, as shown in Fig. 2. At level L_0, the image is considered to compute $h_{0_{[1\times256]}}$. At level L_1, the image is divided in 4^1 sub-regions and $h_{1_{[1\times256\times4]}}$ is computed. At level L_2, the image is divided in 4^2 sub-regions and $h_{2_{[1\times256\times16]}}$ is computed.

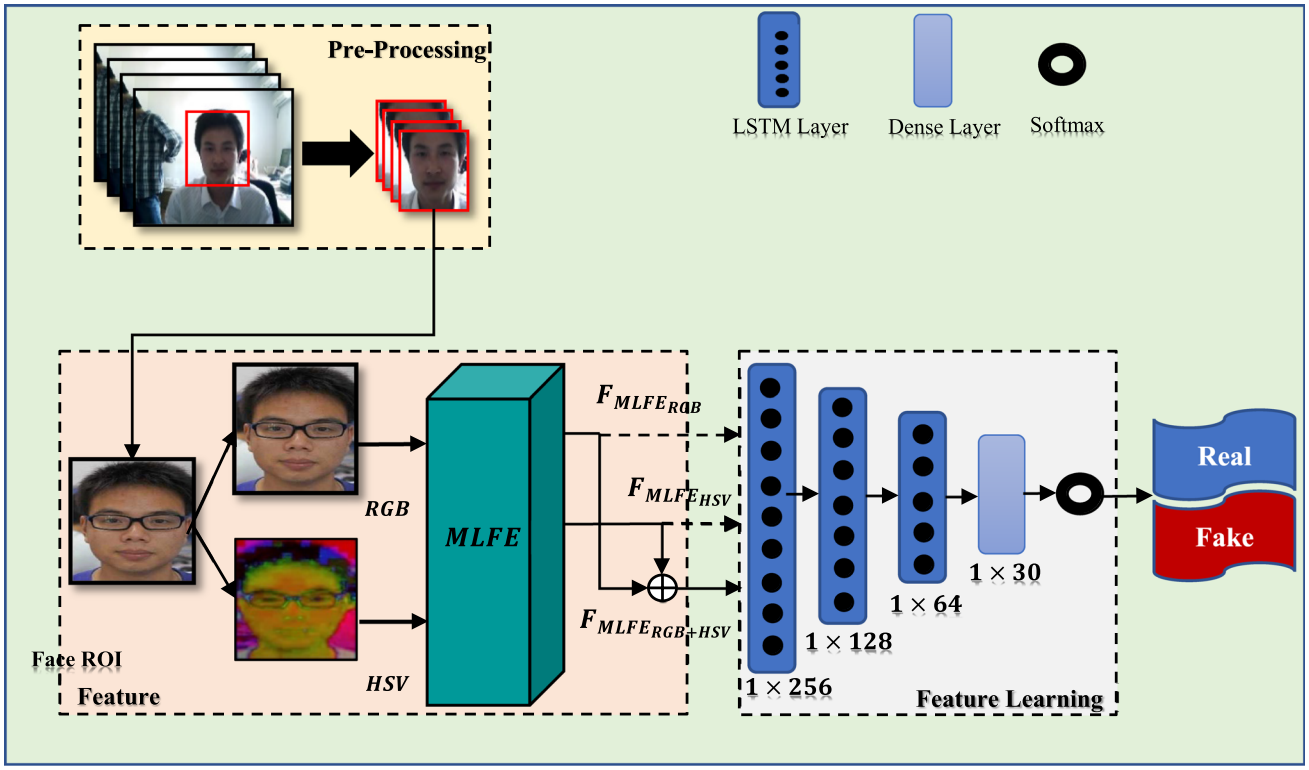


Fig. 1 Proposed deep multi-level ELBP-based anti-spoofing framework

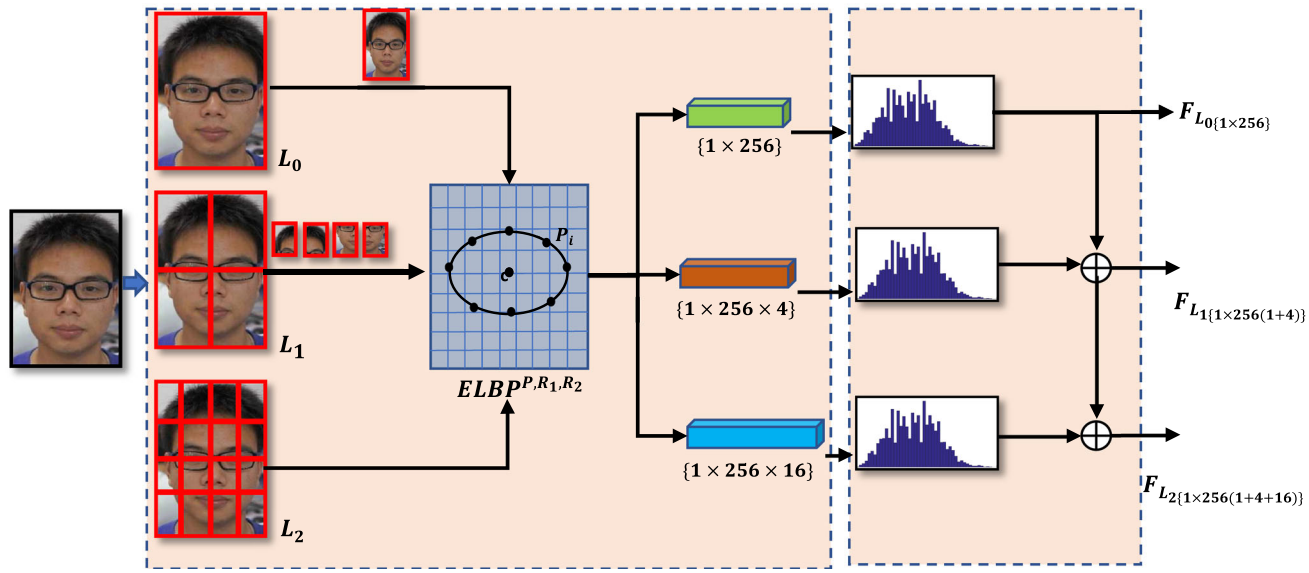


Fig. 2 Illustration of multi-level feature extraction (MLFE) from input image

3.3 Feature concatenation

From MLFE module, three multi-level feature representations are retrieved, given by F_{L_i}, $i \in [0, 2]$, as defined in Eqs. (7)–(9).

$$F_{L_0} = \{h_0\}_{[1\times256]} \tag{7}$$

$$F_{L_1} = \{h_0, h_1\}_{[1\times256(1+4)]} \tag{8}$$

$$F_{L_2} = \{h_0, h_1, h_2\}_{[1\times256(1+4+16)]} \tag{9}$$

As the number of levels increases, the feature vector size also increases. However, including higher levels decreases the performance due to increased computation cost. Thus, our framework is only limited to level-2. Algorithm 1 explains the multi-level-feature extraction (MLFE) method for levels 0, 1, and 2. The concatenation is not only limited to the feature extraction at three different levels F_{L_i}, $i \in (0, 2)$. *M*ulti-level features are further extracted for different color spaces i.e., RGB and HSV color spaces, along with third color space of HSV + RGB.

Algorithm 1 is repeated independently for each color space (RGB, HSV, and RGB + HSV images) as shown in Fig. 1. This increases the feature vector size at different levels for different color space and results in new sizes as $[1 \times 256]$ for L_0, $[1 \times 256(1 + 4)]$ for L_1, and $[1 \times 256(1+4+16)]$ for L_2. Basically, the sizes get increased by twofold as compared to a single-color space.

Finally, MLFE module generates three set of features given by $\mathrm{F}_{\mathrm{MLFE}_{\mathrm{RGB}}}$, $\mathrm{F}_{\mathrm{MLFE}_{\mathrm{HSV}}}$, $\mathrm{F}_{\mathrm{MLFE}_{\mathrm{RGB+HSV}}}$, Eqs. (10)–(12):

$$\mathrm{F}_{\mathrm{MLFE}_{\mathrm{RGB}}} = \{F_{L_0}, F_{L_1}, F_{L_2}\}_{\mathrm{RGB}} \tag{10}$$

$$\mathrm{F}_{\mathrm{MLFE}_{\mathrm{HSV}}} = \{F_{L_0}, F_{L_1}, F_{L_2}\}_{\mathrm{HSV},} \tag{11}$$

$$\mathrm{F}_{\mathrm{MLFE}_{\mathrm{RGB+HSV}}} = \{F_{L_0}, F_{L_1}, F_{L_2}\}_{\mathrm{RGB+HSV}} \tag{12}$$

Table 1 Stacked LSTM: Sequential Model parameters description; Total parameters: 512,732; Trainable parameters: 512,732

Layer	Output shape	Param#
LSTM_1	(21,950, 256, 256)	264,192
LSTM_2	(21,950, 256, 128)	197,120
LSTM_3	(21,950, 64)	49,408
Dense_1	(21,950, 30)	1950
Dense_2	(21,950, 2)	62

Unlike LBP, ELBP descriptors extract more precise spatial features from facial images as they retrieve additional directional features at various angular orientations, including extra micro-patterns. For better uniformity and orientation, we used the angle step of 45° and R_1 value to be 3 and R_2 value to be 2 for all our experiments.

3.4 Feature learning

Instead of performing direct classification based on multi-level ELBP features, the spatial distribution of multi-level ELBP features is learned through three stacked layers of the Long Short-Term Memory (LSTM) model followed by two dense layers and SoftMax Classifier, as mentioned in Table 1. The layers of LSTM units are utilized to learn the most efficient feature correlations for the spoof detection task.

Though LTSM units have been typically applied to temporal sequences, some recent works [8, 9, 37] have also utilized the memory capabilities of LSTMs in image space. In our case, we treat the output vector with size $[1 \times 256 \times n]$, where value of n changes according to the level of abstraction of features. We propose to insert 256,128, and 64 LSTM units after the MLFE module. It reduces the dimensionality of the feature vector in a structured way. Also, at the same time, memory units identify the most useful feature correlations for the task of spoof detection. The first LSTM layer is chosen to be defined with 256 units to align it with the size of the ELBP image features fed directly to it.

Algorithm 1 Computation of Multi-Level ELBP Features

Step 1: Input a frame $I(i,j)$.
Step 2: Apply Viola Jones Algorithm to extract Face ROI from the frame as $ROI(i,j)$
Step 3: Compute Multi-Level ELBP features as follows:

a) **At Level 0:** Calculate the Elliptical Local Binary Patterns (ELBPs) for each central pixel (x_c, y_c) using following formulas.

$$ELBP^{P,R1,R2}(x_c, y_c) = \sum_{i=1}^{P} s(g_i^{P,R1,R2} - g_c)2^{i-1}$$

where P is the no. of neighbouring pixels, $P \in [1,8]$ that lie on an ellipse with radii R_1 and R_2, $R_1 = 3, R_2 = 2$ and i^{th} neighboring pixel coordinates of the central coordinate (x_c, y_c), is calculated as:

$$angle\ step = 2 \times \pi/P,$$
$$x_i = x_c + R1 \times \cos(i-1) \times angle\ step$$
$$y_i = y_c - R2 \times \sin(i-1) \times angle\ step$$

And $s(x)$ is defined as $s(x) = \begin{cases} 1, & if\ x \geq 0 \\ 0, & if\ x < 0 \end{cases}$

$ELBP^{P,R1,R2}$ value for each pixel position is calculated. Obtained ELBP values for the entire image at Level 0, are quantized into 256 (2^8) evenly spaced bins generating a histogram vector h_0 of dimension $[1 \times 256]$.

b) **At Level 1:** Image is divided in four sub-regions $I_{SR1}(i,j), I_{SR2}(i,j), I_{SR3}(i,j), I_{SR4}(i,j)$. A histogram vector for each sub-region is generated, as defined in Level 0, step 3(a). Length of generated histogram at level 1, h_1, is $[1 \times 256 \times 4]$.

c) **At Level 2:** Each sub region $I_{SRi}(x,y)$, $i \in [1,4]$, is further divided into four sub regions. For each sub region a histogram vector is generated as defined in step 3(a). The Length of the histogram h_2 vector generated at level 2 is $[1 \times 256 \times 16]$.

Step 4: A combined feature vector is defined at all three levels as F_{L_i} , $i \in [0,2]$, with dimension given by $[1 \times 256]$, $[1 \times 256\ (1+4)]$, $[1 \times 256(1+4+16)]$ respectively.

4 Experiments

In this section, we evaluate our framework on two popular datasets. There will be a succinct description of these two datasets, followed by the performance analysis of the proposed method with various sought-after baseline methods.

4.1 Dataset

To demonstrate the performance of the proposed method, experiments are conducted on two challenging benchmarks datasets: CASIA-FASD, CASIA-SURF, and OULU-NPU.

CASIA-FASD. In the CASIA-FASD database, there are around 12,657 images for 50 subjects where the spoofs are generated using high-quality records of the real faces. The dataset holds around 5123 real images and 7534 fake images. The images are present in three different quality levels, i.e., low, medium, and high. The spoof images include cut photographs, distorted photograph spoofs, and video spoofs.

CASIA-SURF. CASIA-SURF is a large-scale multi-modal benchmark for face anti-spoofing. This dataset has subjects with modalities RGB, depth, and IR; however, in our approach, we will only be using the RGB images from this dataset. Therefore, we segregated all the RGB images from the dataset. It consists of 29,266 images, where 20,324 are spoofed and 8942 are genuine facial images.

OULU-NPU. The OULU-NPU dataset contains 5940 video sequences capturing real access and attempted attacks under various conditions. These sequences are carefully divided into training, development, and testing subsets. Additionally, the dataset includes four protocols designed to evaluate the effectiveness of face presentation attack (PA) detection algorithms. Protocol I examines the algorithm's ability to generalize across unseen illumination and background conditions, while Protocol II assesses the impact of attacks using unseen printers and display devices. Protocol III employs the leave-one-camera-out technique to

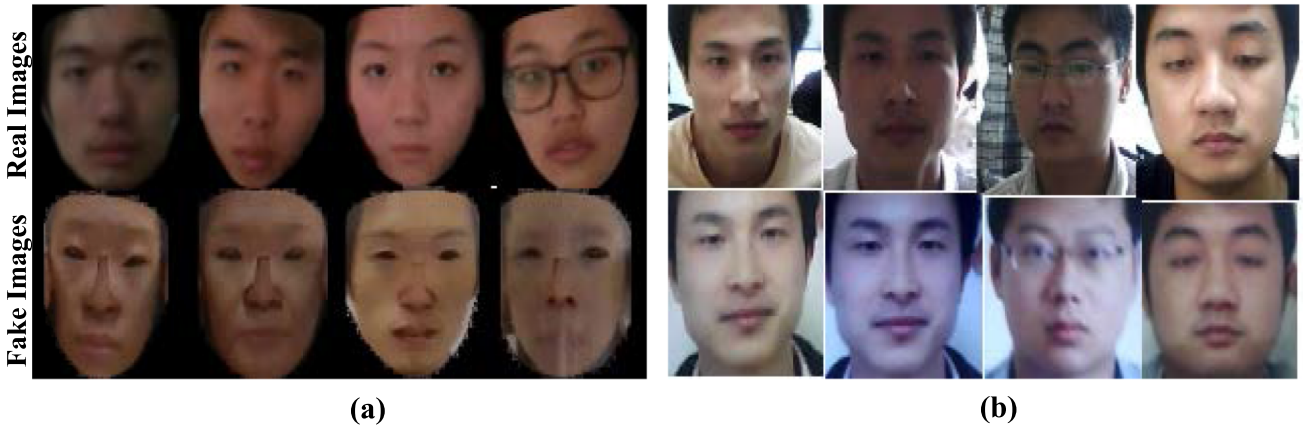


Fig. 3 Sample images **a** CASIA-SURF and **b** CASIA-FASD datasets

generalize the model across sensor variations. Finally, Protocol IV combines all three scenarios—unseen environmental setups, attacks, and sensors—posing a more challenging generalization task. This meticulous dataset design ensures a comprehensive evaluation of our proposed approach.

Figure 3 shows the sample of images from each database. We further split the images in both databases in the ratio of 75% for training and 25% for testing per Real and Fake images.

4.2 Evaluation metric

For CASIA-FASD dataset, the observed parameter is Equal Error Rate (EER), whereas CASIA-SURF dataset is analyzed in three terms, namely, Attack Presentation Classification Error Rate (APCER), Normal Presentation Classification Error Rate (NPCER), and Average Classification Error Rate (ACER) as performance evaluation metrics. EER is the point on a ROC or DET curve where the false acceptance rate equals the false rejection rate. APCER and NPCER are abbreviations for error rates that are used to gauge how well the PAD subsystem is performing. They help to assess the balance between security and convenience by measuring how well a picture is identified as either a live person or a spoof. ACER is the average of APCER and NPCER.

Besides these parameters, we have also obtained Receiver Operating Characteristic (ROC) curves obtained results in FPR and TPR values. We have also shown the Confusion Matrix in terms of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values. The mathematical formulae of the above-mentioned metrics are as follows:

$$\mathrm{FPR} = \mathrm{FP}/(\mathrm{FP} + \mathrm{TN}) \tag{13}$$

$$\mathrm{TPR} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \tag{14}$$

$$\mathrm{APCER} = \mathrm{FP}/(\mathrm{TN} + \mathrm{FP}) \tag{15}$$

$$\mathrm{ACER} = (\mathrm{APCER} + \mathrm{NPCER})/2 \tag{16}$$

$$\mathrm{NPCER} = \mathrm{FN}/(\mathrm{FN} + \mathrm{TP}) \tag{17}$$

$$\mathrm{EER} = (\mathrm{FPR} + \mathrm{TPR})/2 \tag{18}$$

4.3 Training settings

We implemented the architecture using Keras and conducted the experiment on Google Colab pro, utilizing an Nvidia P100 GPU with 16 GB RAM**.** We performed horizontal flips and rotation for data augmentation. During network training, we used the Adam optimizer with a learning rate of 10e−4 and employed a batch size of 32 with random shuffling. The whole architecture was trained for 20 epochs.

4.4 Comparative analysis with other state-of-the-art

In order to demonstrate the effectiveness of our proposed architecture, we have conducted a comparison with other state-of-the-art approaches. The comparison is presented in Tables 2, 3, and 4 for the CASIA-FASD, CASIA-SURF, and OULU-NPU datasets, respectively. For the CASIA-FASD dataset, the previous works [11, 19, 23, 38, 39, 40, 33, 41] used EER as the performance measurement parameter in their experiments. Thus, Table 2 compares the performance based on EER values. The use of Local Binary Patterns (LBPs), such as LBP-TOP [40], LBP+HOOF [39], and Color LBP [19], demonstrated their efficacy in capturing texture details and learning prominent facial features for detecting spoofs. However, these methods achieved comparably high EERs. In contrast, [23] learned CNN features, and

Table 2 Comparison of the proposed method with other state-of-the-art for CASIA-FASD Dataset

Method	EER (%)
LBP-TOP [40]	10.0
LBP + Histogram of Oriented Optical Flow (HOOF) [39]	3.1
CNN [23]	7.4
Color LBP [19]	7.1
Color Texture Markov Feature (CTMF) [33]	8.0
RI-LBP + Speed Up Robust Features (SURF) [41]	4.6
Color SURF [38]	2.8
Color Texture [19]	2.1
RI-LBP (YCbCr + HSV) + CNN [11]	4.4
Adversarial [44]	3.2
Patch CNN [45]	4.8
Antil et. al [10]	2.37
Proposed Method	**1.12 ± 0.8**

Table 3 Comparison of the proposed method with other state-of-the-art for CASIA-SURF Dataset. The values in bold highlight observed error rate is least among other state-of-the-art

Method	APCER (%)	NPCER (%)	ACER (%)
Zhang et al. [42]	1.9	0.1	1.0
Chen et al. [43]	0.92	1.74	1.33
MA-Net [46]	2.4	1.7	2.0
MF^2ShrT with T-Encoder branch [47]	1.6	1.2	1.4
Proposed Method	**0.02**	**0.019**	**0.02**

[11] integrated CNN features with RI-LBP features in the YCbCr+HSV color space resulting in an improved EER of 4.4%. Another work, [41] combined RI-LBP with SURF features and achieved comparable performance with [11], with an EER of 4.6%. However, our proposed anti-spoofing framework outperforms these previous works, achieving an EER of 1.2% by utilizing the multi-level elliptical variant of LBP features with stacked LSTMs.

For the evaluation of the CASIA-SURF dataset, we have used the parameters APCER, NPCER, and ACER. As this dataset is relatively new, only a few studies [42, 43] have utilized it for face anti-spoofing. Zhang et al. [42] used multi-modal data to learn deep features from RGB, IR, and depth frames, achieving an APCER of 1.9%, an NPCER of 0.1%, and an ACER of 1%. On the other hand, Chen et al. [43] utilized depthwise-separable attention-based deep features for anti-spoofing. Our proposed face anti-spoofing framework achieved superior performance, with an APCER of 0.02%, an NPCER of 0.018%, and an ACER of 0.02% for the CASIA-SURF dataset.

Furthermore, we conducted thorough experiments on the OULU-NPU dataset, adhering to all four prescribed protocols, and compared our method with alternative approaches, as depicted in Table 4.Across all protocols, our method consistently outperforms others, demonstrating the effectiveness of utilizing the elliptical variant of LBP-based features learned with stacked LSTM layers as a robust representation for spoof detection.

The performance of the proposed work is also observed in the form of ROC curves and a Confusion Matrix for both datasets in all three color spaces. The ROC curves (Fig. 4) obtained in RGB+HSV color space are observed to approach leftmost corners quickly with increased TP and TN and reduced FN and FP in the confusion matrix (Fig. 5).

4.5 Ablation study

To determine the effectiveness of our proposed approach, we conducted a series of ablation experiments. These experiments allowed us to choose an appropriate model demonstrate the effect of different color spaces on efficient multi-ELBP feature representation. The results of these experiments are discussed below.

4.5.1 Ablation study for MLFE

For the first set of analysis, we perform experiments on the CASIA-FASD dataset and the results are presented in Table 5. The proposed model is evaluated at different levels $\boldsymbol{L}_0$, $\boldsymbol{L}_1$, $\boldsymbol{L}_2$, using input images in the RGB, HSV, and RGB + HSV color spaces. The results in the table demonstrate that at $\boldsymbol{L}_2$ APCER, NPCER, ACER, and EER have significantly decreased in comparison with $\boldsymbol{L}_0$ and$\boldsymbol{L}_1$. Also, multi-level elliptical texture features extracted in RGB + HSV color space detect the spoofs in images with least error rate measured as APCER, NPCER, ACER, and EER. The proposed anti-spoofing framework achieves an EER of 13.8% for the RGB color space, 6.24% EER for HSV color space, and a minimum 1.12% for the combined RGB + HSV color space. It proves that multi-level detailed features extracted in RGB + HSV color space retain discriminant information resulting in enhanced performance with least EER value. The APCER and ACER metrics also support similar observations in the combined RGB + HSV color space.

Similarly, the results obtained from the CASIA-SURF dataset demonstrate that the MLFE-based features at $\boldsymbol{L}_2$ in the RGB + HSV produce the best results. This further reinforces the observation that more detailed features lead to better accuracy.

Table 4 Comparison of the proposed method with other state-of-the-art for OULU-NPU Dataset in terms of ACER (%). The values in bold highlight that observed error rates in all four protocols are least among other state-of-the-art

Method	Protocol 1	Protocol 2	Protocol 3	Protocol 4
LBP + GS-LBP [48]	15.42	11.25	16.60 ± 11.44	36.67 ± 22.56
CCoLBP [49]	12.50	14.86	14.58 ± 11.92	22.92 ± 18.06
ED-LBP + SP y[50]	32.93	5.49	15.14 ± 15.28	41.04 ± 24.53
Proposed Method	**11.6**	**5.20**	**12.3 ± 9.33**	**18.5 ± 16.55**

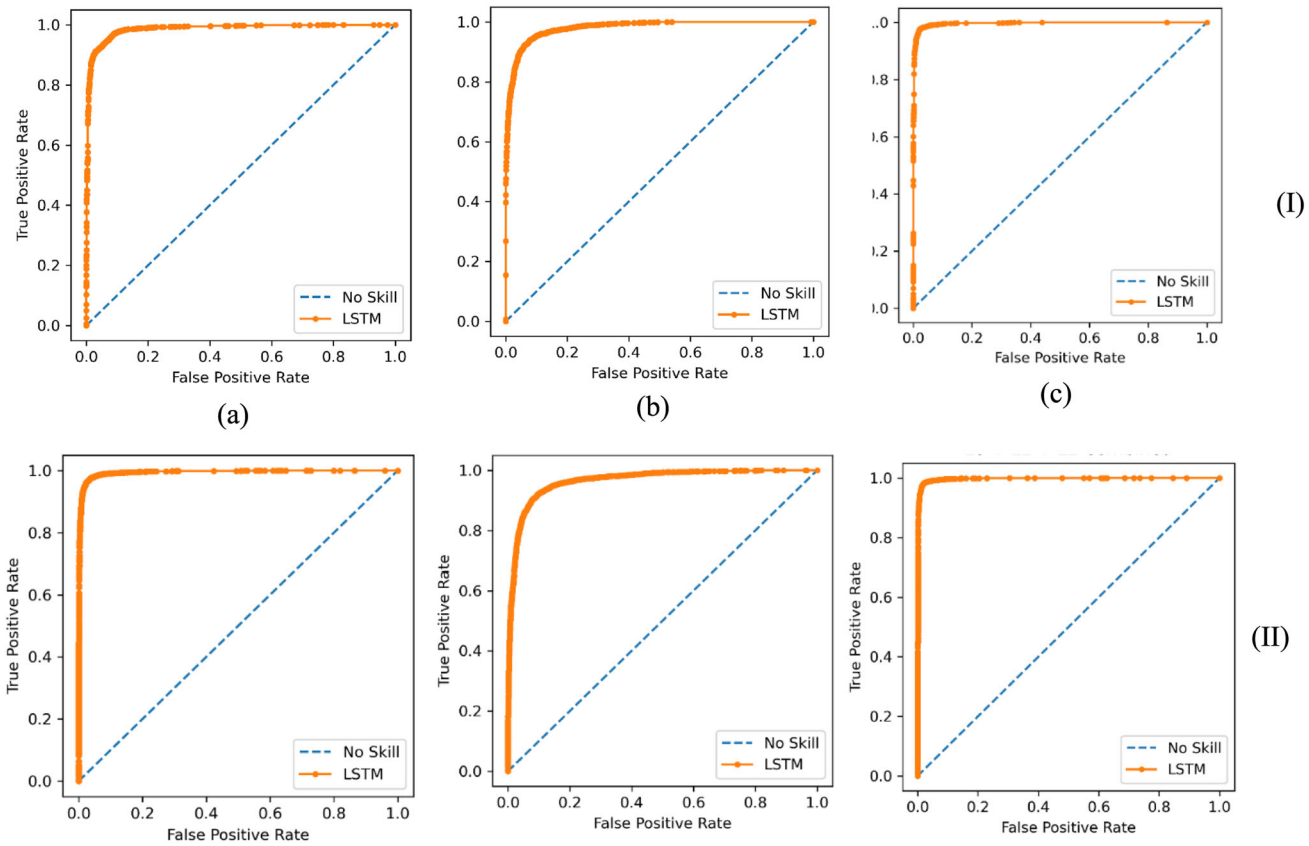


Fig. 4 Performance of proposed anti-spoofing framework in the form of ROC curves at L_2 level extracetd in **a** RGB color space, **b** HSV color space, **c** RGB + HSV color space, using LSTM learning model for (I) CASIA-FASD and (II) CASIA-SURF dataset

In terms of color spaces, it is observed in Table 6; the RGB data outperform the HSV data on the CASIA-SURF dataset with an EER of 4.1% for RGB and 9.08% for HSV. However, the results are opposite in the case of CASIA-FASD. This is due to the fact that CASIA-SURF dataset significantly holds a large set of images as compared to the CASIA-FASD dataset. Nevertheless, the combination of color space RGB+HSV is outperformed by a significant margin by getting an EER of 1.92% for both RGB and HSV color spaces on both datasets.

4.5.2 Ablation study for selection of LSTM layers

An ablation study is conducted on the CASIA-FASD dataset to demonstrate the effectiveness of using stacked LSTM layers to learn the multi-level elliptical variants of LBP face features. The experiment is evaluated for two parameters, namely, the number of stacked layers and the number of neurons in each layer. After a rigorous experimental analysis as detailed in Fig. 6, the size of LSTM layer has been chosen. The optimal results, in terms of the least EER value, are obtained for a stack of three LSTM layers with a gradual change in the number of neurons, using a downsampling factor of 2, i.e., 256, 128, and 64 neurons. As observed in Fig. 6, an increase in the number of neurons in the LSTM-3 layer leads to an increase in EER, from 1.12 to 1.83 and 2.41. Furthermore, increasing the number of neurons in LSTM-1 does not result in a great achievement in terms of a reduced value of EER.

To further analyze the impact of factors such as illumination variations and image quality on our proposed method,

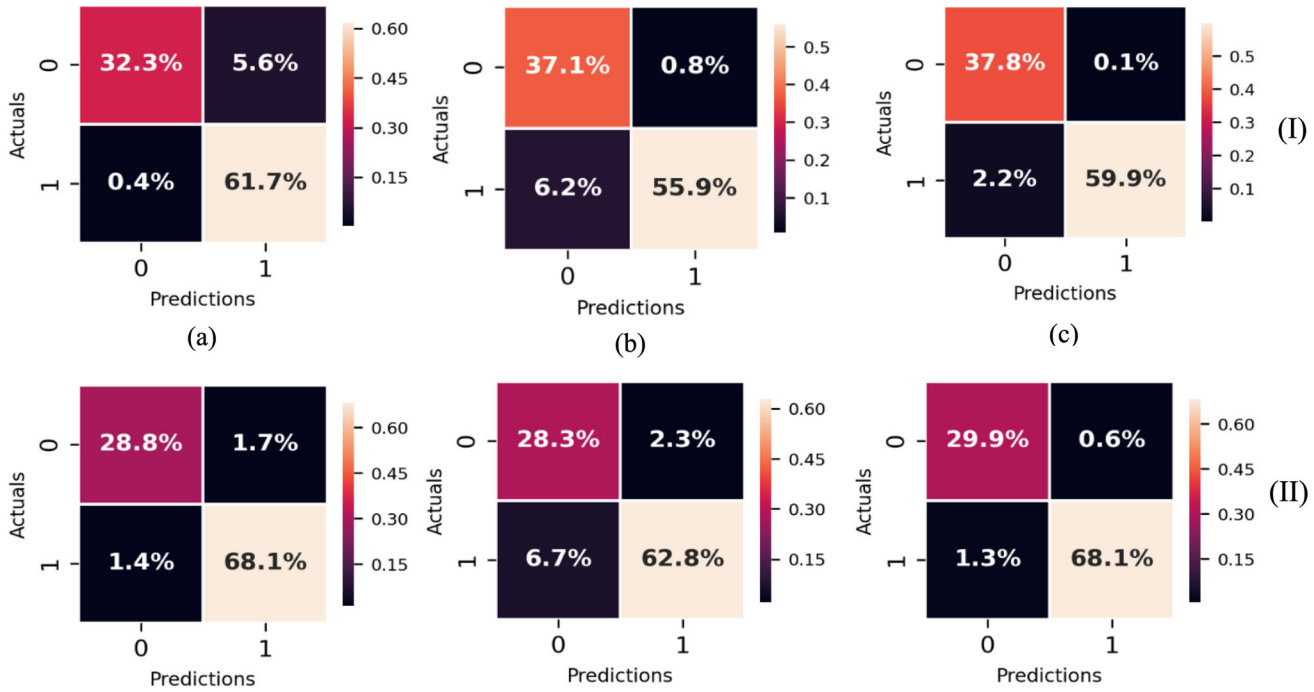


Fig. 5 Confusion Matrix for (I) CASIA-FASD dataset and (II) CASIA-SURF dataset computed at L_2 level in **a** RGB color space, **b** HSV color space, **c** RGB + HSV color space, using LSTM learning model

Table 5 Performance of the proposed framework for CASIA-FASD dataset. The values in bold highlight that for RGB+HSV inputs at L2 proposed framework observed least error rate

	EER RGB	NPCER	APCER	ACER	EER HSV	NPCER	APCER	ACER	EER RGB + HSV	NPCER	APCER	ACER
L_0	50	0	1	0.5	50	0.20	1	0.5	50	0	1	0.5
L_1	50	0	1	0.5	50	0.11	1	0.5	5.28	0.1	0.005	0.05
L_2	13.8	0.017	0.25	0.13	6.24	0.068	0.006	0.03	**1.12**	**0.021**	**0.00088**	**0.01**

Table 6 Performance of the proposed framework for CASIA-SURF dataset. The values in bold highlight that for RGB+HSV inputs at L2 proposed framework observed least error rate

	EER RGB	NPCER	APCER	ACER	EER HSV	NPCER	APCER	ACER	EER RGB + HSV	NPCER	APCER	ACER
L_0	26.5	0.005	0.48	0.26	50	0.25	1	0.5	34.7	0.054	0.66	0.35
L_1	11.09	0.046	0.17	0.11	17.55	0.18	0.16	0.17	4.39	0.03	0.033	0.044
L_2	4.1	0.020	0.0617	0.04	9.08	0.118	0.0635	0.09	**1.92**	**0.018**	**0.02**	**0.02**

we conducted an additional ablation study. Figure 7 displays a sample of illumination variation and image quality applied to the test samples to assess its effect on the architecture. For illumination variation, three different gamma values ($\gamma_1 = 1.20$, $\gamma_2 = 0.45$, and $\gamma_3 = 0.35$) were chosen; additionally, we introduced sample with different resolutions ($R_1 = \left[\frac{n}{2} \times \frac{n}{2}\right]$, $R_2 = \left[\frac{n}{4} \times \frac{n}{4}\right]$, $R_3 = \left[\frac{n}{8} \times \frac{n}{8}\right]$) to test variation in image quality. These perturbations were applied consistently across all test samples to ensure consistency in observations. The resulting observations are recorded in Table 7 for analysis. From the results presented in Table 7, it is evident that extracting Enhanced Local Binary Pattern (ELBP) features from a face image offers a certain degree of robustness against illumination variations and image quality issues. ELBP's ability to effectively capture local texture information contributes to its relative invariance to changes in illumination conditions.

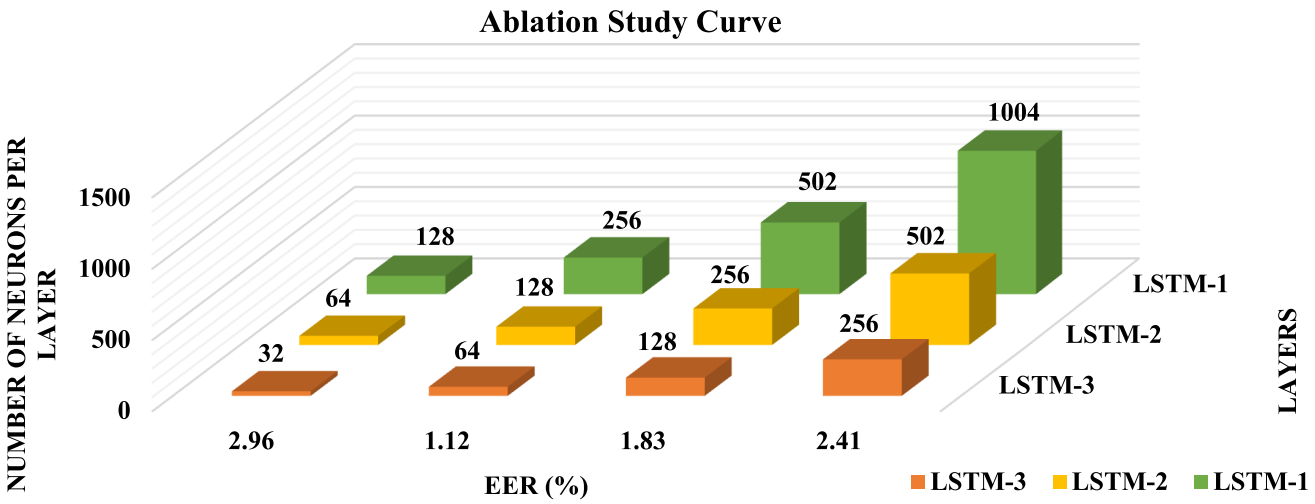


Fig. 6 Ablation study curve between number of neurons in stacked LSTM learning model with three layers and EER (%) observed in the proposed framework

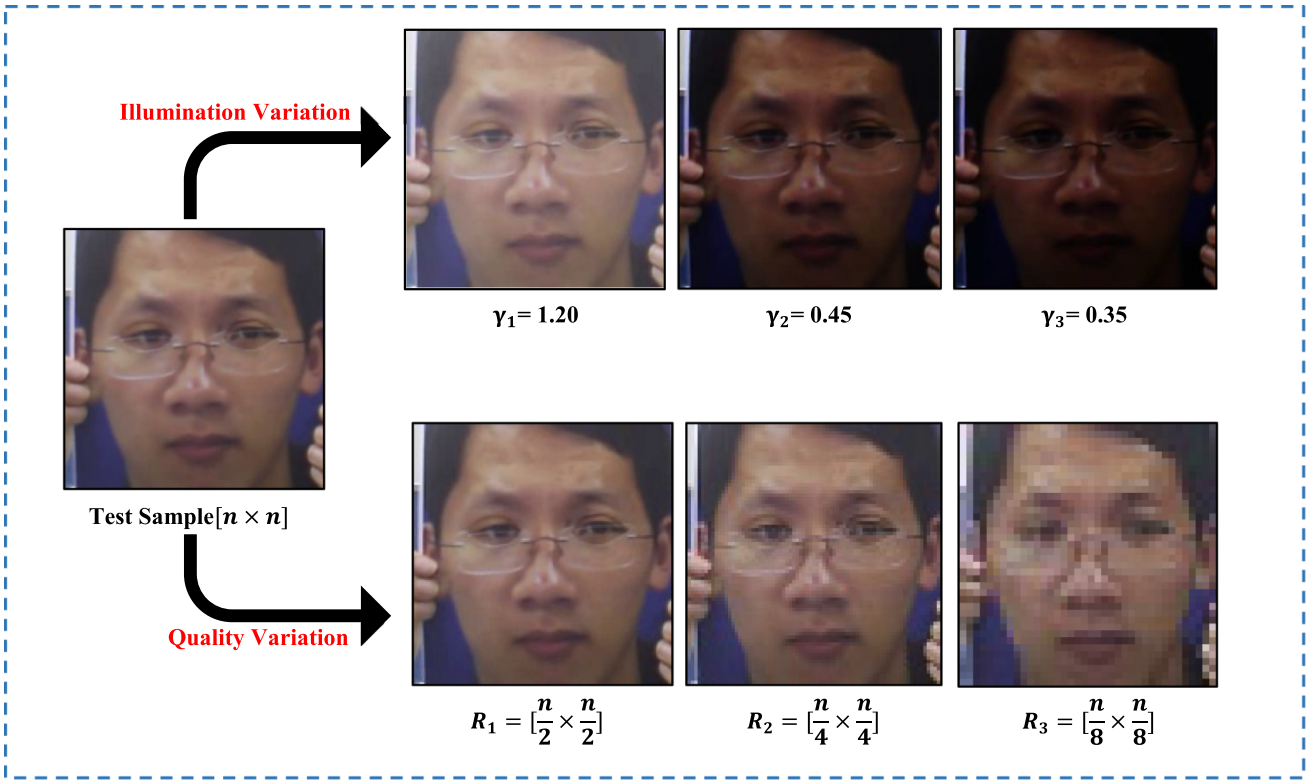


Fig. 7 Visualization of Analyzing Factors: Illumination Variation and Image Quality in CASIA-FASD Test Samples

4.6 Visualization

To gain deeper insights into the dynamics of feature enrichment within our proposed Deep Multi-level ELBP based FAS framework, we used Grad-CAM to generate visual attention maps. These maps provide valuable information about where our architecture directs its focus when analyzing faces. We present the class activation maps for test samples before the classification layer. For live faces, the architecture mainly focuses on the facial region, highlighting authentic features. In contrast, for fake faces, the attention seems scattered, indicating a lack of genuine facial characteristics. These maps provide valuable insights into the architecture's selective attention to facial regions crucial for identifying real faces. As illustrated in Fig. 8a, the architecture's emphasis

Table 7 Ablation study on CASIA-FASD (%)

Effect of impact factors on proposed architecture	Augmentation parameters	EER	NPCER	APCER	ACER
Proposed method		1.12	0.021	0.00088	0.01
Proposed Method + Illumination variation	$\gamma = 1.20$	1.11	0.021	0.0007	0.01
	$\gamma = 0.45$	1.21	0.029	0.0009	0.017
	$\gamma = 0.35$	1.24	0.031	0.0007	0.017
Proposed Method + Image Quality	$\boldsymbol{R}_1 = \frac{n}{2} \times \frac{n}{2}$	1.18	0.027	0.0007	0.01
	$\boldsymbol{R}_2 = \frac{n}{4} \times \frac{n}{4}$	1.21	0.029	0.0009	0.017
	$\boldsymbol{R}_3 = \frac{n}{8} \times \frac{n}{8}$	1.24	0.036	0.0014	0.019

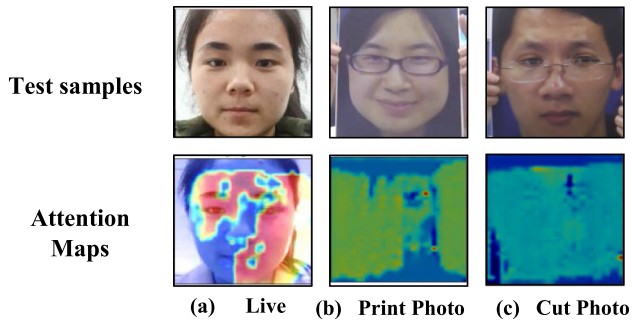


Fig. 8 Visualizing Grad-CAM attention maps for various test samples (**a**) live samples, (**b, c**) fake samples

on these areas underscores its ability to discern subtle differences between authentic and spoofed pixels. In the case of print photograph attacks, as shown in Fig. 8b, the model fails to identify any facial pixels, resulting in attention maps with uniform patterns across the frame. Similarly, for cut-out attacks, illustrated in Fig. 8c, similar activation maps are observed.

Through comprehensive ablation studies and compelling visualizations, we present robust evidence underscoring the effectiveness of our feature enhancement strategy. This comprehensive analysis not only reinforces the robustness of our approach but also sheds light on its pivotal role in advancing face anti-spoofing technology.

5 Conclusion

In this paper, we present a novel approach for detecting face spoof attacks using a Deep Face Spoof Detection Framework that combines multi-level Elliptical Local Binary Pattern (ELBP) and stacked Long Short-Term Memory (LSTM) models. Our method utilizes elliptical variants of Local Binary Patterns (LBPs) at three levels across three color spaces—RGB, HSV, and RGB + HSV, to efficiently learn distinctive features for detecting spoof attacks. By capturing both global and local features, our method achieves superior performance compared to other state-of-the-art methods. The stacked LSTM model is employed to learn the salient distribution of Multi-Level Feature Encoding (MLFE) features for both real and spoof images, generating strong decisions during the classification stage. Additionally, we conducted an ablation study to validate the selection of different parameters in the stacked LSTM model. Our results demonstrate the efficacy of our proposed approach for face spoof detection.

Author contributions Chhavi Dhiman: conceptualization of the proposed work, supervision, Manuscript Draft revision, final draft reviewed Aashania Antil: Experimentation, Ablation study, Manuscript draft revision, final draft reviewed Arnav Anand: Experimentation, Results figure preparation, Manuscript Draft writing, final draft reviewed Sohum Gakhar: Experimentation, Results figure preparation, Manuscript Draft writing, final draft reviewed

Funding Not applicable.

Data availability Data sharing is not applicable to this article as no datasets were generated during the current study.

Declarations

Competing interests The authors declare no competing interests.

Ethical approval Not applicable.

References

1. Rui, Z., Yan, Z.: A survey on biometric authentication: toward secure and privacy-preserving identification. IEEE Access **7**, 5994–6009 (2018)
2. Ramachandra, R., Busch, C.: Presentation attack detection methods for face recognition systems: a comprehensive survey. ACM Comput. Surv. (CSUR) **50**(1), 1–37 (2017)
3. Liu, A., Zhao, C., Yu, Z., Wan, J., Su, A., Liu, X., Tan, Z., Escalera, S., Xing, J., Liang, Y., Guo, G., Lei, Z., Li, S. Z., Zhang, D.: Contrastive context-aware learning for 3D high-fidelity mask face presentation attack detection. http://arxiv.org/2104.06148 [cs.CV] (2021).
4. Gragnaniello, D., Poggi, G., Sansone, C., Verdoliva, L.: An investigation of local descriptors for biometric spoofing detection. IEEE Trans. Inf. Forensics Secur. **10**(4), 849–863 (2015)

5. Boulkenafet, Z., Komulainen, J., Hadid, A.: Face spoofing detection using colour texture analysis. IEEE Trans. Inf. Forensics Secur. **11**(8), 1818–1830 (2016)
6. Kollreider, K., Fronthaler, H., Bigun, J.: Verifying liveness by multiple experts in face biometrics. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (2008)
7. Pan, G., Sun, L., Wu, Z., Lao, S.: Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In: IEEE 11th International Conference on Computer Vision (2007)
8. Bell, S., Zitnick, C.L., Bala, K., Girshick, R.: Inside-outside net: detecting objects in context with skip pooling and recurrent neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA (2016)
9. Byeon, W., Breuel, T.M., Raue, F., Liwicki, M.: Scene labeling with LSTM recurrent neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA (2015)
10. Antil, A., Dhiman, C.: A two stream face anti-spoofing framework using multi-level deep features and ELBP features. Multimed, Syst. **29**(3), 1361–1376 (2023)
11. Chen, F.M., Wen, C., Xie, K., Wen, F.Q., Sheng, G.Q., Tang, X.G.: Face liveness detection: fusing colour texture feature and deep feature. IET Biom. **8**, 369–377 (2019)
12. Chingovska, I., Anjos, A., Marcel, S.: On the effectiveness of local binary patterns in face anti-spoofing. In: BIOSIG-Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG). IEEE, pp. 1–7 (2012)
13. de Freitas Pereira, T., Anjos, A., De Martino, J. M., Marcel, S.: LBP—TOP based countermeasure against face spoofing attacks. In: Asian Conference on Computer Vision. Berlin, Heidelberg (2012)
14. Wang, Y., Nian, F., Lia, T., Meng, Z., Wang, K.: Robust face anti-spoofing with depth information. J. Vis. Commun. Image Represent. **49**, 332–337 (2017)
15. Antil, A., Dhiman, C.: Two stream RGB-LBP based transfer learning model for face anti-spoofing. In: CVIP, India, (2023)
16. Li, J., Wang, Y., Tan, T., Jain, A. K.: ive face detection based on the analysis of Fourier spectra. In: Biometric Technology for Human Identification (2004)
17. Määttä, J., Hadid, A., Pietikäinen, M.: Face spoofing detection from single images using micro-texture analysis. In: International Joint Conference on Biometrics (2011)
18. Tan, X., Li, Y., Liu, J., Jiang, L.: Face liveness detection from a single image with sparse low rank bilinear discriminative model. In: European Conference on Computer Vision, Berlin, Heidelberg (2010)
19. Boulkenafet, Z., Komulainen, J., Hadid, A.: Face anti-spoofing based on color texture analysis. In: IEEE International Conference on Image Processing (ICIP) (2015)
20. Kollreider, K., Fronthaler, H., Faraj, M.I., Bigun, J.: Real-time face detection and motion analysis with application in "liveness" assessment. IEEE Trans. Inf. Forensics Secur. **2**, 548–558 (2007)
21. Shreve, M., Godavarthy, S., Goldgof, D., Sarkar, S.: Macro-and micro-expression spotting in long videos using spatio-temporal strain. In: IEEE International Conference on Automatic Face and Gesture Recognition (2011)
22. Siddiqui, T. A., Bharadwa, S., Dhamecha, T.I., Agarwal, A., Vatsa, M., Singh, R., Ratha, N.: Face anti-spoofing via motion magnification and multifeature videolet aggregation. In: International Conference on Pattern Recognition (ICPR), Cancun, Mexico (2016)
23. Yang, J., Lei, Z., Li, S. Z.: Learn convolutional neural network for face anti-spoofing. http://arxiv.org/1408.5601v2 (2014)
24. Alotaibi, A., Mahmood, A.: Deep face liveness detection based on nonlinear diffusion using convolution neural network. SIViP **11**, 713–720 (2017)
25. Lucena, O., Junior, A., Moia, V., Souza, R., Valle, E., Lotufo, R.: Transfer learning using convolutional neural networks for face anti-spoofing. In: International Conference Image Analysis and Recognition, Cham, (2017)
26. Ren, S., He, K., Girshick, R.: Faster r-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems (2015)
27. Krizhevsky, A., Sutskever, I., Hinton, G. E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems (2012)
28. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
29. Xu, Z., Li, S., Deng, W.: Learning temporal features using LSTM-CNN architecture for face anti-spoofing. In: 3rd IAPR Asian Conference on Pattern Recognition (2015)
30. Tu, X., Zhang, H., Xie, M., Luo, Y., Zhang, Y., Ma, Z.: Enhance the motion cues for face anti-spoofing using CNN-LSTM architecture. arXiv preprint arXiv, (2019)
31. Viola, P., Jones, M. J., Snow, D.: Detecting pedestrians using patterns of motion and appearance. In: International Journal of Computer Vision (2003)
32. Li, L., Feng, X., Xia, Z., Jiang, X., Hadid, A.: Face spoofing detection with local binary pattern network. J. Vis. Commun. Image Represent. **54**(july), 182–192 (2018)
33. Zhang, L.B., Peng, F., Qin, L., Long, M.: Face spoofing detection based on color texture Markov feature and support vector machine recursive feature elimination. J. Visual Commun. Image Represent. **51**, 56–69 (2018)
34. Ojala, T., Pietikäinen, M., Mäenpää, T.: Gray scale and rotation invariant texture classification with local binary patterns. In: European Conference on Computer Vision. Berlin, Heidelberg (2000)
35. Liu, L., Fieguth, P., Guo, Y., Wang, X., Pietikäinen, M.: Local binary features for texture classification: taxonomy and experimental study. Pattern Recognit. **62**, 135–160 (2017)
36. Nguyen, H. T., Caplier, A.: Elliptical local binary patterns for face recognition. In: Asian Conference on Computer Vision (2012)
37. Liang, X., Shen, X., Xiang, D., Feng, J., Lin, L., Yan, S., Shuicheng: semantic object parsing with local-global long short-term memory. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA (2016)
38. Boulkenafet, Z., Komulainen, J., Hadid, A.: Face antispoofing using speeded-up robust features and fisher vector encoding. IEEE Signal Process. Lett. **24**, 141–145 (2016)
39. Siddiqui, T. A., Bharadwaj, S., Dhamecha, T. I., Agarwal, A., Vatsa, M., Singh, R., Ratha, N.: Face anti-spoofing with multifeature videolet aggregation. In: 23rd International Conference on Pattern Recognition (2016)
40. Pereira, T.D.F., Komulainen, J., Anjos, A., Martino, J.M.D., Hadid, A., Pietikäinen, M., Marcel, S.: Face liveness detection using dynamic texture. EURASIP J. Image Video Process. (2014). https://doi.org/10.1186/1687-5281-2014-2
41. Boulkenafet, Z., Komulainen, J., Hadid, A.: On the generalization of color texture-based face anti-spoofing. Image Vis. Comput. **77**, 1–9 (2018)
42. Zhang, S., Liu, A., Wan, J., Liang, Y., Guo, G., Escalera, S., Li, S.Z.: Casia-surf: a large-scale multi-modal benchmark for face anti-spoofing. IEEE Trans. Biom. Behav. Identity Sci. **2**(2), 182–193 (2020)
43. Chen, X., Xu, S., Ji, Q., Ca, S.: A dataset and benchmark towards multi-modal face anti-spoofing under surveillance scenarios. IEEE Access **9**, 28140–28155 (2021)

44. Wang, G., Han, H., Shan, S., Chen, X.: Improving cross-database face presentation attack detection via adversarial domain adaptation. In: ICB, Crete, Greece (2019)
45. Almeida, W.R., et al.: Detecting face presentation attacks in mobile devices with a patch-based CNN and a sensor-aware loss function. PLoS ONE **15**(9), e0238058 (2020)
46. Liu, A., Tan, Z., Wan, J., Liang, Y., Lei, Z., Guo, G., Li, S.Z.: Face anti-spoofing via adversarial cross-modality translation. IEEE TIFS **16**, 2759–2772 (2021)
47. Antil, A., Dhiman, C.: MF2ShrT: multimodal feature fusion using shared layered transformer for face anti-spoofing. ACM Trans. Multimed. Comput. Commun. Appl. **20**(6), 1–21 (2024)
48. Peng, F., Qin, L., Long, M.: Face presentation attack detection using guided scale texture. Multimed. Tools Appl. **77**, 8883–8909 (2017)
49. Peng, F., Qin, L., Min, L.: CCoLBP: chromatic co-occurrence of local binary pattern for face presentation attack detection. In: ICCCN (2018).
50. Shu, X., Tang, H., Huang, S.: Face spoofing detection based on chromatic ED-LBP texture feature. Multimed. Syst. **27**, 161–176 (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.

Environment, Development and Sustainability
https://doi.org/10.1007/s10668-024-04934-6

A logistic regression model to facilitate setting of organic waste composting policy for sustainable waste management

Majed Ibrahim Al-Sari'[1,2] · A. K. Haritash[1]

Received: 19 January 2024 / Accepted: 15 April 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract
Landfilling of solid waste is the common waste management (SWM) option currently in use in the southern West Bank of Palestine, which is unsustainable and causing many environmental and health problems. However, composting is an environmentally friendly waste treatment option, but of low interest due to the limited capacity of the service providers and the absence of effective waste management policies. Understanding the challenges faced by local authorities (LAs), the key stakeholders for SWM, and the factors that affect their attitude towards composting of organic waste is a major contributor to the successful setting and implementation of policies related to composting. Therefore, this study aims to assess the capacities of the LAs, the challenges faced by them, and identify the factors affecting the LAs' attitude towards organic MSW composting in southern West Bank/Palestine, Hebron and Bethlehem governorates, as a prerequisite for setting up a composting policy. For data collection, a structured questionnaire was designed, and the data was collected from all LAs that provide SWM services in the study area. The data analysis showed that 100% of the LAs are suffering from the absence of proper equipment and machinery to conduct composting (the highest challenge), while 22.6% of them are worried about farmers' acceptance of compost produced from MSW (the lowest challenge). The results also showed that LAs' perception of compost contribution to SW reduction, availability of proper place, financial capacity, community awareness, and prevalent SWM bylaws are significant predictors of LAs' attitudes toward composting. Accordingly, setting a composting policy shall consider these influencing factors to ensure the effectiveness of the policy. The results of this study are valid for similar countries and can assist decision-makers in assessing the significant influencing factors. The scientific value added by this research is that the logistic regression modelling (LRM) can be used by other countries as well, as a tool for drafting policies.

Keywords Organic waste · Attitude · Local authorities · Composting · Policy

✉ Majed Ibrahim Al-Sari'
majedsari@gmail.com

[1] Department of Environmental Engineering, Delhi Technological University, Bawana Road, Shahbad Daulatpur, Delhi 110042, India

[2] Present Address: Universal Institute for Applied and Health Research, Nablus, Palestine

1 Introduction

Attitude and behavior and the setting of policies are among the factors that can greatly affect the performance of solid waste management (SWM) systems ((Fan et al., 2021; Yang et al., 2023). Previous studies reported inconsistent application of waste management policies in developing countries (Cetrulo et al., 2018). The lack of resources and technical capacity, lack of public participation, and awareness of waste management practices have proven to be influencing factors and main reason behind the ineffective waste management in many countries (Cetrulo et al., 2018). Understanding the barrier factors is of much importance before setting waste management policies. Trinh and Giao (2023) studied the willingness to pay the cost of electronic waste (e-waste) recycling by young consumers in Vietnam and concluded that those young consumers seriously lack knowledge about e-waste management legislation, policies, and information about the programs of collection and treatment. Budijati et al. (2023) studied the inhibiting factors for the collaboration between formal and informal sectors in handling used cell phones in Indonesia and found that collaboration was significantly hindered by insufficient organization and the absence of infrastructure. In accordance with the Asian Development Bank Institute (ADBI), the existing implications of waste management policies in developing Asia are due to the gaps in the current policies to improve the current SWM practices (ADBI, 2020). However, waste management policies are the main instrument to regulate, promote, and improve waste management practices. Therefore, overcoming the challenges in the waste management sector is a prerequisite to setting and forcing any SWM policy.

In developing countries in general, and Palestine in particular, organic waste represents the largest fraction of municipal solid waste. This waste stream is increasing following population growth and adding more pressure on the environment. It reaches 46% of the total solid waste stream in the southern part of the West Bank/Palestine as reported by the International Finance Cooperation–IFC (IFC, 2012). Improper management of the waste is the major problem behind several environmental and social impacts including surface and groundwater resources contamination (Amano et al., 2021), health risks (Vinti et al., 2023), gaseous emissions to the atmosphere and air quality degradation (Wiedinmyer et al., 2014), social disturbance due to odour release from landfills (Rana et al., 2018), and significant contributor to the climate change because of methane emissions (Sharma et al., 2019).

However, the application of advanced technologies such as green nanotechnology and advanced materials and composites, and nanostructures for improving people's lives, and environmental remediation, sustainability, renewable and clean energy production and storage can significantly contribute to environmental protection and sustainability. For example, $CoFe_2O_4@SiO_2@Dy_2Ce_2O_7$ nanocomposites can be used for removal of water contaminants (Zinatloo-Ajabshir & Salavati-Niasari, 2019), $Zn_3V_3O_8$ nanostructures can be used for electrochemical hydrogen storage (Ghodrati et al., 2019), Z-scheme heterojunction nanocomposite can be used as a photocatalyst for decomposition of nitenpyram insecticide (Hosseinzadeh et al., 2023), $Nd_2Sn_2O_7$ nanostructures produced with applying grape juice as green fuel at 500 °C can be used as photocatalyst for degrading erythrosine (Zinatloo-Ajabshir et al., 2019). Photocatalysis, which is an advanced oxidation process, is a potential environmental-friendly technology to decompose poisonous organic compounds in ambient conditions and under the irradiation of sunlight (Zinatloo-Ajabshir et al., 2024). Lanthanum zirconate ceramic microstructure could be used for electrochemical green energy sources such as hydrogen storage (Zinatloo-Ajabshir et al., 2023). In addition, innovative waste recycling technologies such as the use of machine learning for waste

separation (Fadlil et al., 2022) and recovery of carbon and fiberglass from waste composites for reuse as reinforced building materials (Nistratov et al., 2022) are major contributors to environmental sustainability. Similarly, composting is a green method for sustainable organic waste management. Composting of organic waste was reported as one of the solutions for solid waste management (SWM) in developing countries (Hettiarachchi, et al., 2018), which can eliminate contaminants and mitigate significant environmental impacts. It is a sustainable waste management option as it contributes to the circular economy and combats the depletion of resources (Kadhila et al., 2023). However, composting of organic waste faces several difficulties, especially in low-income countries. Several challenges were reported to affect the attitude towards composting including the ability of the community to participate in composting programs, availability and adequacy of resources (infrastructure, financial and technological), awareness and training of compost-production and use, market availability of the produced compost, support from the government and political leadership, knowledge of national policy and laws, society and business community contribution, and innovative methods for composting (Lalitha & Fernando, 2019).

In Palestine, SWM is within the purview of the local authorities (LA), which include municipalities, village councils (VCs), joint councils for services, planning and development (JCSPD), and joint service councils for solid waste management (JSCs-SWM). The JCSPDs are organizations formed by the VCs to provide services including SW collection to optimize these services. The JSCs-SWM are also organizations formed by the LAs to provide SW collection, transfer, and disposal. In the study area, southern West Bank (Hebron & Bethlehem governorates), the Joint Service Council for Solid Waste Management for Hebron and Bethlehem Governorates (JSC-H&B) is responsible for SW transfer and disposal, which manage SW transfer stations and the regional landfill. The interrelationship between these agencies is presented in Fig. 1. LAs (municipalities and VCs) are responsible for waste collection, which is carried out by its own waste collection vehicles (WCV) or contracted the collection service, fully or partially, to the JSCs-SWM or JCSPDs. Accordingly, these LAs are responsible for fee collection from households and pay the collection, transfer, and disposal fees to the contracting agency (JSCs-SWM or JCSPs) that pay waste transfer and disposal fees to the JSC-H&B, or pay the waste collection fees only to the contracting agency and waste transfer and disposal fees to the JSC-H&B directly. The figure presents the structure of the financial relationship between the LAs in the study area, which reflects the SW flow starting from production to disposal. Any improvement in WM through waste composting will affect this financial relationship.

However, the MoLG classifies the local authorities (municipalities only) into four groups, group "A", group "B", group "C" and group "D". This classification system is based on the local government minister's decision no. 20/4/1998. In accordance with the

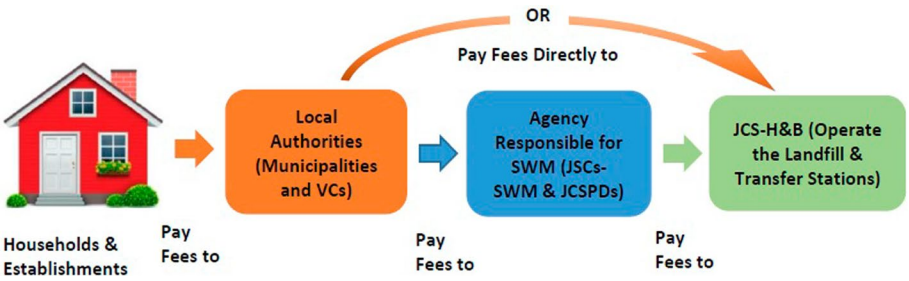


Fig. 1 The current structure of municipal SWM service in the study area

decision, large municipalities that are considered central of governorates are classified as group "A", municipalities that were established before the Palestinian Authority rolled over the West Bank and Gaza and with a population more than 15,000 are classified as group "B", municipalities of the population more than 5000 and less than 15,000 are classified as group "C", and municipalities of population less than 5000 are classified as group "D" (MoLG, 2005). The classification system doesn't provide any other information about the purpose of the classification, and the link between responsibilities and classification.

The current waste management option in the area is collection and landfilling at the engineered sanitary landfill site. The landfill is the major source of social nuisance and environmental problems due to odor emissions and leachate generation. Moreover, the lifespan of the landfill is reducing extensively due to the large quantities of SW delivered to the site, which is around 1200.0 tons daily (JSC-H&B, 2022). The extension of the landfill is currently facing several social and political constraints, which could create waste disposal problems and place the service providers under severe circumstances. Only a few sorting activities are conducted for some recyclables such as plastics, metals, and sometimes wood, but do not exceed 0.5% of the total waste stream. Pilot composting was conducted and succeeded in the production of good quality compost, but stopped because of the absence of technical and financial resources.

On the waste management policy level, the Palestinian Authority has issued the SWM bylaw in accordance with the ministers' decision no. 3 in 2019. The bylaw specified the technical requirements of SWM facilities and the obligations of all stakeholders and parties in SWM. However, waste recycling and composting were mentioned under Article 27, which encourages competent authorities to recycle solid waste and production of composting, but without targets or obligatory standards. Therefore, the bylaw remains ineffective with respect to waste recycling and composting. On the implementation level, many of the previous SW composting programs in Palestine failed. For example, the Palestinian Agricultural Relief Committee (PARK) has established three composting plants in Jericho, Jenin (Di Maria et al., 2017) and Hebron for the benefit of farmers' cooperatives. Both plants in Jericho and Hebron failed due to the high cost of operation and low participation of the local population, while the other in Jenin works during summer only (CESVI, 2019). Another two composting plants were constructed in Jericho and Ramallah (Bytillu village) in 2017 for composting agricultural waste and organic waste collected from hotels (CESVI, 2019). However, both plants stopped working in 2018. In the Gaza Strip, two composting plants have been constructed. One of them is located in the north and belongs to the Gaza and Beit Lahia municipalities, while the other is in Rafah city and was established by the Palestinian Environmental Friend (PEF) NGO, which stopped working (CESVI, 2019). The PEF then obtained funds from the United Nations Development Program (UNDP) in 2011 in coordination with Rafah municipality and constructed a sorting plant and conducted composting for the organic waste separated, but unfortunately, it stopped working in 2017 due to marketing problems (CESVI, 2019). The failure of such composting programs associated with ineffective SWM bylaw indicates a large gap between policy and practice.

On the other hand, the waste management sector was identified as one of the major sectors emitting 751.7 Gg CO_2 eq which is about 23% of total emissions as reported by the State of Palestine's Country Program for the Green Climate Fund (SPCP-GCF, 2019). The national plans and strategies aim to reduce the waste disposal at landfills by 30% and 50% by 2030 and 2040, respectively, and methane emissions by 30% in 2030 and 70% by 2040 (State of Palestine, 2021). To achieve the targets in accordance with these plans, organic waste composting is one of the appropriate options for municipal SWM, and an effective waste management policy is required. However, identifying the attitude of the local authorities towards composting and

overcoming the challenges it faces can facilitate the employment of composting to achieve the country's emissions targets. Previous studies showed that the attitudes of the LA toward organic waste composting were dissatisfactory because 63.5% of the LAs in the study area did not plan for composting (Al-Sari & Haritash, 2023). The study further identified nine factors significantly affecting their attitude including financial capacity, suitable machinery, enough refuse collection trucks for separate collection of solid waste fractions, availability of proper place for composting operations, familiarity with systems of composting, previous experience of the staff in composting, acceptance of rapid composting system, training staff members on compost production, and believing that composting of organic solid waste is within the responsibility of the LA's (Al-Sari & Haritash, 2023).

1.1 Research gap

Based on the context analysis mentioned above, there is a wide gap between the SWM policies and practice in Palestine. Although many studies have been conducted on SW recycling and composting, studies on the attitude of the LAs towards SW recycling in general and composting, in particular, are scarce. The study conducted by Al-Sari' and Haritash (2023) was based on bivariate analysis, which examines the effect of each influencing factor separately, to identify the factors influencing LAs towards composting. However, there is no study focused on using regression analysis, which examines several factors together in one model and identifies the significant ones among them.

1.2 Goal of the study

The main goal of this study is to identify the factors affecting the LAs toward composting using logistic regression analysis. In addition, it aims to assess the current capacities of the LAs and the challenges faced by them concerning organic SW management and composting. The present study bridges that gap and helps the policy-makers better understand the effect of regulating factors in implementation of a policy.

1.3 Conceptual framework

The conceptual framework and scope of this study were designed to provide answers related to the gap between the current policies and practice. Therefore, the research focused on the current situation of the LAs and provided information about their size, and cost-related issues including the cost of waste collection and disposal, number of refuse collection vehicles (RCVs), number of workers in SWM divisions, and collection rate of SW fees. These issues indicate the technical, financial, and institutional capacity of the LAs in SWM. In addition, challenges faced by LAs in organic waste management and composting were also assessed. Furthermore, it assessed the factors that affect their attitude towards the composting of organic waste and inhibit the execution of SWM policies.

2 Research methods

2.1 Research population and sample size

This research focuses on the southern part of the West Bank/Palestine, Hebron and Bethlehem governorates. The population of the study is the local authorities (municipalities, VCs, JSCs-SWM, and JCSPDs). Because the population is limited, the sample size included all the LAs that are currently providing SWM service in accordance with the list obtained from the MoLG. The list showed 85 LAs providing the service, which represents the sample size.

2.2 Survey design and data collection

A structured survey questionnaire was designed to collect the required data. The categories of the questions included information about the service provider of SWM, the quantity of SW generated, tariff, fee collection, cost of SW collection and disposal, community awareness, familiarity with SWM bylaw, compost acceptance by farmers, financial and technical capacity of the LAs to conduct composting, government support to the LAs, composting knowledge, challenges facing the LA's towards composting, and the served population which indicate the size of the LA. The data was collected via the designed survey questionnaire through direct interviews with the key staff upon direction by the LA officials. In some cases, the survey was conducted over the phone and via email at the convenience of the targeted LA. The duration of the survey was two months, which was conducted through the period extending from December 2021 to January 2022.

2.3 Data analysis

The data analysis was conducted using the statistical package for social sciences software (SPSS) version 23, and Microsoft Excel 2016. Descriptive statistics including frequencies and cross tabulation are used to assess the LAs' classifications, population, and challenges faced by them and other related issues. In addition, a binary logistic regression model (LRM) was developed to assess the factors that affect the LAs' attitude towards composting of municipal organic waste. Although the bivariate analysis can be used to assess the relationship between the dependent and independent variables, it can only assess the effect of one single variable in the absence of others, and can't predict the direction of the relationship between the dependent and independent variable. The LRM is widely used to test the effect of a combination of several explanatory variables on response variables in one model to estimate the influence of each explanatory variable in the presence of the other variables. It can identify the significance of the relationship between the dependent and the independent variables and the direction of the relationship as repeated in previous studies (Ali et al., 2012; Al-Khateeb et al., 2017; Al-Madbouh et al., 2019; Al-Sari' et al., 2012; Begum et al., 2006, 2009; Ittiravivongs, 2012). The dependent variable is the LAs' planning to compost organic SW fraction, while the independent variables are presented in Table 1. The compost selling price was not considered as an independent variable because there was no municipal solid waste composting on a commercial scale. The independent variables were selected carefully to avoid multicollinearity occurrence, which means no two independent variables have a correlation of more than 0.7 (Begum et al., 2009). The multicollinearity can cause inaccurate variances and, consequently, incorrect relationships

Table 1 Summary of the independent variables in the LRM

Variable	Description	Definition	Abbreviation in the model
X_1	Monthly SW quantity generated (tons)	1 = ≤ 20; 2 = 21–50; 3 = 51–100; 4 = 101–200; 5 = 201–400; 6 = 401–600; 7 = > 600	Mon. SW generated Qt
X_2	Household SW tariff (NIS)	1 = 10–20; 2 = 21–30; 3 = (2–3)/capita	HH SW tariff
X_3	Fee collection rate (%)	1 = ≤ 20; 2 = 20–40; 3 = 41–50; 3 = 41–60; 4 = 61–80; 5 = 81–100	Fee collection rate
X_4	Perception of composting contribution to reduction in SW	1 = Yes; 2 = Don't know; 3 = NO	Per. of compost reduce SW
X_5	Having an appropriate place for composting	1 = Yes; 2 = NO	Place availability
X_6	Having the financial capacity to conduct composting	1 = Yes; 2 = Yes, but limited capacity; 3 = NO	Financial capacity
X_7	Staff attended training on compost production	1 = Yes; 2 = No	Staff training
X_8	Staff having previous experience in composting	1 = Yes; 2 = No	Previous experience
X_9	Community awareness conducted on organic waste composting	1 = Yes; 2 = NO	Community awareness
X_{10}	Potential community participation in compost through organic SW sorting at the source	1 = Yes; 2 = Yes, but partially; 3 = NO	Pot. Comm. participation
X_{11}	Familiarity with SWM bylaw	1 = Yes; 2 = NO	Familiarity with SWM bylaw

between dependent and independent variables. The LRM used is presented in Eq. (1) as follows:

$$Log\frac{P_i}{1-P_i} = Z_i = \beta_0 + \beta_i X_i + e \tag{1}$$

where *Pi* is the LA's attitude toward organic waste composting. *Pi* = 1 if the LA attitude toward organic waste composting is positive; and Pi = 0 if not; *Xi* is the explanatory (independent) variable which is listed in Table 1; *βo* is a constant term; *βi* is a coefficient of the independent variable; *e* is the error term and *1, 2 …, n* is the number of the independent variables in the model. The sign of the coefficient *βi* determines the direction of the relationship for the dependent variable *Pi* and the independent variable *Xi*.

The model uses the maximum likelihood function to estimate the model coefficients, which is the most widely used method. The logistic regression model estimates the probability of a certain event occurring by calculating the changes in the logarithm of the dependent variable. The likelihood function, as defined in Eq. (2), expresses the values of β in terms of known and fixed values of y (β is related to P) and is derived from the probability distribution of the dependent variable so that the values of β that maximize the output of Eq. 2 are the maximum likelihood estimates (Begum et al., 2009).

$$L\left(\frac{\beta}{y}\right) = \prod_{i=1}^{N} \frac{n_i!}{y_i!(n_i - y_i)!} P_i^{y_i}(1 - P_i)^{(n_i - y_i)} \tag{2}$$

The statistical significance of each coefficient was evaluated using the Wald test (Begum et al., 2009), which is used to determine whether each independent variable in the model is significant or not.

$$W_i = \left(\frac{\beta_i}{S.E_{\beta_i}}\right)^2 \tag{3}$$

where $i = 1, 2 \ldots, n$ and $S.E$ = standard error.

The log-likelihood function, the omnibus test, Cox and Snell R^2, and Naglekerke $\check{R}^2$ were used to evaluate the suitability of the model (Al-Sari' et al., 2012). The log-likelihood function is used to measure the goodness of the data fit of the model and is defined as per Eq. (4) (Begum et al., 2009).

$$Log - likelihood = \sum_{i=1}^{n} [Y_i \ln(\hat{Y}_i) + (1 - Y_i)\ln(1 - \hat{Y}_i)] \tag{4}$$

As the model uses the observations to predict the expectations, Y_i is the actual result (the observation), and $\hat{Y}_i$ the predicted probabilities of the result (the expected), which is quoted as 2log-likelihood since it has approximately chi-squared distribution. The Omnibus test, which is a likelihood-ratio chi-squared test, provides an indication of the goodness of fit if the coefficients of the variables in the model are all jointly equal to zero. Cox and Snell R^2 and Nagelkerke $\check{R}^2$ indicate the variation proportion in the dependent variable explained by the independent variables of the model. Cox and Snell R^2 values cannot achieve a maximum value of 1, therefore Nagelkerke $\check{R}^2$ is used because it is an adjusted version of the Cox and Snell R^2 and its values are in the range from 0 to1, and is often preferred (Bewick et al., 2005).

Figure 1 depicts the methodological approach process of the research starting from the population and sample determination and ending with results presentation.

3 Results and discussion

To understand the situation of the LAs in the study area, institutional, technical, and financial indicators were analyzed to better describe the current situations and challenges faced by these LAs. Then factors of the potential effect of the attitudes of the LAs towards composting were collected in one LRM to assess its degree of effect. The following sections present these results in detail.

3.1 Classification of the LAs

The data analysis showed that 2.4% of the municipalities (2 municipalities) are classified "A", 20.0% (17 municipalities) are classified "B", and 22.4% (19 municipalities) are classified "C", while there is no official classification of VCs, JSCs-SWM, and JCSPDs which represent 55.3% (47 VCs, JSC-SWM & JCSPD) of the total LAs. The LAs and the population served in the study area are shown in Fig. 2. It is worth to mention that the classification of the LAs was based on the number of residents within the border of the LA in accordance with the MoLG rules. The classification system, as discussed in the introduction section, indicates the size of the LA but doesn't provide a link between the size and responsibilities or any other regulations. Previous studies reported that the size of the LA affects the financial performance because large LAs have more resources to generate revenues (Zamzami & Rakhman, 2023).

Figure 2 shows that communities with a population less than 5000 are fully served by VCs, 68.8% of communities with a population of 5001–10000 are served by municipalities and 31.2% of them are served by VCs, 100% of the communities with a population of 10,001–20000 are served by municipalities, 83.3% of the communities with a population of 20,001–30000 are served by municipalities and 16.7% are served by JSCs-SWM and JCSPDs, all of the communities (100%) with population of 30,001–40000 are served by municipalities, 100% of the service is provided by the JSCs-SWM and JCSPDs in the communities with population 40,001–60000, and 60% of the service is provided by municipalities in the communities with population more than 60,000 and 40% is provided by the JSCs-SWM and JCSPDs. The interrelationships between the above-mentioned actors

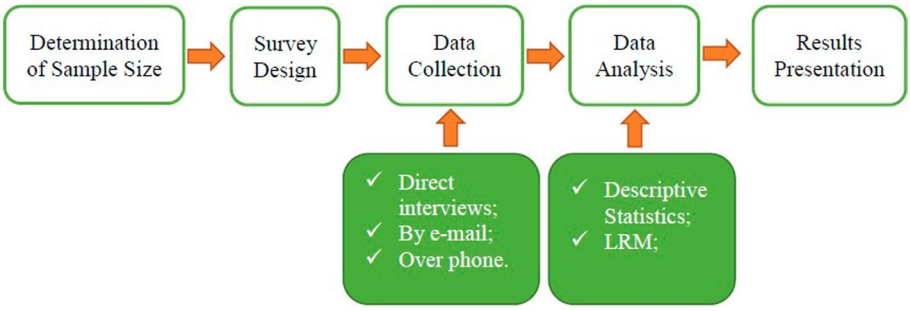


Fig. 2 Research methods process

in SWM are mainly related to the fact that all of them are authorized to provide SWM service and have the authority to contract the service or part of it in case they haven't enough capacity to cope with the waste generated in the community.

3.2 Cost-related issues of municipal SWM

SW collection and disposal cost differs from one LA to another depending on the quantity of waste generated, the number of refuse collection vehicles, and the number of workers. This is dependent on the population and the size of the LA as well. This section highlights the cost of SW collection and disposal and the elements that could influence the costs. The results reveal that the monthly waste collection cost is less than or equal to NIS 10,000.0 for 43.8% of the LAs which are of small size (i.e. the VCs). The monthly waste collection cost is between NIS 10,000.0 and 300,000.0 for the majority of the LAs (51.4%), while the cost exceeds NIS 300,000.0 for a few LAs (5.0%) as presented in Table 2. However, the monthly disposal cost for the majority (59.3%) of the LAs is less or equal to NIS 5000.0, while the disposal cost is between NIS 5000.0 and NIS 30,000.0 for 30.8% of the LAs. Few LAs (9.9%) exceed the cost of NIS 30,000.0 as disposal cost. As the organic waste fraction is the highest, the highest portion of the SW collection and disposal costs is attributed to the organic waste fraction. However, following other options in organic waste management such as composting could, to a large extent, reduce these costs. Jalalipour et al. (2024) concluded that home composting reduces SWM costs and leads to a 5% saving for the municipality of Shiraz/Iran.

Selected factors of potential influence on SWM are presented in Table 3. The results showed that 18.5% of the LAs are generating 20 tons or less of SW monthly, while 14.8% of them are generating more than 600 tons on a monthly basis. The majority of LAs (66.7%) are producing in the range of 50–600 tons of SW per month. For the refuse collection vehicles (RCVs), 67.5% of the LAs don't own RCVs because they contracted the service to the JCSPD, JSC-SWM, or nearby municipality. Around one-fourth of the LAs have one to four RCVs, and a few of them (7.2%) have five or more RCVs. However, more than one-third of the LAs don't have SW workers, which could be attributed to the fact that these are small-size LAs and are contracting the service to another service provider, and they are not conducting street sweeping because these rural areas have low density of population. LAs that have more than 20 SW workers represent 12.0% part of the total LAs in the study area, while 52.9% of them have several SW workers in the range of 1–20. The fees collection showed that 9.9% of the LAs have collection less than or equal to 20%,

Table 2 Monthly cost of SW collection and disposal against a number of LAs in the study area

Monthly cost of SW collection (NIS)	%	Monthly cost of SW disposal (NIS)	%
Less or equal to 10,000.0	43.8	Less or equal to 5000.0	59.3
10,001–30,000.0	21.3	5001.0–10,000.0	12.3
30,001–80,000.0	17.5	10,001.0–20,000.0	12.3
80,001–150,000.0	8.8	20,001.0–30,000.0	6.2
150,001.0–300,000.0	3.8	More than 30,000.0	9.9
More than 300,000.0	5.0		
Total	100.0	Total	100.0

Table 3 Factors of potential effect on SWM at the LAs

Monthly waste generation (tons)	%	RCVs	%	SW workers	%	Fee collection (%)	%
Less or equal to 20	18.5	None	67.5	None	34.9	Less or equal to 20	9.9
21–50	21.0	One	8.4	1–5	32.5	21–40	18.5
51–100	6.2	Two	8.4	6–10	12.0	41–60	25.9
101–200	17.3	Three	7.2	11–20	8.4	61–80	14.8
201–400	16.0	Four	1.2	More than 20	12.0	81–100	30.9
401–600	6.2	Five or more	7.2				
More than 600	14.8						
Total	100	Total	100	Total	100	Total	100

30.9% of them have collection in the range of 81–100%, and the rest of the LAs (60%) in the study area have collection in the range of 21–80%. In general, the fee collection rate is considered relatively low which can in turn affect the performance of the LAs and attitudes toward developing an effective SWM system. Fees are the main source of revenues for the SWM sector (Abu-Qdais et al., 2023) and failure to secure a proper budget for SWM was reported as the root cause of failure of service (Breukelman et al., 2022).

3.3 Challenges in organic SW composting

Composting of the organic municipal SW fraction can attract LAs because it can reduce the transport and disposal costs, and can prolong the life of the landfills (Folz, 1991; Muttamara et al., 1994). However, LAs could face constraints and challenges, including technical, financial, social, and legal, which can affect their decision in selecting the appropriate option. Several challenges were selected in this study as presented in Fig. 3 and the LA opinion was assessed against these challenges. The results revealed that all the LAs (100%) lack proper equipment and machinery to conduct composting; 94.3% of the LAs reported a shortage of RCVs to collect SW; 90.6% of them suffer from lack of financial resources; 86.8% of them lack knowledge and experience in composting, and the SW tariff system

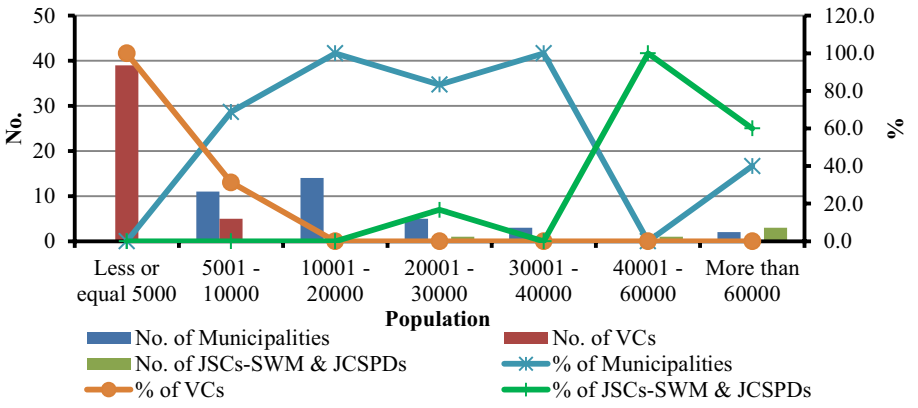


Fig. 3 LAs against the population served for waste management in the study area

was built on the cost of SW collection and disposal only, and any other associated costs with composting (cost for separate bins) were not considered; 75.5% of the LAs consider that SW sorting at the source is difficult to apply due to the lack of awareness; 73.6% of them consider a legal challenge as the compost is not forced by law; 71.7% provided that there is no incentives from the government for composting of organic waste; and 67.9% of them suffer from non-availability of a proper place to carry out composting. However, other challenges listed in Fig. 2 are of less importance in accordance with the LAs as less than 50.0% of the LAs highlighted them as challenges. Typical constraints and challenges being faced by developing countries that affect the development of effective SWM systems include a lack of financial, human, and infrastructure resources (Zhu et al., 2008). Marange et al. (2023) reported several challenges faced by the SWM council in Dangamvura/Zimbabwe which included poor engagement of the community, understaffing, political interference, and shortage of resources. Also, it has been reported that lack of infrastructure and technology, inadequate RCVs, unstandardized waste storage systems, and lack of government incentive policies are the main challenges faced by service providers in Cambodia (Pheakdey et al., 2022).

3.4 Factors affecting the attitude of LAs' towards composting of organic MSW

To evaluate the LAs' attitudes toward composting of municipal organic SW, many factors were selected. The independent variables that better explain the LAs' attitudes included in the model are monthly SW generated quantity by the LA, SW tariff on the household level, the rate of SW fee collection, LA's perception of compost contribution to SW reduction, availability of proper place (area of land) to be used for compost production, financial capacity of the LA to conduct composting, staff attended training on compost production, previous experience in compost production, awareness about compost production from organic SW, potential community participation in SW sorting, and LA's familiarity with the SWM bylaw. The LAs' attitude is estimated based on the LRM Eq. 5.

$$Logit\left(LA's\ attitude\right) = -\ 12.719 + 0.370X_1 + 0.151X_2 + 0.429X_3 + 2.044X_4 + 2.001X_5 + 2.235X_6 + 0.930X_7 + 1.704X_8 - 1.886X_9 + 1.494X_{10} - 2.697X_{11} \quad (5)$$

The results as presented in Table 4 showed that 5 out of 11 factors are significantly influencing the LAs' attitudes toward composting. The data fits the model as presented in Table 5. The significant outcomes drawn from the study showed that the attitude of LAs towards composting was negative for those who didn't perceive that composting of municipal organic SW fraction can contribute to reducing SW generation. LAs that lack a proper place to conduct composting showed a negative attitude toward organic SW composting. This finding is in agreement with Lalitha and Fernando (2019) who found that a limited area of land is a major issue for waste composting and recycling in Sri Lanka, and most of the LAs engaged in composting have failed due to lack of facilities. Al-Sari' and Haritash (2023) found that the availability of area of land to be used for compost production was a significant factor affecting the LAs' attitude toward organic waste composting. Renting an appropriate place for composting activities requires financial resources and can add more overhead on the LA's treasury, which makes it an unfavorable and non-attractive option, and results in a negative attitude towards composting.

The attitude towards organic waste composting is negative for LAs which don't have enough financial resources to cover the cost of organic waste composting. Lack of financial

Table 4 Output of the LRM of the LAs' attitude toward composting of organic municipal SW

Var	Abbreviation	Estimated coefficients (β)	Standard error (S.E)	Wald statistics	Degree of freedom (df)	Significance (*P*-value)
X_1	Mon. SW generated Qt	0.370	0.243	2.324	1	0.127
X_2	HH SW tariff	0.151	0.543	0.078	1	0.780
X_3	Fee collection rate	0.429	0.304	1.991	1	0.158
X_4	Per. of compost reduce SW	2.044	1.042	3.846	1	0.050*
X_5	Place availability	2.001	0.821	5.948	1	0.015*
X_6	Financial capacity	2.235	0.767	8.484	1	0.004**
X_7	Staff training	0.930	0.985	0.890	1	0.345
X_8	Staff prev. experience	1.704	1.096	2.416	1	0.120
X_9	Community awareness	− 1.886	0.925	4.161	1	0.041*
X_{10}	Pot. Comm. participation	1.494	0.807	3.429	1	0.064
X_{11}	Familiarity with SWM bylaw	− 2.697	1.110	5.898	1	0.015*
Constant		− 12.719	4.069	9.770	1	0.002

**significant at $P \leq 0.01$; *significant at $P \leq 0.05$

Table 5 LRM summary and goodness of fit

Test	Results		
Model summary	− 2 Log likelihood	Cox & Snell R^2	Nagelkerke R^2
	53.313	0.413	0.573
Omnibus tests of model coefficients	Chi-square	df	Sig
	38.345	11	0.000

resources is considered a challenge by 90.4% of the LAs as presented in Fig. 4. Financial resources are crucial for the LA's performance, and lack of financial resources can not only affect the attitude toward SW composting but also negatively affect the entire services. Choi (2021) found that financial soundness is significant for the LA's performance, and he reported that the higher the LA revenues the higher the performance level, and LAs with sufficient financial resources are more advantageous for superior performance. Zeidan et al. (2020) pointed out that budget deficit due to lack of financing is the primary obstacle that hinders the strategic plans implementation and local government units' improvements. Financial capacity was found significant factor affecting the attitude of LAs towards composting in the study area (Al-Sari' and Haritash, 2023) (Fig. 4).

In addition, the outcomes showed that LAs that have already conducted awareness in the community about compost production showed a negative attitude toward organic waste composting. LAs which conducted awareness in the community regarding composting of organic waste could have concluded that awareness without practice is not enough to change the community attitude or the awareness campaigns conducted were not enough. Conducting composting experiments for community observation in parallel with awareness could increase the citizens' understanding and knowledge about organic waste composting

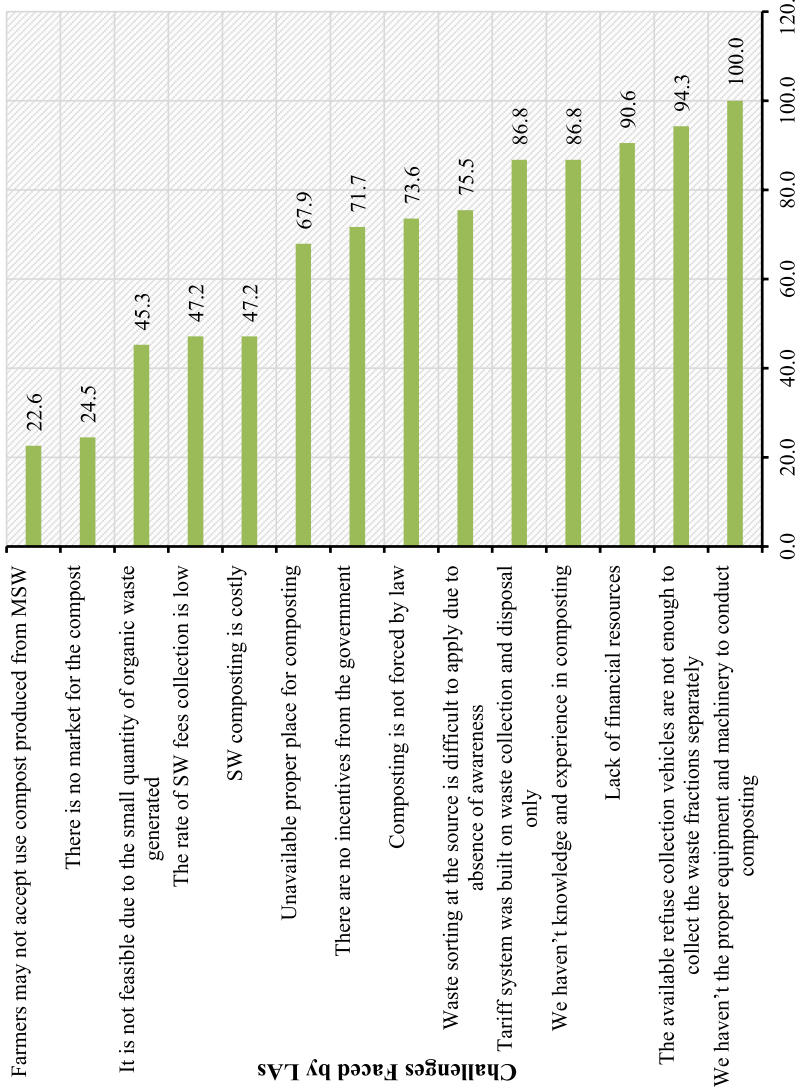


Fig. 4 Challenges faced by LAs in organic municipal SW management and composting

and knowledge, and enhance their participation. Awareness and knowledge of composting have been identified as crucial factors for community cooperation and participation in composting programs (Karkanias & Perkoulidis, 2016; Loan et al., 2019).

Moreover, the outcomes draw that LAs familiar with the SWM bylaw showed a negative attitude toward compost production from municipal organic SW. This could be attributed to the fact that composting of organic waste is not compulsory by law. SWM bylaw was issued in 2019, which encourages compost production from municipal organic SW fraction in article no. 3 and no. 27. Item 7 under article no. 3, encourages the reuse and recycling of the SW as much as possible through compost production and use in agriculture, and this use shall be in compliance with the environmental protection and health requirements. In addition, article no. 27 asks the authorities to take all necessary measures to reduce the waste stream as much as possible through reuse, recycling, and recovery (SWM bylaw, 2019). LAs who are familiar with the SWM bylaw are aware that composting of municipal organic waste fraction lies within their responsibilities, but the bylaw doesn't provide target limits as mandatory to reach, and accordingly, showed a negative attitude toward composting.

Other factors are of less importance because these had insignificant influence. The outcomes of the study showed that, first, the larger the waste quantity generated the lower the attitude toward composting. The reason could be that LA is suffering from a lack of resources such as the availability of enough area of land to conduct composting. Second, the higher the SW tariff group the lower the attitude toward SW composting. Some LAs are using the revenues to cover other services, which means that even if the tariff is high, the revenues are not used to develop the SWM system. However, some LAs have a high tariff system, but the rate of fee collection is low which means that the LA lacks the capacity to conduct composting. Other reasons could be attributed to the fact that the revenues, in some LAs, are not able to cover the cost of the service. The World Bank (2017) also reported that inter-governmental fiscal architecture in Palestine is characterized by local revenue assignments that are insufficient to cover the operational expenditure needs of local government units. Third, the higher the fee collection rate the lower the LA's attitude, which could be ascribed to the fact that LAs are using the revenues to cover other services, or the tariff, is not being able to cover the expenditures. Fourth, LAs with staff members with attained training in compost production showed positive attitudes toward organic waste composting. Training increases knowledge and plays an important role in waste processing, composting process performance, and the quality of the end product. Al-Sari' and Haritash (2023) found that staff training in compost production was a significant factor affecting the LA's attitude towards composting. Fifth, the LAs with staff members having previous experience in compost production showed positive attitudes toward composting. Previous experience enables the staff member to replicate the job successfully, thus reducing the time and effort and accelerating the processing of the waste. Staff's previous experience in compost production was found significant factor influencing the LA's attitude towards composting (Al-Sari' & Haritash, 2023). Finally, the higher the LA's expectation of community participation the higher the positive attitude toward MSW composting. The participation of the community in waste sorting at the source is of high importance because it activates the participatory bottom-up process (De Bonni et al., 2022), facilitates the composting process, and contributes to improving the quality of the compost. This in turn increases the acceptance of compost use and enhances marketing. Previous studies reported that a lack of public participation has the potential to effect attitudes towards composting (Milea, 2009).

The results of this study are compared to the results of the previous publication conducted by Al-Sari' and Haritash (2023). The previous research identified nine significant

factors while this study identified only five significant factors affecting the LAs' attitude towards composting. The comparison is presented in Table 6. The comparison showed two factors were identified were found significant by both methods of analysis, bivariate analysis and LRM. These are financial capacity and availability of place (i.e. piece of land) to be used for composting.

4 Conclusion

LAs in the study area are facing several challenges in SWM and composting. The effect of these challenges is in the range of 22.6–100% of the LAs. It is concluded that five significant factors negatively affect the attitude of LAs toward composting. These include:

1. Perception of compost contribution to SW reduction;
2. Availability of proper place to be used for composting activities,
3. Financial capacity,
4. Community awareness, and
5. Familiarity with prevalent SWM bylaws.

Identifying and overcoming the factors that affect the intention and attitude of LAs towards organic MSW composting is a cornerstone in setting a comprehensive composting policy, and developing and initiation of proper composting programs. This study suggested that a set of mitigation measures should be taken before setting and forcing composting policies to overcome the obstacles that are facing the LAs to shift their attitude towards composting from one negativity to one positivity. These measures include:

- Awareness and training on compost to improve the perception of compost contribution to waste reduction. Competent and regulatory authorities shall conduct intensive training and awareness to increase understanding and perception of the LAs about the benefits of organic waste composting especially that related to the reduction in waste generation, uses of compost, and cost-related impacts on waste management;
- Secure adequate financial resources through external funding, government support, and incentives, which can enable the LAs to cover part of the operation cost and offer a suitable place for composting activities. Also encouraging public–private partnership investment is one of the mechanisms to increase efficiency and optimization;

Table 6 Comparison of findings between Al-Sari' and Haritash (2023) and this study

Factors identified significant by Al-Sari' and Haritash (2023)	Factors identified significant by this study
Financial capacity	Financial capacity
Availability of area of land to be used for composting	Place availability
Familiarity with composting systems	Familiarity with SWM bylaw
Staff's previous experience in compost production	Community awareness
Availability of enough RCV to collect SW fractions separately	Perception of compost can reduce SW
Acceptance of the rapid composting system	
Staff training in compost	
Believe that SW composting is within the LA's responsibility	
Availability of proper machinery	

- Supporting the LAs technically and financially to conduct community awareness to ensure the cooperation of the citizens. Technical reinforcement of the LAs in waste composting is essential to enable LAs to conduct community awareness campaigns for compost promotion among the citizens who can play an important role in waste sorting at the source to facilitate organic waste composting;
- The regulatory authority shall upgrade and promote the SWM bylaw. The bylaw shall include obligatory limits and targets to achieve. These targets shall be in line with the NDCs for emission reduction, which contributes to achieving sustainable development goals (SDGs).

Taking these measures into account will assist in overcoming the influencing factors and can, to a large extent, contribute to the successful implementation of composting programs.

4.1 Applicability of the findings

Given the similarities of the LAs in the West Bank, especially for the current practices in SWM, the findings of this study could apply to the entire LAs in the West Bank and other similar LAs in the region. Policy-makers in Palestine shall consider the outcome of this study to set efficient composting policies to achieve sustainability in SWM.

4.2 Scientific value added of the study

The scientific value-added and broader importance of this research is that the LRM can be used for other SWM options as a tool to facilitate setting the related policies. It can provide the decision-makers and policy-makers with a tool to study the community situations and identify the constraints that can affect the implementation of a policy in advance, and bridge the gap in this concern.

4.3 Novelty of the research

This is the first study in Palestine that focuses on the appropriate way to set an effective composting policy through the assessment of the attitude of the key stakeholders (LAs) in SWM towards composting of organic waste. The study links policy and practice through building a mathematical model to identify the factors that affect their attitude which leads to effective drafting of the policy and better implementation.

4.4 Limitations of the study and future recommendation

This study was conducted on the LAs' located in the southern part of the West Bank/Palestine and focused on composting only. The other stakeholders of environmental protection, SWM regulatory authorities and policy-makers, and private sector companies that are working in the SWM sector, are not included in this study. Moreover, other SW recycling activities were not included. Research institutions are advised to focus future studies on other regions as well, and include other stakeholders including SWM sector regulators, policy-makers, environmental protection agencies, and the private sector. In addition, future studies can include other SW recycling fractions.

Acknowledgements The authors thank the help of different LAs in Palestine for their responses to the questionnaire survey of the study.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability The data used to support the findings of this study are available upon request from the corresponding author.

Declarations

Conflict of interest The authors declare no competing interests.

References

Abu-Qdais, H. A., Shatnawi, N., & Al-Shahrabi, R. (2023). Modeling the impact of fees and circular economy options on the financial sustainability of the solid waste management system in Jordan. *Resources, 12*, 32. https://doi.org/10.3390/resources12030032

ADBI (Asian Development Bank Institute) (2020). Policy brief, solid waste management in developing Asia: Prioritizing waste separation. Kasumigaseki Building 8F, 3-2-5 Kasumigaseki, Chiyoda-ku, Tokyo 100–6008, Japan.

Ali, H., Ali, N., Ahmad, A. R., Ibrahim, M., Ahmad, S., & Yaacob, S. (2012). Solid Waste management and the willingness to pay for improved services towards achieving sustainable living. *Advances in Natural and Applied Sciences, 6*(1), 52–60.

Al-KhateebAl-Sari', A. J. M. I., Al-Khatib, I. A., & Anayah, F. (2017). Factors affecting the sustainability of solid waste management system—the case of Palestine. *Environmental Monitoring and Assessment, 189*, 93. https://doi.org/10.1007/s10661-017-5810-0

Al-Madbouh, S., Al-Khatib, I. A., Salahat, J. I., Jararaa, B. Y. A., & Ribbe, L. (2019). Socioeconomic, agricultural, and individual factors influencing farmers' perceptions and willingness of compost production and use: An evidence from Wadi al-Far'a Watershed-Palestine. *Environmental Monitoring and Assessment, 191*, 209.

Al-Sari, M. I., & Haritash, A. K. (2023). Managing the organic municipal waste in Palestine: Linking policy, practice and stakeholders' attitude towards composting. *Journal of the Air & Waste Management Association, 73*(1), 80–93.

Al-Sari', M. I., Al-Khatib, I., Avraamides, M., & Fatta-Kassinos, D. (2012). A study on the attitudes and behavioural influence of construction waste management in occupied Palestinian territory. *Waste Management and Research, 30*(2), 122–136.

Amano, K. O. A., Danso-Boateng, E., Adom, E., Nkansah, D. K., Amoamah, E. S., & Appiah-Danquah, E. (2021). Effect of waste landfill site on surface and ground water drinking quality. *Water and Environment Journal, 35*, 715–729.

Begum, R. A., Siwar, C., Pereira, J. J., & Jaafar, A. (2006). A Logistic regression analysis of the contractor's awareness regarding waste management. *Journal of Applied Sciences, 6*(9), 1904–1908.

Begum, R. A., Siwar, C., Pereira, J. J., & Jaafar, A. (2009). Attitudes and behavioral factors in waste management in the construction industry of Malaysia. *Resources, Conservation and Recycling, 53*, 321–328.

Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. *Critical Care, 9*(1), 112–118.

Breukelman, H., Krikke, H., & Lohr, A. (2022). Root causes of underperforming urban waste services in developing countries: Designing a diagnostic tool, based on literature review and qualitative system dynamics. *Waste Management & Research, 40*, 1337–1355.

Budijati, S. M., Pujawan, I. N., & Asih, H. M. (2023). Analysis of barrier factors for collaboration in handling used cell phones for second-hand market actors to implement e-waste management. *Civil Engineering Journal, 9*(3), 654–675. https://doi.org/10.28991/CEJ-2023-09-03-011

CESVI (2019). Solid waste management in the Palestinian territory, West Bank including East Jerusalem & Gaza Strip. Overview Report.

Cetrulo, T. B., Marques, R. C., Cetrulo, N. M., Pinto, F. S., Moreira, R. M., Mendizábal-Cortés, A. D., & Malheiros, T. F. (2018). Effectiveness of solid waste policies in developing countries: A case study in Brazil. *Journal of Cleaner Production, 205*, 179–187.

Choi, N. H. (2021). Analyzing local government capacity and performance: Implications for sustainable development. *Sustainability, 13*, 3862. https://doi.org/10.3390/su13073862

De Boni, A., Melucci, F. M., Acciani, C., & Roma, R. (2022). Community composting: A multidisciplinary evaluation of an inclusive, participative, and eco-friendly approach to biowaste management. *Cleaner Environmental Systems, 6*, 100092. https://doi.org/10.1016/j.cesys.2022.100092

Di Maria. F., Lovat, E. & Caniato, M. (2017), Comparing waste management in developed and developing countries: The case study of the umbria region (Italy) and of west bank (Palestine). In *Proceedings Sardinia 2017/sixteenth international waste management and landfill symposium/2-6 October 2017*. S. Margherita di Pula, Cagliari, Italy/CISA Publisher, Italy.

Fadlil, A., Umar, R., & Sunardi, N. A. S. (2022). Comparison of machine learning approach for waste bottle classification. *Emerging Science Journal, 6*(5), 1075–1085. https://doi.org/10.28991/ESJ-2022-06-05-011

Fan, B., Wu, T., Zhuang, Y., Peng, J., & Huang, K. (2021). The development of energy storage in China: Policy evolution and public attitude. *Front. Energy Res., 9*, 838.

Folz, D. H. (1991). Recycling program design, management, and participation: A national survey of municipal experience. *Public Administration Review, 51*(3), 222–231.

Ghodrati, M., Mousavi-Kamazani, M., & Zinatloo-Ajabshir, S. (2019). Zn3V3O8 nanostructures: Facile hydrothermal/solvothermal synthesis, characterization, and electrochemical hydrogen storage. *Ceramics International, 46*, 28894–28902. https://doi.org/10.1016/j.ceramint.2020.08.057

Hettiarachchi, H., Meegoda, J. N., & Ryu, S. (2018). Organic waste buyback as a viable method to enhance sustainable municipal solid waste management in developing countries. *International Journal of Environmental Research and Public Health, 15*, 2483.

Hosseinzadeh, G., Sajjadi, S. M., Mostafa, L., Yousefi, A., Vafaie, R. H., & Zinatloo-Ajabshir, S. (2023). Synthesis of novel direct Z-scheme heterojunction photocatalyst from WO_3 nanoplates and $SrTiO_3$ nanoparticles with abundant oxygen vacancies. *Surfaces and Interfaces, 42*, 103349. https://doi.org/10.1016/j.surfin.2023.103349

IFC (International Finance Corporation) (2012). Solid waste management in Hebron and Bethlehem Governorates, assessment of current situation and analysis of new system. Report, Palestine.

Ittiravivongs A (2012) Household waste recycling behavior in Thailand: The role of responsibility. In *International Conference on Future Environment and Energy* IPCBEE 28.

Jalalipour, H., Haghighi, A. B., Ferronato, N., Bottausci, S., Bonoli, A., & Nelles, M. (2024). Social, economic and environmental benefits of organic waste home composting in Iran. *Waste Management & Research*. https://doi.org/10.1177/0734242X241227377

JSC-H&B (The Joint Service Council for Solid Waste Management for Hebron and Bethlehem Governorates) (2022). Weigh Bridge Records. Bethlehem, Palestine.

Kadhila, T., de Wit, M. P., & Schenck, R. (2023). A conceptual framework for sustainable waste management in small municipalities: The cases of Langebaan, South Africa and Swakopmund, Namibia. *Environmental Science and Pollution Research*. https://doi.org/10.1007/s11356-023-26904-7

Karkanias, C., Perkoulidis, G., & Moussiopoulos, N. (2016). Sustainable management of household biodegradable waste: Lessons from home composting programs. *Waste Biomass Valorization, 7*, 659–665.

Lalitha, R., & Fernando, S. (2019). Solid waste management of local governments in the Western Province of Sri Lanka: An implementation analysis. *Waste Management, 84*, 194–203.

Loan, L. T. T., Takahashi, Y., Nomura, H., & Yabe, M. (2019). Modeling home composting behavior toward sustainable municipal organic waste management at the source in developing countries. *Resources, Conservation and Recycling, 140*, 65–71.

Marange, F., Muteveri, M., Chipfunde, F., & Mapira, J. (2023). Challenges confronting local authorities in solid waste management: the case of Dangamvura residential area, Mutare, Zimbabwe. *European Journal of Social Sciences Studies*. https://doi.org/10.46827/ejsss.v8i5.1477

Milea, A. (2009). Waste as a social dilemma: Issues of social and environmental justice and the role of residents in municipal solid waste management, Delhi, India. Master's thesis, Lund University. Lund, Sweden

MoLG (Ministry of Local Government) (2005). Investigation study on the local authorities' classifications. Arabic document. Palestine.

Muttamara, S., Visvanathan, C., & Alwis, K. U. (1994). Solid waste recycling and reuse in Bangkok. *Waste Management & Research, 12*, 151–163.

Nistratov, A. V., Klimenko, N. N., Pustynnikov, I. V., & Vu, L. K. (2022). Thermal regeneration and reuse of carbon and glass fibers from waste composites. *Emerging Science Journal, 6*(5), 967–984. https://doi.org/10.28991/ESJ-2022-06-05-04

Pheakdey, D. V., Quan, N. V., Khanh, T. D., & Xuan, T. D. (2022). Challenges and priorities of municipal solid-waste management in Cambodia. *International Journal of Environmental Research and Public Health, 19*, 8458. https://doi.org/10.3390/ijerph19148458

Rana, R., Ganguly, R., & Gupta, A. R. (2018). Indexing method for assessment of pollution potential of leachate from non-engineered landfill sites and its effect on ground water quality. *Environmental Monitoring and Assessment, 190*(1), 46. https://doi.org/10.1007/s10661-017-6417-1

Sharma, A., Ganguly, R., & Gupta, A. K. (2019). Characterization and energy generation potential of municipal solid waste from non-engineered landfill sites in Himachal Pradesh, India. *Journal of Hazardous, Toxic, and Radioactive Waste, 23*(4), 04019008. https://doi.org/10.1061/(ASCE)HZ.2153-5515.0000442

SPCP-GCF (State of Palestine's Country Program for the Green Climate Fund) (2019). Climate Resilient Transformation with the Green Climate Fund. Palestine.

State of Palestine (2021). Palestine's First Nationally Determined Contributions – Updated Submission. Palestine.

SWM by Law (Solid Waste Management Bylaw) (2019). Palestinian National Authority (PNA). Ramallah, Palestine.

The World Bank (2017).The performance of the Palestinian local governments: an assessment of service delivery outcomes and performance drivers in the West Bank and Gaza. Report No: ACS22456.

Trinh, L. T., & Giao, N. T. (2023). Evaluating willingness to pay for e-waste recycling in Vietnam. *Civil Engineering Journal, 9*(10), 2399–2410. https://doi.org/10.28991/CEJ-2023-09-10-03

Vinti, G., Bauza, V., Clasen, T., Tudor, T., Zurbrügg, C., & Vaccari, M. (2023). Health risks of solid waste management practices in rural Ghana: A semi-quantitative approach toward a solid waste safety plan. *Environmental Research, 216*, 114728. https://doi.org/10.1016/j.envres.2022.114728

Wiedinmyer, C., Yokelson, R. J., & Gullett, B. K. (2014). Global emissions of trace gases, particulate matter, and hazardous air pollutants from open burning of domestic waste. *Environmental Science and Technology, 48*, 9523–9530.

Yang, Z., Huang, C., Liao, F., Lam, F. I., Li, L., Qin, Q., & Fan, B. (2023). Exploring the characteristics of solid waste management policy in the Guangdong-Hong Kong-Macao greater Bay Area. *Sustainability, 15*, 8160. https://doi.org/10.3390/su15108160

Zamzami, F., & Rakhman, F. (2023). Determinants of local government financial performance in Indonesia. *Academic Journal of Interdisciplinary Studies*. https://doi.org/10.36941/ajis-2023-0148

Zeidan, K., Daas, G., & Awwad, Y. (2020). Municipal bonds as a tool for financing capital investment in local government units in Palestine. *Investment Management and Financial Innovations, 17*(1), 213–226.

Zhu, D., Asnani, P. U., Zurbrugg, C., Anapolsky, S., & Mani, S. K. (2008). *Improving municipal solid waste management in India: A sourcebook for policy makers and practitioners*. The World Bank.

Zinatloo-Ajabshir, S., Esfahani, M. H., Marjerrison, C. A., Greedan, J., & Behzad, M. (2023). Enhanced electrochemical hydrogen storage performance of lanthanum zirconium oxide ceramic microstructures synthesized by a simple approach. *Ceramics International, 49*, 37415–37422. https://doi.org/10.1016/j.ceramint.2023.09.067

Zinatloo-Ajabshir, S., Morassaeib, M. S., & Salavati-Niasarib, M. (2019). Eco-friendly synthesis of $Nd_2Sn_2O_7$–based nanostructure materials using grape juice as green fuel as photocatalyst for the degradation of erythrosine. *Composites, Part B: Engineering, 167*, 643–653. https://doi.org/10.1016/j.compositesb.2019.03.045

Zinatloo-Ajabshir, S., Rakhshani, S., Mehrabadi, Z., Farsadrooh, M., Feizi-Dehnayebi, M., Rakhshani, S., Duˇsek, M., Eigner, V., Rtimi, S., & Aminabhavi, T. M. (2024). Novel rod-like $[Cu(phen)_2(OAc)]\cdot PF_6$ complex for high-performance visible-light-driven photocatalytic degradation of hazardous organic dyes: DFT approach, Hirshfeld and fingerprint plot analysis. *Journal of Environmental Management, 350*, 119545. https://doi.org/10.1016/j.jenvman.2023.119545

Zinatloo-Ajabshir, S., & Salavati-Niasari, M. (2019). Preparation of magnetically retrievable $CoFe_2O_4@SiO_2@Dy_2Ce_2O_7$ nanocomposites as novel photocatalyst for highly efficient degradation of organic contaminants. *Composites, Part B: Engineering, 174*, 106930. https://doi.org/10.1016/j.compositesb.2019.106930

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.

A Passive Floating C-MOS Memristor with Analog Applications

Rahul Kumar Gupta, Research Scholar
Electronics and Communication Engineering
Delhi Technological University,
Delhi, India
Assistant Professor,
JSS Academy of Technical Education, Noida
rahulkumargupta_2k21phdec08@dtu.ac.in

Mahipal Singh Choudhry, Professor
Electronics and Communication Engineering
Delhi Technological University,
Delhi, India
msc_1976@yahoo.com

Varun Saxena, Assistant Professor
School of Engineering
Jawaharlal Nehru University
Delhi, India
varunsaxena@jnu.ac.in

2024 2nd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT) | 979-8-3503-7284-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/DICCT61038.2024.10533143

***Abstract*— A floating fully passive memristor emulator, characterized by six MOSFETs and a capacitor, has been introduced, employing the dynamic threshold MOSFET technique (DTMOS). This proposed memristor emulator circuit underwent simulation using GPDK 180nm CMOS technology in conjunction with Cadence Virtuoso Spectre software. The analysis encompassed transient behavior, hysteresis characteristics, and its nonvolatile properties. Furthermore, the manuscript includes a thorough ex- amination of process corner and temperature variations. Notably, the circuit demonstrated an impressive operating frequency of up to 10MHz. Key advantages of this proposed memristor emulator comprise its streamlined design, minimal power consumption, ability to operate at high frequencies in the MHz range, and dynamic threshold features. The circuit's layout area measures 56.725*57.15µm2. Finally, the study extends its applicability to prospective analog applications, notably in oscillators such as Colpitts chaotic and Chua's oscillators, as well as in a Schmitt trigger circuit, utilizing the proposed memristor emulator.**

Keywords— DTMOS, Memristor, hysteresis loop, nonvolatile, Schmitt trigger, chaotic, colpitts

I. INTRODUCTION

In 1971, Chua introduced a groundbreaking proposal that introduced the memristor as the fourth passive component alongside the conventional trio of resistor, capacitor, and inductor [1]. Chua's work played a pivotal role in establishing the conceptual foundation for memristors [2,3]. In 2008, Hewlett Packard (HP) marked a significant milestone by discovering the world's first nano-scale memristor [4]. Following this breakthrough, in 2010, HP achieved a major feat by successfully creating two stateful logic circuits - a NAND circuit and a material implication circuit - employing memristive switches, a category of non-volatile memristors [5]. Furthermore, volatile memristors can be used to simulate synapses for brain-inspired computing [6]. This is due to the importance of forgetting in the brain, as it is essential for adaptive computing. Without the ability to forget, every piece of information received can quickly overwhelm biological memory. As a result, emulators of volatile memristors, which have a small footprint and operate at high frequencies, have a lot of potential in brain-inspired computing..

The emulation of memristors can be broadly classified into two primary categories: analog memristor emulators [7, 8, 10–15] and digital memristor emulators [16–18]. Analog emulators make use of active components such as voltage differential transconductance amplifiers (VDTA), current conveyor transconductance amplifiers (CCTA), volt- age difference current conveyors (VDCC), differential differ- ence current conveyors (DDCC), differential voltage current conveyor transconductance amplifiers (DVCCTA), operational amplifiers (OPAMP) [10], current backward transconduc- tance amplifiers (CBTA), second-generation current conveyors (CCII), operational transconductance amplifiers (OTA), current feedback operational amplifiers (CFOA), analog multipliers, and passive elements for their implementation. However, it's essential to consider that each passive element consumes a significant amount of real estate on an integrated circuit (IC). Conversely, digital memristor emulators, which rely on field-programmable gate arrays (FPGAs) or application- specific integrated circuits (ASICs), provide several benefits, including hardware flexibility, re-configurability, and straight- forward programming and control [17]. However, it's impor- tant to note that one drawback of digital memristor emulators is their operational frequency, which is often restricted by their clock frequency.

Circuits composed solely of MOS transistors, both with and without the inclusion of external passive components [19–32], rely exclusively on the inherent capacitance of the transistors. These circuits have not used any external passive elements such as resistors, capacitors, and inductors. This approach emphasizes incorporating the most significant intrin- sic capacitances during the circuit design, resulting in high- frequency operational characteristics [26–28, 32]. These circuits also boast a compact chip footprint since intrinsic capacitors offer a higher capacitance density compared to intentionally integrated on-chip capacitors. Several researchers have introduced theoretical frameworks [33, 36], and the validation of memristor theory has been accomplished through Verilog or PSPICE coding [34, 35].

In this context, an innovative floating memristor emulator is introduced, composed of just six MOSFETs N_{1-3} and P_{1-3} and one external capacitor. Specifically, N_1 to N_3 are n- channel MOSFETs, while P_1 to P_3 are p-channel MOSFETs. Various analyses, including transient analysis, hysteresis curve examination, assessment of nonvolatile characteristics, and process corner analysis, have been conducted and presented for this memristor. Notably, this proposed memristor emulator offers several advantages, such as a simpler structure, lower power

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 05:53:41 UTC from IEEE Xplore. Restrictions apply.

consumption, high-frequency operation (reaching up to 10MHz), and dynamic threshold capability. To conclude, the potential applications for this innovative memristor include analog filters, oscillators (both Colpitt and chaotic), and Schmitt triggers.

The structure of this paper can be outlined as follows: Section II is dedicated to presenting the proposed circuit and elucidating its operational principles. Subsequently, Section III offers an in-depth examination of the simulation results. The prospective applications of the emulator circuit under consideration are delineated in Section IV. Ultimately, in Section V, we summarize and provide a comprehensive assessment of the advantages and limitations of the proposed memristor emulators

II. PROPOSED MEMRISTOR EMULATOR

This study introduces a circuit, as illustrated in Fig. 1, comprising six MOSFETs and a single capacitor (C=5pF). Among these components, two MOS transistors, P_1 and N_1, are arranged as diodes, while another pair, P_2 and N_2, is configured as a resistor. Additionally, P_3 and N_3 serve as an inverter circuit, with the output of this inverter circuit providing the biasing voltage to the MOS transistor P_2. It's worth noting that this proposed circuit operates entirely in a passive mode without the need for external bias voltage.

Both transistor N_1 and transistor P_1 in the proposed circuit effectively serve as diodes. When the magnitude of the input signal rises in the positive direction, the NMOS transistor (N_1) goes into forward bias mode, while the pMOS transistor (P_1) assumes a reverse bias configuration. These transistors remain inactive until the input signal magnitude falls below 0.7V. Once the input signal surpasses the 0.7V threshold, transistor N_1 starts conducting, leading to a gradual charging of the capacitor C. Current flows through transistors N_1, N_2, N_3, and the resistor R, causing a rapid increase in the total current passing through resistor R.

Since the transistor N_1 and the transistor P_1 function as a diode in the proposed circuit. When the input signal magnitude is increasing in positive, the transistor NMOS (N_1) is in forward bias and another transistor pMOS (P_1) is reversed. Transistors will not operate until the input signal magnitude less than 0.7V. When the input signal magnitude crosses the 0.7V the transistor N_1 start conducting, and the capacitor C charges slowly. The current flows through the transistors N_{1-3}, and the resistor R. The total current across the resistor R increases tremendously.

When the input signal reaches its maximum positive voltage during the positive half cycle, transistor N_1 enters saturation. At this point, as the input signal's amplitude decreases from its maximum positive level, capacitor C begins to discharge gradually, primarily through the source of transistor N_3. As the input signal's amplitude diminishes to 0.7V, transistor N_1 ceases to active, and capacitor C continues to discharge rapidly. This discharging of the capacitor closely follows the characteristics of the input signal due to the high-resistance state. During the negative half of the input sinusoidal signal's amplitude, transistor P_1 becomes active and conducts. Consequently, capacitor C undergoes a reverse charge cycle, with a negative amplitude, facilitated by the path through P_1, N_3, and resistor R. As the negative half of the input sinusoidal signal experiences a reduction in amplitude, the capacitor commences another discharge phase. This cyclic process of charging and discharging the capacitor occurs in synchronization with every cycle of the input signal, playing a crucial role in shaping the V-I characteristics of this memristor circuit. The composition of N_1 and P_1, called diode connected load because of gate and drain of the transistors are shorted. These transistor behave as a small signal resistor and maintained the saturation region of the transistors N_1 and P_1. The proposed circuit has nonlinear behavior and presents a constricted hysteresis loop with the input signal. The W/L ratio of used MOSFETs in the composition of proposed emulator circuit in the Table I. The present emulator circuit layout has also been depicted in Fig. 2 which engages chip area of $56.725*7.15\mu m^2$.

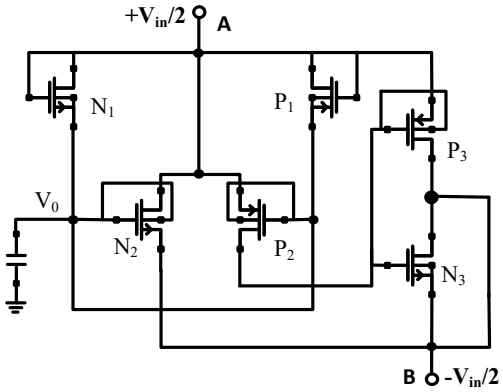


Fig. 1. Circuit diagram of proposed memristor emulator

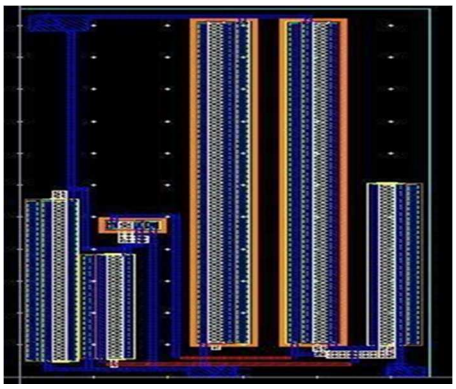


Fig. 2. Layout of proposed memristor emulator

TABLE I
W/L RATIO OF USED MOSFETS

MOSFETs	Type	W/L Ratio
N_1, N_3	NMOS	$25\mu m/1\mu m$
N_2	NMOS	$16\mu m/1\mu m$
P_1, P_3	NMOS	$50\mu m/1\mu m$
P_2, P_3	NMOS	$1\mu m/1\mu m$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 05:53:41 UTC from IEEE Xplore. Restrictions apply.

III. Simulation Results

The memristor emulator circuit performance has been simulated using Cadence Spectre Virtuoso 180nm GPDK CMOS technology. The transient response, illustrated in Fig. 3, displays a output current with respect to the sinusoidal input signal with an amplitude of $1.5V_{pp}$. The response clearly exhibits zero crossings at $V_{in} = 0$ and Iout = 0. Fig. 4 presents the hysteresis response of the proposed memristor emulator at various frequencies (10KHz, 100KHz, 500KHz, 1MHz, and 10MHz), confirming its adherence to memristor characteristics as reported in reference [33]. Fig. 5 demonstrates the scalability behavior, with hysteresis loops observed at different capacitance values (C = 1pF , C = 2pF , and C = 5pF).the scalability means the current depends on capacitor value. The second fingerprint of memristor behavior as defined in reference [33] is portrayed in Fig. 6, highlighting the similarity in the product of frequency and capacitance for memristor characteristics. The significance of the similarity means the capacitance dependence of frequency. The non- volatile characteristics of the memristor emulator circuit is depicted in Fig. 7. Fig. 8 and Fig.9 provide a thorough analysis of process and temperature variations, further establishing the reliability of the memristor emulator.

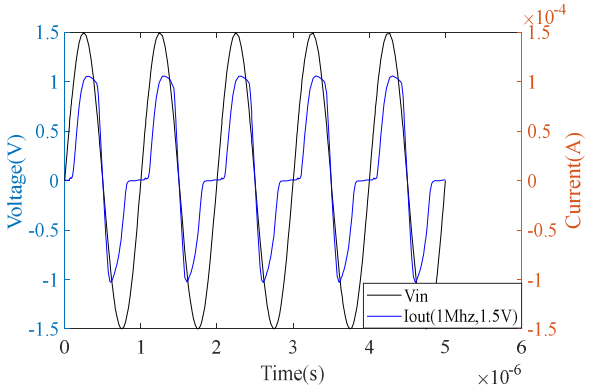


Fig. 3. Response of output current w.r.t. the input signal

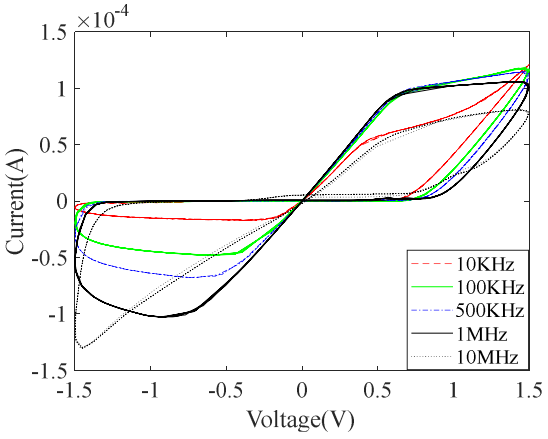


Fig. 4. Hysteresis response for various frequencies

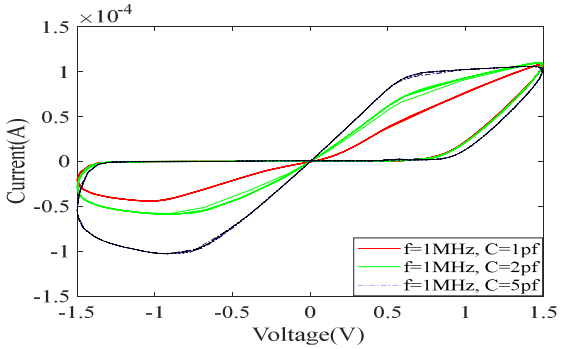


Fig. 5. Hysteresis response for various capacitances to show scalability

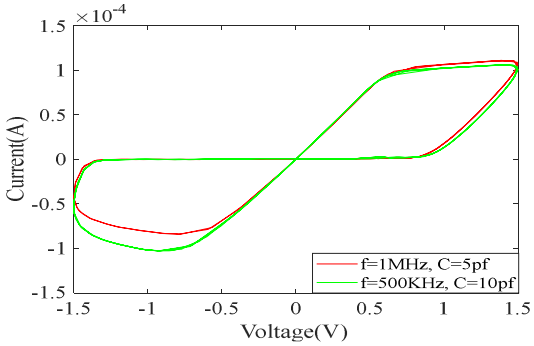


Fig. 6. Hysteresis response for various frequency and capacitances to observe (f * C = Constant) performance

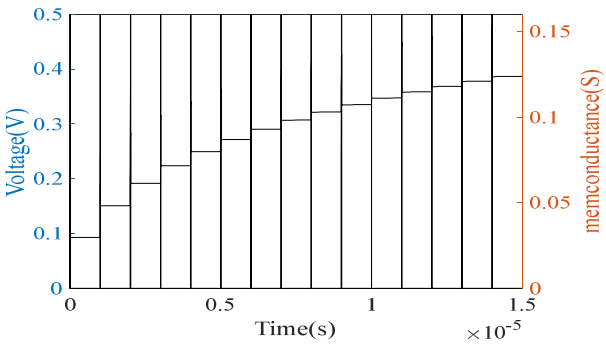


Fig. 7. Non-Volatile characteristics

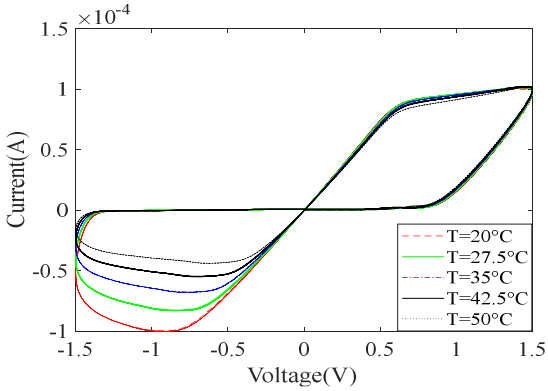


Fig. 8. Hysteresis response for different temperature

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 05:53:41 UTC from IEEE Xplore. Restrictions apply.

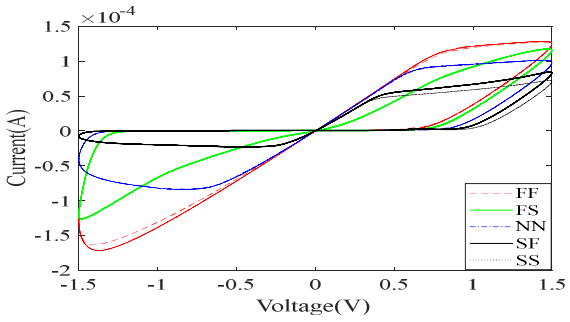


Fig. 9. Process corner analysis of proposed memristor emulator circuit

IV. Application of Memristor Emulator

In this study, a Schmitt trigger, a Colpitts chaotic oscillator, and a Chua's oscillator circuit were introduced. Within these circuits, a resistor was substituted with the proposed memristor circuit, and the circuit's performance was carefully observed. Following the analysis, it was determined that the proposed memristor could effectively replace a resistor in these applications. The integration of this recommended memristor emulator was applied to evaluate a Schmitt trigger circuit featuring an operational amplifier, as illustrated in Fig. 10. This circuit employs positive feedback to achieve regenerative action and is commonly referred to as a regenerative comparator circuit.

The operational amplifier (Op-Amp) is biased with a voltage of ±15Vdc. A sinusoidal input signal at a frequency of 250Hz and an amplitude of 1VPP is employed, and a resistance value of R1 = 20MΩ is selected. To investigate the effect of programming pulses, a series of 1V pulses with a frequency of 1 KHz is applied to observe changes in the threshold voltage. The Schmitt trigger transient analysis can be seen in Fig. 11, and the hysteresis response is depicted in Fig. 12. Schmitt triggers are commonly used in scenarios where an Op-Amp is employed as a simple comparator.

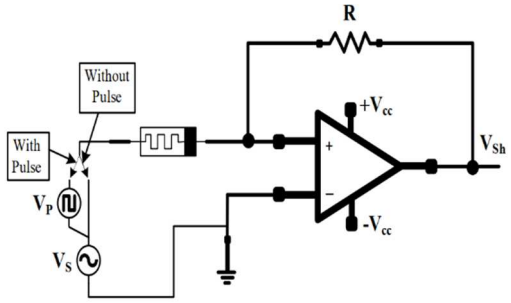


Fig. 10. Non-inverting Schmitt trigger circuit using memristor

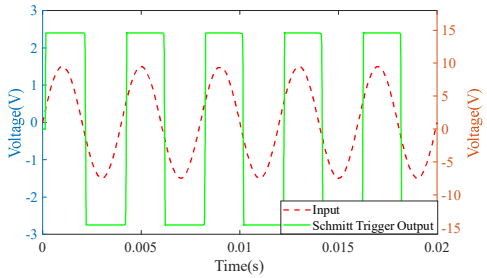


Fig. 11. Transient repsonse of Schmitt Trigger

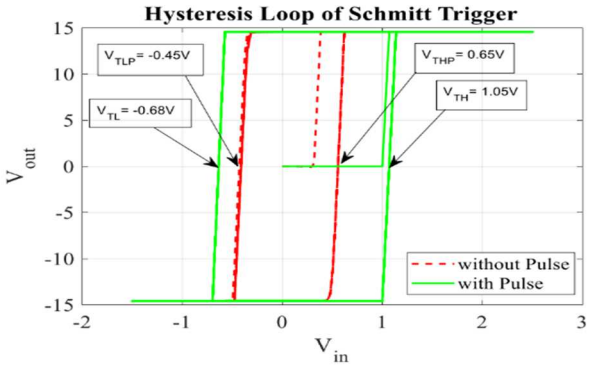


Fig. 12. Hysteresis loop for Schmitt trigger

In the context of this study, an analysis was conducted on a Colpitts oscillator circuit incorporating a memristor [37], wherein a Bipolar Junction Transistor (BJT), serving as a non-linear element, was included. The Colpitts oscillator circuit, depicted in Fig. 13, is composed of capacitors with the values of C_0 = 2µF, C_1 = 2µF, C_2 = 2µF, and C_3 = 33µF, inductor with the value of L = 10mH and Resistors with the values of R_0 = 200Ω, R_1 = 35Ω, and R2 = 1.5kΩ.

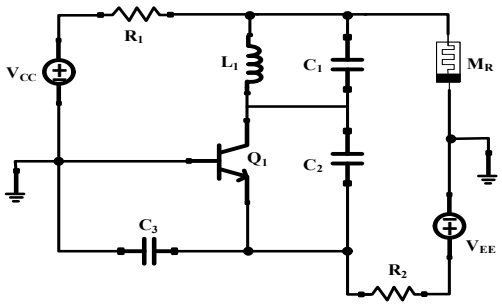


Fig. 13. Colpitts Chaotic Oscillator Circuit using memristor

The power supply is maintained at V_{CC} = 5V and V_{EE} = 5V . A spiral chaotic attractor was plotted in Figs. 14 to 15, initiated with initial values for the five state variables: $V_1(t = 0) = 0.01V$, $V_2(t = 0) = 0.01V$, $V_3(t = 0) = 0V$, $i_L(t = 0) = 0A$, and $V_0(t = 0) = 0V$. Additionally, the Chua's oscillator circuit, characterized by inductors L_0 = 70mH, L_1 = 7mH, capacitors C_0 = 1nF, C_1 = 20nF , and a resistor R_0 = 1M Ω, is illustrated in Fig. 16. The responses of Chua's oscillator have been comprehensively presented in Fig. 17 and Fig. 18.

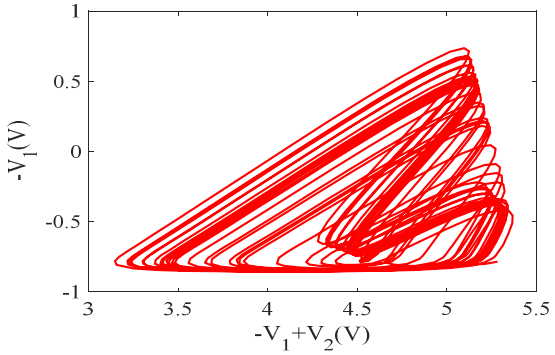


Fig. 14. Colpitts chaotic oscillator circuit performance for $(-V_1 + V_2)$ Vs $-V_1$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 05:53:41 UTC from IEEE Xplore. Restrictions apply.

TABLE II
COMPARISON OF EXISTING MEMRISTORS WITH PROPOSED MEMRISTOR TAKING DIFFERENT PARAMETERS

Ref	Active Devices	Transistors Count	Passive Components Count	Memristor Type	Technology	Operating Frequency	Power Dissipation
[9]	DDCC-1	40	$R \rightarrow 2, C \rightarrow 1$	Grounded	35*nm T SMC*	1*MHz*	-NA-
[29]	-	3	$C \rightarrow 1$	Floating	180*nm T SMC*	13*MHz*	6.725*nW*
[28]	-	3	–	Both	180*nm GPDK*	20*MHz* 30*MHz*	zero
[13]	OTA-1	17	$C \rightarrow 1$	Grounded	180*nm GPDK*	1*MHz*	411.2μW
[7]	VDTA	16	$R \rightarrow 2, C \rightarrow 1$	Floating	180*nm T SMC*	50*MHz*	NA
[21]	-	16	$C \rightarrow 1$	Floating	180*nm T SMC*	*Few Hz*	8.015μW
[30]	-	7	$C \rightarrow 1$	Grounded	180*nm T SMC*	50*MHz*	-NA-
[15]	FCS	6	$C \rightarrow 1$	Floating	180*nm T SMC*	10*Hz*	-NA-
[20]	-	4	$C \rightarrow 1$	Floating	180*nm GPDK*	3*MHz*	8.24μW
[38]	-	5	0	both	180*nm GPDK*	500*MHz*	zero
Proposed Work	-	6	$C \rightarrow 1$	Floating	180*nm GPDK*	10*MHz*	zero

Following this, a comprehensive comparison was undertaken to assess the proposed memristor emulator circuit in relation to existing counterparts [7, 9, 13, 15, 20, 21, 28-30, 38], which encompass fully or partially floating, and grounded configurations. The static power consumption is zero because of no biasing. if there is no applied voltage then power consumption will be zero. The comparison, as outlined in Table II.

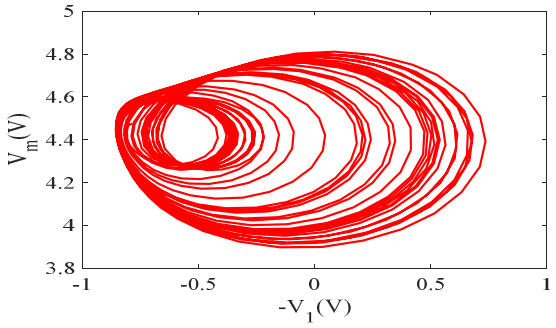


Fig. 15. Colpitts chaotic oscillator circuit performance for $-V_1$ Vs V_m

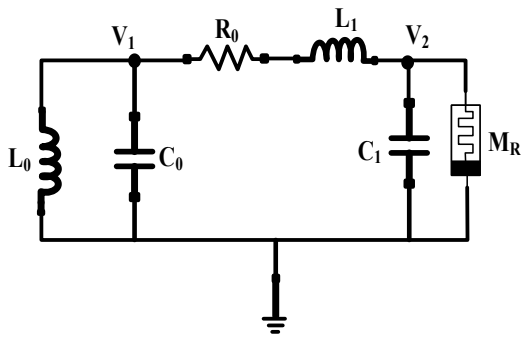


Fig. 16. Chua's Oscillator Circuit using memristor

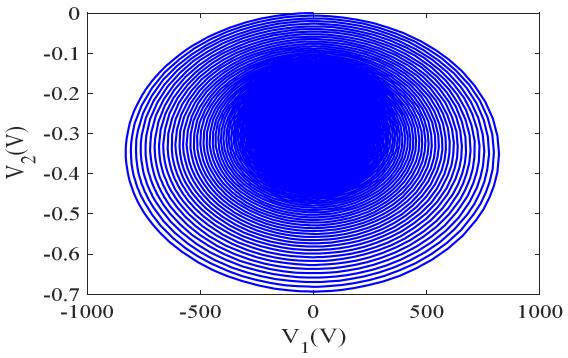


Fig. 17. Chua's oscillator circuit output for V_1 Vs V_2

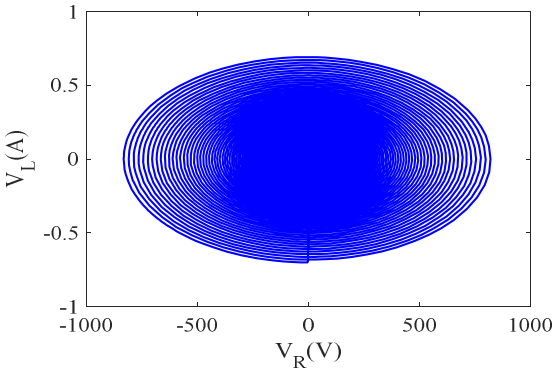


Fig. 18. Chua's oscillator circuit for V_R Vs V_L

V. CONCLUSION

In this research, an alternate approach to emulating a memristor circuit is introduced, consisting of six MOSFETs and one capacitor, devoid of any external biasing. This particular memristor circuit operates on the basis of a completely passive memristor, presenting a less intricate design compared to previously documented circuits. The results of simulations conducted on the suggested circuit

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 05:53:41 UTC from IEEE Xplore. Restrictions apply.

have yielded positive out- comes. Furthermore, this study includes evaluations of process corner and temperature variations, the findings of which are detailed. The publication concludes with the demonstration of a Colpitts chaotic oscillator and Chua's oscillator circuit. Additionally, an analysis and demonstration of a Schmitt trigger circuit as a regenerative comparator circuit are presented. In conclusion, all simulated outcomes are not only satisfactory but also consistent with prior research. Due to its simplified structure, the proposed circuit holds significant promise for future applications.

ACKNOWLEDGEMENT

The authors are thankful to Delhi Technological University for providing access to the simulation tools required for the implementation of the proposed work. Author is also thankful to his supervisors for their continuous support.

REFERENCES

[1] L. Chua, "Memristor-the missing circuit element", IEEE Transactions on Circuit Theory, 18(5), pp.507-519, 1971.

[2] L. O. Chua, and S. M. Kang, "Memristive devices and systems", Proceedings of the IEEE, 64(2), pp.209-223, 1976.

[3] L. O. Chua, "The fourth element", Proceedings of the IEEE, 100(6), pp.1920-1927, 2012.

[4] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found", nature, 453(7191), pp.80-83, 2008.

[5] J. Borghetti, G. Snider, P. Kuekes, et al., "'Memristive' switches enable 'stateful' logic operations via material implication", Nature 464, 873–876, 2010.

[6] G. C. Adam, B. D. Hoskins, M. Prezioso, F. Merrikh-Bayat, B. Chakrabarti, and D. B. Strukov, "3-D memristor crossbars for analog and neuromorphic computing applications", IEEE Transactions on Electron Devices, 64(1), pp.312-318, 2016.

[7] A. Yesil, Y. Babacan and F. Kacar, "Design and experimental evolution of memristor with only one VDTA and one capacitor," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 38, no. 6, pp. 1123-1132, Jun 2019.

[8] R. K. Ranjan, N. Rani, R. Pal, S. K. Paul, and G. Kanyal, "Single CCTA based high frequency floating and grounded type of incremental/decremental memristor emulator and its application", Microelectronics Journal, 60, pp.119-128, 2017.

[9] A. Yesil, Y. Babacan and F. Kacar, "A new DDCC based memris- tor emulator circuit and its applications", Microelectronics Journal, 45(3), pp.282-287, 2014.

[10] A. Yesil, Y. Babacan and F. Kacar "Electronically tunable memristor based on VDCC", AEU-International Journal of Electronics and Communications, 107, pp.282-290, 2019.

[11] J. Vista, and A. Ranjan, "Flux controlled floating memristor employing VDTA: Incremental or decremental operation", IEEE Trans- actions on Computer-Aided Design of Integrated Circuits and Systems, 40(2), pp.364-372, 2020.

[12] Yu, D., Iu, H.H.C., A. L. Fitch, and Y. Liang, "A floating memristor emulator based relaxation oscillator", IEEE Transactions on Circuits and Systems I: Regular Papers, 61(10), pp.2888-2896, 2014.

[13] K. Kumar, B. C. Nagar, and G. Pradhan, "Single OTA-based tunable resistorless grounded memristor emulator and its application", J Comput Electron 22, pp. 549–559, 2023.

[14] N. Raj, V.K. Verma, and R. K. Ranjan, "Electronically Tunable Flux-Controlled Resistorless Memristor Emulator", IEEE Canadian Journal of Electrical and Computer Engineering, 45(3), pp.311-317, 2022.

[15] Y. Babacan, and F. Kacar, "FCS based memristor emulator with associative learning circuit application", IU-Journal of Electrical and Electronics Engineering, 17(2), pp.3433-3437, 2017.

[16] M. T. Abuelma'atti, and Z. J. Khalifa, "A new memristor emulator and its application in digital modulation", Analog Integrated Circuits and Signal Processing, 80, pp.577-584, 2014.

[17] F. Jiang, F. Yuan, and Y. Li, "Design and implementation of XOR logic circuit based on generalized memristor", The European Physical Journal Special Topics, 231(3), pp.481-491, 2022.

[18] H.A. F. Almurib, T. N. Kumar, and F. Lombardi, "Design and evaluation of a memristor-based look-up table for non-volatile field pro- grammable gate arrays. IET Circuits, Devices Systems, 10(4), pp.292- 300, 2016.

[19] Y. Abdullah, B. Yunus, "Tunable memristor employing only four transistors", AEU - International Journal of Electronics and Communications, Volume 169, pp. 1434-8411, 2023.

[20] Y. R. Ananda, N. Raj, and G. Trivedi, "A MOS-DTMOS Implementation of Floating Memristor Emulator for High-Frequency Applications. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2022.

[21] Y. Babacan, and F. Kacar, "Floating memristor emulator with subthreshold region", Analog Integrated Circuits and Signal Processing, 90, pp.471-475, 2017

[22] Y. Babacan, A. Yesil, and F. Gul, "The fabrication and MOSFET- only circuit implementation of semiconductor memristor", IEEE Transactions on Electron Devices, 65(4), pp.1625-1632, 2018

[23] Y. Babacan, Y., "Memristor: Three MOS Transistors and One Capacitor", In IEEE Conference Paper, 2017.

[24] M. Ghosh, A. Singh, S. S. Borah, J. Vista, A. Ranjan, and S. Kumar, "MOSFET-Based Memristor for High-Frequency Signal Processing", IEEE Transactions on Electron Devices, 69(5), pp.2248-2255, 2022.

[25] F. Gul, "Circuit implementation of nano-scale TiO 2 memristor using only metal-oxide-semiconductor transistors", IEEE Electron Device Letters, 40(4), pp.643-646, 2019.

[26] R. K. Gupta, M. S. Chaudhary, S. Taran and V. Saxena, "A Passive Grounded MOSCAP- Memrisor Emulator," 2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, pp. 1-5, 2022.

[27] P. Kumar, P. Srivastava, R. K. Ranjan and M. Kumngern, "New Zero Power Memristor Emulator Model and Its Application in Memristive Neural Computation," in IEEE Access, vol. 11, pp. 5609-5616, 2023.

[28] P. Srivastava, R.K. Gupta, R.K. Sharma, and R. K. Ranjan, "MOS-only memristor emulator", Circuits, Systems, and Signal Processing, 39, pp.5848-5861, 2020.

[29] J. Vista and A. Ranjan, "A Simple Floating MOS-Memristor for High-Frequency Applications," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 27, no. 5, pp. 1186-1195, May 2019.

[30] A. Yesil, "A new grounded memristor emulator based on MOSFET-C", AEU-International Journal of Electronics and Communi- cations, 91, pp.143-149, 2018.

[31] Z. Lei, W. Chune, Q. Huibin, W. Qianjin, "A 300 MHz MOS- only memristor emulator", AEU - International Journal of Electronics and Communications, Volume 162 pp. 1434-8411, 2023.

[32] R. K. Gupta, M. S. Chaudhary, V. Saxena and S. Taran , "A Single MOS-Memristor Emulator Circuit". Circuits Syst Signal Process 2023. https://doi.org/10.1007/s00034-023-02500-5

[33] S. P. Adhikari, M. P. Sah, H. Kim and L. O. Chua, "Three Fingerprints of Memristor," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 60, no. 11, pp. 3008-3021, Nov. 2013.

[34] A. S. Elwakil, M. E. Fouda and A. G. Radwan, "A Simple Model of Double-Loop Hysteresis Behavior in Memristive Elements," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 60, no. 8, pp. 487-491, Aug. 2013.

[35] D. Batas and H. Fiedler, "A Memristor SPICE Implementation and a New Approach for Magnetic Flux-Controlled Memristor Modeling," in IEEE Transactions on Nanotechnology, vol. 10, no. 2, pp. 250-255, March 2011.

[36] E. Gale, "The Memory-Conservation Theory of Memristance," 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, UK, 2014, pp. 599-604.

[37] B. Muthuswamy, and P. P. Kokate, "Memristor-based chaotic circuits" IETE Technical Review, 26(6), pp.417-429, 2009.

[38] P. Srivastava, R. K. Sharma, R. K. Gupta, F. Kacar and R. K. Ranjan, "New DTMOS Based High Frequency Memristor Emulator and Its Nonlinear Applications," in IEEE Access, vol. 12, pp. 9195-9205, 2024.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 05:53:41 UTC from IEEE Xplore. Restrictions apply.

A Survey on Artificial Intelligence-based Cyber Security in IoT Networks

Akshit Tyagi
Dept. of Computer Engineering
Delhi Technological University
New Delhi, India
tyagiakshit1810@gmail.com

Aniket Singh
Dept. of Computer Engineering
Delhi Technological University
New Delhi, India
anikets2002@gmail.com

Amit Yadav
Dept. of Computer Engineering
Delhi Technological University
New Delhi, India
yadavamit92454@gmail.com

Pawan Singh Mehra
Dept. of Computer Engineering
Delhi Technological University
New Delhi, India
pawansinghmehra@gmail.com

2024 2nd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT) | 979-8-3503-7284-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/DICCT61038.2024.10533050

***Abstract*—With the growth of network technologies and the increase in data transmission, cyber threats have become more complicated. Artificial Intelligence allows automated detection of cyber-attacks, which is important for IoT security. To achieve this, this research examines the application of deep learning and machine learning methods to DoS protection strategies for the IoT. The study uses performance metrics, and validation methods to stimulate further research on IoT security and intrusion detection systems using ensemble learning. The objective of this study was to explore and enhance the present state of knowledge regarding IoT Security, as well as to produce intrusion detection systems employing ensemble learning. We conducted a comprehensive survey of existing literature by various authors, without undertaking any original experimentation.**

Keywords—Cyber Security, Artificial Intelligence, Deep Learning , Machine Learning, IoT

I. INTRODUCTION

Today's networked world has never been so simple or efficient thanks to IoT, but there are also major cybersecurity threats. In the current era of inexpensive sensor devices and cloud computing, IoT is seen as a disruptive force that facilitates the integration of intelligent virtual interfaces with physical objects. Each layer comes with a unique security measure within its architecture [1], leaving IoT facing a complex maze of security challenges. Additionally, Internet of Things (IoT) systems are extensive and highly complex. Internet of Things attacks are among the most harmful types of attacks since they prevent individuals who have paid for a particular service from using it. Additionally, hackers these days are smarter such that they design malware and zero-day exploits that avoid detection and sometimes go undetected for quite some time. These systems consist of many components and are highly prone to attacks [2] hence, they require comprehensive security approaches. On the other hand, there exist numerous stand-alone techniques that can detect attacks as well as prevent security breaches or intrusions into these IoT systems; these include firewalls, antivirus programs, and IDSs/IPSs. Furthermore, this advantage given to adversaries by the fact that resources with security demands only need one weak point to be identified is large. To further improve the security stance of any defense responsibility, data science and machine learning can be employed to link events, identify trends, and identify abnormal behavior. With the growing prevalence of IoT devices, so too do the vulnerabilities and attack vectors that must be addressed by robust cybersecurity measures. Two general aspects of IoT security deserve special mention: malware detection and intrusion detection systems. In addition, malware threatens the integrity and secrecy of information in IoT ecosystems while the launch of intrusion may disrupt their normal operations. This comprehensive survey will focus on these key aspects by examining shallow and deep learning algorithms as well as methodologies and techniques. Additionally, this study will evaluate AI-based countermeasures' performance to understand their efficacy in protecting IoT networks. Our analysis encompasses a wide array of IoT threats and innovative AI solutions designed to mitigate these risks.

The following is a list of this survey's major contributions:

- Countermeasure Methods: Exploring shallow and deep learning algorithms.
- Performance Assessment: Evaluating AI-based countermeasures.
- Comprehensive Analysis: Analyzing IoT threats and AI solutions.

A. Types of cyber attacks

An increase in the amount of internet activity is being witnessed in every field. The quantity and type of cyberattacks also rise as a result of this development. Consequently, it is well known that cyber dangers today influence enterprises more than they did in the past. So much so that, as of late, cyberattacks are employed as a weapon [3]. Two typical cyberattacks are described here. We'll talk about the security measures for these assaults later.

1) Malware: Malware is software that, without the user's awareness, compromises or grants illegal access to networks, devices, or websites for identity theft, data breaches, and espionage. The program may act differently depending on the attacker's goals, carrying out tasks like data deletion, encryption, hijacking, and unauthorized computer activity monitoring. To determine whether computer software is malicious or not, one can apply static analysis or dynamic analysis approaches [4]. On the other hand, both methods are inadequate against zero-day attacks. These days deep learning and machine learning approaches are gradually taking their place.

2) DoS attacks: The main agenda behind a DoS attack is prohibiting a system or network so that users who rely on

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:39:55 UTC from IEEE Xplore. Restrictions apply.

it cannot utilize it. DoS attacks achieve their target by flooding that particular system with a lot of traffic or by offering data that causes the target system to crash. DDoS attacks are those attack systems that carry out a synchronized attack against a sole target. The primary distinction is that the target is attacked from various points in contrast with just one [20]. Worldwide research is being done extensively on applying intrusion detection to resolve this issue.

B. *Intrusion detection system (IDS)*

Network IDS are required to shield a network from unauthorized or malicious activities. Three sorts of IDSs are recognized based on the type of detection employed: signature, anomaly, and specification. *For papers with more than six authors:* Add author names horizontally, moving to a third row if needed for more than 8 authors.

1) Signature-based IDS: Network traffic patterns are compared to assault-style (signatures) that have been entered in the database of a signature-based intrusion detection system. If a similar is detected, a warning is sounded. Although this system is highly accurate and has a low false alarm rate, it fails to detect new threats.

2) Specification-based network IDS: To identify malicious activity, a detailed-based network intrusion detection system (IDS) contrasts traffic patterns (parameters) to a predetermined set of guidelines and standards. A security expert manually specifies these requirements.

3) Anomaly-based IDS: Any departure from the typical network profile in network traffic is constantly monitored by a system for detecting intrusions based on anomalies. A warning is triggered to indicate the detection of an attack if the deviation is greater than the threshold. The standard attributes of networks are acquired through machine learning (ML) approaches. The capacity of an intrusion detection system that relies on anomalies to find novel threats makes it superior to signature and specification-based intrusion detection systems; nevertheless, this comes at a high false alarm rate.

II. Related Works

Numerous surveys have examined IoT security, often focusing on encryption, authentication, and access control [5]. However, in the past, some have personally explored topics such as IoT communication security and intrusion detection. However, few surveys have mentioned machine learning (ML) techniques for securing IoT through wireless sensor networks [7]. Our research stands out by delving into IoT safeguarding with special emphasis on ML and DL techniques. Misra and his colleagues [8] suggested a detailed IDS based on Learning Automata for protecting IoT networks from distributed DoS attacks. Meanwhile, Lee et al. [9] presented a DoS-based innovative IDS. Diro et al. 's team [10] introduced an IDS that uses deep learning techniques to defend against DoS attacks on IoT networks. They validated their model using NSL-KDD datasets. Furthermore, Primartha et al. [11] conducted a performance analysis on Random Forest (RF)--based IDS by considering accuracy and false alarm rate. The authors leveraged the UNSW-NB15, GPRS, and NSL-KDD datasets to train and test their model. Paper [12], worked on CPU register information, memory file system, and opcodes to classify malware into families, using SVM and K-NN models. However, the approach's limitation arises from significant variations within malware samples that are of the same family, which can decrease the model's overall accuracy. In the paper [13], the author utilized strings and PE file characteristics as features, with a larger number of malicious samples compared to benign ones. Neural networks were used for classifying malware as malicious or benign. Nonetheless, a limitation of this approach is the potential inaccuracies in labeling the training dataset.

III. DL And ML Methods In IoT Security

The ability of DL and ML algorithms to tackle complicated issues through experience-driven advancement has led to their considerable importance in real-world applications. The rise of low-computation-cost algorithms, The development of novel approaches, and the availability of large datasets are all credited with the algorithms' quick improvement. DL approaches best serve pattern analysis, which use multi-layer non-linear processing to abstract and alter the features, while ML methods usually rely on designed features. By using training data and experience, these algorithms seek to improve task performance.

Decision Trees (DT) are useful for identifying network intrusions because they provide an easy-to-understand paradigm. Support Vector Machines (SVM) help detect malware and breaches by creating hyperplanes to categorize data. To detect intrusions, Naive Bayes (NB) requires the least amount of data by computing probability for characteristics. k-Nearest Neighbors (KNN) is a common intrusion and anomaly detection algorithm that uses nearby data for categorisation. Decision trees are combined in Random Forests (RF) to provide strong predictions that help detect anomalies and illegal devices. k-Means Data may be grouped via clustering, useful for identifying Sybil assaults. Convolutional Neural Networks (CNNs) provide strong virus protection for Internet of Things devices by reducing data utilizing common parameters. Neural networks with recurrent architectures (RNNs) perform exceptionally well with sequential data, improving network traffic categorisation. Last but not least, in Internet of Things contexts, Restricted Boltzmann Machines (RBMs) are useful for anomaly and intrusion detection since they extract characteristics through unsupervised learning. These machine learning techniques offer many tools to tackle security issues in the constantly changing IoT system environment.

Table 1 outlines various Machine Learning (ML) algorithms commonly used in Intrusion Detection Systems (IDS) for enhancing security in IoT (Internet of Things) environments.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:39:55 UTC from IEEE Xplore. Restrictions apply.

TABLE I. COMPARING ALGORITHM FOR THEIR APPLICATIONS IN IOT SECURITY

Algorithm	Basic principle	Benefits	Drawbacks	Possible Use in Security for IoT
DT	Establishes a model using decision trees (DT) from training samples, making predictions based on branches and leaves	Clear, uncomplicated, and simple technique	Needs a lot of storage, easier to understand with a few DTs	Detection of questionable traffic sources and intrusions
SVM	Forms a hyperplane to maximize the distance between sample points of different classes in feature space	Generalization capability, suitable for large feature datasets	Difficult kernel selection, complex model interpretation	Detection of intrusion, malware, smart grid attacks
NB	It applies the theory, which computes the posterior probability to predict potential for a certain feature set.	due to the Bayesian theorem, it is simple. Training dataset requirements are low.	works autonomously on features. unable to get hints from the relationship and interplay of features.	Intrusion detection in networks with issues related to dimensionality
KNN	A sample point is classified by KNN using the votes from neighboring samples.	Popular, effective for intrusion detection systems	Hard to select optimal k-value	Detection of intrusions, anomalies
RF	Combines multiple decision trees to establish a robust prediction model	Robust to overfitting, minimal feature selection, few input parameters	Impractical for large training datasets, time-consuming	Identification of abnormalities, DDoS assaults, and unapproved IoT devices
K-Means	Unsupervised clustering based on feature similarities	Suitable for data without labels, no need for labeled data	Practically less accurate and successful than supervised learning techniques.	Detection of Sybil, data anonymization
CNNs	uses equivariant representations, parameter sharing, and sparse interactions to reduce the parameters of the data.	Robust supervised DL method, scalable, improved training time complexity	High computational cost, challenging implementation on devices with restricted resources	End-to-end assurance strategy for IoT devices, malware detection
RNNs	handles sequential data by integrating a temporal layer; provides outstanding performance in several applications.	Achieves excellent performance in sequence-based threats	Gradients that disappear or explode provide a problem.	Classifying network traffic and enhancing IoT system security
RBMs	Deep generative models for unsupervised learning, extract vital features through unsupervised approach	Extraction of important features, unsupervised learning approach	High computational cost, challenging for resource-constrained devices	intrusion detection, Network anomaly detection

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:39:55 UTC from IEEE Xplore. Restrictions apply.

IV. RESULTS

The results of the study encompass an extensive evaluation of deep learning and machine learning techniques in connection to IoT security. These techniques, applied to various cyber-security focus areas, demonstrated impressive accuracy and performance. Diverse datasets, coupled with feature selection methods, were instrumental in optimizing IDS.

A. Intrusion detection systems using machine learning algorithms

Table 2 presents the performance of various intrusion detection models on different datasets. Notably, XGBoost achieved the highest accuracy of 98.69% on the NSL-KDD dataset, with precision and recall both exceeding 98%. The Grasshopper optimization algorithm with Support Vector Machine(GOA-SVM) also demonstrated strong performance on the KDDCUP99 dataset, achieving an accuracy of 98.52% with precision and recall both over 95%. These results highlight the effectiveness of these models in accurately detecting intrusions in network traffic data.

TABLE II. COMPARISON OF ML MODELS

Model	Dataset	Accuracy	Precision	Recall	F1 Score
GNB [14]	KDDCUP99	0.8328	0.90	0.83	0.85
GOA-SVM[15]	KDDCUP99	0.9852	0.9563	0.9752	0.976
GA-RF [16]	UNSW-NB15	0.928	0.616	0.630	0.623
XGBoost [17]	NSL-KDD	0.9869	0.9841	0.9911	0.9876
C4.5 [18]	NSL-KDD	0.941	0.945	0.941	0.943

B. Detection systems including intrusion using deep learning algorithms

This section explores diverse DL methods for intrusion detection. The results obtained by different authors are shown in table 3. Wang et al. [19] applied a CNN similar to LeNet-5, achieving 99.17% accuracy on CTU-13 and IXIA datasets. Staudemeyer [20] used an RNN on KDD-1999 with 100% detection rate and 2.30% FPR. Diro and Chilamkurti [22] used a DNN on NSL-KDD, reaching 99.20% accuracy, 99.27% TPR, and 0.85% FPR. Kang and Kang [23] utilized a DBN on in-vehicle networks with 97.80% accuracy. Maimó et al. [24] combined a two-stage DBN and RNN approach, achieving 81.26% precision and 99.34% recall in the first, and 68.63% precision and 70.95% recall in the second.

C. Malware detection systems using machine learning algorithm

An assessment of four distinct classification algorithms: Logistic Regression, SVM, GA-KNN, and PSO-KNN, using a provided dataset is shown in fig 1. The analysis revealed that Logistic Regression achieved the highest accuracy, standing at 99.36%, closely followed by SVM with an accuracy rate of 99.08%. Notably, GA-KNN showcased exceptional precision at 99.49%, while PSO-KNN excelled in terms of recall, reaching 99.16% [25-28].

TABLE III. OVERVIEW OF DL MODELS

Authors	Dataset	Model Description	Accuracy	TPR	FPR
Wang et al. [19]	CTU-13 & IXIA (USTC-TFC2016)	CNN (Similar to LeNet-5)	99.17%	100 %	Varies
Staudemeyer [20]	KDD-1999	RNN	100% (Detection Rate)	-	2.30%
Chawla [21]	IoT	DBN (Each hidden layer from autoencoder)	95.03%	95.40 %	-
Diro and Chilamkurti [22]	NSL-KDD	DNN (3 hidden layers)	99.20%	99.27 %	0.85%
Kang and Kang [23]	In-vehicle network (CAN packets)	Deep Belief Network (DBN) with RBMs	97.80%	-	1.60%
Maimó et al. [24]	CTU dataset	Stage 1: DBN (Two hidden layers), Stage 2: RNN	Precision: 81.26%, Recall: 99.34% (Stage 1)	Precision: 68.63 %, Recall: 70.95 % (Stage2)	-

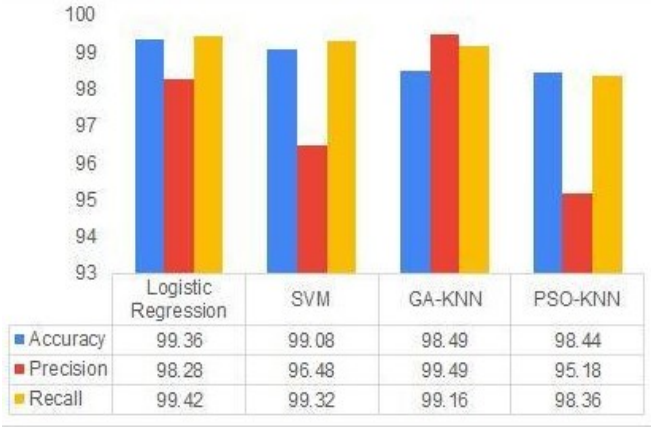


Fig. 1. Comparison of machine learning Algorithms

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:39:55 UTC from IEEE Xplore. Restrictions apply.

D. Malware detection systems using DL algorithms

Figure 2 lists the outcomes of using DL algorithms for malware identification. The analysis encompassed four key approaches: Hybrid CNN, LSTM CNN, DNN, and CNN. Notably, Hybrid CNN and LSTM CNN exhibited high accuracy rates of 99.36% and 99.08%, respectively [25]. While DNN and CNN achieved slightly lower accuracy at 90.4% and 96.6%, they demonstrated good precision and recall values [26].

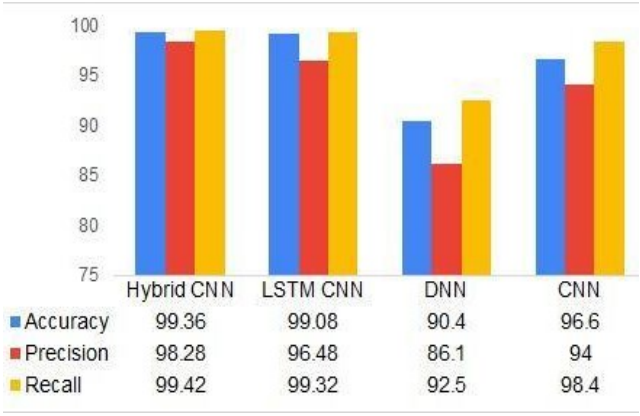


Fig. 2. Comparison of Deep learning Algorithms

V. CONCLUSION

In cybersecurity, privacy and security are essential, especially in IoT scenarios such as smart homes and hotels. This study evaluates the effectiveness of Artificial Intelligence (AI) in analyzing malware and detecting intrusions for IoT security. We analyze accuracy, methodologies, datasets, and outcomes. The research highlights AI's high performance in detecting and classifying attacks, emphasizing its value in enhancing IoT security measures.

A comprehensive survey on malware analysis and intrusion detection in IoT systems conducted from 2013 to 2023 yielded noteworthy findings. Machine learning algorithms demonstrated exceptional performance, with XGBoost emerging as the top performer in intrusion detection, achieving an AUC score of 98.76%. Deep learning techniques, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), also exhibited strong capabilities. CNNs achieved an impressive accuracy of 99.17%, while RNNs showcased a remarkable 100% detection rate with a low false positive rate. Furthermore, traditional machine learning methods like SVM and logistic regression delivered excellent classification performances, surpassing 99% accuracy. Hybrid models that combined CNNs and LSTM CNNs excelled in malware identification, reaching accuracy rates of 99.36% and 99.08%, respectively.

The study has shown that Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Extreme Gradient Boosting (XGBoost) are all effective methods for analyzing malware and detecting intrusions in security systems for the Internet of Things.

REFERENCES

[1] H. Nagdewani and P. S. Mehra, "A Complete Internet of Things based Home Security System," In 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA), pp. 1–5, 2022.

[2] D. Chawla and P. S. Mehra, "A roadmap from classical cryptography to post-quantum resistant cryptography for 5G-enabled IoT: Challenges, opportunities and solutions," Internet of Things, vol. 24, p. 100950, Dec. 2023.

[3] R. Saxena, E. Gayathri, "Cyber threat intelligence challenges: Leveraging blockchain intelligence with possible solution," Mater. Today: Proc. 51, pp. 682–689, 2021.

[4] M. Goyal, R. Kumar, "A survey on malware classification using machine learning and deep learning," Int. J. Comput. Netw. Appl. Vol. 8, pp. 758–775, 2021.

[5] C. Garg, D. Kumar Mishra, D. Raj, and P. Singh Mehra, "A Survey on Integration of Blockchain and IoT (BIoT): Open Issues, Challenges & Solution," Proceedings of the 2022 3rd International Conference on Intelligent Computing, Instrumentation and Control Technologies: Computational Intelligence for Smart Systems, ICICICT 2022, pp. 880–886, 2022.

[6] B. B. Zarpelão, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga, "A survey of intrusion detection in Internet of Things," Journal of Network and Computer Applications, vol. 84, pp. 25-37, 2017.

[7] Rodríguez E, Otero B, Canal R, "A Survey of Machine and Deep Learning Methods for Privacy Protection in the Internet of Things," Sensors (Basel), vol. 23(3), p. 1252, 2023.

[8] Misra, S., Krishna, P. V., Agarwal, H., Saxena, A., & Obaidat, M. S., "A learning automata-based solution for preventing distributed denial of service in Internet of Things. In IEEE, 4th international conference on cyber, physical and social computing, Internet of Things (ithings/cpscom), pp. 114–122, 2011.

[9] Lee, T. H., Wen, C. H., Chang, L. H., Chiang, H. S., & Hsieh, M. C., "A lightweight intrusion detection scheme based on energy consumption analysis in 6lowpan," Advanced technologies, embedded and multimedia for human-centric computing Dordrecht: Springer, pp. 1205–1213, 2014.

[10] Diro, A. A., & Chilamkurti N., "Distributed attack detection scheme using deep learning approach for Internet of Things," Future Generation Computer Systems, vol. 82, pp. 761–768, 2018.

[11] Primartha, R., & Tama, B. A., "Anomaly detection using random forest: A performance revisited," In 2017 International conference on data and software engineering (ICoDSE), pp. 1–6, 2017.

[12] Deguang Kong and Guanhua Yan, "Discriminant malware distance learning on structural information for automated malware classification," In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1357–1365, 2013.

[13] Joshua Saxe and Konstantin Berlin, "Deep neural network-based malware detection using two-dimensional binary program features," In 2015 10th International Conference on Malicious and Unwanted Software (MALWARE), pp. 11–20, 2015.

[14] B. Zhang, Z. Liu, Y. Jia, J. Ren, X. Zhao, "Network intrusion detection method based on PCA and Bayes algorithm," Secur. Commun. Netw. 2018.

[15] S.S. Dhaliwal, A.A. Nahid, R. Abbas, "Effective intrusion detection system using XGBoost," Information (Switzerland) vol. 9, 2018.

[16] J. Ren, J. Guo, W. Qian, H. Yuan, X. Hao, H. Jingjing, "Building an effective intrusion detection system by using hybrid data optimization based on machine learning algorithms," Secur. Commun. Netw. 2019.

[17] Y. Zhou, G. Cheng, S. Jiang, M. Dai, "Building an efficient intrusion detection system based on feature selection and ensemble classifier," Comput. Netw., vol. 174, p. 107247, 2020.

[18] Dwivedi, Shubhra & Vardhan, Manu & Tripathi, Sarsij, "Defense against distributed DoS attack detection by using intelligent evolutionary algorithm. International Journal of Computers and Applications, vol. 44, pp. 1-11, 2020

[19] W Wang, M. Zhu, X. Zeng, X. Ye, Y. Sheng, "Malware traffic classification using convolutional neural network for representation

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:39:55 UTC from IEEE Xplore. Restrictions apply.

learning," In Proceedings of the IEEE 2017 International Conference on Information Networking (ICOIN), pp. 712–717, 2017.
[20] Staudemeyer, R.C., "Applying long short-term memory recurrent neural networks to intrusion detection," S. Afr. Comput. J., vol. 56, pp. 136–154, 2015.
[21] P. S. Mehra, "LBECR: load balanced, efficient clustering and routing protocol for sustainable internet of things in smart cities," J Ambient Intell Humaniz Comput, Jan. 2022.
[22] Diro, A.A., Chilamkurti, N., "Distributed attack detection scheme using deep learning approach for internet of things," Future Gener. Comput. Syst., vol. 82, pp. 761–768, 2018.
[23] M.J. Kang, J.W. Kang, "Intrusion detection system using deep neural network for in-vehicle network security," PLoS ONE, vol. 11, p. e0155781, 2016.
[24] D. Chawla and P. S. Mehra, "A Survey on Quantum Computing for Internet of Things Security," Procedia Comput Sci, vol. 218, pp. 2191–2200, Jan. 2023.
[25] S. Riaz, S. Latif, S.M. Usman, S.S Ullah, A.D. Algarni, A. Yasin, A. Anwar, H. Elmannai, S. Hussain, "Malware Detection in Internet of Things (IoT) Devices Using Deep Learning," Sensors, vol. 22, p. 9305, 2022.
[26] E.S. Alomari, R.R. Nuiaa, Z.A.A. Alyasseri, H.J. Mohammed, N.S. Sani, M.I. Esa, B.A. Musawi, "Malware Detection Using Deep Learning and Correlation-Based Feature Selection," Symmetry, vol. 15, p. 123, 2023.
[27] R. Chanana, A. K. Singh, R. Killa, S. Agarwal, and P. S. Mehra, "Blockchain Based Secure Model for Sensor Data in Wireless Sensor Network," in 2020 6th International Conference on Signal Processing and Communication (ICSC), pp. 288–293, Mar. 2020.
[28] A. Kumar, P. S. Mehra, G. Gupta, and M. Sharma, "Enhanced block Playfair Cipher," Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, vol. 115, pp. 689–695, 2013.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:39:55 UTC from IEEE Xplore. Restrictions apply.

2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)
Amity University, Noida, India. Mar 14-15, 2024

2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) | 979-8-3503-5035-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICRITO61523.2024.10522110

A Survey on Blockchain for Rental Lease Management

Anushka Jain
Department of CSE
Delhi Technological University Delhi,
jainanushka523@gmail.com

Aparna Gangwar
Department of CSE
Delhi Technological University Delhi,
aparnagangwar2001@gmail.com

Chandni Kumari
Department of CSE
Delhi Technological University Delhi,
chandnidtucoe@gmail.com

Pawan Singh Mehra
Department of CSE
Delhi Technological University Delhi,
pawansinghmehra@gmail.com

***Abstract*—The application of blockchain technology is causing a revolution in the rental management industry by improving the security, privacy, traceability and visibility of any rental commodity. Complex tasks like commodity listings, lease agreements, payment methods, identity verification, and authentication are performed with unparalleled integrity, ultimately reducing the risk of fraud and disputes. The main objective of this survey is to review and analyse peer-reviewed literature that discusses previous works that present solutions to traditional rental lease management using blockchain technology. We start by elucidating the foundational concepts and traditional rental management industry. We then evaluated a few papers, examining the technologies used and their advantages, key focuses, and limitations. This survey aims to determine the potential of blockchain technology in improving transaction efficiency, transparency, cost reduction, and other beneficial aspects of rental management. Blockchain technology is a workable alternative to traditional rental lease management, streamlining processes, building confidence, and asserting the dynamics between landlords and tenants. This paper addresses a survey in different domains of application of blockchain technology and some suggestions for future research.**

Keywords— Blockchain, IoT, Smart Contracts, Rental lease management, Survey

I. INTRODUCTION

Blockchain technology is leading the way in this revolutionary shift that will drastically alter the rental management industry. Traditional Rental Lease Management has long served as the foundation for property rentals, yet inherent inefficiencies have plagued it, including a lack of transparency and substantial reliance on intermediaries. In this system, landlords and tenants engaged in direct negotiations, often involving the laborious process of paper-based lease agreements, where responsibilities, rental terms, and payment schedules were outlined [1,3]. However, it has been criticised for its reliance on manual paperwork, which can be time-consuming and prone to errors and communication between landlords and tenants conducted through in-person meetings, phone calls, or physical mail, leading to delays and misunderstandings. Additionally, there was a lack of transparency in the process, which led to disputes frequently arising from a lack of easily accessible records and the involvement of intermediaries like property managers and lawyers, who are essential in handling transactions, facilitating lease agreements, and resolving disputes but also increased the overall cost and complexity of the rental management process. The industry saw a revolutionary phase of digitalisation to address these issues. The rental lease management industry adopted digital records, online listing platforms, and lease management software. This change was intended to increase digital efficiency, reduce paperwork, and enhance communication. Landlords boosted productivity by simplifying property management tasks thanks to the growing popularity of property management software. Internet listing services made rental properties more visible and easier for tenants and landlords to access. Intermediaries nevertheless added to the overall complexity and expense of the rental management process despite these advancements. Incorporating blockchain technology into rental lease management marked the beginning of the next wave of change. The decentralised, immutable ledger of blockchain technology ushered in a new era of efficiency, transparency, and security. It allowed for peer-to-peer transactions between landlords and tenants, eliminating the need for middlemen. Cryptographic techniques ensured data security while reducing the risk of fraud and unauthorised access. In addition to redefining the landlord-tenant dynamic and resolving the problems with traditional RLM, this technology has also reduced expenses, boosted trust, and expedited procedures, all of which have improved the security and efficiency of the rental management process.

An in-depth analysis of the blockchain's uses, difficulties, and prospects is the goal of this survey, which focuses primarily on the technology's implications and uses in the rental management industry. The demand for innovative solutions is growing due to the long-standing issues with conventional property leasing processes, such as their inefficiencies, lack of transparency, and dependence on intermediaries. This motivation prompted us to study blockchain in the rental lease sector. As we embark on this investigation, we will examine how blockchain technology may revolutionise the rental market by empowering landlords and tenants and enabling more efficient and transparent transactions. This paper also evaluates the knowledge, opinions, and firsthand accounts of different parties on blockchain-enabled rental leasing programs in smart cities, such as landlords, tenants, and local government representatives. The survey results will provide valuable insights crucial for guiding the development and implementation of these advanced systems, ensuring that they meet smart city residents' diverse needs and expectations. This paper explores how blockchain technology can enhance rental lease administration in smart cities.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:16:01 UTC from IEEE Xplore. Restrictions apply.

II. RELATED WORK

The results of the Rental Lease Management System survey are shown in Table 1 about smart cities that are utilising blockchain technology.

TABLE I. RELATED WORK

References and Year	Title	Objective	Description	Methodology	Focus	Limitations
[1] 2019	Land Records on Blockchain for Implementation of Land Titling in India	Overcoming the problem of traditional land records using blockchain networks.	Implementing the nation's land titling system using blockchaintechnology.	Remove minimal transparency, Responsibility, and Data setsthat are incoherent	Block chain techno logy	Lack of proper land demarcation.
[2] 2019	Cryptober: A Blockchain-based Secure and Cost-Optimal CarRental Platform	An in-depth analysis of how smart contracts and blockchain technology may be utilised to develop a car- rental network that benefits both car owners and renters.	A Blockchain and smart contract -powered car rental platform to optimize cost .	Centralised automobile rental services. Requirement for the same kind of service that is easily accessible. Time and money would be saved.	Blockchain, PoS, PoW , Hyper Ledger Fabric	Scalability issues
[4] 2021	Blockchain for IOT-based smartcities: Recent advances, requirements, and future challenges	Evolution of Blockchain technologies, examine various Blockchain-enabled smart applications, and implementation in smart cities in the real world	Advancements in consensus algorithms and the establishment of several blockchain platforms were observed in the growth of blockchain technology.	Improving security, transparency, and scalability to enable the seamless integration of IoT devices and data in urban settings	Blockchain, IoT, Smart Contract	Problems with scalability, identity and privacy issues
[5] 2021	A secure Blockchain-based housing rental platform	Provide a safe rental housing transaction platform based on blockchain technology.	Use of homomorphic encryption algorithms and zero-knowledge proofs to create a secure housing rental platform	Use of Decentralized ledger and zero-knowledge proofs	Blockchain, homomorphic encryption, Zero Knowledge Proof	Data Storage andProtection
[6] 2021	Housing rental system based on Blockchain technology	Construct a house leasing system using Alliance Chain Network	Using Blockchain, this paper constructs a housing leasing Alliance chain.	Use Blockchain Architecture for record maintenance and remove mediation charge	Blockchain	Security issues
[7] 2022	Towards Blockchain-IoT- based shared mobility: Car-sharing and leasing as a casestudy	A high-level design for a shared mobility network based on blockchain and the Internet of Things that combines automobile leasingand sharing.	P2P Car leasing Platform with keyless authentication, IoT-based car data collection and crypto-based payments	Blockchain and IoT technologies are driving the evolution of automobile leasing and sharing. Basic design concepts needed for blockchain-enabled Internet of Things vehicle leasing and sharing.	Blockchain, RFID, Raspberry Pi3 , HyperLedger Fabric	Data Silos, Authenticity ofIoT devices
[8] 2022	Cutting Out the Middleman - Automating Business Processes with Blockchain andsmart contacts.	a creative method for interested parties to share data and automate contracts.	In the financial markets, cryptocurrencies—of which Bitcoin is the most well-known—have grown significantly in importance.	Application of blockchain and distributed ledger technology in smart contracts. Smart contracts' benefits for industry. Examining a few fascinating application cases?	Blockchai n, IoT, Ethereum	Problems with cryptocurrency concepts, Blockchain-based wallets pose unique challenges and have caused significant harm to some cryptocases.
[9] 2023	Adoption of Blockchain technology in the real estate sector toward the improvement of smart cities	What is a smart city, why IOT-based smart cities are necessary, what challenges facing the real estate sector and how is blockchain technology being applied it to build sustainable smart cities	Describe why there is a need for IOT-based smart cities and the challenges faced. Also described various potential applications.	Describe how Blockchain and IoT in the real estate sector helps to improve smart cities.	Blockchai n, IoT, Smart Contract, Ethereum	Expenses connected with smart contracts and scalability of network
[10] 2023	Implementation of Property Rental Website Using Blockchain with Soulbound Tokens for Reputation and Review System.	A decentralised profile reputation system based on soul-bound tokens and digital currency is being introduced.	To reduce the number of rental frauds, transactions in rental listing websites should be made.	The authenticity of listed profiles and user profiles and verification of documents before being uploaded on a chain. Housing and lease scams commonly target individuals who want to rent a property.	Blockchain , IPFS, NFT, Soulbound Tokens	Lack of rules and regulations to protect users

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:16:01 UTC from IEEE Xplore. Restrictions apply.

The survey aimed to determine public opinion and the effectiveness of blockchain-integrated rental leasing systems in smart cities that offer security and transparency for real estate rentals. The survey data is divided into sections in the table, each concentrating on a distinct aspect. This table's data is based on nine different research papers. This survey table aims to improve the paper's usability and accessibility by making it easier for readers to rapidly understand its key points, background, and possible discussion topics. It seeks to simplify the reading experience for the reader and make it easier for them to comprehend the document's main points in Table 1. The survey table provides an in-depth look at how the general public feels about blockchain-enabled smart city rental leasing systems. This data will significantly influence the evolution of these systems in the future. City authorities and property management organisations may make well-informed decisions to increase transparency, security, and the general efficacy of their lease management systems by gaining insight into user attitudes, measuring blockchain trust levels, and finding areas for development. More studies and action plans based on the survey results will be created to enhance and broaden the integration of blockchain technology in the smart city property rental processes. This will facilitate the continued development and expansion of the global smart city movement.

III. PRELIMINARIES

A. Blockchain

Blockchain is a decentralised digital ledger that records network transactions. A group of individuals using Satoshi Nakamoto's pseudonym created it in 2009 as a framework for the peer-to-peer electronic cash system Bitcoin [11,12]. In 1990, Nick Szabo proposed Bit Gold, which was the precursor of blockchain technology. The idea was to create a decentralised system for generating and verifying digital currencies. Various blockchain systems, including Ethereum and Hyperledger Fabric, have later implemented the technology introduced by Nick Szabo. Transparency stands as one of blockchain's primary characteristics. A transaction is accessible to every network member after it has been logged. Because of this openness, which guarantees responsibility and trust, it is practically hard to change or tamper with recorded data without the network's approval. Therefore, it is the best technology for use in accounting and trust applications, including voting systems, supply chain management, and financial transactions[14,15]. Security is another crucial element in blockchain. A consensus method, like proof of stake (pos) or proof of work (pow), connects each block to its precedent securely using unbreakable hash codes. The smart contract refers to a computer program stored in the blockchain, and it is run when specific terms are fulfilled. A smart contract state is recorded on the distributed ledger whenever it is executed. New blocks storing state changes for the state are appended to form an irreversible record representing contract state changes. Unlike traditional contracts, smart contracts eliminate the need for intermediaries like solicitors or notaries. The contract operates automatically, with the code enforcing the agreement's terms once the requirements are fulfilled. The architecture of the blockchain system was designed to create a decentralised and trustless network that is secure and transparent. A blockchain comprises a dispersed network of computers or nodes that collaborate to keep the blockchain secure from censorship and single points of failure. Blockchain consists of mainly two components:-

- A peer-to-peer network- This architecture enables direct communication and interaction between nodes, eliminating the need for a central intermediary. It distributes the blockchain ledger across all nodes, creating a decentralised and trustless system. Each node contributes computational resources to the network, such as processing power, storage capacity, and bandwidth.
- Distributed Ledger- Blockchain is a distributed ledger between a chain of connected members. The ledger is duplicated at each node, and so is transaction verification. Network nodes reach a consensus on adding a new transaction by solving a mathematical puzzle or adopting a proof of work procedure. The entry of a transaction into a ledger makes it inerasable and ensures that the ledger remains unaltered. This architecture ensures blockchain technology's security and transparency.

B. IoT in Smart Cities

IoT is one innovative technological paradigm that has gained much attention lately. The acronym for the Internet of Things is IoT. It is a network of actual physical objects, devices, cars, buildings, and other things that are embedded with network connectivity, software, and sensors to allow them to exchange and gather data over the internet. There are countless IoT applications. Through smartphones, it is possible to control smart lighting, surveillance cameras, and programmable heaters. In healthcare, wearable technology tracks vital signs and transmits information directly to medical professionals for real-time evaluation. IoT sensors are employed in agriculture, where they monitor crop health, weather changes, and soil conditions, aiding farmers in making informed decisions to enhance their harvests. The main idea behind the Internet of Things is to allow these objects to speak with centralised systems and each other via the Internet [13]. This will enable them to share and acquire information, make wise decisions, and carry out a variety of tasks without the assistance of a human. The smart-cities concept of development involves employing and applying ICTs to improve cities' sustainability, livability, and efficiency. Smart cities use data, connectivity, and the latest technologies like IoT to solve complex urban issues. Smart cities leverage collected information from various sources, including cameras, smart sensors, and other electronic devices, to optimize traffic systems, minimise energy consumption, and enhance the quality of public services, among other benefits. Residents gain from more safety, less traffic, and better mobility. The concept of smart cities aims to improve the living conditions of its population by creating more sustainable, intelligent and linked urban systems. The extensive applications of IoT in various domains make it a critical factor in advancing smart cities. The field of traffic management is one where the Internet of Things has significantly contributed to smart cities. Real-time data is gathered by IoT devices, including cameras and sensors, from roads, intersections, and moving cars that are examined to optimise traffic flow by modifying traffic signals dynamically. As a result, travel times are reduced, traffic is lessened, and urban mobility is enhanced overall. Another area where IoT is transforming urban operations is waste management. It can become more effective by keeping an eye on trash cans and streamlining collection routes. This contributes to a more sustainable and effective waste

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:16:01 UTC from IEEE Xplore. Restrictions apply.

management system by lowering operating costs and minimising the environmental impact of needless truck trips. Regarding energy management, Internet of Things (IoT)technologies—such as smart grids—have revolutionised how cities receive and use power. It helps utilities reduce energy waste by better balancing supply and demand. This will benefit the city's economy and its commitment to environmental sustainability. Though there are clear benefits to IoT in smart cities, important issues also need to be resolved. The most pressing of these are security and privacy concerns. The vast amounts of data that IoT devices generate and exchange could be the target of hackers and cyberattacks. IoT system breaches can result in data leaks, city service interruptions, and even safety risks. These risks can be reduced by ensuring robust security policies and encryption for confidential information and citizens' privacy.

C. Rental Lease Management(RLM)

Any management system made to transform the complex process of managing rental leases of any particular commodity into an efficient and smooth process is called Rental Lease Management (RLM). It is a complete system that provides support and convenience for commodity managers, owners, and tenants. It includes management ease and provides the facility for creating and enforcing rental agreements. It makes the process of renting any commodity accessible for ordinary people as it contains all the legally binding documents that are necessary to understand the terms of leased commodity, rules that protect their rights as a landlord and tenant, information about payments and lease obligations, and the length of the tenancy agreement. A system that provides a solution for renting any commodity, from lease negotiation to the termination or reinstatement of your lease. The objective behind having any system that performs all this is to have a clearly defined lease management strategy, as it helps the landlords manage and hold a proper record of their commodities, which promotes accountability and reduces the risk of confusion and conflicts. Implementation of Rental lease management systems is revolutionised because of digital platforms and software solutions as they have secured the databases, an authentication feature provides access to rental agreements to the landlord and tenant, regular inspection of reports and financial transactions helps increase security and authenticity and decreases paperwork. RLM also performs ethical responsibility by conducting an automatic periodic check and releasing notifications promptly to reduce litigation costs and penalties. Tenants get many benefits from using RLM systems, like having access to simplified and documented rental records, which otherwise were hard to maintain access to lease-related info, pay rent electronically, and submit maintenance requests directly to RLM, improving the overall rental experience. All the above simplifications that RLM has offered to the current rental lease system make it an essential part of the leasing industry as it protects the interests of all parties involved, such as landlords, property managers, and tenants. RLM offers the entire traditional leasing process with transparency, accountability, security, and efficiency. Regarding the role of Blockchain technology in the Rental Lease Management process, it holds the potential to revolutionise the entire leasing agreement management process. Due to features like a decentralised and immutable ledger, security will ultimately decrease the probability of fraud, disputes and unauthorized lease notifications. All participants in the rental process will have instant access to the shared ledger due to the decentralised structure, which will ultimately increase trust and openness between the landlords, tenants, and managers. The transparency offered by this technology will go a long way as it reduces conflicts and improves the overall efficiency of the management system. The prime objective of Blockchain technology and smart contracts is to automate processes such as rent payments to streamline lease enforcement and eliminate the need for intermediaries such as banks and the ministry's payment processing. This will not just simplify the payment process but also help in reducing costs and associated delays. Storing money in blockchain-based smart contracts and ensuring that it will only be released when the conditions outlined in the lease are met helps us manage security deposits and users. An easy process of getting an account allows us to avoid relying on third-party resellers, reducing costs and increasing security for everyone. Also, with features like securely storing and exchanging data, including rental history, credit scores, and references, this technology can facilitate tenant screening and verification. Therefore, blockchain technology can revolutionise traditional rental management, adding improved security, transparency, and efficiency, benefiting all parties involved in the leasing process.

IV. FINDINGS AND DISCUSSIONS

The survey results revealed that blockchain technology[16-18] is integrated into rental management in various ways. Blockchain has demonstrated a noteworthy ability to augment trust and transparency within the rental sector. Using a decentralised and unchangeable ledger significantly lowers the possibility of fraud and disagreements, which promotes more confidence between landlords and tenants. Applications such as property advertisements, leasing agreements, payment processing, identification verification, and maintenance management are all covered by this transparency. Real-world case studies and applications in the rental management industry demonstrated blockchain's usefulness. These examples showed how technology may facilitate real estate transactions, reduce fraud, and enhance tenant-landlord relations. It is essential to recognise the obstacles that come with integrating blockchain technology. One significant barrier that has surfaced in the rental management sector is regulatory compliance due to the complex and often region-specific legal and regulatory systems that govern it. These regulations encompass various aspects of the rental market, including tenant rights, property standards, lease agreements, and rent control. Finding a balance between innovation and following the rules is essential for the sector's success. In addition, scalability problems were found in the blockchain ecosystem. Enormous-scale rental businesses require solutions that address the technology's limited capacity to manage enormous amounts of data and transactions. Furthermore, the rental management sector has proven challenging to change because of the industry's firmly rooted traditional practices and established intermediaries. Industry participants must be open to innovation, make the necessary infrastructural investments, and start a cultural revolution using blockchain technology. To address these issues, the survey report offers several practical suggestions. These include creating transparent regulatory frameworks, funding blockchain infrastructure and education, and encouraging industry cooperation. Stakeholders in the rental management industry

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:16:01 UTC from IEEE Xplore. Restrictions apply.

may negotiate the revolutionary terrain and realise the full potential of blockchain technology by tackling these obstacles and putting these suggestions into practice, eventually altering the future of rental management.

V. CONCLUSION AND FUTURE DIRECTIONS

The review of blockchain technology's uses in rental management concludes by highlighting the sector's enormous potential for positive change. By examining its core ideas and valuable applications, we have shown how blockchain technology may improve trust, efficiency, and transparency in real estate rental operations. This survey study, which focuses on rental management, attempts to provide an in-depth analysis of blockchain technology and offers insights into its possible applications, challenges, and barriers. Case studies from the real world have given concrete proof of its advantages, from cutting down on conflicts and fraud to streamlining intricate processes. However, there are difficulties with this transition. Significant challenges still include industry-wide adoption, scalability issues, and regulatory compliance. Despite the apparent advantages, overcoming these obstacles requires coordinated efforts by regulators, innovators, and stakeholders. Blockchain's promise for rental management is vital. It presents the possibility of improved security, simpler operations, and redefining the relationship between property owners and tenants. So, we intend to integrate IoT and Blockchain to prevent IoT data threats. The complete potential of blockchain in rental management is yet unknown as long as the industry and technology don't stop evolving. Our survey study will prove to be a valuable tool for stakeholders and enthusiasts alike, helping them navigate the rapidly changing field of blockchain-driven rental management as the industry continues to embrace this revolutionary technology.

REFERENCES

[1] V. Thakur, M. N. Doja, Y. K. Dwivedi, T. Ahmad, and G. Khadanga, "Land records on Blockchain for implementation of Land Titling in India," International Journal of Information Management, vol. 52, p. 101940, Jun. 2020, doi: 10.1016/j.ijinfomgt.2019.04.013.

[2] V. Hassija, M. Zaid, G. Singh, A. Srivastava, and V. Saxena, "Cryptober: A Blockchain-based Secure and Cost-Optimal Car Rental Platform," -, Aug. 2019, doi: 10.1109/ic3.2019.8844943.

[3] K. M. Alam, J. M. A. Rahman, A. Tasnim, and A. Akther, "A blockchain-based land title management system for Bangladesh," Journal of King Saud University - Computer and Information Sciences,vol. 34, no. 6, pp. 3096–3110, Jun. 2022, doi: 10.1016/j.jksuci.2020.10.011.

[4] U. Majeed, L. U. Khan, I. Yaqoob, S. M. A. Kazmi, K. Salah, and C. S. Hong, "Blockchain for IoT-based smart cities: Recent advances, requirements, and future challenges," Journal of Network and Computer Applications, vol. 181, p. 103007, May 2021, doi: 10.1016/j.jnca.2021.103007.

[5] B. Gami, M. Agrawal, D. K. Mishra, D. Quasim, and P. S. Mehra, "Artificial intelligence-based blockchain solutions for intelligent healthcare: A comprehensive review on privacy preserving techniques," Transactions on Emerging Telecommunications Technologies, vol. 34, no. 9, p. e4824, Sep. 2023, doi: 10.1002/ETT.4824R. Yu, Z. Wang, C. Zhang, and S. Guan, "A secure blockchain-based housing rental platform," -, Jun. 2021, doi: 10.1109/imcec51613.2021.9482058.

[6] Q. Xue, Z. Hou, H. Ma, H. Zhu, X. Ju, and Y. Sun, "Housing rental system based on blockchain Technology," Journal of Physics, vol. 1948, no. 1, p. 012058, Jun. 2021, doi: 10.1088/1742-6596/1948/1/012058.

[7] S. Auer, S. Nagler, S. Mazumdar, and R. R. Mukkamala, "Towards blockchain-IoT based shared mobility: Car-sharing and leasing as a case study," Journal of Network and Computer Applications, vol. 200, p. 103316, Apr. 2022, doi: 10.1016/j.jnca.2021.103316.

[8] Dr. N. Jesse, "'Cut out the middleman' – Automating Business Processes with Blockchains and Smart Contracts," IFAC-PapersOnLine, vol. 55, no. 39, pp. 352–357, 2022, doi: https://doi.org/10.1016/j.ifacol.2022.12.079.

[9] D. Kumar, D. Kashyap, D. Kumar Mishra, and P.S.Mehra, "Integrating Blockchain and 5G: Advantages and Challenges of Decentralization," Proceedings of the 2022 3rd International Conference on Intelligent Computing, Instrumentation and Control Technologies: Computational Intelligence for Smart Systems, ICICICT 2022, pp. 866–873, 2022, doi: 10.1109/ICICICT54557.2022.9917645

[10] M. Sankar, S. Pachiyappan, S. Paramasivan, and A. Srivastava, "Adoption of blockchain technology in the real estate sector toward the improvement of smart cities," in Elsevier eBooks, 2023, pp. 187–198. doi: 10.1016/b978-0-323-95407-5.00002-5.

[11] S. Sharma, A. Kumar, N. Sengar, and A. Kaushik, "Implementation of Property Rental Website Using Blockchain with Soulbound Tokens for

[12] R. Chanana, A. K. Singh, R. Killa, S. Agarwal, and P. S. Mehra, "Blockchain Based Secure Model for Sensor Data in Wireless Sensor Network," in 2020 6th International Conference on Signal Processing and Communication (ICSC), IEEE, Mar. 2020, pp. 288–293. doi: 10.1109/ICSC48311.2020.9182776

[13] Q. Chen, R. Ye, and F. Lin, "A Blockchain-based Housing Rental System," -, Jan. 2019, doi: 10.5220/0008097201840190.

[14] Z. Zheng, S. Xie, H.-N. Dai, X. Chen, and H. Wang, 'An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends', 06 2017.

[15] D. Chawla and P. S. Mehra, "A Survey on Quantum Computing for Internet of Things Security," Procedia Comput Sci, vol. 218, pp. 2191–2200, Jan. 2023, doi: 10.1016/J.PROCS.2023.01.195.

[16] C. Garg, D. Kumar Mishra, D. Raj, and P.S. Mehra, "A Survey on Integration of Blockchain and IoT (BIoT): Open Issues, Challenges & Solution," Proceedings of the 2022 3rd International Conference on Intelligent Computing, Instrumentation and Control Technologies: Computational Intelligence for Smart Systems, ICICICT 2022, pp. 880–886, 2022, doi: 10.1109/ICICICT54557.2022.9917632

[17] H. Arasteh et al., 'Iot-based smart cities: A survey', in 2016 IEEE 16th international conference on environment and electrical engineering (EEEIC), 2016, pp. 1–6.

[18] J. Sharma and P. S. Mehra, "Secure communication in IOT-based UAV networks: A systematic survey," *Internet of Things*, vol. 23, p. 100883, Oct. 2023, doi: 10.1016/J.IOT.2023.100883

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:16:01 UTC from IEEE Xplore. Restrictions apply.

2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)
Amity University, Noida, India. Mar 14-15, 2024

2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) | 979-8-3503-5035-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICRITO61523.2024.10522341

A Survey on Blockchain in Financial Institutions

Anshu Tiwari
Department of Computer Engineering
Delhi Technological University
New Delhi, India
anshow77777@gmail.com

Aryan Kanjolia
Department of Computer Engineering
Delhi Technological University
New Delhi, India
aryankanjoliacp22@gmail.com

Bhishek Kumar
Department of Computer Engineering
Delhi Technological University
New Delhi, India
bhishek21403@gmail.com

Pawan Singh Mehra
Department of Computer Engineering
Delhi Technological University
New Delhi, India
pawansinghmehra@gmail.com

***Abstract*— This paper examines the transformative impact of blockchain technology in the financial sector, exploring the advantages, challenges, and future prospects of adopting blockchain and emphasising its role in enhancing efficiency and reducing transactional costs. Also, a comparative analysis is done across various blockchain technologies that are being used in the current FinTech sector. AI-powered chatbots streamline banking, while AI is implemented in blockchain for reducing operational costs, increasing scalability and AI in decentralised finance. Also, The decentralised method of operations across the FinTech sector is explored by disrupting the intermediaries across the network and the integration of UPI in regulating compliances is also discussed. Furthermore, the research highlights ongoing efforts to address scalability, interoperability, reducing security concerns, underscoring the potential for continued evolution and innovation in financial systems.**

Keywords*— *Blockchain, Financial Institutions, Consensus Algorithms, Smart Contracts, Cryptocurrency Trading, Artificial Intelligence.

I. INTRODUCTION

Increasing demand for digital usage and e-commerce has led to a significant increase in financial services, especially during the pandemic, as more individuals migrate online [1]. Blockchain turned into a complete game-changer in the domain of Fintech, as it reduces the cost of transactions, ensures security concerns, enables international payment methods and enhances regulation & auditing.

Blockchain's capacity to initiate, process, authorise, record, and report on every transaction conducted across diverse nodes has significantly impacted the realms of accounting and auditing. The advent of this groundbreaking technology has brought about a transformation in the back office, efficiently replacing the laborious tasks associated with financial audits and tax preparation.

The banking and financial sector has seen enormous developments due to the combination of blockchain and AI technologies. While AI has increased compliance and safety through autonomous verification methods, blockchain's distributed architecture has improved effectiveness and decreased transaction costs. Despite regulatory compliance and scalability challenges, the trend toward blockchain usage is nevertheless encouraging. Artificial intelligence (AI)-powered chatbots greatly streamline banking procedures by providing cheap transaction processing, detection verification, and customer service.

Future research will focus on using AI-driven solutions to enhance blockchain's scaling, connectivity, and security.

Interoperability between blockchains, artificial intelligence in decentralised banking, and enhancing security and privacy protocols are important fields. This multifaceted strategy highlights how blockchain and AI can work jointly to drive constant enhancements in financial systems. Combining blockchain technology with artificial intelligence (AI) provides more development and innovation in the financial sector as experts investigate prediction models and algorithms.[2]

Despite high-level assessments of blockchain technology, there remains a notable absence of a systematic comparison of platforms utilising blockchain technology, specifically within the financial services domain. Current research lacks a comprehensive examination of the technical implementation and applications of blockchains and distributed ledger technology across various FinTech categories in the financial services sector. While some studies have offered theoretical overviews and plans to implement in fintech platforms[4], others have explored digital finance from an operational function perspective[5].

Through our research paper we tried to delved into FinTech advancements such as analysis of various blockchain technologies and their use cases, distributed ledger[6] and the use of AI in blockchain technology. Our research's exclusive focus is to delineate blockchain-based FinTech applications, providing a dedicated exploration into this niche area.

II. EFFECTS ON FINANCIAL INSTITUTIONS

The culture of the financial world depends on centralised intermediaries, but blockchain broke the tradition by allowing a decentralised substitute. It helps reduce the efficiency and transaction costs. Centralised intermediaries boosted the unauthenticity, due to cybersecurity scams, and even steven in finances.[7]

Fundamental Banks having a feeling of threat when thinking of blockchain, some competed with markets using it. Fragile investments compressed in information technology make for cost-efficient and innovative prompts.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:14:02 UTC from IEEE Xplore. Restrictions apply.

TABLE I. USE OF BLOCKCHAIN IN DIFFERENT FINANCIAL INSTIUTIONS

Financial Sector	Blockchain Types	Key companies/Projects	Use cases	Advantages	Challenges/Risks
Cryptocurrencies and Exchanges	Bitcoin (BTC), Ethereum (ETH)	Bitcoin: Decentralized digital currency Ethereum: Smart contracts and decentralized applications	Peer-to-peer transactions, decentralized applications, programmable contracts	Decentralization, Security, Anonymity	Regulatory Uncertainty, Volatility, Scalability Issues
Cross-Border Payments	Ripple (XRP), Stellar (XLM)	Ripple (XRP): RippleNet Stellar (XLM): Stellar Network	Facilitate fast, low-cost cross-border transactions, connecting financial institutions for efficient international transfers	Speed, Cost Efficiency, Reduced Intermediaries	Regulatory Compliance, Adoption Challenges, Network Effects
Supply Chain Finance	IBM Blockchain: Hyperledger Fabric VeChain (VET): VeChainThor Blockchain	Real-time tracking, transparency, and authentication of products in the supply chain, reducing fraud and improving traceability	Traceability, Efficiency	Integration Challenges, Standardization,	Privacy Concerns
Trade Finance	WeTrade, Marco Polo (R3 Corda)	Built on Hyperledger Fabric Marco Polo: R3 Corda	Digitizing and automating trade processes, reducing paperwork, enhancing transparency in trade finance	Efficiency, Transparency, Reduced Paperwork	Standardization, Interoperability, Legal Frameworks
Smart Contracts and Insurance	Ethereum (ETH), Chainlink (LINK), AXA	AXA: Exploring blockchain for smart contracts Chainlink (LINK): Oracle network for smart contracts	Automated and transparent insurance claims processing through smart contracts, improving efficiency and trust in insurance processes	Automation, Trust, Transparency	Legal and Regulatory Challenges, Security Concerns, Complexity
Digital Identity Verification	Evernym, Sovrin (Hyperledger Indy)	Evernym: Decentralized identity solutions Sovrin (Hyperledger Indy): Permissioned blockchain for digital identity	Secure and verifiable digital identities, improving privacy and control over personal information	Privacy, Security, User Control	Adoption Barriers, Interoperability, Regulatory Compliance
Asset Tokenization	, Polymath (POLY) Securitize	Tokenization platform Polymath (POLY): Security token issuance platform	Representing traditional assets (e.g., real estate, company shares) as digital tokens, facilitating fractional ownership and trading	Liquidity, Fractional Ownership, Accessibility	Regulatory Compliance, Market Maturity, Security Issues
AML and KYC Compliance	JPMorgan Chase, Chainalysis, Elliptic	JPMorgan Chase: Exploring blockchain for AML and KYC Chainalysis, Elliptic: Blockchain analytics for AML and compliance	Using Blockchain to improve the efficiency and accuracy of KYC and AML procedures	Efficiency, Transparency, Security	Privacy Concerns, Data Accuracy, Regulatory Frameworks

Based on the insights captured from the table, transactions across the ledgers are prioritised through bitcoin and ripple, Ethereum signifies smart contracts, IBM blockchain targets the enterprise solutions, WeTrade streamlined trading monetary aspects, Evernym emphasizes on self-sovereign identity, and Polymath specializes in security tokenization.

Each of them employs unique consensus mechanisms, security aspects and asset management across the blockchain network.

On the contrary, bitcoin uses PoW, Ripple uses a unique consensus algorithm and Ethereum transitions to PoS, WeTrade builds and hyper ledger fabric and Ethereum gets leveraged by Polymath.

These blockchain technologies diversify the needs depending upon the services they are pioneered for, they reflects varying goals in privacy, decentralisation and functioning across the network.[3]

Blockchain technology can also affect the ledger maintaining system (AIS or LMS) [8] by digitising research-studied processes, securely storing accounting data, and improving transaction efficiency. It simplifies operations, reduces settlement time, minimises fraud, and enhances regulation and liquidity.

In summary, the adoption of blockchain in banking and finance has led to significant changes in the industry, with both advantages and challenges, and its trend appears to be positive and increasing since 2008. However, there is a need to address potential limitations and Threats.

Transforming the technically advanced banking system was not easy without facing obstacles, blockchain technology to be trusted and primitive adjustments to be made for smooth involvement. [10]

III. CHALLENGES IN BLOCKCHAIN APPLICATIONS

The FinTech sector face challenges in adopting Blockchain including the uncertainties in regulating compliances, scalability issues, interoperability concerns, security and data privacy also imposes obstacles, and in order to overcome these hurdles collaborative efforts, regulatory clarity and tech advancements are required.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:14:02 UTC from IEEE Xplore. Restrictions apply.

A. Challenges associated with using UPI and implementation of Blockchain

The culture of the financial world depends on centralized intermediaries, but blockchain broke the tradition by allowing a decentralized substitute. It helps in the efficient and transaction costs reduction. Centralized intermediaries boosted the unauthenticity, due to cybersecurity scams, and even steven in finances[3].

Traditional banking processes often involve complex and expensive intermediaries, resulting in significant cost inefficiencies. However, with the integration of UPI and blockchain technology, these inefficiencies can be minimized or even eliminated.

Blockchain eliminates the need for intermediaries by providing a decentralized and trustless network. This reduces the transaction fees and operational costs associated with traditional banking systems. Additionally, UPI enables instant and seamless transactions, further optimizing the efficiency of the banking sector.[13]

B. Regulatory Compliances

Regarding the banking sector, regulatory compliance is of utmost importance. Integrating UPI (Unified Payments Interface) with blockchain technology ensures enhanced regulatory compliance. Blockchain technology is transforming several industries, and the banking industry is no exception. The combination of blockchain technology and artificial intelligence (AI) is one of the major developments in this sector. Combining the capabilities of these two cutting-edge technologies has greatly benefited the banking industry.

Moreover, blockchain technology's automation and smart contract capabilities free up valuable resources and reduce manual errors. This not only saves time and effort but also decreases costs that were previously incurred due to human intervention. Integrating UPI with blockchain brings significant cost savings and operational efficiencies to the banking sector.[14]

Fundamental Banks having a feeling of threat when thinking of blockchain, some competed with markets using it. Fragile investments compressed in information technology make for cost-efficient and innovative prompts.

Transforming the technically advanced banking system was not easy without facing obstacles, blockchain technology to be trusted and primitive adjustments to be made for smooth involvement.

The risk factor and stabilisation made positive gains with this blockchain technology. Some information states vice-versa. Despite the benefits, there are acknowledged limits and potential threats associated with blockchain adoption in banking, such as scalability, security, energy consumption, and privacy concerns[9].

In summary, the adoption of blockchain in banking and finance has led to significant changes in the industry, with both advantages and challenges and its trend appears to be positive and increasing since 2008. However, there is a need to address potential limitations and threats.

IV. AI IN BLOCKCHAIN TECHNOLOGY

Blockchain technology is transforming several industries, and the banking industry is no exception. The combination of blockchain technology and artificial intelligence (AI) is one of the major developments in this sector. Combining the capabilities of these two cutting-edge technologies has greatly benefited the banking industry. AI in blockchain introduces intelligent algorithms that enhance the efficiency and security of transactions. It enables automated verification processes and improves fraud detection, ultimately contributing to a more reliable and transparent banking system. By leveraging AI, banks can streamline operations and provide customers with a seamless and secure banking experience.[11]

AI has significantly transformed the landscape of blockchain technology. It enhances security through real-time fraud detection and identity verification, improving transaction trust. Smart contracts are optimised and automated, and supply chain management benefits from AI's data analysis capabilities. Decentralised Autonomous Organizations (DAOs) are more efficient with AI-driven decision-making, and content filtering is enhanced. The tokenisation of assets is streamlined, and AI aids in market prediction and trading. AI-driven data analysis on the blockchain yields valuable insights. While it fosters innovation, AI integration poses challenges, like concerns about data privacy, job displacement, and ethical considerations, requiring careful navigation in the blockchain space.

Implementing Artificial Intelligence (AI) involves integrating advanced algorithms and machine learning into various sectors, offering automation, data analysis, personalisation, and predictive capabilities. There could be a lot of advantages of it, and at the same time disadvantages as well such as it is responsible for the replacement of human jobs, thus increasing unemployment, data security concerns, ethical concerns, etc. But there might be technical complexity alongwith the lack of desired resources and capital to implement it. A successful AI driven system implemented blockchain technology has to be cost efficient to be feasible and overcome it’s possible drawbacks.

V. AI DRIVEN CHATBOTS

Deep learning (DL), machine learning (ML), and natural language processing (NLP) are some of the methods used to create AI-driven chatbots. The procedures that are followed to develop it are as follows:

Collection of Data: Firstly the data is collected and it has to be preprocessed to train the chatbot to meet our requirement.

Implementing Machine Learning: ML algorithms can train the chatbot on the data that has been collected and preprocessed. One of the method is through supervised learning in which the chatbot gets trained on the labeled data, meanwhile in unsupervised learning the chatbot is trained on the clustering and topic modeling.[16]

Dialog Management: These techniques are implanted in order to manage the conversation. It involves interpreting the user’s conversation and generating possible responses.

Integration with Blockchain: After the development of the chatbot, it gets integrated with blockchain technology. This integration can be done in various ways, depending on the use case.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:14:02 UTC from IEEE Xplore. Restrictions apply.

Secure Data Transmission and Private Environment: Implementing the Blockchain Fintech thus enhances the security of the data used to drive the chatbot, through various cryptographic algorithms used for the its encryption.

Continuous Improvement: These AI-driven chatbots ain't robust and thus it can be further enhanced through user feedbacks (reinforcement learning) thus it would improve it's user experience over the course of time.

There could be various purposes where we can use these AI-driven chatbots in blockchain technology-

Customer Support: It can be used to provide a one-on-one automated user interaction, thus providing a 24/7 facility to the user, thus decreasing the man power needs for the support system.

Transaction Processing: They can coordinate across various AI-logs available through the simultaneous network, thus emphasizing the transactions and can be used to trade assets.[19]

Identity Verification: It can be used to provide a user authentication system, allowing only the legitimate user to fetch the info, whose info is being fetched from the blockchain network which is highly encrypted on a distributive network.

Analytics: It can be used to analyze the data provided by the users and provide them with the relevant data as per the user demands.

Overall, AI-driven chatbots can be efficiently used for a better user experience, which is automated and thus cost efficient.

VI. FUTURE RESEARCH THROUGH ANALYSING BLOCKCHAIN MARKET TRENDS THROUGH AI

Security algorithm techniques can be potentially ultimate in the capital storage sectors, and there is still much ground to be covered. As we look towards the future, several areas of development and improvement can be anticipated.

- Scalability: Blockchain networks must address scalability concerns, especially when handling a massive number of transactions. Changes are constant in the tech used to improve higher-scale usage, such as sharding and layer-two solutions.
- Interoperability: Networks connected through these algorithms and blocks should be focused. Banks are exploring ways to ensure seamless communication and data sharing across multiple blockchain platforms, enabling efficient interbank transactions and collaborations.
- Security Enhancements: While blockchain is already highly secure, continuous efforts are being made to enhance its security features further. Banks are investing in cutting-edge encryption algorithms and advanced consensus mechanisms to fortify the security of their blockchain systems.
- Integration with IoT: The integration of blockchain and the IoT could be significantly proven to be beneficial for the banking sector. Enabling secure and transparent smart contract-based transactions, automate supply chain finance, and enhance data security in IoT devices.[20]

In conclusion, the above mentioned points have been proven significant in future advancements, for example the integration of UPI with blockchain technology has brought significant advancements to the banking sector. The utilization of AI in blockchain enhances efficiency and ensures regulatory compliance. The cost inefficiencies oftraditional banking systems are minimized, and the future holds even more exciting possibilities for blockchain in the banking industry. As we continue to explore the potential of blockchain technology, the banking sector is poised for a transformative journey towards a more secure, efficient, and customer-centric landscape.

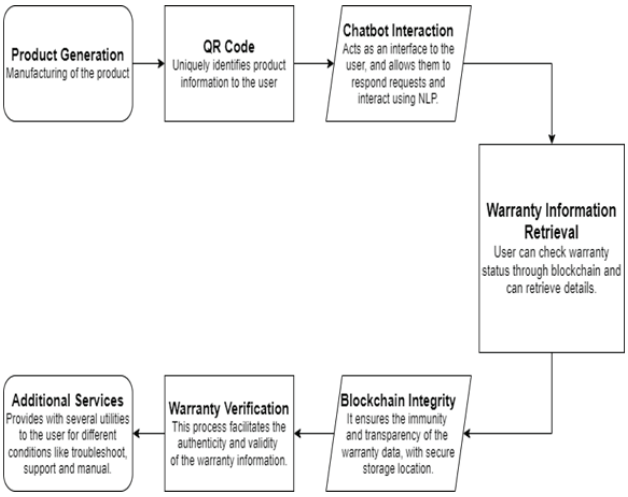


Fig. 1. This flowchart represents the sequential processes on which an AI generated chatbot works

Analyzing market trends in the context of blockchain technology with the integration of Artificial Intelligence (AI) is an evolving and highly promising field that offers both insights and challenges.

A. Current State --

Blockchain technology is responsible for revolutionizing various sectors including trading platforms, cryptography, finance, etc. The blockchain market is characterized by plethora of cryptocurrencies, decentralized applications (dApps), and diverse blockchain platforms. AI is increasingly employed to decipher complex market.

B. AI-driven tools are used for-

- Sentiment Analysis: NLP algorithms scan news, social media, and community forums to gauge market sentiment, helping investors make informed decisions.
- Predictive Analytics: Machine learning models are applied to historical data to predict token prices, market volatility, and trading patterns.
- Fraud Detection: AI assists in identifying fraudulent transactions and protecting the integrity of blockchain networks.[12]
- Smart Contract Auditing: Automated audits of smart contracts using AI to uncover vulnerabilities and ensure their reliability.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:14:02 UTC from IEEE Xplore. Restrictions apply.

- Consistency & Loss-less Flow: AI streamlines compliance processes and helps in tracking and adhering to evolving cryptocurrency regulations.
- Market Risk Management: Further exploration of AI's role in risk management within the blockchain market, considering factors like network upgrades and forks.
- Sustainability and Green Blockchain: Investigating AI tools for assessing the environmental impact of blockchain networks and optimizing their energy consumption.
- Regulatory Technology (RegTech): AI-powered RegTech solutions to assist compliance with evolving blockchain-related regulations.

Blockchain evolving on outskirts, AI and this technology better understand the market trends and significantly increase the growth and efficient transformation. High speed upgradation and opportunities coming with unique challenges, building future updates.



Fig. 2. In the above Figure an AI generated indicator is being used to trade in cryptocurrencies

C. *Future Research Directions* --

- AI in Decentralized Finance (DeFi): Investigating the role of AI in assessing risk and improving yield optimization in the growing DeFi space, as well as addressing security concerns.
- Interoperability and Cross-Chain Analysis: Research can delve into AI solutions that facilitate interoperability between diverse blockchains and assess the broader implications of such integration on the market.
- Security and Privacy: Development of AI-based security solutions for protecting blockchain networks and enhancing privacy features.[14]

VII. CONCLUSION

This elicited research work providing an examined report of the multidimensional role of Blockchain in fin-tech sector. Blockchain impacts applications, laws, intermediation, future prospects, and other range of topics in this sector. Diving in the core of the blockchain, consisting of consensus algorithms and smart contracts, and other platforms are covered in the paper.

Furthermore, the paper emphasizes on the transitions of blockchain on accounts and audits, explaining blockchain technology enhances data transparency, auditability, and trusted access in financial data.

Artificial Intelligence combined with Blockchain for the betterment of ecosystem, projecting the collaborative capability of it.

Research also shows how the blockchain made the amendment in the culture of financial institutions, switching from centralized to decentralized intermediaries. Subsequently affecting business sectors and financial models.

Furthermore, this written work highlighted the crypto algorithmic trading world, facing different challenges and strategies to endure option trading and crypto space. Mining cryptocurrency along with the Blockchain technology and sustainable ecology of the financial accounts.

Blockchain technology on integrating with financial systems, leads to everyday evolution process. Regular reports being made for academic year and enthusiastic professionals moving towards understanding the financial sector with blockchain factor. For the future approach, influence and opportunities developing predictive models, AI impacting on decentralized assets, high-scale problems, algorithms helping in the development of Blockchain.

Financial instrumentation in upgrading the potentially collaborative blockchain approach in the further research.

REFERENCES

[1] Demirguc-Kunt, A.; Klapper, L.; Singer, D.; Ansar, S.; Hess, J. The Global Findex Database 2017. In Measuring Financial Inclusion and the FinTech Revolution; World Bank: Washington, DC, USA, 2018. ISBN 1464812594

[2] Gartner. The CIO's guide to blockchain. 2022. Available online: https://www.gartner.com/smarterwithgartner/the-cios-guideto-blockchain (accessed on September 2022).

[3] Romānova, I.; Kudinska, M., Banking and Fintech: A challenge or opportunity? In Contemporary Issues in Finance: Current Challenges from Across Europe; Emerald Group Publishing Limited: Bingley, UK, 2016; ISBN 1786359070.

[4] Ng, E.; Tan, B.; Sun, Y.; Meng, T. The strategic options of fintech platforms.

[5] Md.Javaid, A.Haleem, RaviPSingh, Rajiv Sumanand Shahbaz Khan,"A review of Blockchain Technology applications for financial services" Institute of Business Management, GLA University, 2022

[6] Yi Cao & Jia Zhai (2022),"A survey of AI in finance", Journal of Chinese Economic and Business Studies, 20:2, 125-137, DOI:10.1080/14765284.2022.2077632

[7] Hongdan Han, Radha K. Shiwakoti, Robin Jarvis, Chima Mordi, David Botchie, "Accounting and auditing with blockchain technology and artificial Intelligence: A literature review", Brunel Business School, Brunel University London, UB8 3PH, United Kingdom, 2023

[8] RashikalaWeerawarna, Shah J. Miah, Xuefeng Shao, "Emerging advances of blockchain technology in finance: a content analysis", Newcastle Business School, University of Newcastle, Newcastle, NSW, Australia, 2023

[9] Eyal, I.; Gencer, A.E.; Sirer, E.G.; Renesse, R.V., {Bitcoin-NG}: A scalable blockchain protocol. In Proceedings of the 13th USENIX symposium on networked systems design and implementation (NSDI 16), Santa Clara, CA, USA, 16–18 March 2016; pp. 45–59; ISBN 978-1-931971-29-4.

[10] Ritesh Patel, Milena Migliavacca and Marco E. Oriani, "Blockchain in banking and finance: A bibliometric review", Research in International Business and Finance 62 (2022), 101718

[11] B. Gami, M. Agrawal, D. K. Mishra, D. Quasim, and P. S. Mehra, "Artificial intelligence-based blockchain solutions for intelligent healthcare: A comprehensive review on privacy preserving techniques," Transactions on Emerging Telecommunications Technologies, vol. 34, no. 9, p. e4824, Sep. 2023, doi: 10.1002/ETT.4824

[12] Bandara, E.; Shetty, S.; Mukkamala, R.; Liang, X.; Foytik, P.; Ranasinghe, N.; Zoysa, K. Casper: A blockchain-based system for

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:14:02 UTC from IEEE Xplore. Restrictions apply.

efficient and secure customer credential verification. J. Bank. Finance. Technol. 2022, 6, 43–62.

[13] Malhotra D, Saini P, Singh AK (2021) How blockchain can automate KYC: Systematic review. Wirel Pers Commun 122:1987–2021

[14] C. Garg, D. Kumar Mishra, D. Raj, and P. Singh Mehra, "A Survey on Integration of Blockchain and IoT (BIoT): Open Issues, Challenges & Solution," Proceedings of the 2022 3rd International Conference on Intelligent Computing, Instrumentation and Control Technologies: Computational Intelligence for Smart Systems, ICICICT 2022, pp. 880–886, 2022, doi: 10.1109/ICICICT54557.2022.9917632.

[15] Hau, H., Y. Huang, H. Shan, and Z. Sheng, 2018. "Fintech Credit, Financial Inclusion and Entrepreneurial Growth". Unpublished working paper.

[16] Bianchi, D., M. Büchner, A. Tamoni, and S. Van Nieuwerburgh. 2021. "Bond Risk Premiums with Machine Learning." The Review of Financial Studies 34 (2): 1046–1089. doi:10.1093/rfs/hhaa062

[17] D. Kumar, D. Kashyap, D. Kumar Mishra, and P. Singh Mehra, "Integrating Blockchain and 5G: Advantages and Challenges of Decentralization," Proceedings of the 2022 3rd International Conference on Intelligent Computing, Instrumentation and Control Technologies: Computational Intelligence for Smart Systems, ICICICT 2022, pp. 866–873, 2022, doi: 10.1109/ICICICT54557.2022.9917645.

[18] Guo, Y.; Liang, C. Blockchain application and outlook in the banking industry. Financ. Innov. 2016, 2, 1–12. https://doi.org/10.1186/s40854-016-0034-9.

[19] A. Bansal, Arju, Esha, and P. S. Mehra, "A Post-Quantum Consortium Blockchain Based Secure EHR Framework," 2023 International Conference on IoT, Communication and Automation Technology (ICICAT), pp. 1–6, Jun. 2023, doi: 10.1109/ICICAT57735.2023.10263717.

[20] Gomber, P.; Koch, J.-A.; Siering, M. Digital Finance and FinTech: Current research and future research directions. J. Bus. Econ. 2017, 87, 537–580.

[21] R. Chanana, A. K. Singh, R. Killa, S. Agarwal, and P. S. Mehra, "Blockchain Based Secure Model for Sensor Data in Wireless Sensor Network," in 2020 6th International Conference on Signal Processing and Communication (ICSC), IEEE, Mar. 2020, pp. 288–293. doi: 10.1109/ICSC48311.2020.9182776.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:14:02 UTC from IEEE Xplore. Restrictions apply.

2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA) | 979-8-3503-4922-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/AIMLA59606.2024.10531571

A Survey on Skin Lesion Detection and Classification using Machine Learning

1st Rashmi Yadav
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
rashmiyadav.1512@gmail.com

2nd Dr. Aruna Bhat
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
aruna.bhat@dtu.ac.in

The research explores how dermatologists use machine learning to quickly and accurately identify and classify skin injury. Conventional diagnosis techniques depend on visual examination, but are subjective and have different interpretations. The ability to analyze data and recognize patterns is a possible remedy for ML. The paper examines the current technology of machine learning, focusing on validation in the real world, and deals with data set variability. Although the research acknowledges the fruitfulness of deep learning (DL), the research emphasizes the benefits of traditional ML techniques regarding interpretation and processing performance. Methods for automating the analysis of skin lesions, such as feature engineering, rule-based techniques and traditional ML algorithms, have been studied. The study suggests using advanced transfer learning techniques, integrating genetic and clinical data, and refining the way artificial intelligence (AI) is explained in order to get over barriers. Intending to enhance the accessibility and correctness of skin lesion identification, the future requires collaboration between dermatology and machine learning to develop real-time diagnostic tools. By offering scalable solutions for rapid diagnosis of lesions and improved patient outcomes, this combination of medical expertise and ML capabilities has the power to revolutionize dermatology. The future of automated dermatological diagnosis is expected to be shaped by collaboration between machine learning experts and dermatologists, enabling more personalized treatment for patients.

Keywords— skin lesion, machine learning, survey, deep learning, CNN

I. INTRODUCTION

Skin injuries, from benign to malignant, are an essential component of dermatology diagnosis [3]. Accurate and prompt identification is necessary for the timely recognition and effective medicaments of skin lesions. To successfully cured some incurable diseases like skin cancer, melanoma can be cured only at primary stage [10]. The recent advancement in imaging technology have enabled high resolution skin images which gives opportunity to machine learning algorithms to step in dermatology. This work is driven by a growing tendency in the medical field to apply ML for skin lesion identification and categorization in order to enhance diagnosis. Traditional diagnosis techniques are time consuming and slow to changes in human capabilities, which motivates researchers to explore automated, effective, and unbiased evaluations by machine learning [26].

In the decade, the machine learning algorithms becomes popular and also they help in improving diagnosis [16]. Dermatologists mainly diagnose skin infections by visual examination in traditional procedures, which can lead to a variety of diagnoses and subjective interpretations. By analyzing huge amounts of data and analyzing complex patterns of identity, ML provides creative solutions to these problems [19]. High-quality dermatological imaging data sets have exponentially expanded, and the development of machine learning methods has allowed for robust and accurate models. Through methods such as DL, Convolutional Neural Networks (CNNs) [21], and learning transfer, scientists have been able to automatically recognize and catergorize skin disorders with remarkable effectiveness.

In this work, the advanced with traditional approaches for machine learning based on skin lesion identification and categorization are reviewed. It explores the traditional machine learning techniques with the recent advancement, publicly available datasets with evaluation metrics and at the end future scope and challenges are discussed.

The paper also addresses how machine learning models in dermatological applications could perform better if other modalities, such as dermoscopy and clinical data, were integrated. The integration of ML skills with medical knowledge has the possibility to transform dermatological practices as we traverse the intricacies of skin lesion detection. Machine learning has more possibilities to influence the healthcare industry by providing a scalable, dependable, and efficient solution that will help with early identification, better patient outcomes, and more efficient use of resources [9] [17].

A. Background

Skin disorders, which include a wide range of lesions and dermatological illnesses, are a major global health concern. Skin disorders are becoming more common, which highlights the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:41:01 UTC from IEEE Xplore. Restrictions apply.

critical need for advanced and effective diagnostic tools. The machine learning models' proven accuracy and consistency in image analysis, along with their ability to manage the increasing amount of skin image datasets, serve as further sources of incentive. Beyond just increasing productivity, the survey aims to further dermatological personalized care by customizing treatment regimens based on unique skin features. This survey attempts to shed light on the traditional as well as advanced machine learning-based skin lesion identification by combining current research trends, methodology, and findings. By doing so, it hopes to uncover gaps in the field and open up new avenues for future research that will directly affect patient care.

B. Motivation

First of all, this survey gives an encyclopedic overview of the current state-of-art methodologies which helps researchers, practitioners and dermatologists with the understanding of current research and also future advancement of the field. This survey serves as a guide for future research projects by highlighting current challenges and limitations and directing efforts towards the area of innovation and improvement. The comparative analysis of various machine learning models allows us to evaluate their strengths and weaknesses. In addition, this research helps to explore the recent practical applications of machine learning in skin lesion detection. Analysing the previous studies aids in the understanding of data diversity and quality with ethical consideration. This survey serves as a road map, shaping the direction of future research and providing a brief yet significant view of the problem.

II. Literature Survey

Skin lesion recognition and categorization has made substantial use of ML techniques and DL techniques in our study. For skin lesion identification and categorization usually, researchers go for ML and DL techniques, but there are other methods like Rule based approach like ABCD method [2], CASH method [27], three-point checklist [25], seven-point checklist [27] etc popular. The dependency of choosing any technique will dependent on the size of the dataset. In this work, we are focusing on ML techniques as they provide more interpretable results and ease of implementation compared to DL models. Additionally, these models perform better on smaller datasets, are computationally lighter and are less prone to overfitting when the data is limited. ML model involves manual feature engineering which is its limitation, whereas DL model extracts more features which results in better accuracy. Many strategies have been used, from conventional techniques to sophisticated deep learning architectures. The following are a few popular machine-learning methods for analyzing skin lesions:

A. Traditional Image Processing [22]

Color-based segmentation is a popular technique that divides skin lesions from the surrounding tissue according to unique color features. This technique helps isolate lesions by using color differences to identify regions of interest. Another important component is texture analysis, which examines the patterns of skin lesions to identify variations that may be suggestive of particular illnesses.

B. Classical Machine Learning Algorithms

The broad range of methods known as classical machine learning algorithms has been used for several purposes, such as the identification and categorization of skin lesions. Several important approaches are covered by classical ML algorithms in the context of skin lesions:

- Support Vector Machine (SVM) [16]: For binary classification, commonly engrossed for benign and malignant tumours [7] from the skin lesion, SVM is popular.
- Random Forest (RF) and Decision Tree (DT) [17] [20]: For feature selection and classification, ensemble techniques like random Forests and decision trees are used. They are flexible enough to accommodate a wide range of data kinds since they can handle both continuous and categorical information.
- K-Nearest Neighbour(KNN) [9]: In KNN technique uses closeness to classify data items according to the class their neighbours belong to. KNN can be used in skin lesion analysis to find commonalities between lesions to classify them.
- Naïve Bayes [28] : The naïve Bayes classifiers is based on using the Bayes theorem with strong assumptions about feature independence. They are easy to use and effective, particular for smaller datasets.

C. Feature Engineering [3]

It is taking statistical measurements like mean and variance or pertinent aspects like size and shape descriptors and extracting them from the photos. To differentiate benign and malignant tumours, these characteristics are essential.

D. Deep Learning Architectures

For identification and categorization, CNNs [21] shows one of the most notable DL architectures that have transformed the skin lesions. CNNs easily identify structural features from dermatological images to achieve astonishing results. In this process pretrained model through transfer learning like VGG16 [21] and ResNet [10] are included. Recurrent neural networks (RNNs) [13] are used to ensnare time related dependencies in skin lesion patterns, for sequential data processing. Interpretability is enhanced by attention mechanisms like self-attention models that focus on specific or relevant areas of images. One of the popular techniques can be used for data-augmentation is GANs which will be discussed briefly in the next section. However, this will help significantly improve the productivity and precision for diagnosis.

E. Autoencoders [26]

For the study of skin lesion, autoencoders are a key element of deep learning architectures. They are experts in feature extractions and unsupervised learning by flatting the regions of dermatological images. During encoding process, autoencoders recreate the input images by dimension and noise reduction. Additionally, they may be used for data augmentation, creating synthetic data that helps to make model

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:41:01 UTC from IEEE Xplore. Restrictions apply.

more robust. Autoencoders are essential as they provide a productive means of extracting meaningful representations from a complicated image for skin lesion identification and categorization in a automated dermatological diagnosis.

F. Generative Adversarial Networks (GANs) [19]

GANs are highly useful for skin lesion identification and classification, when data augmentation is used. A discriminator along with generator network makeup a GAN. A discriminator network and a generator network work together to produce artificial images much like real skin lesion. GANs help dermatology by improving generalization to previously unexplored data, enhancing model training, and expanding the scope of constrained datasets. GANs deal with the problem of insufficiently labeled data by creating realistic images of artificial skin lesions. The advancement of technology significantly boosts the robustness and efficacy of algorithms for recognition and categorization of skin lesions. The ability of GANs to create realistic images similar to actual skin lesions shows that GANs can be an important utility in the creation of automated dermatological diagnostic models.

III. DATA AND EVALUATION METRICS

To further the field of skin lesion identification and categorization, it is essential to have access to a variety of carefully selected datasets. To promote research collaborations, test machine learning algorithms, and ultimately aid in the development of reliable diagnosis models for dermatological applications, several publicly available datasets have become essential. Some of them are listed in Table 1. Various measures are employed to evaluate the models' performance and determine whether they are performing better or not. We refer to these measurements as performance metrics. Some common and significant outcome measurements, which vary depending on the kind of model, are covered in the section below [8] [14].

- Accuracy: The performance of the model is evaluated based on its accuracy. It is calculated as the proportion of all occurrences to the all instances. It is represented as

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

- Precision: A model's precision indicates how accurate its positive predicts are, It is defined as the proportion of actual positive predictions to all positive predictions the model produced.

$$\frac{TP}{TP+FP} \quad (2)$$

- Sensitivity (Recall): The recall of a classification model measures how well it can find every relevant instance in a dataset. As a percentage of all false negative and true positive forecasts, it represents the percentage of true positive cases or predications.

$$\frac{TP}{TP+FN} \quad (3)$$

- Specificity: Specificity play a crucial statistic for assessing classification for binary classification. This assesses a model's capacity to accurately recognize negative examples.

$$\frac{TN}{TN+FP} \quad (4)$$

- F1 Score: The in-general performance is assessed using F1-score for the classification model. It's the precision and recall harmonic mean.

$$2 * \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

$$(TP: True\ Positive, TN: True\ Negative,$$
$$FP: False\ postive, FN: False\ Negative)$$

TABLE 1. LIST OF AVAILABLE DATASETS [29]

Name of the Dataset	Year	Number of Images	Image Dimensions	Image Format	Disease(s) present
DermIs [4]	Not Available	69	2067*1555	.jpg	Melanoma, Nevusor or Atypical nervus
DermNet [5]	1998	23000	540*722 to 4499*6748	.bmp	23 distinct types
PH2 [18]	2013	200	766*576	.bmp and .jpg	Melanoma, Nevusor or Atypical nervus, Common Nevus

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:41:01 UTC from IEEE Xplore. Restrictions apply.

ISIC Archives [11]	2016 to 2020	2094 to 33,126	variable	.jpg	Melanoma, Nevusor or Atypical nervus, Actinic Keratosis, Seborrheic Keratosis, Basal Cell Carcinoma
DermoFit (Paid) [6]	Not Available	1300	380*380	.jpg	Melanoma, Nevusor or Atypical nervus, Seborrheic Keratosis, Basal Cell Carcinoma, Actinic Keratosis
EDRA [1]	2000	1011	768*512	.jpg	Melanoma, Seborrheic Keratosis, Nevusor or Atypical nervus, Common Nevus, Basal Cell Carcinoma
Skin Cancer: Malignant Vs Benign [7]	2018	1800	224*244	.jpeg	Benign and Malignant skin mole
HAM10000 (Human Against machine) [15]	2018	10015	400*600	.jpg	Basal Cell Carcinoma, Benign Keratosis like Lesions, Actinic Keratosis, Melanoma, Common Nevus, Dermatofibroma, Vascular Lesions
ISBI 2016 [23]	2016	900	768*500	.jpg	Common Nevus, Seborrheic Keratosis, Melanoma
ISBI 2017 [12]	2017	2000	4500*6000	.jpg	Common Nevus, Seborrheic Keratosis, Melanoma
ISBI 2018 [24]	2018	2594	4500*6000	.jpg	Dermatofibroma, Vascular Lesions, Melanoma Actinic Keratosis, Basal Cell Carcinoma, Benign Keratosis like Lesions,

IV. Challenges and Future Scope

Numerous obstacles affect the effectiveness and usability of the field of machine learning-based skin lesion identification and categorization. Training models that generalize across different skin conditions are hampered by limited access to balanced and diverse datasets. Even using DL algorithms, it is still a challenging task to extract information from the images of skin lesions that may utlized for clinical interpretation . Robust algorithm development is further complicated by the clinical heterogeneity in skin types, lighting settings, and presentation styles. To achieve responsible implementation in clinical settings, ethical factors like as biases in algorithm predictions and patient privacy need to be carefully considered.

To improve accuracy, this field's future scope calls on the integration of several modalities, including genetic and clinical data. The interpretability of models will increase with the development of explainable AI techniques, and feature learning can be further enhanced by continuing research into transfer learning strategies. The future landscape will need to include real-time diagnostic tools, online learning tactics, and joint machine learning and dermatologist efforts to overcome obstacles and improve skin lesion detection technologies' usability, precision, and global impact.

V. Conclusion

In conclusion, this research study clarifies the significant progress made in the use of ML in the field of dermatology to identify and categorize skin lesions. This work explores the synergistic interaction between advanced machine learning techniques and high-quality dermatological datasets, realizing the critical importance of early detection of cutaneous abnormalities. Although the study duly notes the achievements of DL, also highlights the advantages of traditional ML algorithms in terms of interpretability, processing performance, with simplified application, especially in situations where datasets are limited. Various techniques, including rule-based strategies, traditional algorithms such as RF and SVM, and feature engineering, were examined for their effectiveness in automating the analysis of skin lesions.

Looking ahead, this field's direction anticipates overcoming obstacles like the need for different datasets and ethical problems. It suggests exploring advanced transfer learning algorithms, improving explainable AI techniques, and integrating genetic and clinical data. The development of real-time diagnostic instruments and joint efforts between the fields of dermatology and machine learning are anticipated to be essential elements in enhancing the accuracy, availability, and worldwide influence of skin lesion detection technologies. Considering skin conditions diagnostics, combining medical knowledge with machine learning expertize has the potential to completely transform dermatological practices by offering scalable and effective solutions for the quick identification of lesions, better patient outcomes, and economical use of resources. The field of individualized patient care is expected to benefit greatly from the joint efforts of machine learning researchers and dermatologists, as they are anticipated to play an essential role in defining the future of automated dermatological diagnostics.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:41:01 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] Giuseppe Argenziano et al. “Interactive atlas of dermoscopy”. In: (2000).

[2] Saptarshi Chatterjee et al. “Dermatological expert system implementing the ABCD rule of dermoscopy for skin disease identification”. In: Expert Systems with Applications 167 (2021), p. 114204..

[3] Priya Choudhary, Jyoti Singhai, and JS Yadav. “Skin lesion detection based on deep neural networks”. In: Chemometrics and Intelligent Laboratory Systems 230 (2022), p. 104659.

[4] Dermatology Information System. https://www. dermis . net / dermisroot / en / home / index . htm/. DermIS Dataset, Accessed November 29, 2023.

[5] DermNet-All about Skin. https : / / dermnetnz . org/. DermNet Dataset, Accessed November 29, 2023. 1998.

[6] DERMOFIT Project Datase. https : / / homepages . inf . ed . ac . uk / rbf / DERMOFIT / datasets . htm. DermoFit Dataset, Accessed November 29, 2023.

[7] CLAUDIO FANCONI. Skin Cancer: Malignant Vs Benign. https : / / www. kaggle . com / datasets/fanconic/skin- cancer- malignant- vs-benign. Skin Cancer, Accessed November 29, 2023. 2018.

[8] Evgin Goceri. “Diagnosis of skin diseases in the era of deep learning and mobile technology”. In: Computers in Biology and Medicine 134 (2021), p. 104458.

[9] Mustafa Qays Hatem. “Skin lesion classification system using a K-nearest neighbor algorithm”. In: Visual Computing for Industry, Biomedicine, and Art 5.1 (2022), pp. 1–10.

[10] Syed Inthiyaz et al. “Skin disease detection using deep learning”. In: Advances in Engineering Software 175 (2023), p. 103361.

[11] ISIC Challenge. https://challenge.isic-archive. com/. ISIC 2016 to 2020 Dataset, Accessed November 29, 2023. 2016.

[12] Muhammad Attique Khan et al. “Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework”. In: Pattern Recognition Letters 143 (2021), pp. 58–66.

[13] Muhammad Attique Khan et al. “Pixels to classes: intelligent learning framework for multiclass skin lesion localization and classification”. In: Computers & Electrical Engineering 90 (2021), p. 106956.

[14] Ling-Fang Li et al. “Deep learning in skin disease image recognition: A review”. In: IEEE Access 8 (2020), pp. 208264–208280.

[15] K SCOTT MADER. Skin Cancer MNIST:HAM10000. https://www.kaggle.com/ datasets/kmader/skin-cancer-mnist-ham10000. Skin Cancer, Accessed November 29, 2023. 2018.

[16] Sarmad Maqsood and Robertas Damaseviˇciusˇ “Multiclass skin lesion localization and classification using deep learning based features fusion and selection framework for smart healthcare”. In: Neural Networks 160 (2023), pp. 238–258.

[17] K Meena et al. “A novel method for prediction of skin disease through supervised classification techniques”. In: Soft Computing 26.19 (2022), pp. 10527–10533.

[18] Teresa Mendonc¸a et al. “PH 2-A dermoscopic image database for research and benchmarking”. In: 2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE. 2013, pp. 5437–5440.

[19] Sufiyan Bashir Mukadam and Hemprasad Yashwant Patil. “Skin Cancer Classification Framework Using Enhanced Super Resolution Generative Adversarial Network and Custom Convolutional Neural Network”. In: Applied Sciences 13.2 (2023), p. 1210.

[20] A Murugan et al. “Diagnosis of skin cancer using machine learning techniques”. In: Microprocessors and Microsystems 81 (2021), p. 103727.

[21] Ahmad Naeem et al. “SCDNet: A Deep Learning-Based Framework for the Multiclassification of Skin Cancer Using Dermoscopy Images”. In: Sensors 22.15 (2022), p. 5652..

[22] Şaban Öztürk and Umut Özkaya. “Skin lesion segmentation with improved convolutional neural network”. In: Journal of digital imaging 33 (2020), pp. 958–970.

[23] Nurullah Şahin, Nuh Alpaslan, and Davut Hanbay. “Robust optimization of SegNet hyperparameters for skin lesion segmentation”. In: Multimedia Tools and Applications 81.25 (2022), pp. 36031–36051.

[24] Pufang Shan et al. “Automatic skin lesion classification using a new densely connected convolutional network with an SF module”. In: Medical & Biological Engineering & Computing 60.8 (2022), pp. 2173–2188.

[25] P Varalakshmi et al. “Enhanced dermatoscopic skin lesion classification using machine learning techniques”. In: 2021 sixth international conference on wireless communications, signal processing and networking (WiSPNET). IEEE. 2021, pp. 68–71.

[26] Chao Xin et al. “An improved transformer network for skin cancer classification”. In: Computers in Biology and Medicine 149 (2022), p. 105939.

[27] Zhen Yu et al. “Convolutional descriptors aggregation via cross-net for skin lesion recognition”. In: Applied Soft Computing 92 (2020), p. 106281.

[28] Balaji, V. R., et al. "Skin disease detection and segmentation using dynamic graph cut algorithm and classification through Naive Bayes classifier." In: Measurement 163 (2020): 107922.

[29] Yadav, R et al. “A systematic literature survey on skin disease detection and classification using machine learning and deep learning”. In: Multimedia Tools and Applications, Springer (2024), pp. 1-32.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:41:01 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Engineering Management. 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/TEM.2024.3400401

TEM-23-0876.R3

Adoption of Metaverse in Healthcare Sector: An Empirical Analysis of its Enablers

Vipulesh Shardeo, Bishal Dey Sarkar, Umar Bashir Mir, Prerna Kaushik

***Abstract*— The emergence of novel innovations has disrupted multiple service sectors, and healthcare is one such sector that has seen significant changes since the inception of Internet. The incorporation of information communication and technology (ICT) in the healthcare sector has been introduced previously; however, the recent COVID-19 pandemic has exponentially accelerated the use of ICT in the healthcare sector and the demand for contactless health services. The use of technology for online consultation, robot-assisted surgeries, training, and education are some of the common use cases of tech-driven healthcare. Metaverse, the latest innovation that is the amalgam of multiple technologies like artificial intelligence (AI), augmented reality (AR), virtual reality (VR) and robotics, holds massive potential to deliver high-quality healthcare services (HS) to the masses, eliminate bottlenecks between healthcare providers and patients and eliminate the need of in-person interaction for availing HS. This study aims to identify the factors that drive the adoption of metaverse in the healthcare industry. Following a mixed-method approach, the study combines secondary academic literature data with primary data from domain experts. A total of thirteen enablers are identified for adopting metaverse in the healthcare domain. Subsequently, a hierarchical model reflecting upon the relationship among these factors based on their significance is developed using Grey-Ordinal Priority Approach (G-OPA) and modified TISM methodology (mTISM). The proposed hierarchical model provides valuable insights into these factors' dependencies and relative importance which could empower healthcare organizations and policymakers in making informed decisions and strategic planning for an effective and efficient healthcare system.**

***Index Terms*—** Metaverse, Healthcare, Information and Communication Technology, Technology Management, modified TISM, Grey-OPA

I. Introduction

The innovations in the field of Information and Communication Technology (ICT) have transformed the way humans consume and deliver products and services. One such technology that has been in focus lately is metaverse. It is evolving rapidly and the interest in its application across domains is growing exponentially [1]. Hackl [2] defines metaverse as an evolving virtual universe in which people can create their own representation of the world based on their experiences from real world [2]. Koss [3] refers to metaverse as an immersive 3-D virtual space that enables people to connect with the other people, shop, and work in the real time. Metaverse is the amalgamation of technologies like artificial intelligence, digital twining, augmented and virtual reality that facilitate interactions between people and their digital counter parts in the virtual environments [4]. Integrating different technologies in the metaverse holds significant potential to revolutionize various sectors, including healthcare [1].

The application of digital technologies in the healthcare system is not new. Healthcare providers have adopted digital alternatives like mobile apps, and telehealth to offer various services but are primarily laggards [5], [6], [7]. Such systems were mostly for disseminating healthcare information among the stakeholders. However, metaverse has emerged as an innovative technology, and the approach to adopting it in the healthcare sector is proactive and growing. It is anticipated to transform the structure of the healthcare sector, enhancing the experience of patients and healthcare providers [8], [9]. Medical professionals are already utilizing components of metaverse for delivering HS. For example, doctors are helping patients with phobias by putting them into virtual environments where they must face their fears. Similarly, medical practitioners are also using VR to treat patients with mental diseases like social anxiety and obsessive-compulsive disorder[10]. The potential benefits of using metaverse in the healthcare sector could be ensuring anonymity, integrated health records, and providing wider reach and virtual experience. Metaverse has shifted the HS from hospitals to homes and from information-driven healthcare to experience-driven healthcare, significantly contributing to its market in the healthcare sector. As per the recent report [11], the current market for metaverse in healthcare sector is around $6 billion and is estimated to reach $71 billion by 2030. Because metaverse offers highest possibility of virtual interaction, immersive experience, duality, and the support for interoperability across platforms, it has the potential to disrupt the healthcare sector [5], [12], [13].

Despite the potential benefits of metaverse in general and healthcare in particular, there is still minuscule research in this direction. This could be because metaverse is a novel innovation still in its infancy stage. The advantages that metaverse can offer to the healthcare industry are immense and could lead to significant changes, enhancing patient engagement and experience through high levels of interaction and increasing accessibility to HS in far-flung areas [5]. Its potential applications across sectors are still being explored, and historically, the healthcare sector does not fall under early adopters of the innovations [14], [15]. However, as more people become familiar with metaverse, its adoption is expected to grow multi-fold. The success of any digital platform depends on its acceptance and use by its target audience [16]. Unlike other sectors, healthcare is a critical domain; thus, any new technology is adopted only after it has been tested, validated,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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 Engineering Management. 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/TEM.2024.3400401

and approved by the concerned authorities. The use of metaverse in healthcare has received greater emphasis, and many researchers have recommended examining how the metaverse can contribute to the healthcare sector [17], [18]. The ultimate adoption of any technology is driven by its enablers. These enablers are the factors that motivate concerned stakeholders to accept and use the technology. In the case of healthcare, enablers are the factors that could facilitate patients and healthcare providers to use metaverse for providing and receiving various HS. With this background, the study attempts to answer the following research questions:

- RQ1: What are the key enablers driving the adoption of metaverse in the healthcare sector?
- RQ2: What are the priority rankings of enablers in the context of emerging economies?
- RQ3: How are these enablers interlinked?

By answering the above questions, the study contributes to the metaverse applications and technology adoption literature in emerging economies. Precisely, the study identifies the thirteen key enablers impacting the adoption of metaverse in the healthcare sector. Further, by determining the priorities of each enabler in the context of emerging economies, the study addresses an important gap in current academic literature, contextualizing the utility of metaverse to resource-constrained geographies. Lastly, by identifying the interrelationships between enablers, the study highlights the complex dynamics of metaverse technology and its integration with the healthcare system in the rural communities of emerging economies. This combination of identifying enablers, prioritizing them in terms of emerging economies, and investigating interrelationships across enablers provides useful information for the use of metaverse in healthcare for emerging economy rural communities.

The remainder of the paper is structured as follows: Section 2 presents the literature review on the adoption of ICT in the healthcare sector, metaverse in the healthcare sector and research gaps. Section 3 outlines the research methodologies used. The results and findings of the research are presented in Section 4, and the study's implications are discussed in the Section 5. Finally, Section 6 concludes the study and offers suggestions for future research.

II. Literature Review

The term “Metaverse” was first coined by Neal Stephenson in 1992 in his novel Snow Crash, wherein it reflects upon the integration of the internet and augmented reality technology to provide real-world experience in the digital space – virtual reality [6], [19], [20], [21]. Currently, online immersive platforms are proliferating and are expected to go beyond just reflecting the real world, which is limited by boundaries and strict rules [22]. It is believed that current devices carried to access the metaverse will be replaced by the devices worn by the users, thereby better integrating our physical world with the virtual world. Considering the potential of such futuristic technologies, tech giants are spending around 5% -20% of their R&D funds on these technologies.

Healthcare is another sector that has seen massive growth, especially after the COVID-19 pandemic. The impact of COVID-19 resulted in health expenditures experiencing the most rapid growth in the past twenty years during the year 2020. Experts believe that the metaverse could impact the healthcare sector significantly because of its potential to integrate three major technological innovations, i.e., telepresence, digital twining and blockchain [19], [23], [24]. With this background, we provide a focused literature review covering two significant dimensions of the technology adoption and healthcare sector. First, we review major developments related to the adoption of traditional ICT in healthcare, and second, we review the contributions made towards the metaverse for the healthcare sector.

A. Adoption of Information and Communication Technologies in Healthcare Sector

Healthcare is divided into several areas, each using technology differently [25]. Healthcare organizations, practice settings, and nursing units use technology differently [26], [27]. For example, in nursing ICT play an important role in maintaining real-time connectivity among healthcare professionals, patients and the management. Apart from communication, it facilitates accurate record-keeping, efficient information organization, and seamless retrieval processes. Similarly, doctors use it in research, tele-consultation, effective management and communication with other healthcare professionals[27], [28], [29]. Broadly, healthcare professionals, encompassing doctors, nurses, technicians, pharmacists, and allied practitioners, harness the power of ICT to access vital information, streamline administrative tasks, elevate patient care standards, and stay updated on novel advancements within their specialized domains [12], [28], [30]. In general, ICT facilitates better knowledge management, a phenomenon extensively investigated in diverse studies [25], [31], [32]. Earlier studies have reflected upon the enhanced use of various tools, and techniques to address multiple healthcare related issues and improve the overall quality of life [33]. However, it is still challenging to introduce new technologies in the healthcare sector considering the challenges involved. Within this context, healthcare sector has diverse set of technologies at its disposal including but not limited to metaverse, blockchain, Internet of Things, augmented reality, artificial intelligence, and robotics. The strategic integration of these technologies has the potential to transform the overall healthcare industry [34], [35], [36], [37]. The need for such transformation is further accelerated by the recent COVID-19 pandemic [38]. Organizations in the healthcare industry have quickly embraced digital solutions and cutting-edge technological tools in response to the pandemic's initial phase [12], [13], [39]. To lessen the impact of COVID-19 on people and health systems, the study conducted by Golinelli et al., [40] provides the digital solutions that have been documented in the literature. Popa et al., [41] present the use of digital twins in the healthcare sector from a socio-ethical risks perspective. Attaran [42] presents the adoption of blockchain in the healthcare sector. Dhar et al., [43] explained how adopting

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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 Engineering Management. 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/TEM.2024.3400401

augmented reality in medical education will help students and enhance their performances. By adopting these cutting-edge technologies, the healthcare sector may transform and build a safe, secure, and stable digital ecosystem for the management and improvement of a better, more individualized client experience [44]. To help the general populace of their economies, developed and developing countries have recently started to implement digital transformation in the healthcare sector. Nnaji and Tsai [45], [46] discusses various barriers in the adoption of these technologies, including whether the technology is safe, how it will be used, and how it will affect patients and healthcare professionals. Due to worries about the expense and a lack of knowledge with the technology, many hospitals take their time implementing new technologies [47], [48]. There are several barriers, such as the fact that much medical personnel do not know how to use the technology or how it will impact their daily work [37]. Because their current systems are so convenient and easy to use, users are reluctant to adopt new technology [28], [33]. Also, the security or efficacy of new technologies may raise concerns in clinics and hospitals. Awad et al., [49] stated that adopting new technology is a gradual and ultimately beneficial process. While the practicality of ICT adoption has improved, it remains imperative to identify the underlying reasons for low adoption rates and actively work towards resolving these challenges.

B. Metaverse in Healthcare Sector

The metaverse can be seen as the immersive follow-up to the text-and-picture-based Internet of today. The metaverse promises to deliver online interactions that are more immersive and dynamic, a seamless fusion of both the digital and physical worlds, by utilizing cutting-edge technology [50]. Technologies like artificial intelligence, augmented and virtual reality, blockchain, and 6G network constitute advanced technological innovations within the metaverse, exerting significant impact on the healthcare sector. Huynh stated that the adoption of these technologies' exposes people to the novel, much less expensive therapy delivery methods, improving patient outcomes [19]. According to Park & Kim [4], the healthcare system in the realm of the metaverse offers recreational, immersive, and interactive experiences that are tailored to each patient's needs .Metaverse is a universe that uses the internet to create an alternative reality experience, recreating patients 'emotions and gestures while encompassing the social and economic structures of the virtual and real worlds. Petrigna & Musumeci [51] explain how healthcare professionals could adopt the metaverse for disease diagnosis and effective planning. The recent pandemic has generated the need for remote HS and metaverse has the potential to provide a greater experience than the classical videoconferencing-based telemedicine systems [15]. Further, Huh [52] emphasized that Medical education and training could be revolutionized by the metaverse. Instead of distributing conceptual understanding, using AR creates a favourable setting for explaining actual techniques. Many academicians have investigated the possibility of the metaverse in various domains like, the study conducted by Yang [18] explored the application of metaverse in medical IoT using AR/VR. Bhattacharya [24] highlighted the potential of Healthcare 5.0 for renders personalized care, virtual clinics, wellness monitoring, telesurgery, telemedicine etc. Thomason [53] addressed the metaverse's potential role in enriching and revolutionizing healthcare regarding wellness, monetization, clinical care, education, and collaborative working. In another study, Krishnan [33] studied the application of metaverse in healthcare from a service provider's perspective and highlighted its relevance in conducting training, education and offering consultation in a more realistic environment. However, the study concludes that it is difficult to prove that there will be significant improvements in healthcare by using metaverse. Yang et al., [54] highlighted the role of intelligent healthcare systems, where many examples of data science and AI applications in healthcare have been studied in the context of hospital management in the metaverse. The analysis of the examples reveals challenges in the metaverse applications in healthcare sector. This dimension of metaverse adoption is important, as emphasized in multiple studies. According to Henz [55]adopting the metaverse at a large scale could reduce human interaction and increase human-to-technology interactions in the future. This could seriously affect users' well-being, leading to depression and loneliness [56]. Also, the cost associated with the adoption and access of metaverse technology pose additional challenges. Prior studies [38], [57], [58] have reflected upon the digital divide as the potential implication of metaverse adoption because of the economic inequalities present in the society, thus compromising on the wellbeing-related benefits offered through metaverse. To ensure the safe and effective usage of metaverse, various researchers have emphasized on the need to explore metaverse technology from consumers' perspectives [12], [38], [56], [59] and the motivations that drive its adoption in healthcare sector [5], [20], [56], [57], [58]. With this background, the study attempts to explore the key factors behind the adoption of the metaverse in the healthcare sector, with a primary emphasis on patient-doctor interactions in emerging economies. Emerging economies have unique characteristics in terms of literacy rate, socio-economic wellbeing, and technology adoption and contribute 85% to the global population [60]. Most of these countries suffer from inadequate healthcare infrastructure [61]. Metaverse could bridge this gap and enable better healthcare facilities to these most vulnerable populations.

C. Research Gaps

Metaverse has emerged as a significant technology and it is anticipated that the adoption of metaverse in the healthcare sector will enhance the healthcare system in developed economies for multiple purposes including disease prevention, treatment, promoting healthcare programs, training, and education [14], [55]. Yet, there is a perceived lack of effectiveness in the adoption of metaverse in the healthcare sector for the developing economies[20]. In literature, many studies have identified the factors/critical success factors for metaverse adoption [62], [63], [64], [65]. These studies have explored the barriers or enablers and provided typologies or taxonomies without taking interdependencies specific to healthcare into the consideration. To the best of our knowledge, we did not find any study specifically focused on identifying the key enablers in adopting metaverse technology in the healthcare sector by considering domain experts—most of the existing research papers in this area are related to the issues and challenges faced during the implementation phase. However,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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 Engineering Management. 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/TEM.2024.3400401

few research papers specifically developed a comprehensive framework for identifying and studying the key enablers for metaverse adoption. Additionally, prior studies have looked mainly at the technological aspects of metaverse adoption in the healthcare sector which is only one aspect of the adoption [66], [67]. Adopting novel technologies and digital platforms depends on the combination of tech and non-tech factors [16]. There is no consensus on how to classify the factors that make the adoption of the metaverse possible [5]. Further, comprehensive studies exploring the adoption of metaverse in delivering HS in the developing economies is largely missing. In this sense, an academic gap exists in identifying and analyzing the key enablers of metaverse adoption in healthcare sector in the context of emerging economies. Academic gaps must be addressed before the full potential of metaverse technology in the healthcare sector can be realized. In comparison with the prior studies, this study presents a managerial perspective on the adoption of metaverse in the healthcare industry in the context of developing economies. Because potential adopters lack access to adequate information to make informed judgements, these research gaps are impeding the widespread implementation of metaverse technologies in the healthcare sector. The study develops precise research questions to bridge the research gaps, as mentioned in section 1.

III. Research Methodology

The main objective of this research is to identify and prioritize the enablers that facilitate the adoption of metaverse technology in the healthcare sector and investigate how these enablers are interrelated. The study develops a methodological framework to fulfil this objective by integrating various theories and methodologies like Critical Success Factor Theory, Grey-OPA and modified-TISM. This study initially identified multiple enablers from the available literature on metaverse adoption, specifically in healthcare sector. Further, selected enablers have been finalized with the consultation of domain-related experts and by adopting the critical success factor theory. Furthermore, G-OPA is utilized to assess finalized enablers to obtain their importance weights. Here, OPA is preferred over other MCDM techniques for its efficiency and reliability [68]. Uncertainty is common in healthcare, especially with cutting-edge technologies like the metaverse. Integrating grey system theory can resolve this ambiguity, allowing for a more precise analysis despite scant or confusing data. This makes it more preferable than other soft computing methods [69]. The importance weights obtained by G-OPA are then used to rank the enablers based on their importance for metaverse adoption—the next sub-objective deals with investigating interrelationships among these enablers. Modified Total Interpretive Structural Modelling is used to meet this. The m-TISM is preferred over other similar methods like ISM, DEMATEL, etc. because (i) it involves simultaneous transitivity check, which reduces the requirement of pairwise comparisons, and (ii) it provides an interpretation matrix of interrelationships of enablers. The methodological flowchart representing the research flow is presented in Fig. 1. The brief details, including the steps for adopted methodologies, are explained in subsequent sub-sections.

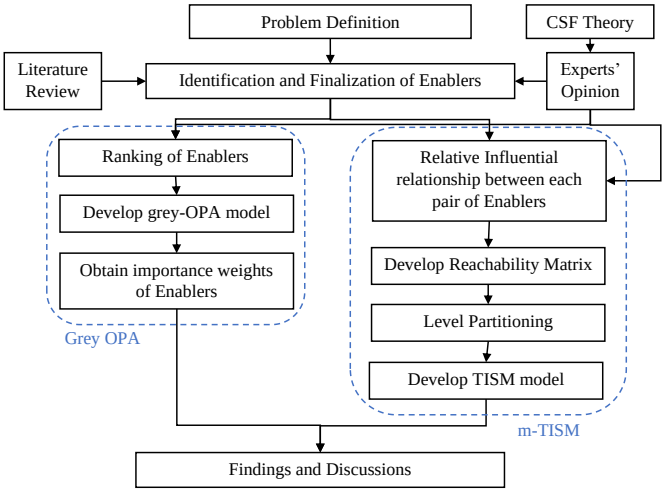


Fig. 1. Methodological Flowchart

A. *Critical Success Factor Theory & Selection of Experts*

Critical Success Factors (CSFs) were first discussed by Rockart [70] as part of his investigation into the function of information systems in organizations. His contributions set the groundwork for comprehension of the idea of critical success factors and their function in the management of businesses. Rockart [71], [72] has recommended a two-step method for interviews to pinpoint the success factors. In a series of open interviews, experts are initially questioned about the factors they believe will contribute most to the entity's success. Based on these, a preliminary list of criteria is developed, and in a future round, they are rated according to an essential factor.

This study used the CSF theory to identify the enablers that could facilitate the adoption of metaverse in the healthcare sector. Enablers are identified through a two-step process.

TABLE I
EXPERTS' PROFILE

Expert No.	Designation	Educational Qualification	Age	Gender	Field of Expertise	Work Exp.
Expert 1	Sr. Manager	Post Graduate	42 Yr	Male	Innovation Management	13 Yr
Expert 2	Medical Officer	Post Graduate	41 Yr	Male	Healthcare Services	17 Yr
Expert 3	Sr. Medical Officer	Post Graduate	47 Yr	Female	Healthcare Services	19 Yr
Expert 4	Doctor	Post Graduate	36 Yr	Female	Healthcare Services	11 Yr
Expert 5	Surgeon	Post Graduate	46 Yr	Male	Healthcare Services	18 Yr
Expert 6	Sr. Technical Officer	Graduate	37 Yr	Male	Technology Management	15 Yr
Expert 7	Professor	Doctorate	54 Yr	Male	Technology Management	21 Yr

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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 Engineering Management. 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/TEM.2024.3400401

Initially, the authors compiled a preliminary list of enablers from the literature. Subsequently, the list of enablers was shared with the experts to finalize them. The experts chosen for this study are professionals in the Indian context of technology management, HS, and innovation management. India, with 1.4 billion people, is the most populous country in the world. Around 65% of its population living in rural areas [61], [73], highlights the country's unique demographic and geographic diversity making it a compelling case for investigating the phenomena of metaverse in healthcare. Additionally, hospitals in rural India are often not equipped with state-of-the-art medical equipment, suffer from a lack of sufficient clinical expertise among local doctors and necessitate rural patients to visit major hospitals located outside rural areas for better treatment [74], [75], [76], [77]. Metaverse could be a timely solution to address these issues and improve the overall quality of the healthcare system in India [78]. Moreover, experts also believe that addressing any issue in India means addressing it for over 40% of the global population [79].

To ensure that their perspectives are sufficient, the experts chosen for this study are above mid-level executives[80]. Initially, 15 experts have been consulted. Out of them, seven experts confirmed their participation in this study. The experts consulted in this study are mostly from technology management (28.6%), HS (57.1%) and innovation management (14.3%) domains. The experts' profile is presented in TABLE I. Further, the enablers were explained to the experts, who were then tasked with evaluating whether or not their interpretations of the enablers were consistent and relevant to adopting metaverse in the healthcare sector in light of their knowledge and experience. The experts gathered afterward for a brainstorming session to reach a consensus and compiled the final set of significant enablers that are important for the implementation of the metaverse in the healthcare sector.

B. Grey Ordinal Priority Approach

Ataei et al. [68] proposed Ordinal Priority Approach as a MCDM technique to solve complex decision-making problems. OPA is favoured over other MCDM techniques such as, ANP, AHP, BWM, etc. due to providing flexibility to the decision makers [68]. OPA does not require pairwise comparisons, normalization or decision matrices. Also, it allows decision-makers to either rank the factors/elements same rank or can be kept blank, if decision-maker is unsure about it. This further provides flexibility and better efficiency in decision-making process. However, few decision-making problems are bit complex and contains subjective opinion of the decision-makers. To cater this, we have integrated the grey numbers with OPA as done by Mahmoudi [81]. The steps of G-OPA are as follows:

Step 1: Determine the factors. The first step begins with the identification of factors (enablers in our case) to be analysed. In this study, enablers of metaverse adoption in healthcare sector have been extracted from the literature survey which is presented in TABLE III.

Step 2: Assessment of the factors. Once the factors are identified, the experts are asked to rank those factors using linguistic terms as mentioned in TABLE II.

Step 3: Develop the model. The next step relates to model building. In order to obtain grey weights of each factor, linear programming model can be built as follows:

$$Max \otimes Z$$
$$s.t.$$

TABLE II
GREY NUMBERS AND ITS CORRESPONDING LINGUISTIC TERMS (SOURCE: [82])

Linguistic Variables	Grey Numbers	Rank
Very Low Importance (VLI)	(6.5, 7.5)	7
Low Importance (LI)	(5.5, 6.5)	6
Medium Low Importance (MLI)	(4.5, 5.5)	5
Medium Importance (MI)	(3.5, 4.5)	4
Medium High Importance (MHI)	(2.5, 3.5)	3
High Importance (HI)	(1.5, 2.5)	2
Very High Importance (VHI)	(0.5, 1.5)	1

$$\otimes Z \leq \otimes e\left(\otimes j\left(\otimes W_{ej}^{r} - \otimes W_{ej}^{r+1}\right)\right) \quad \forall e, j \text{ and } r \qquad (1)$$
$$\otimes Z \leq \otimes e\left(\otimes j\left(\otimes W_{ej}^{m}\right)\right) \quad \forall e \text{ and } j$$
$$\sum_{e=1}^{E}\sum_{j=1}^{J} \otimes W_{ej} = [0.8, 1.2]$$

Where, $\otimes Z$ is unrestricted in sign; $\otimes e$ & $\otimes j$ are the grey ranks of the experts and enablers, respectively.

Step 4: Obtaining the crisp weights. Solving equation (1), the grey weights of the enablers can be obtained. Further, obtained grey weights can be converted to crisp weights using equation (2) as follows:

$$C_j = \frac{1}{2}\left(\overline{\otimes}\, j + \underline{\otimes}\, j\right) \qquad (2)$$

C. Modified Total Interpretive Structural Modelling

MCDM approaches are being utilized more frequently to settle new conflicts and analyze different scenarios in light of many competing evaluation criteria. It is regarded as a decision-supporting and operational assessment technique that is best suited for complex problems with many viewpoints, different interests, competing aims, and high levels of decision-making uncertainty. Some of the common MCDM techniques includes AHP, ANP, TISM, SMART, TOPSIS, ERA, DEA etc [72], [82], [83], [84], [85]. Recently, Sushil [86] developed a new MCDM technique and termed it as a modified TISM (m-TISM). The Total Interpretive Structural Modelling approach, which is used to create a hierarchical model of identified factors of interest [72], [87], [88], is extended using this technique. Although TISM could have been used, m-TISM was chosen because it enables simultaneous transitivity checks, unlike traditional TISM where transitivity is checked sequentially [89]. M-TISM not only accelerates the entire process but also

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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 Engineering Management. 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/TEM.2024.3400401

reduces the risk of errors originating from redundant pairwise comparisons, as seen in the case of TISM [90]. In order to make it easier to develop a hierarchical model of the modified TISM technique addresses three fundamental concerns linked to "what," "why," and "how." Regarding the final hierarchical model, there is no significant difference between the model developed using traditional TISM and m-TISM. When contrasting m-TISM with other research methodologies, the "Systematic Literature Review (SLR)" can offer pertinent literature in the field of study and assist in gathering various constructs for study, but it is unable to produce a systematic hierarchy of the model of those constructs. The SLR also falls short in giving guidance and an explanation of how the components that were detected are linked. Literature shows that researchers have used m-TISM in various domains. For an instance, Rajan & Dhir [91] has used m-TISM for technology management in organizations [91]. Meena & Dhir [92] uses m-TISM for studying the factors of automotive industry in India. Shardeo [93] has integrated m-TISM with fuzzy-ANP method for analyzing the critical success factors of blockchain adoption in transportation domain. Further, Sharma, [94] has used m-TISM for studying the impact of artificial intelligence on tourist firms and Sindhwani [95] has used m-TISM in the domain of logistics 4.0[95]. The steps of m-TISM are as follows:

Step 1: *Identify the factors and the nature of their relationships.* The first step begins with the identification of factors (enablers in our case). In our study, 13 enablers have been identified which is presented in TABLE II. Furthermore, we are interested in exploring the influential relationship of enablers.

Step 2: *Pairwise comparison.* The pairwise comparison of factors is used to investigate the link between the factors. In this phase, respondents are requested to compare the two factors and highlight their relationship, as well as provide justification. For example, if the first component effects the second, the expert will record the link as 1 – 2. Similarly, the process continues until all pairs of factors will be compared. Unlike TISM, the modified TISM questionnaire checks transitivity with logic checks, reducing the number of pairwise comparisons.

Step 3: *Develop the reachability matrix*. The following step is the development of a reachability matrix. For instance, the influential relationship between the two factors is represented by 1; otherwise, it is represented by 0. If a relationship is transitive, it will be represented as 1*. Similarly, the entire reachability matrix is developed.

Step 4: *Level Partitioning.* To begin the level partitioning procedure, reachability, antecedent, and intersection sets must be defined. All row-wise factors that have value 1 or 1*must be included in the reachability set. In contrast, the column-wise factors with the value 1 or 1* should be included in the antecedent set. The intersection set includes the factors that are common in reachability and the antecedent set. If the intersection set of any factor equals the reachability set of that factor, that factor is assigned the level, and the factors of that level's intersection set are excluded from the rest of the matrix for further iterations. The process is repeated until all of the factors have been assigned a level.

Step 5: *To develop the digraph using the final reachability matrix:* The digraph includes nodes representing the enablers and directed links representing the direction of the relationship among the enablers. Each enabler in the digraph is assigned a level obtained in the previous step. Important relationships are only shown in the final digraph with black lines and indirect relationships with a dotted line.\

Step 6: *Develop the TISM model.* The final step is building the TISM model. In this step, the factor(s) from the last level are considered to be the fundamental factor(s) upon which the model is built. The reachability matrix, on the other hand, can be used to determine the relationships between the factors.

IV. Results

This section consists of the findings from an analysis of the enablers of metaverse adoption in the healthcare sector. The final sub-section contains discussions of the findings of the analysis.

A. Identifying and finalizing the enablers

Initially, various articles in the field of technology adoption in the healthcare sector were assessed, as shown in section 2.1. An initial list of enablers has been compiled based on a review of the literature. The list has also been shared with the experts (TABLE I) for feedback. Following an agreement with domain experts, 13 enablers were finalized. TABLE III presents the final list of enablers of metaverse adoption in the healthcare sector.

TABLE III
LIST OF ENABLERS OF METAVERSE ADOPTION IN HEALTHCARE SECTOR

Description	Source
E1: Ability to link metaverse with existing IT systems that deal with other patient related information like steps walked, diet etc.	[14], [96]
E2: Ability to provide virtual presence with real-like experience.	[6], [21]
E3: Ease with which people can use a metaverse system for HS.	[9]
E4: People are comfortable in adopting digital innovations.	[14], [97]
E5: Ability to tailor patient profile based on his/her needs.	[50], [51]
E6: Impact of individuals or groups on users' attitudes, beliefs toward use of metaverse based healthcare.	[17], [98]
E7: Users' data are confidential, & users have clear control over it.	[5], [99]
E8: Cost to avail HS via metaverse returns the value out of it.	[66], [98]
E9: Capacity to process and analyze large amounts of complex data.	[50],[100]
E10: Rules/regulations/recommendations/laws specific to use of metaverse for healthcare.	[14], [66]
E11: Reward given for using metaverse based healthcare system.	[66],[101]
E12: Hardware & software for operational metaverse system.	[14], [66]
E13: Ability to get treatment without disclosing real identity. For example, STD and other diseases.	[102]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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 Engineering Management. 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/TEM.2024.3400401

B. Determining importance weights of enablers

The first objective of this study was to identify the potential enablers of metaverse adoption in healthcare sector. To achieve this, literature review is conducted as mentioned in section 2.1. and list of enablers has been finalized after consultation with the experts. As mentioned earlier, the experts chosen for this study are mostly healthcare professionals and technologists. Due to complexities involved in MCDM problems, too much sample size may lead to high inconsistency [103], [104]. As a result, the sample size of seven respondents is taken in this study which meets the sample size requirements of MCDM problems [105], [106]. The finalized list of enablers along with its brief description is presented in TABLE III. Furthermore, in order to achieve the second objective i.e., to obtain the importance weights of the enablers, grey OPA is employed. The brief description of steps of grey OPA is explained in section 3.1. Following the steps of grey OPA, the experts are asked to rank the enablers based on its importance using linguistic variables mentioned in TABLE II. The experts' opinion regarding the ranking of enablers on the basis of its importance is presented in TABLE IV. After collecting the data from all experts, the inputs were converted into corresponding grey numbers. For an instance, if an expert ranked Enabler 1 as "Medium High Important" then the corresponding grey number assigned is {2.5, 3.5}. Similarly, the experts' inputs have been converted into grey numbers. Then using equation (1), grey OPA model is developed as a linear programming (LP) model. The LP model is run on Python using *gekko* package. This resulted in an optimal solution of upper and lower grey values of importance weights of enablers. Further, the upper and lower grey values of importance weights assigned by each expert is summed up to obtain aggregated grey weights of enablers. Then, using equation (2), the crisp weights of the enablers are obtained. The importance weights and ranks of enablers of metaverse adoption in healthcare sector is presented in TABLE V. It can be seen from the TABLE IV, "Telepresence (E2)", "Anonymization (E13" AND, "Familiarization (E4)" are found to be three most important enablers. Whereas, "High-powered Analytics (E9)" is found relatively least important. This might be due to entrance of multiple high-powered technologies which further shifted the users' attention towards its ease of use than its powering capabilities. The importance weights of the enablers are also presented through Fig. 2.

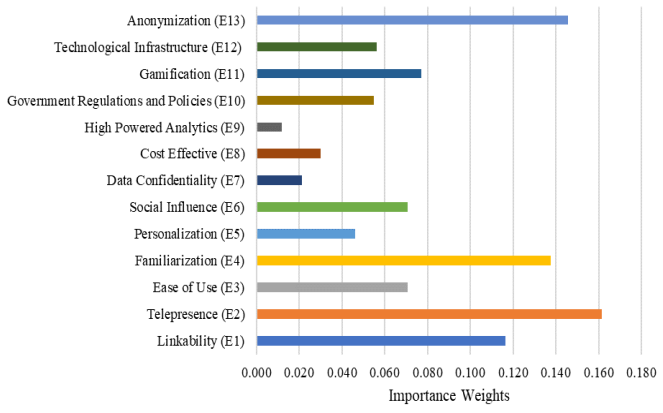


Fig. 2. Enablers and its ranking based on importance weights

TABLE IV
EXPERTS' OPINION REGARDING RANKING OF ENABLERS

Enablers	Experts						
	E1	E2	E3	E4	E5	E6	E7
Linkability (E1)	MI	VHI	HI	MHI	HI	HI	MI
Telepresence (E2)	VHI	VHI	HI	HI	VHI	MHI	MI
Ease of Use (E3)	MHI	MHI	MHI	MI	MLI	MHI	HI
Familiarization (E4)	HI	HI	VHI	HI	VHI	MLI	HI
Personalization (E5)	MI	MI	VLI	MHI	MI	MLI	MI
Social Influence (E6)	HI	MI	MLI	HI	MI	MHI	MLI
Data Confidentiality (E7)	LI	LI	MLI	MLI	MI	LI	VLI
Cost Effective (E8)	MLI	MLI	HI	LI	MI	VLI	LI
High Powered Analytics (E9)	VLI	VLI	LI	MLI	VLI	MLI	LI
Government Regulations and Policies (E10)	HI	HI	MLI	MHI	MI	LI	MI
Gamification (E11)	HI	MHI	MLI	MHI	MHI	HI	HI
Technological Infrastructure (E12)	MLI	MHI	MI	MHI	HI	MHI	MI
Anonymization (E13)	VHI	VHI	MI	HI	MHI	MI	VHI

TABLE V
IMPORTANCE WEIGHTS OF IDENTIFIED ENABLERS

Enablers	Importance Weights	Rank
Linkability (E1)	0.116	4
Telepresence (E2)	0.162	1
Ease of Use (E3)	0.071	7
Familiarization (E4)	0.137	3
Personalization (E5)	0.046	10
Social Influence (E6)	0.071	6
Data Confidentiality (E7)	0.021	12
Cost Effective (E8)	0.030	11
High Powered Analytics (E9)	0.012	13
Government Regulations and Policies (E10)	0.055	9
Gamification (E11)	0.077	5
Technological Infrastructure (E12)	0.056	8
Anonymization (E13)	0.146	2

C. Interaction among enablers

The subsequent objective of the study was to highlight the interaction of the enablers. To accomplish this objective, we employed m-TISM to develop the interpretive structural model. The same experts were consulted for collecting data for the m-TISM exercise. The pairwise comparison of enablers was conducted based on group discussion and maintaining consensus among them. As recommended by Creswell and Poth, [107] the sample size requirement for qualitative study should be 5 to 25 which supports the consideration of 7 experts in this study. It can also be justified further by prior studies used fewer sample sizes to conduct TISM analysis [106], [108]. Further, the experts were asked to mark their opinion on the influence of one enabler on another, along with justification. Based on the data collected, a reachability matrix is created that contains the concurrent transitive relationships, i.e., the strength of m-TISM, as mentioned in Section 3.2. Appendix I shows the reachability matrix with simultaneous transitive links.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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 Engineering Management. 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/TEM.2024.3400401

Furthermore, the reachability set and antecedent set are constructed depending on the enablers that appear in the reachability matrix's columns and rows, respectively. The intersection set is then produced, which contains the enablers that are common by the reachability set and the antecedent set. If the intersection set equals the reachability set, the level is partitioned, and the enablers from the Intersection set are deleted from the table for the following iteration. Similarly, the process is repeated until all enablers are assigned to different levels. Appendix II depicts the iteration-wise level partitioning. As can be seen, level partitioning takes 7 iterations, resulting in 7 levels. The TISM model can be constructed based on the level assigned to the enablers. Fig. 3. depicts the TISM model of metaverse enablers in the healthcare sector.

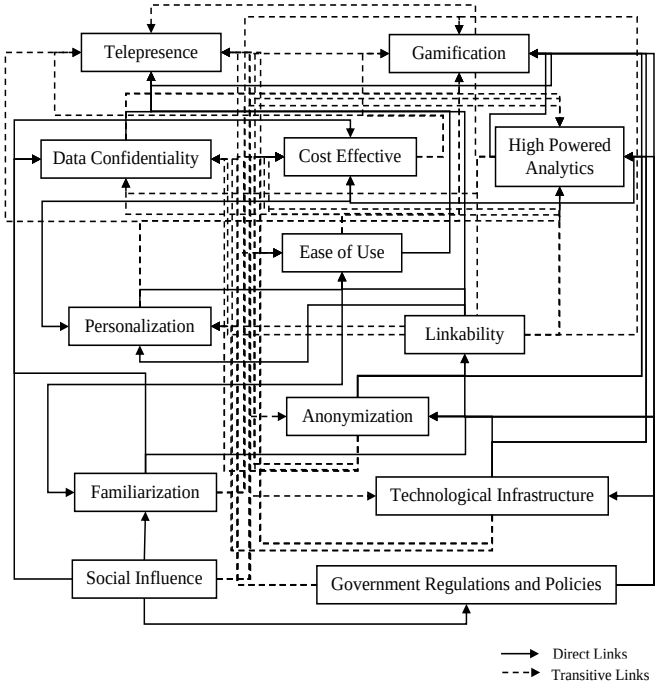


Fig. 3. Interpretive Structural Model of the Enablers

As shown in Fig. 3., the enablers "Social Influence" and "Government Regulations and Policies" are positioned at the bottom level and constitute the foundation for metaverse adoption in the healthcare sector. These two enablers have significant influence over other enablers and are critical to elevating metaverse adoption in the healthcare sector. Similarly, the model depicts the influential interrelationship of the enablers. The interpretations of the interrelationships among enablers are presented in TABLE VI. It is worth noting that the mTISM model focuses on influential relationships, whereas the G-OPA analysis reported in the preceding section underlined the importance of enablers in terms of priority weights perceived by experts. The following section provides a more extensive discussion of the importance and influence of enablers.

D. Discussions

The study aimed to explore the enablers contributing to adopting a metaverse in healthcare. Metaverse is still in its infancy stage, and it is still unclear how this technology will unfold in the future. However, researchers, business leaders and technology giants have shown their confidence and expressed keen interest in the metaverse technology and its potential applications across sectors. Broadly, the metaverse can be used in healthcare system for multiple purposes including disease prevention, treatment, promoting healthcare programs, training, education, sharing research to name a few. Thus, it becomes necessary to examine the enablers so that we can better understand the impact and their role in adopting metaverse in healthcare.

The findings showed that thirteen enablers comprise technology-related factors like telepresence, anonymization, ease of use, linkability, infrastructure, personalization, and non-technology-related factors, including social influence, familiarization, government regulations and policies, among others. Furthermore, the first aspect we focused upon is to assess the importance of enablers of metaverse adoption in healthcare sector. The results of Grey-OPA analysis indicated that the “Telepresence (E2)” is the most important enabler for metaverse adoption. Telepresence enables users to stay connected, making physical distance irrelevant. Thus, it would be relevant in contagious diseases or pandemic situations like COVID-19 outbreak[66]. However, it has been argued that telepresence could also possess socio-ethical issues due to lack of physical presence. “Anonymization (E13)” is identified as second most important enabler for metaverse adoption. Anonymization enables patients to receive medical services without disclosing their real identity. In the offline space, ensuring patients' anonymity is extremely difficult and still not considered a priority. It is one of the primary reasons many patients suffering from different life-threatening diseases, avoid consulting doctors, especially in rural areas [105]. Adopting metaverse can support the anonymity of the patients, enabling socially conscious patients to consider virtual consultation and result in building trust among patients and healthcare providers thus promotes the responsible use of data in the metaverse. Technology adoption, in general, people with sufficient knowledge about innovations are interested in experiencing them and thus seek to use such systems, leading to increased trust in the technologies. Likewise, the successful adoption of any new technology requires education, training, and awareness. By conducting training programs, workshops, and demonstrations, healthcare organizations can familiarize their staff with the metaverse, its benefits, and potential use cases. Such efforts can help overcome scepticism, reduce resistance to change, and promote the adoption of metaverse solutions in healthcare sector.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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 Engineering Management. 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/TEM.2024.3400401

TABLE VI: INTERPRETIVE MATRIX FOR THE ENABLERS

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13
E1	-	It enhances telepresence through expanded interaction	Seamless connections simplify user navigation	-	It offers personalized experiences through user profiles	-	It increases the risk of unauthorized access	It may necessitate extra infrastructure investment	Interlinked data enables powerful analytics	-	Seamless interaction across linked platforms enhances gamification	-	-
E2	-	-	-	-	-	-	-	-	-	-	Enhances the likelihood of telepresence and vice-versa	-	-
E3	-	Simpler systems reduce cognitive load, boosting user focus and enhancing telepresence	-	Familiar people would find the system easy and vice-versa	-	-	Ease of use fosters user familiarity, reinforcing data privacy awareness.	It drives data confidentiality awareness, demanding stronger security measures and higher costs	It raises security concerns, potentially limiting investment in high-powered analytics	-	It requires higher investment, potentially impacting effective gamification	-	-
E4	-	Reduces cognitive load, boosting telepresence by focusing users on the virtual environment	People would find the system easy and vice-versa	-	-	-	Digital civilized people are likely to be aware of data confidentiality	It with data privacy practices would impact cost considerations.	User familiarity streamlines data collection and cleaning, improving analytics quality	-	Familiar users confidently embrace gamification, fostering deeper engagement.	-	It enables awareness to anonymization and increase its likelihood
E5	-	It boosts telepresence through user interaction, deepening immersion in the VE.	It simplifies platform usage by tailoring services to user preferences	-	-	-	-	Improved experience leads to long-term cost-effectiveness	It enhances user experience but could raise costs	-	Customized rewards and challenges amplify user engagement based on preferences	-	-
E6	-	It can enhance the feeling of telepresence	It ensures virtual platform familiarity, improving ease of use.	It raises tech awareness, fostering familiarity	-	-	Conscious society would enforce data confidentiality	Privacy-conscious society demands stricter data confidentiality, affecting cost	It impacts high-powered analytics by restricting data access	It shapes societal norms	It indirectly impacts gamification design	Influences government policies and regulations	It encourages users to willingly engage
E7	-	Data confidentiality in telepresence builds trust, enhancing virtual security	-	-	-	-	-	It requires robust security measures, which leads to cost-effectiveness.	It demands strong security, raising costs for high-powered analytics	-	It limits analytics, impacting gamification	-	-
E8	-	Cost-effective solutions may compromise data confidentiality.	-	-	Cost-cutting may limit system's personalization features	-	Cost-effectiveness can weaken security, risking data confidentiality	-	It requires costly computational resources	-	Cost focus may limit high-powered analytics, hindering gamification	-	-
E9	-	It enhances telepresence with personalized experiences	-	-	It drives personalized user experiences	-	Trading insights for cost cuts can compromise data confidentiality.	It requires costly computational resources	-	-	High-powered analytics enable personalized interactions.	-	-
E10	It impacts platform linkage, affecting connectivity	It enhances telepresence by fostering user trust	It promotes user-centric design for user-friendly technology	It shapes tech norms, influencing public platform familiarity	It impacts personalization by setting boundaries for data	It shapes societal norms, influenced by public pressure	It mandates data security, enhancing user data confidentiality	Regulated data analytics may require specialized tools	It would drive the type of analytics allowed on users' data	-	Regulated gamification builds user confidence, attracting users	It establishes infrastructure standards	Policies on anonymization demand suitable infrastructure
E11	-	It increases telepresence by increasing user engagement	-	-	-	-	-	-	-	-	-	-	-
E12	It ensures seamless connections	It underpins engaging gamification	It facilitates seamless user interaction towards easy use	-	It enables user data analysis, facilitating personalized experiences	-	It ensures confidentiality	It impacts high-powered analytics	Technological support would be decisive for type of analytics offered	-	It is required to develop gamification process	-	Enabling anonymization feature in the system
E13	It restricts the system to link with other platforms	It enhances user engagement and telepresence satisfaction	It requires additional steps, impacting ease of use	-	It hinders personalization by obscuring user data	-	Ensuring data confidentiality becomes easier	It reduces the value gained from high-powered analytics,	Limited user details, could hinder in deriving analytics	-	Lack of personalized rewards limit the yield of gamification.	-	-

Among others, another significant advantage of metaverse is its ability to link with other emerging technologies such as IoT, AR/VR, Blockchain, to name a few. This ability is termed as "Linkability (E1)". It reduces the friction in the system for both healthcare providers and patients. In the healthcare sector, linkability can facilitate seamless integration of patient records, medical devices, and other data sources, enabling comprehensive and holistic care. Linkability, from the perspective of patients, provides a means for sharing data between platforms and institutions without any additional effort and aids in the accumulation and diffusion of knowledge, therefore contributing to the convenience of the healthcare system. As a result, it can further enhance care coordination, improve efficiency, and enable better clinical decision-making. Since metaverse leads to the convergence of virtual and physical worlds, which presents both opportunities as well as challenges that must be addressed properly. It has been argued that the metaverse adoption would be slower due to clinical trials before using it as a realistic component [99]. Among others, the major challenge highlighted by the literature is "Data Confidentiality (E7)". As metaverse relies on digital interactions and the exchange of personal health information. In such a case, it makes patients worrisome regarding confidentiality of their data thus, ensuring robust data privacy and security measures is essential[14], [101] . Implementing strong encryption, authentication protocols, and secure data

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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 Engineering Management. 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/TEM.2024.3400401

storage systems can be helpful in building trust and confidence among healthcare providers and patients. Another way to address such issues is to clearly communicate with the patients regarding the control of sharing the data to third parties. Additionally, the "Government Regulations and Policies (E10)" should also address such issues including patient privacy, data protection, ethical considerations, and professional guidelines. Establishing clear regulations and policies will provide a solid foundation for the responsible and ethical use of the metaverse in healthcare. Moreover, government-approved and endorsed platforms enhance the credibility and legitimacy of the overall healthcare system. The influence of "Government Regulations and Policies (E10)" can be seen from the Fig. 3. where it acts as a foundational enabler directly impacting other significant enablers and this result is also in accordance with the study by Petrigna & Musumeci [51].

Healthcare professionals and patients need intuitive and user-friendly interfaces that make it easy to navigate and interact within the metaverse. Complex or cumbersome user experiences can create barriers to adoption, as users may feel overwhelmed or discouraged. The Ease of Use (E3) can drive the likelihood of individuals to embrace it and integrate it into their daily workflows or health management routines. At the same time, its negative experience may discourage people from using metaverse for HS. Therefore, metaverse solutions should prioritize simplicity, accessibility, and seamless user experiences. Providing comprehensive tutorials, clear instructions, and responsive technical support can help users overcome initial challenges and feel confident in utilizing metaverse technologies for healthcare purposes. Implementing a reward strategy is another aspect of the adoption that is not directly linked to the metaverse technology but is related to the adoption policy followed. The ability to reward users upon using a system is termed as "Gamification (E11)" and could be instrumental in adopting the system, especially at the initial stages, wherein fewer people are aware of the platform or are reluctant to try the new system. It can make the metaverse more engaging, interactive, and enjoyable for both healthcare professionals and patients [101]. Along with the gamification, Social Influence (E6) i.e., the impact of society on users' attitudes and beliefs toward using metaverse-based healthcare could also drive the adoption. This result is in accordance with the study by Arpaci et al. [98]. Tech-friendly community reflects a higher trust level in the system, which has a ripple effect on others who are natives. People often get motivated by their local natives more as compared to others, where the trust level is not that strong. The model shown in Fig. 3. also demonstrated the influencing power of social influence. Peer testimonials, success stories, and recommendations from trusted sources can create a sense of social proof and credibility, reducing uncertainties and increasing the willingness to try metaverse solutions. Healthcare organizations can leverage social influence through targeted marketing campaigns, user testimonials, and community-building efforts to accelerate metaverse adoption in the healthcare sector.

The metaverse generates vast amounts of data, ranging from patient interactions and virtual simulations to real-time monitoring and diagnostics. High-powered analytics (E9) can process and analyse this data to extract valuable insights. Analytics-driven metaverse applications can assist healthcare professionals in making data-informed decisions, optimizing treatment plans, and improving patient outcomes [109]. This further enhance the effectiveness and impact of the metaverse in healthcare. However, high-powered analytics, such as AI and machine learning, heavily rely on advanced technological infrastructure. Technological infrastructure (E8), core for working metaverse-based healthcare systems, amalgamates state-of-the-art hardware and software components. Analysing vast amounts of data, running complex algorithms, and generating real-time insights require High-speed internet connectivity, continuous supply of electricity, substantial computing power and storage capabilities. It would be expensive to have required technological infrastructure for healthcare organizations. These enablers are interrelated and assuring such technological infrastructure could be expensive for short-term but provides greater return [14]. For instance, through data-driven insights, healthcare organizations can make informed decisions that improve operational efficiency, reduce costs, and enhance patient care. Thus, Cost-Effective (E8) solutions enable healthcare organizations to acquire the required infrastructure at affordable prices, making metaverse adoption financially viable and accessible. Moreover, the government policies should also incorporate subsidies or tax-rebates for setting up technological infrastructure resulting in faster adoption rate of metaverse.

Summarily, the enablers discussed above drive the adoption of the metaverse in the healthcare sector by leveraging technological infrastructure, enhancing collaboration and communication. As a result, it would potentially empower telemedicine and remote care, medical training and education, patients' experience, accelerating research and development, ensuring data privacy and security. No matter how novel an innovation is, it will ultimately fail if it is not easy for people to use. It is also worth noting that, like any other technology, the metaverse can lead to addiction. This may result in one of the most well-known psychological and behavioural maladaptation, Fear of Missing Out (FOMO). Alternatively, a lack of physical presence may contribute to socio-ethical difficulties. Providing HS via metaverse must be easy and safe for both healthcare providers and consumers for its successful adoption. At the end, we observed that adoption of metaverse in the healthcare sector is dependent on the intertwining of tech and non-tech factors. The enablers collectively contribute to improved healthcare delivery, research, and patient outcomes through the transformative potential of the metaverse.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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 Engineering Management. 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/TEM.2024.3400401

V. Implications of the Study

This section emphasizes on the explanation of the value delivered from this study from theoretical and managerial perspective.

A. Theoretical implications

The healthcare industry is anticipating a similar paradigm shift from metaverse as ICT did. Therefore, the implementation of the metaverse requires scientists and researchers in the healthcare sector to find relevant research issues from various perspective, prioritizing the specific needs and neglected populations within emerging economies. This includes infrastructure limitations, access to quality HS, and lack of trained practitioners.

The healthcare industry has access to numerous interesting new opportunities because of the metaverse. The impact of the metaverse on the healthcare sector is still a subject of many interesting research issues. Prior studies have looked at the technological aspects of metaverse adoption in the healthcare sector [97], [110]. In comparison with the prior studies, this study provides the viewpoint on the adoption of metaverse in the healthcare sector, particularly for neglected populations in emerging economies from the management perspective. As a result, the current study emerges promising in bridging important research gaps in the context of technology adoption in emerging economies. The study used CSF theory to extract the enablers and mixed-method approach to analyse those enablers that drive metaverse adoption in the healthcare industry in emerging economies. These enablers contribute to the existing knowledge of the factors for IT-enabled healthcare delivery, primarily emphasising virtual healthcare. The application of CSF theory not only aids in understanding the factors contributing to the adoption of metaverse technologies but also offers practical implications for organizational success. By identifying and addressing these key enablers, healthcare organizations can streamline their adoption processes, optimize resource allocation, and ultimately enhance their operational efficiency and effectiveness. The hierarchical model developed for the study seems affirm in terms of providing a foundation for future studies to comprehend the connections between various enablers. The outcomes of this study further highlight the dynamic nature of enablers of technology adoption in healthcare and why these enablers need to evolve for continued success in embracing innovations like metaverse. Additionally, the enablers investigated in this study can serve as the foundational variables to develop research hypotheses or develop questionnaire to further explore the domain.

B. Managerial Implications

Businesses have several hurdles to satisfy their customers in the current economic environment. The healthcare sector is no exception. The adoption of innovative technology like metaverse may catalyse improved customer experience. Our research reveals several critical enablers, such as telepresence, personalization, anonymization, government regulations and linkability, among others, that could impact the use of metaverse in healthcare and aid healthcare providers and policymakers to leverage metaverse to address rural healthcare challenges especially of emerging economies, ultimately leading towards healthy societies. One of the main takeaways for managers is to recognize the significance of telepresence in addressing healthcare challenges faced in rural communities. It has emerged as the most significant enabler in adopting metaverse in providing HS to rural communities of emerging economies. Telepresence facilitates the core HS through remote monitoring, virtual consultation, education, medical training and virtual environments for medical research. Managers must prioritize anonymization in the adoption of metaverse. As it enables patients to receive medical services without disclosing their real identity. In the offline space, ensuring patients' anonymity is extremely difficult and still not considered a priority. It is one of the primary reasons many patients suffering from different life-threatening diseases, avoid consulting doctors, especially in rural areas. Anonymization is important, especially in cases where patients suffer from health issues considered undesirable in society, like HIV, abortion, Psychiatric consultation etc. From a managerial perspective, Technology adoption, in general, is high among tech-savvy people who make themselves familiar with innovations and are often among early adopters. Although familiarization is not something metaverse as technology provides, it has emerged as the third most important enabler in adopting metaverse in healthcare.

Another key implication of this study is the need to recognize the importance of linkability. It reduces the friction in the system for both healthcare providers and consumers through effective use of knowledge management. Manual record-keeping of consultations and prescriptions from different healthcare departments poses a substantial challenge in rural healthcare settings, where the literacy rate is usually low, and people have limited resources at their disposal to live a healthy life. Seamless integration of the metaverse system with existing systems enables healthcare institutions to adopt the metaverse with minimal changes in the existing working system, removing the burden of manual record-keeping from the patients. Additionally, the government should promote the seamless integration and encourage best practices throughout the healthcare sector. Managers should focus on technological infrastructure, as it is the core for working metaverse-based healthcare systems. High-speed internet connectivity, continuous supply of electricity, immersive technologies like AR/VR, interoperability and data security and privacy play an important role in building a robust technological infrastructure for a metaverse-based healthcare system that could solve many problems in rural areas of emerging economies and safeguard them from any potential exploitations. Moreover, a collaborative regulation system comprising healthcare organizations and governments should be in place to increase customer trustworthiness. Further, decision-makers should set up an effective strategy, which is essential if you want to use the metaverse to revolutionize the entire healthcare sector. Overall, the proposed model would help practitioners and policymakers in better metaverse adoption decision making.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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 Engineering Management. 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/TEM.2024.3400401

VI. Conclusion

Metaverse has potential to transform the existing healthcare practices through its variety of affordances. However, its adoption in healthcare sector is still in its initial stage and thus requires thorough examination of its potential challenges and opportunities. In line with this, the present study examined enablers of metaverse adoption in healthcare sector. The study adopted integrated theories and methodologies combining critical success factor theory, grey-OPA and modified TISM to meet the research objectives. The integrated methodology aided in investigating the significance and important relationships of the enablers. Chosen enablers from the literature review, based on CSF theory, were presented to experts for feedback and finalization. Grey-OPA was employed to determine priority weights, and the enablers were subsequently ranked accordingly. The mTISM was utilized to examine interconnections between the finalized enablers. This resulted in a hierarchical model that represented the connections between the enablers. In terms of importance from the practitioners' perspective, the study identified thirteen significant enablers with Telepresence (E2), Anonymization (E13) and Familiarization (E3) as the topmost important enablers for successful adoption of metaverse while High Powered Analytics (E9), Data Confidentiality (E7) and Cost (E8) are perceived to have comparatively lower impact on the adoption. Based on the TSIM analysis, Government Regulations and Policies (E10) and Social Influence (E6) are identified as leading drivers and play a crucial role in the adoption as depicted in the m-TISM model which highlighted that these two enablers are building blocks of the hierarchical model. As a result, healthcare organizations should focus on social influence to accelerate metaverse adoption in the healthcare sector through focused marketing campaigns, user testimonials, and community-building efforts. The research contributes significantly to the literature by presenting a hierarchical model for metaverse adoption in the healthcare sector.

The present study has few limitations too. To begin with, the present study involved seven experts in total from an Indian healthcare and technology innovation set up. Although, the experts were carefully chosen but the lack of first-hand experience and the limited understanding about the metaverse technology could not be ruled out and thus could limit the generalizability of the findings. Additionally, the scope of the current study is limited to the adoption of metaverse in healthcare sector, primarily focusing on the doctor-patient relationship. Second, the study did not cover the ethical issues, or the challenges involved with the metaverse which could have significant impact on adoption of metaverse. Lastly, the inputs from the potential consumers of the healthcare sector i.e., patients have not been covered in this study. These limitations could be addressed by the researchers building on the findings of this study. Precisely, future research is needed to understand the metaverse adoption exclusively from the patient's perspective by keeping ethical practices at the centre. Additionally, the stakeholders' rights, and legal concerns regarding metaverse adoption in the healthcare sector should also be investigated.

Appendix

Appendix I: Reachability Matrix with Transitive Links

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13
E1	1	1	1	0	1	0	1*	1*	1*	0	1*	0	0
E2	0	1	0	0	0	0	0	0	0	0	1	0	0
E3	0	1	1	1	0	0	1*	1*	1*	0	1*	0	0
E4	0	1*	1	1	0	0	1	1*	1*	0	1*	0	1
E5	0	1*	1	0	1	0	0	1	1*	0	1*	0	0
E6	0	1*	1*	1	0	1	1	1*	1*	1	1*	1*	1*
E7	0	1	0	0	0	0	1	1	1*	0	1*	0	0
E8	0	1*	0	0	1	0	1	1	1	0	1*	0	0
E9	0	1*	0	0	1*	0	1*	1	1	0	1	0	0
E10	1	1*	1*	1*	1*	1	1*	1*	1	1	1	1	1*
E11	0	1	0	0	0	0	0	0	0	0	1	0	0
E12	1	1*	1*	0	1*	0	1*	1*	1	0	1	1	1
E13	1	1*	1*	0	1*	0	1*	1*	1	0	1	0	1

Appendix II: Level Partitioning (Iteration 1-7)

Enablers	Reachability Set	Antecedent Set	Intersection Set	Level
E2	2, 11	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	2, 11	I
E11	2, 11	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	2, 11	I
E7	7, 8, 9	1, 3, 4, 6, 7, 8, 9, 10, 12, 13	7, 8, 9	II
E8	5, 7, 8, 9	1, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13	5, 7, 8, 9	II
E9	5, 7, 8, 9	1, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13	5, 7, 8, 9	II
E3	3, 4	1, 3, 4, 6, 10, 12, 13	3, 4	III
E1	1	1, 10, 12, 13	1	IV
E5	-	1, 10, 12, 13	-	IV
E13	13	6, 10, 12, 13	13	V
E4	-	6, 10	-	VI
E12	12	6, 10, 12	12	VI
E6	6, 10	6, 10	6, 10	VII
E10	6, 10	6, 10	6, 10	VII

ACKNOWLEDGMENT

The authors would like to thank the editorial team of the journal and anonymous reviewers for improving the manuscript. The authors would also acknowledge the infrastructural and technical support provided by FORE School of Management, Delhi, India.

References

[1] A. Agarwal and S. Alathur, "Metaverse revolution and the digital transformation: intersectional analysis of Industry 5.0," *Transforming Government: People, Process and Policy*, vol. 17, no. 4, pp. 688–707, Nov. 2023, doi: 10.1108/TG-03-2023-0036.
[2] C. Hackl, "Defining The Metaverse Today," Forbes.
[3] H. Koss, "What Is the Metaverse, Really?," builtin.
[4] S.-M. Park and Y.-G. Kim, "A Metaverse: Taxonomy, Components, Applications, and Open Challenges," *IEEE Access*, vol. 10, pp. 4209–4251, 2022, doi: 10.1109/ACCESS.2021.3140175.
[5] Y. K. Dwivedi *et al.*, "Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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 Engineering Management. 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/TEM.2024.3400401

practice and policy," *Int J Inf Manage*, vol. 66, Oct. 2022, doi: 10.1016/j.ijinfomgt.2022.102542.

[6] Wiederhold Brenda K. and Giuseppe Riva, "Metaverse creates new opportunities in healthcare," *Annua1 Review of Cybertherapy and Telemedicine*, vol. 20, pp. 3–7, 2022.

[7] M. A. Z. Chudhery, S. Safdar, J. Huo, H.-U. Rehman, and R. Rafique, "Proposing and Empirically Investigating a Mobile-Based Outpatient Healthcare Service Delivery Framework Using Stimulus–Organism–Response Theory," *IEEE Trans Eng Manag*, vol. 70, no. 8, pp. 2668–2681, Aug. 2023, doi: 10.1109/TEM.2021.3081571.

[8] J. Lee and K. H. Kwon, "Future value and direction of cosmetics in the era of metaverse," *J Cosmet Dermatol*, vol. 21, no. 10, pp. 4176–4183, Oct. 2022, doi: 10.1111/jocd.14794.

[9] A. Almarzouqi, A. Aburayya, and S. A. Salloum, "Prediction of User's Intention to Use Metaverse System in Medical Education: A Hybrid SEM-ML Learning Approach," *IEEE Access*, vol. 10, pp. 43421–43434, 2022, doi: 10.1109/ACCESS.2022.3169285.

[10] P. M. G. Emmelkamp and K. Meyerbröker, "Virtual Reality Therapy in Mental Health," *Annu Rev Clin Psychol*, vol. 17, no. 1, pp. 495–519, May 2021, doi: 10.1146/annurev-clinpsy-081219-115923.

[11] InsightAce, "Metaverse in healthcare market size, share and trend analysis," Feb. 2023.

[12] G. Bansal, K. Rajgopal, V. Chamola, Z. Xiong, and D. Niyato, "Healthcare in Metaverse: A Survey on Current Metaverse Applications in Healthcare," *IEEE Access*, vol. 10, pp. 119914–119946, 2022, doi: 10.1109/ACCESS.2022.3219845.

[13] Wiederhold Brenda K., "Metaverse Games: Game Changer for Healthcare?," *Cyberpsychol Behav Soc Netw*, vol. 25, no. 5, pp. 267–269, May 2022, doi: 10.1089/cyber.2022.29246.editorial.

[14] R. Chengoden *et al.*, "Metaverse for Healthcare: A Survey on Potential Applications, Challenges and Future Directions," *IEEE Access*, vol. 11, pp. 12765–12795, 2023, doi: 10.1109/ACCESS.2023.3241628.

[15] A. Garavand and N. Aslani, "Metaverse phenomenon and its impact on health: A scoping review," *Inform Med Unlocked*, vol. 32, p. 101029, 2022, doi: 10.1016/j.imu.2022.101029.

[16] U. B. Mir, S. Sharma, A. K. Kar, and M. P. Gupta, "Critical success factors for integrating artificial intelligence and robotics," *Digital Policy, Regulation and Governance* , no. March, 2020, doi: 10.1108/DPRG-03-2020-0032.

[17] O. J. Gupta, S. Yadav, M. K. Srivastava, P. Darda, and V. Mishra, "Understanding the intention to use metaverse in healthcare utilizing a mix method approach," *Int J Healthc Manag*, pp. 1–12, Mar. 2023, doi: 10.1080/20479700.2023.2183579.

[18] Y. Yang, K. Siau, W. Xie, and Y. Sun, "Smart Health," *Journal of Organizational and End User Computing*, vol. 34, no. 1, pp. 1–14, Aug. 2022, doi: 10.4018/JOEUC.308814.

[19] T. Huynh-The *et al.*, "Blockchain for the metaverse: A Review," *Future Generation Computer Systems*, vol. 143, pp. 401–419, Jun. 2023, doi: 10.1016/j.future.2023.02.008.

[20] M. Brahma *et al.*, "Learning impact of recent ICT advances based on virtual reality IoT sensors in a metaverse environment," *Measurement: Sensors*, vol. 27, p. 100754, Jun. 2023, doi: 10.1016/j.measen.2023.100754.

[21] K. Bhugaonkar, R. Bhugaonkar, and N. Masne, "The Trend of Metaverse and Augmented & Virtual Reality Extending to the Healthcare System," *Cureus*, Sep. 2022, doi: 10.7759/cureus.29071.

[22] Ficklin Jared and Rolston Mark, "The metaverse will not look the way Facebook imagines it," Venturebeat.

[23] S. Ramzan, A. Aqdus, V. Ravi, D. Koundal, R. Amin, and M. A. Al Ghamdi, "Healthcare Applications Using Blockchain Technology: Motivations and Challenges," *IEEE Trans Eng Manag*, vol. 70, no. 8, pp. 2874–2890, Aug. 2023, doi: 10.1109/TEM.2022.3189734.

[24] P. Bhattacharya, M. S. Obaidat, D. Savaliya, S. Sanghavi, S. Tanwar, and B. Sadaun, "Metaverse assisted Telesurgery in Healthcare 5.0: An interplay of Blockchain and Explainable AI," in *2022 International Conference on Computer, Information and Telecommunication Systems (CITS)*, IEEE, Jul. 2022, pp. 1–5. doi: 10.1109/CITS55221.2022.9832978.

[25] R. Manne and S. C. Kantheti, "Application of Artificial Intelligence in Healthcare: Chances and Challenges," *Current Journal of Applied Science and Technology*, pp. 78–89, Apr. 2021, doi: 10.9734/cjast/2021/v40i631320.

[26] E. Broadbent *et al.*, "Attitudes towards health-care robots in a retirement village," *Australas J Ageing*, vol. 31, no. 2, pp. 115–120, Jun. 2012, doi: 10.1111/j.1741-6612.2011.00551.x.

[27] R. Ganiga, M. R. Pai, M. M. M. Pai, and R. K. Sinha, "Cloud Enabled Standard Electronic Health Record Architecture for Indian Healthcare Sector," *Indian J Public Health Res Dev*, vol. 8, no. 4, p. 554, 2017, doi: 10.5958/0976-5506.2017.00398.9.

[28] N. Upadhyay, A. Kamble, and A. Navare, "Virtual healthcare in the new normal: Indian healthcare consumers adoption of electronic government telemedicine service," *Gov Inf Q*, vol. 40, no. 2, p. 101800, Apr. 2023, doi: 10.1016/j.giq.2022.101800.

[29] M. S. Rahman and M. Ko, "Toward Systematic Identification of Stakeholders for Healthcare Information Systems: A Feature-Based Method," in *Proceedings of the Nineteenth Americas Conference on Information Systems*, Chicago, Illinois, 2013.

[30] George M. Gatero, "Utilization of ICTs for Accessing Health Information by Medical Professionals in Kenya: A Case Study of Kenyatta National Hospital," *Journal of Health Informatics in Developing CountriesGeorge M. GATERO*, vol. 5, no. 1, Dec. 2010.

[31] L. Iaia, C. Nespoli, F. Vicentini, M. Pironti, and C. Genovino, "Supporting the implementation of AI in business communication: the role of knowledge management," *Journal of Knowledge Management*, Apr. 2023, doi: 10.1108/JKM-12-2022-0944.

[32] X. Zhang, T. Xu, X. Wei, J. Tang, and P. Ordonez de Pablos, "The establishment of transactive memory system in distributed agile teams engaged in AI-related knowledge work," *Journal of Knowledge Management*, Jun. 2023, doi: 10.1108/JKM-10-2022-0791.

[33] G. Krishnan, "Metaverse and Healthcare A Clinician's Perspective," *Apollo Medicine*, vol. 19, no. 4, pp. 256–261, 2022.

[34] K. Bhugaonkar, R. Bhugaonkar, and N. Masne, "The trend of Metaverse and augmented & virtual reality extending to the healthcare system," *Cureus*, vol. 14, no. 9, 2022.

[35] S. Siripurapu, N. K. Darimireddy, A. Chehri, and P. AV, "Technological Advancements and Elucidation Gadgets for Healthcare Applications: An Exhaustive Methodological Review-Part-II (Robotics, Drones, 3D-Printing, Internet of Things, Virtual/Augmented and Mixed Reality)," *Electronics (Basel)*, vol. 12, no. 3, p. 548, 2023.

[36] L. Soltanisehat, R. Alizadeh, H. Hao, and K.-K. R. Choo, "Technical, Temporal, and Spatial Research Challenges and Opportunities in Blockchain-Based Healthcare: A Systematic Literature Review," *IEEE Trans Eng Manag*, vol. 70, no. 1, pp. 353–368, 2023, doi: 10.1109/TEM.2020.3013507.

[37] S. Siripurapu, N. K. Darimireddy, A. Chehri, S. B., and P. A.V., "Technological Advancements and Elucidation Gadgets for Healthcare Applications: An Exhaustive Methodological Review-Part-II (Robotics, Drones, 3D-Printing, Internet of Things, Virtual/Augmented and Mixed Reality)," *Electronics (Basel)*, vol. 12, no. 3, p. 548, Jan. 2023, doi: 10.3390/electronics12030548.

[38] N. Omrani, N. Rejeb, A. Maalaoui, M. Dabic, and S. Kraus, "Drivers of Digital Transformation in SMEs," *IEEE Trans Eng Manag*, pp. 1–14, 2022, doi: 10.1109/TEM.2022.3215727.

[39] Y. K. Dwivedi *et al.*, "Exploring the Darkverse: A Multi-Perspective Analysis of the Negative Societal Impacts of the Metaverse," *Information Systems Frontiers*, Jun. 2023, doi: 10.1007/s10796-023-10400-x.

[40] D. Golinelli, E. Boetto, G. Carullo, A. G. Nuzzolese, M. P. Landini, and M. P. Fantini, "Adoption of Digital Technologies in Health Care During the COVID-19 Pandemic: Systematic Review of Early Scientific Literature," *J Med Internet Res*, vol. 22, no. 11, p. e22280, Nov. 2020, doi: 10.2196/22280.

[41] E. O. Popa, M. van Hilten, E. Oosterkamp, and M.-J. Bogaardt, "The use of digital twins in healthcare: socio-ethical benefits and socio-ethical risks," *Life Sci Soc Policy*, vol. 17, no. 1, p. 6, Dec. 2021, doi: 10.1186/s40504-021-00113-x.

[42] M. Attaran, "Blockchain technology in healthcare: Challenges and opportunities," *Int J Healthc Manag*, vol. 15, no. 1, pp. 70–83, Jan. 2022, doi: 10.1080/20479700.2020.1843887.

[43] P. Dhar, T. Rocks, R. M. Samarasinghe, G. Stephenson, and C. Smith, "Augmented reality in medical education: students' experiences and learning outcomes," *Med Educ Online*, vol. 26, no. 1, Jan. 2021, doi: 10.1080/10872981.2021.1953953.

[44] T. Saba, K. Haseeb, I. Ahmed, and A. Rehman, "Secure and energy-efficient framework using Internet of Medical Things for e-healthcare," *J Infect Public Health*, vol. 13, no. 10, pp. 1567–1575, Oct. 2020, doi: 10.1016/j.jiph.2020.06.027.

[45] C. Nnaji and A. A. Karakhan, "Technologies for safety and health management in construction: Current use, implementation benefits and limitations, and adoption barriers," *Journal of Building Engineering*, vol. 29, p. 101212, May 2020, doi: 10.1016/j.jobe.2020.101212.

[46] C. H. Tsai, A. Eghdam, N. Davoody, G. Wright, S. Flowerday, and S. Koch, "Effects of Electronic Health Record Implementation and Barriers to Adoption and Use: A Scoping Review and Qualitative Analysis of the Content," *Life*, vol. 10, no. 12, p. 327, Dec. 2020, doi: 10.3390/life10120327.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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 Engineering Management. 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/TEM.2024.3400401

[47] Hasanain Rihab Abdulaziz and Cooper Helen, "Solutions to Overcome Technical and Social Barriers to Electronic Health Records Implementation in Saudi Public and Private Hospitals," *J Health Inform Dev Ctries*, vol. 8, no. 1, 2014.
[48] S. S. Furusa and A. Coleman, "Factors influencing e-health implementation by medical doctors in public hospitals in Zimbabwe," *SA Journal of Information Management*, vol. 20, no. 1, Jun. 2018, doi: 10.4102/sajim.v20i1.928.
[49] A. Awad *et al.*, "Connected healthcare: Improving patient care using digital health technologies," *Adv Drug Deliv Rev*, vol. 178, p. 113958, Nov. 2021, doi: 10.1016/j.addr.2021.113958.
[50] A. Agarwal and S. Alathur, "Metaverse revolution and the digital transformation: intersectional analysis of Industry 5.0," *Transforming Government: People, Process and Policy*, vol. 17, no. 4, pp. 688–707, Nov. 2023, doi: 10.1108/TG-03-2023-0036.
[51] L. Petrigna and G. Musumeci, "The Metaverse: A New Challenge for the Healthcare System: A Scoping Review," *J Funct Morphol Kinesiol*, vol. 7, no. 3, p. 63, Aug. 2022, doi: 10.3390/jfmk7030063.
[52] S. Huh, "Application of computer-based testing in the Korean Medical Licensing Examination, the emergence of the metaverse in medical education, journal metrics and statistics, and appreciation to reviewers and volunteers," *J Educ Eval Health Prof*, vol. 19, p. 2, Jan. 2022, doi: 10.3352/jeehp.2022.19.2.
[53] Thomason Jane, "MetaHealth - How will the Metaverse Change Health Care?," *Journal of Metaverse*, vol. 1, no. 1, 2021.
[54] D. Yang *et al.*, "Expert consensus on the metaverse in medicine," *Clinical eHealth*, vol. 5, pp. 1–9, Dec. 2022, doi: 10.1016/j.ceh.2022.02.001.
[55] P. Henz, "The societal impact of the metaverse," *Discover Artificial Intelligence*, vol. 2, no. 1, p. 19, Oct. 2022, doi: 10.1007/s44163-022-00032-6.
[56] S. Puntoni, R. W. Reczek, M. Giesler, and S. Botti, "Consumers and Artificial Intelligence: An Experiential Perspective," *J Mark*, vol. 85, no. 1, pp. 131–151, Jan. 2021, doi: 10.1177/0022242920953847.
[57] S. S. Usmani, M. Sharath, and M. Mehendale, "Future of mental health in the metaverse," *Gen Psychiatr*, vol. 35, no. 4, p. e100825, Aug. 2022, doi: 10.1136/gpsych-2022-100825.
[58] D. B. Yaden, J. C. Eichstaedt, and J. D. Medaglia, "The Future of Technology in Positive Psychology: Methodological Advances in the Science of Well-Being," *Front Psychol*, vol. 9, Jun. 2018, doi: 10.3389/fpsyg.2018.00962.
[59] Cheng Xusen, Mou Jian, Shen Xiao-Liang, Vreede Triparna, and Alt Rainer, "Special issue "Opportunities and challenges in the Metaverse," *Internet Research*, 2022.
[60] R. (Raj) G. Javalgi and D. A. Grossman, "Aspirations and entrepreneurial motivations of middle-class consumers in emerging markets: The case of India," *International Business Review*, vol. 25, no. 3, pp. 657–667, Jun. 2016, doi: 10.1016/j.ibusrev.2015.10.008.
[61] A. Karan *et al.*, "Size, composition and distribution of health workforce in India: why, and where to invest?," *Hum Resour Health*, vol. 19, no. 1, p. 39, Dec. 2021, doi: 10.1186/s12960-021-00575-2.
[62] A. Athar, S. M. Ali, M. A. I. Mozumder, S. Ali, and H.-C. Kim, "Applications and Possible Challenges of Healthcare Metaverse," in *2023 25th International Conference on Advanced Communication Technology (ICACT)*, IEEE, Feb. 2023, pp. 328–332. doi: 10.23919/ICACT56868.2023.10079314.
[63] Z. Teng, Y. Cai, Y. Gao, X. Zhang, and X. Li, "Factors Affecting Learners' Adoption of an Educational Metaverse Platform: An Empirical Study Based on an Extended UTAUT Model," *Mobile Information Systems*, vol. 2022, pp. 1–15, Aug. 2022, doi: 10.1155/2022/5479215.
[64] N. Mkedder and M. Das, "Metaverse integration challenges: An in-depth ISM and MICMAC analysis," *Journal of Retailing and Consumer Services*, vol. 77, p. 103684, Mar. 2024, doi: 10.1016/j.jretconser.2023.103684.
[65] I. A. Akour, R. S. Al-Maroof, R. Alfaisal, and S. A. Salloum, "A conceptual framework for determining metaverse adoption in higher institutions of gulf area: An empirical study using hybrid SEM-ANN approach," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100052, 2022, doi: 10.1016/j.caeai.2022.100052.
[66] A. Koohang *et al.*, "Shaping the Metaverse into Reality: A Holistic Multidisciplinary Understanding of Opportunities, Challenges, and Avenues for Future Investigation," *Journal of Computer Information Systems*, vol. 63, no. 3, pp. 735–765, May 2023, doi: 10.1080/08874417.2023.2165197.
[67] R. Aria, N. Archer, M. Khanlari, and B. Shah, "Influential Factors in the Design and Development of a Sustainable Web3/Metaverse and Its Applications," *Future Internet*, vol. 15, no. 4, p. 131, Mar. 2023, doi: 10.3390/fi15040131.
[68] Y. Ataei, A. Mahmoudi, M. R. Feylizadeh, and D.-F. Li, "Ordinal Priority Approach (OPA) in Multiple Attribute Decision-Making," *Appl Soft Comput*, vol. 86, p. 105893, Jan. 2020, doi: 10.1016/j.asoc.2019.105893.
[69] V. Shardeo, J. Madaan, and F. T. S. Chan, "An empirical analysis of freight mode choice factors amid the COVID-19 outbreak," *Industrial Management & Data Systems*, vol. 122, no. 12, pp. 2783–2805, Nov. 2022, doi: 10.1108/IMDS-03-2022-0133.
[70] G. B. Davis, "An Opinion... 'Comments on the Critical Success Factors Method for Obtaining Management Information Requirements in Article by John F. Rockart, "Chief Executives Define Their Own Data Needs," Harvard Business Review, March-April, 1979,'" *MIS Quarterly*, vol. 3, no. 3, pp. 57–58, 1979.
[71] Christine, V., J. Bullen, F., and Rockart, "A Primer on Critical Success Factors," *The Rise of Management Computing*, 1981, [Online]. Available: https://www.researchgate.net/publication/5175561_A_primer_on_critical_success_factors
[72] U. B. Mir, A. K. Kar, Y. K. Dwivedi, M. P. Gupta, and R. S. Sharma, "Realizing digital identity in government: Prioritizing design and implementation objectives for Aadhaar in India," *Gov Inf Q*, vol. 37, no. 2, p. 101442, 2020, doi: 10.1016/j.giq.2019.101442.
[73] World Bank, "Rural Population (% of total population - India," World Bank.
[74] D. C. Sharma, "India still struggles with rural doctor shortages," *The Lancet*, vol. 386, no. 10011, pp. 2381–2382, Dec. 2015, doi: 10.1016/S0140-6736(15)01231-3.
[75] Venkatesh Viswanath, Sykes Tracy, and Zhang Xiaojun, "ICT for Development in Rural India: A Longitudinal Study of Women's Health Outcomes," *MIS Quarterly*, vol. 44, no. 2, pp. 605–629, 2020.
[76] M. Mohanan, K. Hay, and N. Mor, "Quality Of Health Care In India: Challenges, Priorities, And The Road Ahead," *Health Aff*, vol. 35, no. 10, pp. 1753–1758, Oct. 2016, doi: 10.1377/hlthaff.2016.0676.
[77] Kasthuri Arvind, "Challenges to Healthcare in India - The Five A's," *Indian J Community Med*, vol. 43, no. 3, pp. 141–143, 2018.
[78] PWC, "The Future of Healthcare in the Metaverse," PWC.
[79] NITI Aayog, "National Strategy for Artificial Intelligence," 2018.
[80] I. J. Orji, S. Kusi-Sarpong, and H. Gupta, "The critical success factors of using social media for supply chain social sustainability in the freight logistics industry," *Int J Prod Res*, vol. 58, no. 5, pp. 1522–1539, Mar. 2020, doi: 10.1080/00207543.2019.1660829.
[81] A. Mahmoudi, X. Deng, S. A. Javed, and N. Zhang, "Sustainable Supplier Selection in Megaprojects: Grey Ordinal Priority Approach," *Bus Strategy Environ*, vol. 30, no. 1, pp. 318–339, Jan. 2021, doi: 10.1002/bse.2623.
[82] C. Gayathri, V. Kamala, M. S. Gajanand, and S. Yamini, "Analysis of operational and financial performance of ports: an integrated fuzzy DEMATEL-TOPSIS approach," *Benchmarking: An International Journal*, vol. 29, no. 3, pp. 1046–1066, Mar. 2022, doi: 10.1108/BIJ-03-2020-0123.
[83] B. D. Sarkar and R. Shankar, "Understanding the barriers of port logistics for effective operation in the Industry 4.0 era: Data-driven decision making," *International Journal of Information Management Data Insights*, vol. 1, no. 2, p. 100031, Nov. 2021, doi: 10.1016/j.jjimei.2021.100031.
[84] B. D. Sarkar, R. Shankar, and A. K. Kar, "Severity analysis and risk profiling of port logistics barriers in the Industry 4.0 era," *Benchmarking: An International Journal*, vol. 30, no. 9, pp. 3253–3280, Dec. 2023, doi: 10.1108/BIJ-03-2022-0153.
[85] S. Onut, U. R. Tuzkaya, and E. Torun, "Selecting container port via a fuzzy ANP-based approach: A case study in the Marmara Region, Turkey," *Transp Policy (Oxf)*, vol. 18, no. 1, pp. 182–193, Jan. 2011, doi: 10.1016/j.tranpol.2010.08.001.
[86] Sushil and Sushil, "Interpretive Matrix: A Tool to Aid Interpretation of Management and Social Research," *Global Journal of Flexible Systems Management*, vol. 6, no. 2, pp. 27–30, 2005.
[87] R. Rajesh, "Technological Forecasting & Social Change Technological capabilities and supply chain resilience of fi rms : A relational analysis using Total Interpretive Structural Modeling (TISM)," *Technol Forecast Soc Change*, vol. 118, pp. 161–169, 2017, doi: 10.1016/j.techfore.2017.02.017.
[88] Sushil, "Incorporating polarity of relationships in ISM and TISM for theory building in information and organization management," *Int J Inf Manage*, vol. 43, pp. 38–51, 2018.
[89] Sushil, "Modified ISM/TISM Process with Simultaneous Transitivity Checks for Reducing Direct Pair Comparisons," *Global Journal of Flexible Systems Management*, vol. 18, no. 4, pp. 331–351, Dec. 2017, doi: 10.1007/s40171-017-0167-3.
[90] Sushil and K. K. Dinesh, "Structured Literature Review with TISM Leading to an Argumentation Based Conceptual Model," *Global Journal of Flexible Systems Management*, vol. 23, no. 3, pp. 387–407, Sep. 2022, doi: 10.1007/s40171-022-00309-w.
[91] R. Rajan, S. Dhir, and Sushil, "Technology management for innovation in organizations: an argumentation-based modified TISM approach,"

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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 Engineering Management. 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/TEM.2024.3400401

Benchmarking: An International Journal, vol. 28, no. 6, pp. 1959–1986, Jun. 2021, doi: 10.1108/BIJ-01-2020-0019.
[92] A. Meena, S. Dhir, and Sushil, "An analysis of growth-accelerating factors for the Indian automotive industry using modified TISM," *International Journal of Productivity and Performance Management*, vol. 70, no. 6, pp. 1361–1392, Jun. 2021, doi: 10.1108/IJPPM-01-2019-0047.
[93] V. Shardeo, A. Patil, and J. Madaan, "Critical Success Factors for Blockchain Technology Adoption in Freight Transportation Using Fuzzy ANP–Modified TISM Approach," *Int J Inf Technol Decis Mak*, vol. 19, no. 06, pp. 1549–1580, Nov. 2020, doi: 10.1142/S0219622020500376.
[94] K. Sharma, M. Jain, and S. Dhir, "Analysing the impact of artificial intelligence on the competitiveness of tourism firms: a modified total interpretive structural modeling (m-TISM) approach," *International Journal of Emerging Markets*, vol. 17, no. 4, pp. 1067–1084, May 2022, doi: 10.1108/IJOEM-05-2021-0810.
[95] R. Sindhwani, A. Behl, A. Sharma, and J. Gaur, "What makes micro, small, and medium enterprises not adopt Logistics 4.0? A systematic and structured approach using modified-total interpretive structural modelling," *International Journal of Logistics Research and Applications*, pp. 1–26, May 2022, doi: 10.1080/13675567.2022.2081672.
[96] Rania. A. Mohamed, K. K. Mohammed, A. Darwish, and A. E. Hassanien, "The Use of Metaverse in the Healthcare Sector: Analysis and Applications," 2023, pp. 77–91. doi: 10.1007/978-3-031-29132-6_5.
[97] C. W. Lee, "Application of Metaverse Service to Healthcare Industry: A Strategic Perspective," *Int J Environ Res Public Health*, vol. 19, no. 20, p. 13038, Oct. 2022, doi: 10.3390/ijerph192013038.
[98] I. Arpaci, K. Karatas, I. Kusci, and M. Al-Emran, "Understanding the social sustainability of the Metaverse by integrating UTAUT2 and big five personality traits: A hybrid SEM-ANN approach," *Technol Soc*, vol. 71, p. 102120, Nov. 2022, doi: 10.1016/j.techsoc.2022.102120.
[99] B. Bhatia and S. Joshi, "Applications of Metaverse in the Healthcare Industry," in *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, IEEE, Mar. 2023, pp. 344–350. doi: 10.1109/ICIDCA56705.2023.10099897.
[100] A. M. Aslam, R. Chaudhary, A. Bhardwaj, I. Budhiraja, N. Kumar, and S. Zeadally, "Metaverse for 6G and Beyond: The Next Revolution and Deployment Challenges," *IEEE Internet of Things Magazine*, vol. 6, no. 1, pp. 32–39, Mar. 2023, doi: 10.1109/IOTM.001.2200248.
[101] T. F. Tan *et al*., "Metaverse and Virtual Health Care in Ophthalmology: Opportunities and Challenges," *Asia-Pacific Journal of Ophthalmology*, vol. 11, no. 3, pp. 237–246, May 2022, doi: 10.1097/APO.0000000000000537.
[102] G. Kang, J. Koo, and Y.-G. Kim, "Security and Privacy Requirements for the Metaverse: A Metaverse Applications Perspective," *IEEE Communications Magazine*, pp. 1–7, 2023, doi: 10.1109/MCOM.014.2200620.
[103] P. K. Tarei, J. J. Thakkar, and B. Nag, "A hybrid approach for quantifying supply chain risk and prioritizing the risk drivers," *Journal of Manufacturing Technology Management*, vol. 29, no. 3, pp. 533–569, Mar. 2018, doi: 10.1108/JMTM-10-2017-0218.
[104] K. F. Pun and I. K. Hui, "An analytical hierarchy process assessment of the ISO 14001 environmental management system," *Integrated Manufacturing Systems*, vol. 12, no. 5, pp. 333–345, Sep. 2001, doi: 10.1108/EUM0000000005711.
[105] J. A. Batsis, S. N. Pletcher, and J. E. Stahl, "Telemedicine and primary care obesity management in rural areas – innovative approach for older adults?," *BMC Geriatr*, vol. 17, no. 1, p. 6, Dec. 2017, doi: 10.1186/s12877-016-0396-x.
[106] V. Shardeo and B. D. Sarkar, "Adoption of hydrogen-fueled freight transportation: A strategy toward sustainability," *Bus Strategy Environ*, Jun. 2023, doi: 10.1002/bse.3482.
[107] Creswell John W. and Poth Cheryl N., *Qualitative Inquiry and research design : choosing among five approaches*. 2017.
[108] A. Singh and A. R. Singla, "Modelling and analysis of factors for implementation of smart cities: TISM approach," *Journal of Modelling in Management*, vol. 17, no. 4, pp. 1587–1622, Nov. 2022, doi: 10.1108/JM2-07-2020-0192.
[109] G. Dicuonzo, F. Donofrio, A. Fusco, and M. Shini, "Healthcare system: Moving forward with artificial intelligence," *Technovation*, vol. 120, p. 102510, Feb. 2023, doi: 10.1016/j.technovation.2022.102510.
[110] G. Wang *et al*., "Development of metaverse for intelligent healthcare," *Nat Mach Intell*, vol. 4, no. 11, pp. 922–929, Nov. 2022, doi: 10.1038/s42256-022-00549-6.



Vipulesh Shardeo is an Assistant Professor in the area of Quantitative Techniques and Operations Management at FORE School of Management, Delhi, India. He has completed his Ph. D. in Operations from Department of Management Studies, IIT Delhi, India. He has done M. Tech. majoring Industrial Engineering and Management from IIT(ISM) Dhanbad, India. His research has been published in various journals of international repute including Industrial Management & Data Systems, International Journal of Information Technology and Decision Making, International Journal of Logistics Management, Sustainable Production and Consumption and so forth. He has also presented his research works in various international conferences or repute including POMS, INFORMS, IEOM and so on.



Bishal Dey Sarkar is working as an Assistant Professor at Symbiosis Institute of Operations Management, Symbiosis International (Deemed University), Pune, India. He earned his Doctorate from Indian Institute of Technology Delhi in 2021. His research appeared in journals like IEEE Transactions on Engineering Management, Business Strategies, and the Environment, Benchmarking: An International Journal, Journal of Modelling in Management, Journal of Global Operations and Strategic Sourcing, Journal of Information Management: Data Insights.



Umar Bashir Mir is an Assistant Professor in the area of Information Systems at Lal Bahadur Shastri Institute of Management, Delhi (LBSIM), India. He has done his PhD in Information Systems Area at the Department of Management Studies, Indian Institute of Technology, Delhi. He has more than seven years of experience in academia and industry. Dr Mir is an active researcher. His research focuses on the topics related to ICT, e-Governance, Artificial Intelligence, Multi-Criteria Decision Making, Social Media, and Digital Identity Systems. His research has been published in various ABDC / ABS category journals and international conference proceedings. He is also a reviewer in various highly reputed journals.



Prerna Kaushik was born in Delhi, India on 12th November, 2002 and is a under graduate with a B. Tech in Engineering Physics from Delhi Technological University. Her academic background is complemented by strong analytical, research and technical skills, including proficiency in Python, SQL, and data visualization tools like Power BI and Tableau. She possesses a diverse experience record, having interned at PricewaterhouseCoopers (PWC).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 05:04:14 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.

2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE 2024)

An Analytical Comparison of Deep Learning Frameworks for 2D Image-Based Hatha-Yoga Pose Identification

Debashree Debalaxmi
Department of Information Technology, Delhi Technological University
Bawana Road, Delhi-110042, India
debashreedebalaxmi@gmail.com

Virender Ranga
Department of Information Technology, Delhi Technological University
Bawana Road, Delhi-110042, India
drvirender.ranga@gmail.com

Dinesh Kumar Vishwakarma
Biometric Research Laboratory, Department of Information Technology, Delhi Technological University
Bawana Road, Delhi-110042, India
dvishwakarma@gmail.com

2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE) | 979-8-3503-8328-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICWITE59797.2024.10503024

***Abstract*— The comprehensive practice of yoga seeks to improve one's mental, bodily, and spiritual well-being. Among the many forms of yoga, hatha yoga is a conventional style that uses physical postures and breath exercises to balance the body and mind. We can reap its greatest health benefits by adopting the ideal postures and adhering to the recommended techniques and sequencing. However, adopting improper postures while practising yoga can result in a number of health issues such as short-term chronic issues or acute muscle discomfort. Therefore, there is a need for scientific evaluation of yoga posture recognition in order to help people practise yoga effectively. To support self-learning, we'll provide a model for classifying poses using posture detection in this paper. Here we have taken a 2D image dataset of 20 famous hathayoga poses and we comparatively evaluated the performance of several cutting-edge deep learning architectures and DenseNet201 yields the highest accuracy of 90.4%.**

Keywords*— Pose recognition, yoga, deep learning, computer vision.

I. INTRODUCTION

With yoga's worldwide popularity growing at an exponential rate, an auto-assisted model that can support and promote properly taught and self-guided yoga practises is urgently needed. The identification of the individual body parts moving in a certain direction to create the desired position is the focus of this issue statement. Yoga positions that are incorrect or incorrect can cause more harm than good because they can result in injuries like muscle pulls, stiff necks, ankle sprains, etc.

Numerous attempts have been made to circumvent these limitations. Software that acts as an AI system trainer has been created using computer vision and data science methods. The benefits of the stance are explained by this model. It also conveys information regarding the performance's correctness. With the aid of this programme, one can practise yoga independently of a trainer. A large number of image datasets including 20 hatha yoga pose features ('Trikonasana', 'ardha matsyendrasana', 'adho mukha svanasana', 'bakasana', 'dhanurasana', 'bhujangasana', 'kapotasana', 'matsyasana', 'paschimottanasana', 'salabhasana', 'savasana', 'setu bandha sarvangasana', 'tadasana', 'uttanasana', 'balasana', 'vajrasana', 'virabhadrasana i', 'virabhadrasana ii', 'virasana', 'vriksasana') have been developed. With the use of computer vision techniques and the tf-position algorithm, these datasets have made it possible to extract important posture information. The precision and effectiveness of the system are increased by harvesting data points, such as coordinates and joint angles, and using them as features for deep learning models. Several deep learning models have been rigorously tested in order to validate the correctness of the model. Through this iterative process, the system's performance and dependability are improved, guaranteeing that practitioners may rely on its direction for the best possible and secure yoga experience.

II. RELATED WORK

The literature contains a large number of papers pertaining to the yoga expert system and posture classification. Of these, Microsoft Kinect is used for pose detection in the majority of the work, and other data mining techniques including SVM, Random Forest, Decision Trees and other Deep Learning frameworks based on CNN. Using a Kinect camera in [1], a self-training technique was created for yoga posture identification. A total of 300 films with five yoga practitioners executing 12 different positions five times each were gathered. Initially, the body map was taken from the movies and the body contour was retrieved. Next, a fast skeletonization method using a star skeleton was employed to derive a human stance descriptor. There was a 99.33% accuracy rate in the system. Edwin et al. [2] have suggested a yoga recognition system that uses Kinect v2 and Ada-boost classification to recognize six asanas with an accuracy score of 94.78%. However, the method takes more processing time for more complicated models. They used a camera that is based on a depth sensor, which is frequently out of reach for customers. A system that could recognize and adjust yoga positions by comparing the practitioner video with the reference video using the SURF algorithm was presented by Fazil et al. [3].

Using Mask RCNN, the researchers turned the dataset into keypoints. A CNN & LSTM model produced the best accuracy of 0.9996. The Yoga-82 dataset was used by Verma et al. [4] to perform keypoint recognition. The dataset originally included a variety of yoga forms that were gathered from the internet. The postures were manually cleaned and grouped into a superclass hierarchy. For

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:51:33 UTC from IEEE Xplore. Restrictions apply.

network learning, this hierarchical annotation offered rich data that included pose details, spine effects, and body positions. Three hierarchical layers, containing 6, 20, and 82 classes at each level, were used to structure their dataset. DenseNet-201 performed remarkably well, attaining 74.91% top-1 classification accuracy. Kothari [5] conducted a thorough investigation on the classification of yoga postures using deep learning algorithms and human pose estimate. They used OpenPose to pre-process the "Yoga Vid Collected" dataset in order to extract posture keypoints. A comparison of deep learning and machine learning techniques for estimating yoga poses was carried out as part of the study. In terms of misclassifications, the hybrid CNN-LSTM model outperformed the SVM and standalone CNN models, showing the fewest errors with validation accuracy of 0.9987.

In [6], the researchers developed a computer-aided system that recognizes yoga poses and offers feedback for correction. They used the hybrid model using the same video dataset, which consisted of six yoga positions performed by fifteen people divided as 60% for training set, 20% for testing set, and last 20% for validation sets. Their approach obtained an exceptional 0.9941 frame-by-frame validation accuracy. Agrawal et al. [7] present a YOGI dataset of around 5500 pictures. Using the "tf-posture," the angles and joints from these image frames were extracted. These characteristics were then input into a number of machine learning models. At 99.04%, the random-forest classifier had the best accuracy. Long et. al. [8] used a dataset gathered from eight volunteers to develop a yoga coaching system that made use of 6 transfer learning models. They divided the dataset into train, validation, and test sets at a ratio of 3:1:1 in order to train their model. They also used strategies for data augmentation on the dataset. The best model was determined to be the TL-MobileNet-DA model as evidenced by its 98.43% total accuracy

Shubham et al. (2022) [9] suggested a CNN-based approach for classifying yoga poses in a different research. They used the MediaPipe library in their study to pinpoint important spots and make a skeletal diagram. The researchers downloaded a dataset from Kaggle that was made accessible to the public, and they obtained the greatest accuracy of 97.09%. A deep learning-based yoga pose estimation method was presented in[10] . It uses 12 identified joints compressed into 13 vectors to evaluate posture accuracy and extracts 18 keypoints. According to their analysis, the MLP classifier had an impressive testing accuracy of 0.9958. Aman et al. [11] presented the three-stage Y-PN-MSSD paradigm for yoga posture recognition. First, it uses an open-source dataset with seven postures and records the poses of four users to collect data. During training, it then connects key body components to extract characteristics. This model allows for 99.88% accuracy in real-time monitoring of users in yoga poses. However, implementation still faces difficulties because of problems like occlusion and uneven illumination.

Using a variety of datasets, these studies illustrate the several methods and strategies used in the categorization of yoga poses. Out of a wide range of categories, the majority of study has been conducted on the five most popular hatha yoga postures (‘Tree’,’Warrior-2’,’Plank’,’Goddess’ and ’Downdog’), which are relatively limited in number. It is necessary to expand the number of categories while retaining high precision and accuracy.

III. Proposed Approach

The process of creating a yoga posture identification system includes gathering a variety of picture datasets, preprocessing information, choosing models that have already been trained, fine-tuning, assessing performance, optimizing hyperparameters, deploying the best model as shown in fig 1.

A. Dataset Collection

There is a publicly accessible dataset for 107 yoga poses in Kaggle [12] from which we have taken 20 different hatha yoga poses each containing 50 to 90 images. There are 1248 photos in our compiling color image dataset. The dataset includes images of several people executing different yoga positions from various perspectives. Path, pixel values, and the associated label (asana index) are kept for every picture.

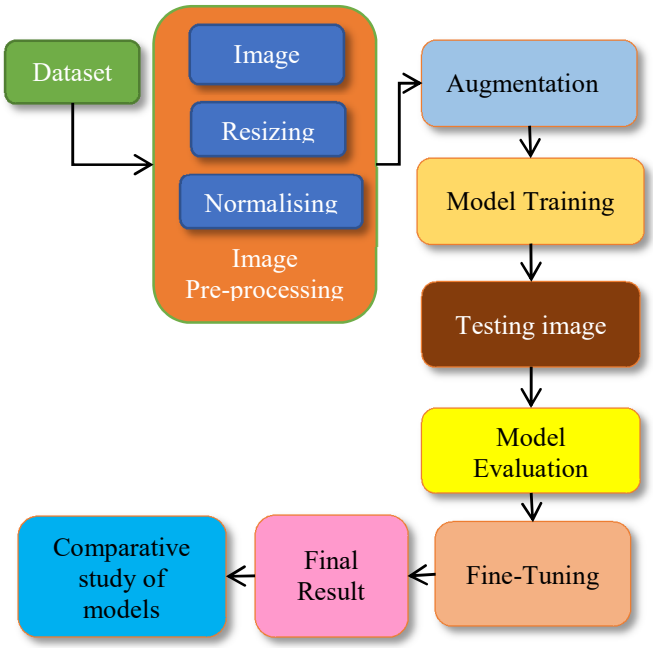


Fig. 1 Workflow

B. Data Processing and Augmentation

To guarantee consistency, images are loaded using OpenCV (cv2) and scaled to a standard size of 224x224 pixels. Each picture is given a label according to the asana to which it belongs, and the target variable is created by one-hot encoding these labels. To create a random sequence for the photos and labels, the dataset is shuffled. Then input data is normalized using the standard normalization values [-1, 1]. Then we created training and testing sets from the dataset. Around 80% of the time is usually spent on training and 20% on testing. Data augmentation is carried out using Keras' ImageDataGenerator. Techniques for augmentation include rotating, moving, zooming, and flipping both horizontally and vertically. Data augmentation exposes the model to different picture modifications, which improves its ability to generalize.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:51:33 UTC from IEEE Xplore. Restrictions apply.

C. Model Architecture and Training

Pre-learned weights from models trained on massive datasets such as ImageNet were used in transfer learning. As a result, the models were able to use information gleaned from generic picture attributes. Feature extraction is the process of extracting hierarchical features from input photos by means of the convolutional layers of pre-trained models. In order to decrease spatial dimensions and concentrate on important features, Global Average Pooling (GAP) layers were added. To build a custom classifier for yoga pose classification, dense layers with suitable activation functions, dropout, and batch normalization were included. Models were assembled using appropriate evaluation metrics (accuracy), loss functions (categorical crossentropy), and optimizers (e.g., Adam). Models were first trained using the yoga pose dataset for a predetermined number of epochs, early stopping to avoid overfitting. The best performing model was captured.

D. Model Evaluation and Fine-Tuning

We assessed the performance of every model on a different test set. For quantitative analysis, we created accuracy, confusion matrices, and classification reports. We also found out which pose is the most accurately detected for each model. To determine convergence, overfitting, and generalization abilities, we compared metrics from training and testing. Next, in order to improve alignment with the dataset, we unfroze and fine-tuned certain layers. Model architecture and performance play a role in the selection of the layers. After making adjustments, we recompile the models and keep practicing for more epochs.

E. Comparative Study

Taking into account training speeds, convergence patterns, and final metrics, a comparative analysis allowed for a more nuanced understanding of each model's performance. This rigorous process resulted in the selection of the best-performing model, whose weights were saved for later deployment and which had the highest testing accuracy.

IV. Experimental Analysis

A. Model Architecture

Residual learning was first shown by ResNet50, a 50-layer deep convolutional neural network. Deep networks may be trained by using shortcut connections to overcome the vanishing gradient problem. The skip connections, or shortcuts, in ResNet50 allow the model to learn identity mappings, improving the training and optimization of deeper networks. The model has 48 convolutional layers, completely connected layers, and global average pooling.

VGG16 and VGG19's combined architecture, which has max-pooling and 13 convolutional layers, has a similar topology. For classification, both models employ 3x3 filters and stack fully linked layers at the end. VGG19 is marginally more difficult than VGG16 because it features an additional completely linked layer. VGG16 has three fully connected layers, each with 4096 units. For efficient feature learning, the design places a strong emphasis on max-pooling and tiny filters.

EfficientNetB0 makes use of squeeze-and-excitation modules, depthwise separable convolutions, EfficientNet blocks with inverted residuals, and stem convolutions. For the representation of global features, global average pooling is employed first, and then a dense layer is utilized for classification logits. This architecture is adaptable for computer vision jobs because it strikes a compromise between performance and efficiency.

Depth-wise separable convolution layers are present in MobileNet. A pointwise convolution comes after a depthwise convolution, offering an effective substitute for conventional convolutions. In order to reduce computational complexity, depthwise separable convolutions seek to record spatial and channel-wise dependence independently.

The Inception module, which combines convolutions of various kernel sizes, is used by InceptionV3. The 1x1, 3x3, and 5x5 convolutions included in each module allow the model to collect information at different sizes. The training gradient flow is improved with auxiliary classifiers, and factorized convolutions lower computing costs.

The convolutional blocks in Xception's architecture are detachable according to depth. These blocks enable the model to capture spatial relationships more effectively by substituting depthwise separable convolutions for ordinary convolutions. The hierarchical application of the depthwise separable convolutions highlights the need of precise spatial information.

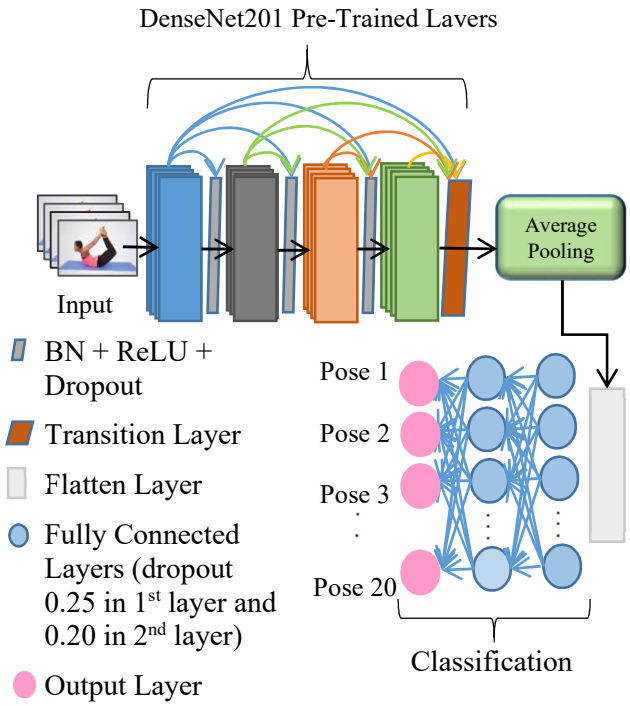


Fig. 2 Densenet201 Customized Model

DenseNet121 uses four dense blocks with skip connections that are densely linked convolutional blocks. By tweaking the DenseNet201 (fig 2) architecture, adding more layers for fine-tuning, and freezing the top layers, the model is modified for yoga posture categorization. These extra layers consist of batch normalization, dropout layers (at 25% and 20% rates), dense layers with ReLU activation (1024 and 512 units), and so on. The twenty yoga positions

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:51:33 UTC from IEEE Xplore. Restrictions apply.

are represented by the output layer with softmax activation. During fine-tuning, certain layers are unfrozen in order to capture pose-specific characteristics. The loss function used in the model's compilation is categorical cross-entropy, and the Adam optimizer (lr: 0.001).

B. Requirements and Specification

Python version 3.7 was used throughout the careful execution of each experiment. Nine different deep learning models were trained on the HATHAYOGA dataset, which consists of 20 different classes that correspond to different yoga positions. These models require the following hardware specifications: Tesla T4, 2 x vCPU, 32GB RAM, and a K80 GPU from Google Colab Pro.

C. Result

Table 1 Benchmark accuracies of the models

Model Name	Testing Accuracy		
	Before fine-tuning	After fine-tuning	
		Top-1	Top-5
ResNet50	72.0%	84.4%	96.8%
VGG16	62.4%	80.0%	96.2%
VGG19	57.6%	82.4%	97.6%
EfficientNetB0	74.4%	86.8%	**98.4%**
MobileNetV2	66.4%	80.0%	95.6%
InceptionV3	54.4%	80.8%	96.0%
Xception	60.4%	81.2%	94.8%
DenseNet121	74.0%	85.2%	97.6%
DenseNet201	79.2%	**90.4%**	97.2%

Table 2 Precision, Recall, F1-Score of the models

DL Model	Precision	Recall	F1-Score
ResNet50	0.86	0.84	0.84
VGG16	0.82	0.80	0.80
VGG19	0.84	0.83	0.82
EfficientNetB0	0.88	0.87	0.87
MobileNetV2	0.82	0.80	0.79
InceptionV3	0.82	0.82	0.80
Xception	0.82	0.81	0.81
DenseNet121	0.86	0.85	0.85
DenseNet201	0.92	0.90	0.90

We carried out a thorough analysis of many deep learning models for yoga position identification in our study. Table 1's accuracy results demonstrated DenseNet201's better performance, with the model attaining the maximum accuracy of 90.4% for top-1 classification out of all the models. Among the best performers, EfficientNetB0 stood out with a 98.4% accuracy rate for top-5 categorization. Table 2 presents information on each model's F1 score, precision, and recall.

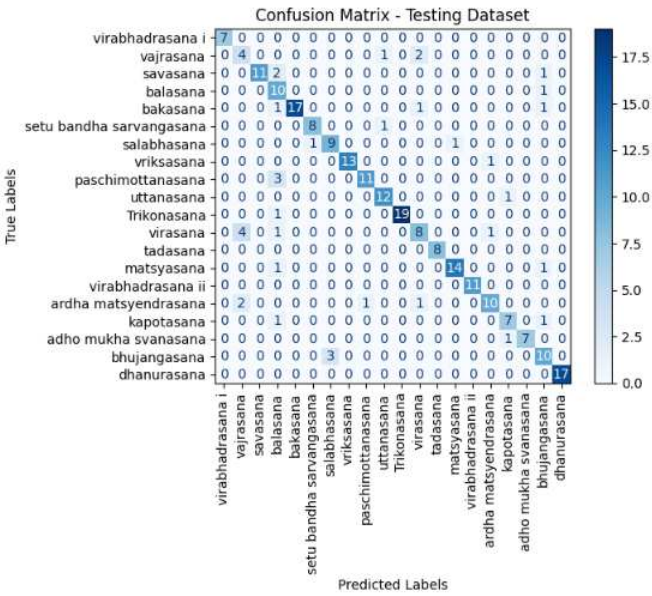


Figure 3 Confusion Matrix of DenseNet201 model

Subsequent investigation concentrated on DenseNet201, our selected model, for which we provided a thorough confusion matrix in Figure 3 and AUC-ROC curve in Figure 4. We have added Table 3 to give a more detailed picture of posture-specific accuracy using DenseNet201 architecture. The poses known as 'Virabhadrasana i', 'Salabhasana', 'Vriksasana', 'Uttanasana', 'Trikonasana', 'Matsyasana', 'Virabhadrasana ii' and 'Adho mukha svanasana' have been discovered most accurately detected with accuracy measuring 100%.

Table 3 Pose-wise accuracy of DenseNet201 Model

Pose	Accuracy
Virabhadrasana i	100.0%
Vajrasana	81.82%
Savasana	50.0%
Balasana	91.66%
Bakasana	88.88%
Setu bandha sarvangasana	76.47%
Salabhasana	100.0%
Vriksasana	100.00%
Paschimottanasana	77.77%
Uttanasana	100.00%
Trikonasana	100.0%
Virasana	88.88%
Tadasana	91.66%
Matsyasana	100.00%
Virabhadrasana ii	100.0%
Ardha matsyendrasana	88.88%
Kapotasana	86.67%
Adho mukha svanasana	100.00%
Bhujangasana	88.23%
Dhanurasana	80.00%

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:51:33 UTC from IEEE Xplore. Restrictions apply.

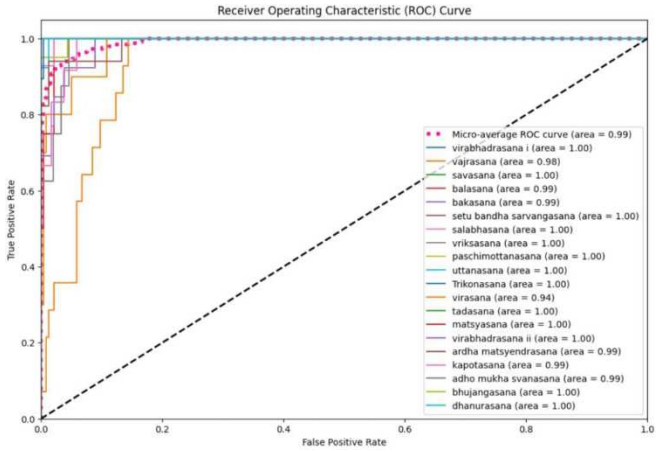


Fig. 4 AUC-ROC Curve for DenseNet201 Model



Fig. 5 Actual vs Predicted pose of some test images

In figure 5, some test images of different yoga poses are taken and their true vs predicted labels are shown.

V. Conclusion

In summary, for the purpose of classifying yoga postures, we exhaustively investigated many cutting edge deep learning architectures. The models' strengths varied according to the findings; DenseNet201 and DenseNet121 performed better than the others, while EfficientNet and MobileNet showed efficiency in terms of processing resources. These findings emphasize the trade-offs between accuracy and computing economy and offer vital insights for choosing suitable models in practical circumstances. Our work provides a baseline and important direction for further research in the area of yoga position categorization.

A. Future work

Future study should prioritize the investigation of real-time applications and deployment of the established yoga posture identification algorithm. This will enable its incorporation into user-friendly platforms or mobile applications that may offer immediate feedback during yoga practice. Furthermore, conducting an examination of the model's efficacy on a wide range of datasets, which include a more extensive array of yoga postures and practitioners with different degrees of expertise, would improve its capacity to apply to different situations and its overall strength.To enhance the model, it is necessary to tackle obstacles such occlusion and fluctuations in lighting conditions, in order to enhance its capacity to adapt to real-world situations. Incorporating user input and preferences may enhance the yoga training experience by making it more customized and effective. Finally, by examining the possible combination of biomechanical knowledge and health measurements[13], [14], the model's usefulness may be enhanced as it would be able to offer personalized suggestions to individuals according to their physical ability and wellness objectives.

VI. References

[1] H.-T. Chen, Y.-Z. He, C.-C. Hsu, C.-L. Chou, S.-Y. Lee, and B.-S. P. Lin, 'Yoga Posture Recognition for Self-training', in *MultiMedia Modeling*, vol. 8325, C. Gurrin, F. Hopfgartner, W. Hurst, H. Johansen, H. Lee, and N. O'Connor, Eds., in Lecture Notes in Computer Science, vol. 8325. , Cham: Springer International Publishing, 2014, pp. 496–505. doi: 10.1007/978-3-319-04114-8_42.

[2] E. W. Trejo and P. Yuan, 'Recognition of Yoga Poses Through an Interactive System with Kinect Device', in *2018 2nd International Conference on Robotics and Automation Sciences (ICRAS)*, Jun. 2018, pp. 1–5. doi: 10.1109/ICRAS.2018.8443267.

[3] F. Rishan, B. De Silva, S. Alawathugoda, S. Nijabdeen, L. Rupasinghe, and C. Liyanapathirana, 'Infinity Yoga Tutor: Yoga Posture Detection and Correction System', in *2020 5th International Conference on Information Technology Research (ICITR)*, Dec. 2020, pp. 1–6. doi: 10.1109/ICITR51448.2020.9310832.

[4] M. Verma, S. Kumawat, Y. Nakashima, and S. Raman, 'Yoga-82: A New Dataset for Fine-grained Classification of Human Poses', in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA: IEEE, Jun. 2020, pp. 4472–4479. doi: 10.1109/CVPRW50498.2020.00527.

[5] S. Kothari, 'Yoga Pose Classification Using Deep Learning', *Masters Proj.*, May 2020, doi: https://doi.org/10.31979/etd.rkgu-pc9k.

[6] S. K. Yadav, A. Agarwal, A. Kumar, K. Tiwari, H. M. Pandey, and S. A. Akbar, 'YogNet: A two-stream network for realtime multiperson yoga action recognition and posture correction', *Knowl.-Based Syst.*, vol. 250, p. 109097, Aug. 2022, doi: 10.1016/j.knosys.2022.109097.

[7] Y. Agrawal, Y. Shah, and A. Sharma, 'Implementation of Machine Learning Technique for Identification of Yoga Poses', in *2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)*, Apr. 2020, pp. 40–43. doi: 10.1109/CSNT48778.2020.9115758.

[8] C. Long, E. Jo, and Y. Nam, 'Development of a yoga posture coaching system using an interactive display based on transfer learning', *J. Supercomput.*, vol. 78, no. 4, pp. 5269–5284, Mar. 2022, doi:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:51:33 UTC from IEEE Xplore. Restrictions apply.

10.1007/s11227-021-04076-w.
[9] S. Garg, A. Saxena, and R. Gupta, 'Yoga pose classification: a CNN and MediaPipe inspired deep learning approach for real-world application', *J. Ambient Intell. Humaniz. Comput.*, Jun. 2022, doi: 10.1007/s12652-022-03910-0.
[10] V. Anand Thoutam *et al.*, 'Yoga Pose Estimation and Feedback Generation Using Deep Learning', *Comput. Intell. Neurosci.*, vol. 2022, p. e4311350, Mar. 2022, doi: 10.1155/2022/4311350.
[11] A. Upadhyay, N. K. Basha, and B. Ananthakrishnan, 'Deep Learning-Based Yoga Posture Recognition Using the Y_PN-MSSD Model for Yoga Practitioners', *Healthcare*, vol. 11, no. 4, Art. no. 4, Jan. 2023, doi: 10.3390/healthcare11040609.
[12] 'Yoga Pose Image classification dataset'. Accessed: Dec. 09, 2023. [Online]. Available: https://www.kaggle.com/datasets/shrutisaxena/yoga-pose-image-classification-dataset
[13] A. Verma, V. Ranga, and D. K. Vishwakarma, 'A novel approach for forecasting PM2.5 pollution in Delhi using CATALYST', *Environ. Monit. Assess.*, vol. 195, no. 12, p. 1457, Nov. 2023, doi: 10.1007/s10661-023-12020-z.
[14] A. Verma, V. Ranga, and D. K. Vishwakarma, 'Forecasting of Satellite Based Carbon-Monoxide Time-Series Data Using a Deep Learning Approach', in *2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT)*, Feb. 2023, pp. 1–4. doi: 10.1109/ICITIIT57246.2023.10068609.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:51:33 UTC from IEEE Xplore. Restrictions apply.

Eurasian Economic Review
https://doi.org/10.1007/s40822-024-00275-x

ORIGINAL PAPER

An empirical examination of the effect of domestic monetary policy on external commercial borrowings to India

Virender Kumar[1] · **Poonam**[2]

Received: 13 October 2023 / Accepted: 20 March 2024
© The Author(s) under exclusive licence to Eurasia Business and Economics Society 2024

Abstract
This paper empirically examines the effect of domestic monetary policy shock on external commercial borrowings (ECBs) to India for the period 2000:Q2 to 2019:Q4. The impact of domestic monetary policy shock on ECBs is assessed using two variables: interest rate differential and domestic money supply growth. The paper employs a structural VAR model and utilizes impulse response functions and forecast error variance decomposition analysis to derive empirical results. The empirical results based on impulse response functions indicate that the interest rate differential has a positively significant effect on external commercial borrowings to India, while the effect of domestic money supply growth is found to be negative. The forecast error variance decomposition analysis indicates that the interest rate differential explains about 14.6% and domestic money supply growth explains about 2% of the total variation in external commercial borrowings, implying that domestic monetary policy variables together explain about 16.6% of the total variation in ECBs to India. The results suggest that the central bank, through monetary policy interventions, can influence and manage the flow of external commercial borrowings to India. Furthermore, the study finds that the exchange rate, domestic industrial activity, domestic output growth, domestic country's creditworthiness, and domestic macroeconomic instability are other significant determinants of external commercial borrowings.

Keywords Monetary Policy · External Commercial Borrowings · Interest rate differential · SVAR · India

✉ Virender Kumar
virender0202@gmail.com

Poonam
poonam1110s@gmail.com

[1] Delhi Technological University, Delhi, India

[2] Shri Ram College of Commerce, University of Delhi, Delhi, India

Published online: 25 April 2024

JEL Classification E52 · F21 · C32 · F41

1 Introduction

India is a major emerging economy that needs a constant flow of external funds to support its investment requirements. External commercial borrowings (ECBs) by domestic corporates are an important source of external funds to India. ECBs comprise commercial loans in the form of bank loans, suppliers' credit, buyers' credit, securitized instruments such as floating rate notes, securitized commercial borrowings, and fixed rate bonds from non-resident lenders. One of the major advantages of ECBs is that they provide an additional source of funds for Indian corporates to expand their existing capacity and finance new investment ventures, thereby contributing to the country's economic growth (Sur et al., 2019; Verma & Prakash, 2011). However, due to their significant share in India's total external debt,[1] ECBs have sparked concerns about debt repayments, raising fears of widespread crisis and threatening financial stability, particularly in the event of external shocks (Pradhan & Hiremath, 2019). Moreover, volatility in ECB flows, as observed in Fig. 1, has the potential to disrupt India's balance of payments position and create significant fluctuations in the foreign exchange market. Maran (2023) discusses the need for macroprudential policies to address such financial risks.

Figure 1 illustrates the considerable volatility in net ECB flows to India over the years. Despite the volatility, the total outstanding stock of ECBs in India has consistently grown. ECBs constitute a substantial proportion of India's total external debt and this proportion has been continuously increasing over the years, increasing from 13.7% in 1991–92 to 20.3% in 1990–00, 27.8% in 2007–08, 33.5% in 2013–14 and further to 37.9% in 2018–19 (Table 1). Given the volatility and the growing significance of ECBs in India, it is essential to understand what determines ECB flows in general and how a domestic monetary policy shock affects such flows, in particular, to help domestic policymakers, including the central bank, formulate appropriate policies to manage them. Hence, the current study aims to empirically examine the effect of the domestic monetary policy shock and other domestic factors on ECB flows to India.

In this study, the impact of the domestic monetary policy shock on external commercial borrowings to India is explored using two variables: interest rate differential and domestic money supply growth. The empirical analysis is conducted through the application of the structural VAR model. This methodology allows for an examination of the dynamic behaviour of external commercial borrowings in response to a domestic monetary policy shock, and it also helps determine the relative contribution of the domestic monetary policy shock in explaining external commercial borrowings to India.

[1] Approximately 30% of the total external debt to India over the period 1990–91 to 2018–19 is in the form of external commercial borrowings by domestic corporations, with the share further rising to around 35% in recent years. Refer to Table 1.

Table 1 Share of External Commercial Borrowings in India's Total External Debt. Source: Reserve Bank of India

Year	Total External Debt	Total ECBs Outstanding	Total ECBs Outstanding (% of External Debt)
1990–91	83,801	10,209	12.2
1991–92	85,285	11,715	13.7
1992–93	90,023	11,643	12.9
1993–94	92,695	12,363	13.3
1994–95	99,008	12,991	13.1
1995–96	93,730	13,873	14.8
1996–97	93,470	14,335	15.3
1997–98	93,531	16,986	18.2
1998–99	96,886	20,978	21.7
1999–00	98,263	19,943	20.3
2000–01	101,326	24,408	24.1
2001–02	98,843	23,320	23.6
2002–03	104,914	22,472	21.4
2003–04	112,653	22,007	19.5
2004–05	134,002	26,405	19.7
2005–06	139,114	26,452	19
2006–07	172,360	41,443	24
2007–08	224,407	62,334	27.8
2008–09	224,498	62,461	27.8
2009–10	260,935	70,726	27.1
2010–11	317,891	100,476	31.6
2011–12	360,766	120,136	33.3
2012–13	409,374	140,125	34.2
2013–14	446,178	149,375	33.5
2014–15	474,675	180,295	38
2015–16	484,791	180,480	37.2
2016–17	471,012	172,045	36.5
2017–18	529,278	201,821	38.1
2018–19	543,135	205,804	37.9

The existing studies on ECB flows to India (Dev, 2014; Misra et al., 2022; Patnaik et al., 2016; Sur et al., 2019; Pradhan & Hiremath, 2019; Pradhan & Hiremath, 2020; Singh, 2007; Singh, 2009) have mainly focused on analysing the trends and determinants of ECBs. The role of domestic monetary policy in affecting ECB flows to India is not well researched in the literature. By examining the role of the domestic monetary policy shock and a wider set of other domestic factors in influencing ECB flows to India, this study makes an effort to bridge this gap. Furthermore, by employing the structural VAR model, the current study is able to determine how much proportion of the total variation in ECBs to India

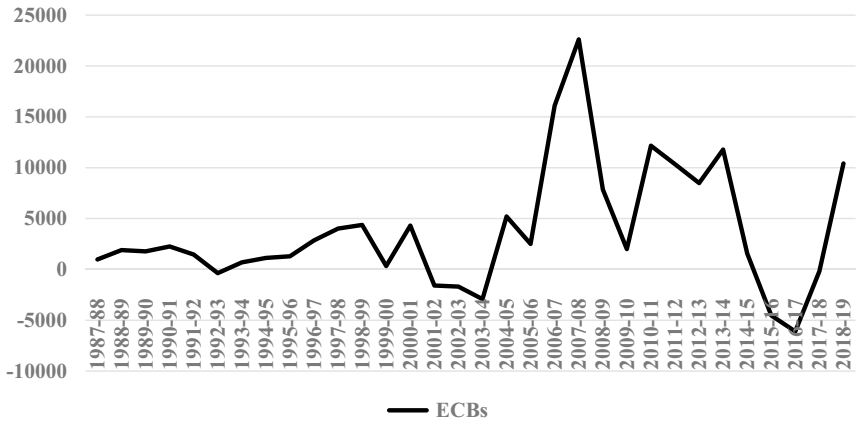


Fig. 1 Net External Commercial Borrowings to India, US$ Million. Source: Reserve Bank of India (RBI)

is explained by the domestic monetary policy shock, something that previous research has not explored.

The following section discusses the trends in external commercial borrowings to India and how it relates with domestic monetary policy shocks. Section 3 reviews the existing literature and discusses the theoretical outline. Section 4 details the empirical model, data, and methodology. Section 5 presents and examines the empirical results. At last, Sect. 6 concludes.

2 Trends in external commercial borrowings to India and how it relates to domestic monetary policy shocks

External commercial borrowings are an additional source of funds for Indian corporates to expand their existing capacity and finance new investment ventures, taking into account interest rate differential between domestic and global markets as well as associated market risks (Verma & Prakash, 2011). The ECB policy in India has been designed to provide Indian corporates with flexibility in borrowings while maintaining judicious limits for total external borrowings, subject to conditions on maturity, cost, and end-use restrictions. Unlike FPI and FDI flows, ECB flows in India are modulated based on the overall cycle of capital flows, through the use of price-based measures (such as linking interest costs to LIBOR) and administrative measures (such as end-use restrictions for ECBs).

The use of ECBs began in the 1970s but remained modest due to the dominance of external aid from multilateral and bilateral entities. In the 1980s, as external aid started to decline and the trade deficit widened, the importance of ECBs in meeting India's foreign exchange and investment requirements increased. The bulk of the increase in ECB flows was observed during the second half of the 1980s when there was a significant surge in demand for foreign currency due to a widening trade

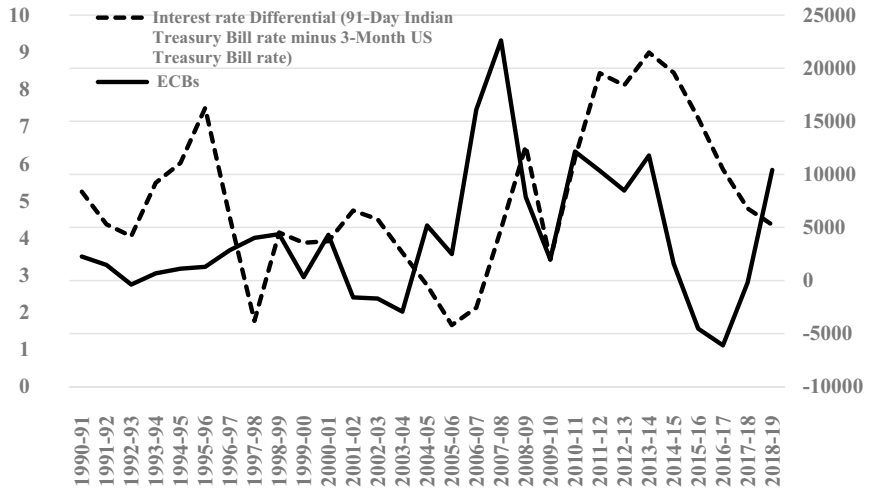


Fig. 2 Trends in net external commercial borrowings to India and interest rate differential (91-Day Indian Treasury Bill rate minus 3-Month US Treasury Bill rate). Source: Reserve Bank of India and Federal Reserve

deficit caused by higher oil prices, increased import demand, and a sharp depreciation of the domestic currency against the dollar. This led to a substantial rise in India's external liability by 1990, which eventually culminated into the balance of payment crisis of 1991. Following the crisis, external commercial borrowings dried up in response to a fall in India's credit rating. The Indian government also adopted a conservative ECB policy to address the external debt situation, imposing an annual ceiling on additional external borrowings, particularly on short-term debt, and placing restrictions on the end-use of ECBs. ECBs were mainly permitted for importing capital goods, with priority given to the infrastructure sector.

Figure 2 shows the trends in Net ECB flows and domestic monetary policy shocks, as represented by the change in 91-Day Indian Treasury Bill rate relative to 3-Month US Treasury Bill rate, i.e., the interest rate differential. We observe that net ECB flows to India exhibited a declining trend in the initial years of liberalization from 1990–91 to 1992–1993, accompanied by a decrease in the interest rate differential. However, net ECB flows began to rise thereafter and continued to do so till 1998–99. This continuous rise in ECBs was the result of improved credit ratings, enhanced domestic demand, and the liberalization of ECB policy with respect to maturity, amount and eligible borrowers in the mid-1990s. As shown in Fig. 2, the interest rate differential also experienced an upward trend during this period till 1995–96, suggesting that, if allowed, domestic corporates tend to borrow more from external sources when borrowing costs in the domestic country are higher relative to the foreign country. The turbulence in global financial markets triggered by the East Asian crisis, economic sanctions on India due to Pokhran nuclear test, global slowdown in the early 2000s, highly volatile credit markets, and lower investor confidence, interrupted developments in India's capital account, leading to fluctuations and a decline in ECB flows until 2003–04. Figure 2 shows that this period was also

Table 2 Net external commercial borrowings to India, US$ Million. Source: Reserve Bank of India

Year	ECBs	Year	ECBs
1987–88	976	2003–04	– 2925
1988–89	1894	2004–05	5194
1989–90	1777	2005–06	2508
1990–91	2254	2006–07	16,103
1991–92	1462	2007–08	22,610
1992–93	– 366	2008–09	7861
1993–94	686	2009–10	2000
1994–95	1124	2010–11	12,160
1995–96	1284	2011–12	10,344
1996–97	2856	2012–13	8485
1997–98	4010	2013–14	11,777
1998–99	4367	2014–15	1570
1999–00	333	2015–16	– 4529
2000–01	4303	2016–17	– 6102
2001–02	– 1585	2017–18	– 183
2002–03	– 1692	2018–19	10,416

characterized by a declining trend in the interest rate differential, which may mean that, ceteris paribus, domestic corporates tend to borrow less from external sources when borrowing costs in the domestic country fall relative to the foreign country.

The period from 2004–05 to 2007–08 proved to be golden years for global capital flows as well as for domestic GDP expansion, reflected in an increased net flow of ECBs from $5194 million in 2004–05 to an all-time peak of $22,610 million in 2007–08 (Table 2). The period witnessed an important policy change in the form of the issuance of a new set of policy guidelines on ECBs on January 31, 2004. The objective was to establish "more transparent and simplified policies and procedures" on external commercial borrowings. As shown in Fig. 2, this period was also marked by an upward trend in the interest rate differential up to 2008–09 as the central bank followed a contractionary monetary policy to control speculation and inflation, making borrowings from external sources cheaper. Encouraged by cheaper international liquidity, simplified policies and procedures, increasing domestic demand, and a favorable exchange rate, many Indian firms increasingly raised funds through ECBs during this period.

There was a sharp reversal in net ECB flows in 2008–09 due to the global financial crisis whose impact was felt all over the world. Net ECBs fell in 2009–10 as well. Although ECB flows recovered strongly in 2010–11, they continued to fluctuate around that level until 2013–14 before following a downward trend for the next three years, reaching their lowest point in 2016–17. The interest rate differential also exhibited a similar pattern during 2008–09 to 2016–17. The following year, 2017–18, also witnessed a net outflow of external commercial borrowings. Nevertheless, due to reform measures taken by the government, further liberalization of the ECB policy, and the huge demand for cheaper funds by Indian corporates to support their expansion, there was a sharp recovery in net ECB flows in 2018–19.

Overall, it can be stated that ECB flows to India have demonstrated substantial volatility, particularly in the aftermath of the 2008–09 financial crisis. Additionally, as shown in Fig. 2, a rise in ECB flows has often been associated with monetary policy shocks causing an increase in the interest rate differential, whereas a fall in ECB flows has been associated with monetary policy shocks causing a decrease in the interest rate differential. Hence, to explore this further, it is necessary to examine the effect of domestic monetary policy shocks on ECB flows empirically.

3 Review of literature

3.1 Empirical literature

One of the initial studies focusing on analysing external commercial borrowings to India in the post-liberalization period was conducted by Singh (2007). Using Johansen co-integration and VECM mechanisms, the study revealed that long-term external borrowings by Indian corporates were influenced by the pace of domestic real activity, interest rate differentials, and credit conditions. In terms of magnitude, it was the real variable, i.e. domestic real activity, that had the greatest impact in attracting ECB flows to India compared to price variables such as interest rate differential. Singh (2009) also concluded that external commercial borrowings to India were primarily driven by domestic real economic activity, interest rate differential, and credit constraints in the global market. The study found a strong co-movement between overseas borrowings and domestic real activity. It also observed that during normal periods, ECBs were predominantly driven by domestic real demand shocks, while during the periods of crisis, global credit shocks played a more significant role in affecting ECBs.

Verma and Prakash (2011) employed the Granger causality and Johansen co-integration methods to examine the impact of interest rate differential and other potential determinants on ECB flows to India. The co-integration analysis revealed a positive and statistically significant impact of interest rate differentials on ECBs. The granger causality test also indicated that ECB flows were granger caused by interest rate differential. The study further identified domestic currency appreciation, improvement in the current account balance, and domestic industrial activity as the other important factors attracting ECB flows to India. Dev (2014) also applied the Johansen co-integration and Granger causality to investigate the ECB flows to India and its relationship with imports, exports, interest rate differential, exchange rate, the index of industrial production, and foreign investment. The study found a significant long-term positive relationship between ECBs and the index of industrial production, exchange rates, and interest rate differentials. The granger causality test indicated a unidirectional causality running from these variables to ECB flows. Through variance composition analysis, the study observed that most of the variation in ECB flows was caused by interest rate differential, followed by the index of industrial production. On similar lines, Sur et al. (2019) analysed the determinants of ECB flows to India using Johansen cointegration and vector error correction techniques and revealed that both domestic and external factors played a role in ECB

inflows to India. The study highlighted that regulatory changes related to the capital account had the most significant impact on ECB flows. Additionally, the study identified exchange rates, India's sovereign rating, interest rate differential, and international credit conditions as other significant determinants of ECB flows to India.

Parnaik et. al (2016) talked about existing regulations, reviewed the outcomes and discussed areas of concern and recent policy changes related to foreign currency borrowings by Indian firms. They concluded that in an emerging country with incomplete markets and a managed floating exchange rate, there are systemic risks associated with foreign currency borrowings if left unhedged by large firms, hence emphasizing for a new policy framework. Pradhan and Hiremath (2019) also talked about the risks and financial stability concerns associated with foreign currency borrowings by Indian firms and recommended implementing a targeted policy approach that includes a focus on the domestic bond market, effective hedging mechanisms, and appropriate regulatory decisions regarding the cap and level of ECBs. Misra et al. (2022) also emphasized on hedging against risks associated with foreign currency borrowings and suggested firms to employ hedging strategies to cover a predefined minimum portion of their currency risk while obtaining funds via ECBs. Pradhan and Hiremath (2020) investigated reasons for external borrowings by Indian firms using a dynamic panel regression approach and concluded that both macroeconomic and firm specific factors are important in explaining external borrowings by Indian firms. Thus, we note that existing studies have mainly focused on analysing the trends and determinants of ECBs. None of the studies has specifically focused on determining the impact of domestic monetary policy shock on ECB flows to India, particularly when we want to know how much of the total variation in ECB flows to India is explained by the domestic monetary policy shock.

3.2 Literature on theoretical linkage between domestic monetary policy and ECBs

According to the open market portfolio balance theory (Branson, 1972; Kouri & Porter, 1974; Mckinnon & Oates, 1966), international debt flows such as external commercial borrowings respond to interest rate differential. Therefore, an Indian company's decision to borrow from the foreign market as compared to the domestic market primarily depends on the interest rate differential between the domestic and foreign markets.

The higher the interest rate differential, the greater the incentive for domestic firms to access low-cost funds from the international market. Singh (2007), Verma and Prakash (2011), Dev (2014) and Sur et al. (2019) have empirically observed the positive association between interest rate differential and ECB flows to India. Therefore, a domestic monetary policy shock (measured through shocks in domestic money supply growth and interest rate differential) is expected to have a significant influence on external commercial borrowings. While the interest rate differential is expected to positively impact ECB flows, domestic money supply growth, by inversely influencing the interest rate differential, is likely to have a negative impact.

However, ECB flows are also likely to be influenced by other domestic and external variables (Dev, 2014; Singh, 2007, 2009; Sur et al., 2019; Verma & Prakash, 2011). It is important to control for these variables while investigating the role of the domestic monetary policy shock. Accordingly, domestic variables that we control for include domestic output growth, foreign exchange reserves to import ratio representing domestic country's creditworthiness, domestic fiscal deficit to GDP ratio representing domestic macroeconomic instability, exchange rate defined as the rupee price of a US dollar, index of industrial production representing domestic industrial activity, and the degree of openness. External variables that are controlled include foreign output growth and volatility in the global market.

4 Empirical model, data and methodology

Drawing from the theoretical idea presented in the previous section, we obtain the following empirical model for external commercial borrowings:

$$ECB_t = \lambda + \lambda_1(i_t - i_t*) + \lambda_2 Mg_t + \lambda_3 y_t + \lambda_4 rm_t + \lambda_5 (fd/y)_t + \lambda_6 ex_t + \lambda_7 IIP_t + \lambda_8 open_t + \lambda_9 y_t* + \lambda_{10} globalVol_t + \mu_t \quad (1)$$

where ECB_t: External Commercial Borrowings, $(i_t - i_t*)$: Interest Rate Differential, Mg_t: Domestic Money Supply Growth, y_t: Domestic Output Growth, rm_t: Foreign Exchange Reserves to Import ratio representing Domestic Country's Creditworthiness, $(fd/y)_t$: Domestic Fiscal Deficit to GDP ratio representing Domestic Macroeconomic Instability, ex_t: Exchange Rate defined as the Rupee price of a US Dollar, IIP_t: Index of Industrial Production representing Domestic Industrial Activity, $open_t$: Degree of Openness, y_t*: Foreign Output Growth, $globalVol_t$: Volatility in the Global Market.

The study applies the structural VAR (SVAR) model to empirically examine the effect of domestic monetary policy variables (*i-i*and Mg*) and other explanatory variables included in the empirical model on external commercial borrowings to India. Additionally, the study includes two dummy variables, specifically EUcrisis and UScrisis, to control for the effects of the Eurozone debt crisis and the US crisis on the flow of external commercial borrowings to India. Considering India's status as a small open economy, certain variables such as foreign output growth (y_t*), global market volatility ($globalVol_t$), and the aforementioned dummy variables (EUcrisis and UScrisis) are regarded as exogenous factors within the system. Degree of openness ($open_t$) is also assumed to be exogenous. *Thus, we estimate the empirical model (1) with eight endogenous variables, namely, interest rate differential (i-i*), domestic money supply growth (Mg), foreign exchange reserves to import ratio (rm), domestic fiscal deficit to GDP ratio (fd/y), domestic output growth (y), exchange rate (ex), index of industrial production (IIP) and external commercial borrowings (ECB) after controlling for all exogenous variables in the framework of Structural VAR.*[2] As discussed in the previous section, the interest rate differential

[2] Structural VAR model includes endogenous variables only.

is expected to positively impact ECB flows, i.e., $\lambda_1 > 0$, and domestic money supply growth, by inversely influencing the interest rate differential, is likely to have a negative impact, i.e., $\lambda_2 < 0$. The expected signs of coefficients associated with other endogenous explanatory variables are $\lambda_3 > 0$, $\lambda_4 > 0$, $\lambda_5 < 0$, $\lambda_6 < 0$, $\lambda_7 > 0$, which we have discussed below:

- *Domestic Output Growth (y)* Higher domestic growth indicates rapidly growing domestic economic activity and a larger market size, which creates demand for funds by local firms both for expansion and new investment ventures. It also boosts the confidence of foreign lenders by making them feel safe about their investments. Thus, domestic output growth is expected to have a positive influence on ECB flows.
- *Foreign Exchange Reserves to Import Ratio (rm)* A country with greater foreign reserves relative to its import requirements can easily repay its external liabilities and hence is considered more creditworthy. A high level of creditworthiness lowers the sovereign credit risk of the domestic country, boosts the confidence of foreign creditors, and thus allows domestic firms easier access to international debt markets. Therefore, a rise in foreign exchange reserves to import ratio is likely to attract greater ECB flows to the domestic country.
- *Domestic Fiscal Deficit to GDP Ratio (fd/y)* A higher domestic fiscal deficit to GDP ratio indicates greater macroeconomic instability in the domestic country, which increases investment risk and thereby discourages domestic corporates from expanding their existing capacity or undertaking new investment ventures. It also discourages foreign lenders from supplying more debt capital to the domestic (host) country. Therefore, the domestic fiscal deficit to GDP ratio is expected to have a negative effect on ECBs.
- *Exchange Rate (ex)* The depreciation of the domestic currency is likely to discourage external borrowings by Indian corporates. Depreciation of the rupee means a higher repayment burden in terms of rupee and a higher effective cost of external borrowings for Indian corporates. Thus, the depreciation of the rupee is expected to negatively influence ECB flows to India.
- *Index of Industrial Production (IIP)* Index of industrial production represents the level of industrial activity in the domestic country. An increasing domestic industrial activity involves an increasing need for funds to finance new investments or expansion of existing capacities, and this, in turn, induces greater external borrowings by domestic corporates. Thus, index of industrial production is expected to have a positive influence on ECB flows.
- *Degree of Openness (open)* A highly open domestic economy means greater flexibility for domestic corporates to access funds internationally at a lower cost. It also means fewer restrictions on all types of capital inflows, including debt inflows. Further, foreign lenders feel more secure about their money in an open market. Thus, a high degree of openness is likely to facilitate greater external commercial borrowings.

4.1 Model specification and empirical strategy

This study applies the structural VAR model to empirically examine the effect of domestic monetary policy shock on external commercial borrowings to India. Innovation accounting analysis under the structural VAR model (impulse response function and forecast error variance decomposition function) is utilized to derive empirical results. The impulse response function (IRF) enables tracking the dynamic path of one variable in response to one standard error shock in another variable. The forecast error variance decomposition (FEVD) function indicates the proportion of overall variation in one variable that can be attributed to shocks from other variables in the model. These two applications together assist in determining the impact and the relative contribution of domestic monetary policy shock in explaining external commercial borrowings to India.

To perform innovation analysis, we follow Han and Kim (2023), wherein we first identify the contemporaneous relation matrix (C_0).[3] This helps recover structural shocks (ε) from the reduced form shocks (e), which as per Sims (1986) and Blanchard and Watson (1986) requires imposition of at least $(n^2\text{-}n)/2$ identification restrictions on the matrix C_0 of structural VAR model.[4] In the structural VAR model, such restrictions come from the economic theory. Once C_0 suitably is restricted, the SVAR model is labelled identified, and the structural shocks (ε_t) are recovered from the reduced form shocks (e) using the following equation: $\varepsilon_t = C_0\, e_t$. Additionally, in this study, Monte Carlo Integration is employed for innovation accounting analysis. Monte Carlo Integration, a Bayesian method, is utilized to statistically determine the significance of impulse responses by computing confidence bands around them (Doan, 1990).

As discussed earlier, the empirical model has 8 endogenous variables, interest rate differential (i-i*), domestic money supply growth (Mg), foreign exchange reserves to import ratio (rm), domestic fiscal deficit to GDP ratio (fd/y), domestic output growth (y), exchange rate (ex), index of industrial production (IIP) and external commercial borrowings (ECB). The vector of variables (X) for the structural VAR model is therefore given as:

$$\mathrm{X} = [Mg, (i - i*), y, ex, (fd/y), rm, IIP, ECB]$$

With n = 8, a minimum of $(8^2–8)/2 = 28$ restrictions must be imposed on matrix C_0 in order to recover or identify the structural shocks. The required identification is achieved using the equation $\boldsymbol{\varepsilon_t = C_0\, e_t}$ by implementing the following restrictions on matrix C_0:

[3] C_0 is a matrix having ones in the diagonal, which represents contemporaneous relationships among variables (variables in vector X) included in the structural VAR model.

[4] The formal structural VAR methodology is explained in the appendix.

$$\begin{bmatrix} \varepsilon_{Mg} \\ \varepsilon_{i-i*} \\ \varepsilon_{y} \\ \varepsilon_{fd/y} \\ \varepsilon_{rm} \\ \varepsilon_{IIP} \\ \varepsilon_{ECB} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma_{21} & 1 & \gamma_{23} & \gamma_{24} & 0 & 0 & 0 & 0 \\ 0 & \gamma_{32} & 1 & \gamma_{34} & 0 & 0 & 0 & 0 \\ 0 & \gamma_{42} & \gamma_{43} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_{53} & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \gamma_{63} & \gamma_{64} & 0 & 1 & 0 & 0 \\ 0 & \gamma_{72} & \gamma_{73} & 0 & 0 & 0 & 1 & 0 \\ \gamma_{81} & \gamma_{82} & \gamma_{83} & \gamma_{84} & \gamma_{85} & \gamma_{86} & \gamma_{87} & 1 \end{bmatrix} \begin{bmatrix} e_{Mg} \\ e_{i-i*} \\ e_{y} \\ e_{fd/y} \\ e_{rm} \\ e_{IIP} \\ e_{ECB} \end{bmatrix}$$

In the above specification, it is assumed that shocks to other variables in the model do not contemporaneously affect domestic money supply growth. This is because domestic money supply growth is assumed to be determined exogenously by the central bank, which consequently produces following restrictions: $\gamma_{12}=\gamma_{13}=\gamma_{14}=\gamma_{15}=\gamma_{16}=\gamma_{17}=\gamma_{18}=0$. On the contrary, shocks to all variables are assumed to have a contemporaneous impact on ECB flows to India. Hence, none of the gammas (γ's) in the eighth row are zero. The zeroes in the matrix presented above indicate the restrictions imposed based on the theoretical considerations. As there are 36 zeros, the SVAR model has a total of 36 restrictions, indicating that the model is over-identified.

4.2 Data and variables definition

The study uses quarterly data from 2000:Q2 to 2019:Q4 for the analysis.[5] The Reserve Bank of India, OECD database and the Federal Reserve Economic Data are the main data sources. The study defines external commercial borrowings (ECB) broadly as the sum of net external commercial borrowings and net short-term credit to India, both denominated in US$ million. The difference between yield on Indian Treasury Bills (91-Day) and yield on US Treasury Bills (3-Month) is taken as the interest rate differential (***i-i****). Domestic money supply growth (***Mg***) is given by rate of growth of broad money supply in India. Domestic output growth (***y***) is defined by growth rate of GDP in India. Foreign exchange reserves to import ratio (***rm***) representing domestic creditworthiness is defined as the ratio of total foreign exchange reserves and imports, both denominated in US$ million. Domestic fiscal deficit to GDP ratio (***fd/y***) representing domestic macroeconomic instability is defined as the ratio of gross fiscal deficit and GDP.

The rupee price of a US dollar is taken as the exchange rate (***ex***). Index of industrial production (***IIP***) for India is used to measure domestic industrial activity. The degree of openness (***open***) is assessed through two factors: trade openness and financial openness. Trade openness is captured by calculating the ratio of trade to GDP, while financial openness is captured using financial openness index for India by Chinn and Ito (2008). Foreign output growth (***y****) is measured through output

[5] To remove the impact of the unexpected shock associated with COVID, the study has only considered the pre-COVID period for analysis.

growth of OECD countries. The CBOE volatility index (VIX) for the United States is utilized to capture the level of volatility in global market (***globalVol***).

5 Empirical results

5.1 Unit root results

The unit root tests reveal that external commercial borrowings and the degree of financial openness are stationary variables. On the other hand, the remaining variables, namely domestic money supply growth, foreign exchange reserves to import ratio, interest rate differential, domestic fiscal deficit to GDP ratio, domestic output growth, exchange rate, index of industrial production, degree of trade openness, volatility in the global market, and foreign output growth, are found to be non-stationary,[6,7]

Typically, when estimating Structural VAR (or VAR) models, it is necessary for all variables to exhibit stationarity. However, this study has incorporated all variables in levels, even though most of them are level non-stationary. This is because the study utilizes Monte Carlo Integration for innovation analysis, as mentioned earlier. Since Monte Carlo Integration is a Bayesian technique, the presence of non-stationary variables does not influence the statistical inferences, as established by Sims and Uhlig (1991), and Kim and Roubini (2000). Consequently, all variables in the SVAR model have been incorporated in levels.

5.2 Structural VAR results

We estimate the structural VAR model at lag length two,[8,9] In Sect. 5.2.2, it was established that the structural VAR model for ECB is over-identified. To assess the validity of the imposed restrictions, we employ the likelihood ratio test of over-identifying restrictions. The outcome of the test is shown in Table 3. We observe that the likelihood ratio test does not reject the validity of the restrictions at a significance level of 1%, thus marginally confirming the validity of the imposed restrictions.

[6] The study has employed three unit root tests, KPSS, DF-GLS and PP, to check for the unit root, the results of which are available in the appendix.

[7] The descriptive statistics of all these variables has also been presented in the Appendix.

[8] Optimal lag length is chosen using the lag length criteria under VAR.

[9] Note that the structural VAR is estimated only with endogenous variables (interest rate differential, domestic money supply growth, foreign exchange reserves to import ratio, domestic fiscal deficit to GDP ratio, domestic output growth, exchange rate, index of industrial production, and external commercial borrowings) after controlling for all the exogenous variables (UScrisis dummy, foreign output growth, EUcrisis dummy, degree of openness, volatility in global market).

Table 3 Likelihood Ratio Test of Over-identifying Restrictions

	ECB
Log Likelihood	– 1166.884
Log Likelihood Unrestricted	– 1150.779
Chi-Squared	32.208
Significance level	0.022

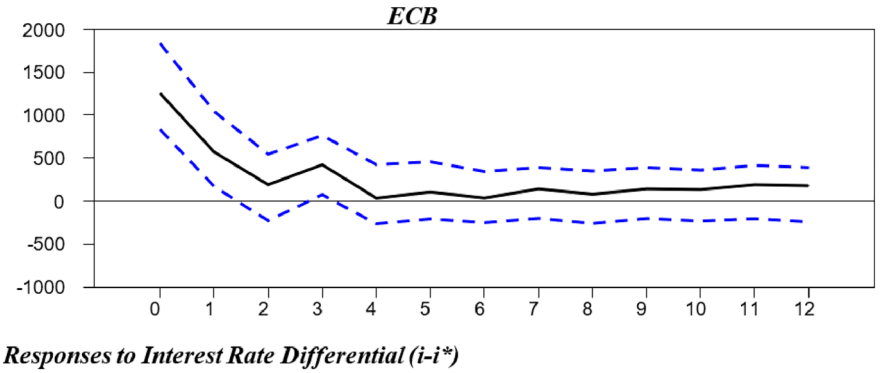


Fig. 3 Impulse response of external commercial borrowings (ECB) to one standard error shock to interest rate differential (i-i*)

5.2.1 Impulse response function

The impulse response function (IRF) is an econometric tool that traces the dynamic movement in one variable when there is a unit standard error shock in another variable. In this study, it helps trace the dynamic (over time) response of external commercial borrowings where there is one standard error shock to monetary policy variables and other variables in the SVAR model. Figure 3 shows that the impulse response of external commercial borrowings to one standard error shock to Interest rate differential. As expected, the response is positive and statistically significant[10] for up to three quarters, indicating that an increase in the interest rate differential has a positive and statistically significant influence on external commercial borrowings to India. The impulse response of external commercial borrowings to a unit standard error shock to domestic money supply growth is shown in Fig. 4. It should be noted that while the response is negligible on impact, it turns negative and statistically significant in the first quarter, which is as expected. However, the response becomes insignificant thereafter.

The results in Fig. 3 and Fig. 4 suggest that domestic monetary policy shock has a statistically significant influence on external commercial borrowings to India. In particular, interest rate differential has a positive, and the

[10] 90% confidence interval based on Monte Carlo Integration is used to check the significance of the impulse responses.

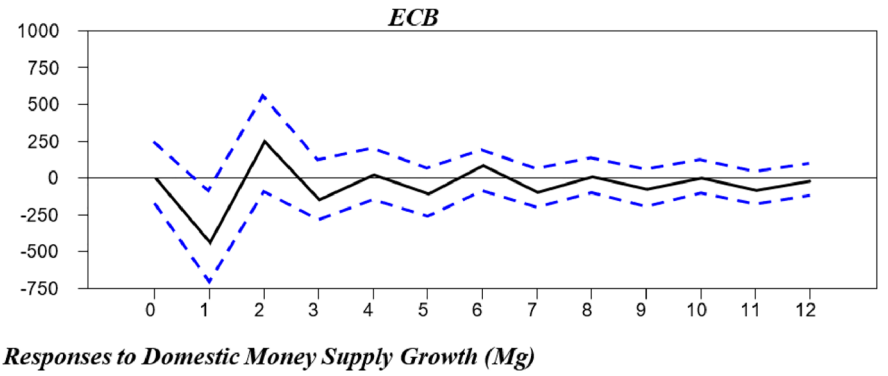


Fig. 4 Impulse response of external commercial borrowings (ECB) to one standard error shock to domestic money supply growth (Mg)

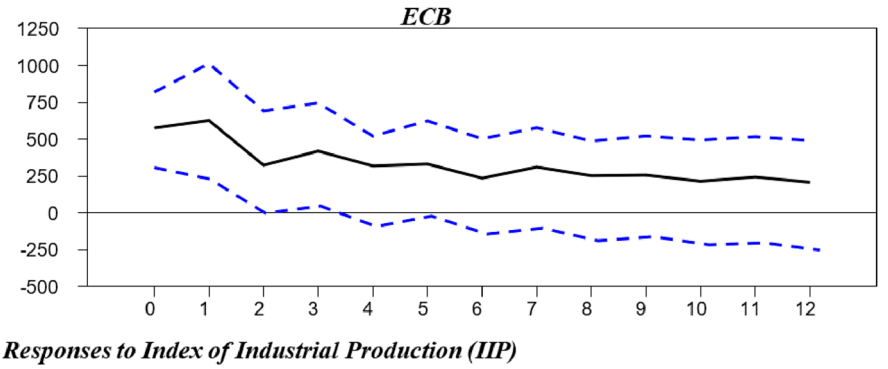


Fig. 5 Impulse response of external commercial borrowings (ECB) to one standard error shock to index of industrial production (IIP)

domestic money supply growth has a negative influence on external commercial borrowings.

The impulse response of external commercial borrowing to one standard error shock to index of industrial production representing domestic industrial activity is depicted in Fig. 5. Note that the response is positive and statistically significant for about three quarters and continues to be positive after that. This positive response indicates that an increasing domestic industrial activity induces greater external borrowings by domestic corporates to finance new investments or expand existing capacities. The response of external commercial borrowings to one standard error shock to foreign exchange reserves to import ratio representing domestic country's creditworthiness is positive and statistically significant on impact and in the first quarter (Fig. 6). We also find that the response has become persistently positive from the third quarter. This positive response indicates that a highly creditworthy country

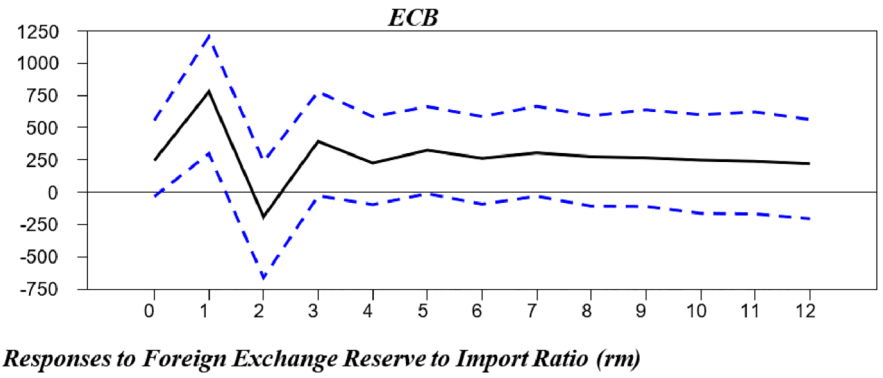


Fig. 6 Impulse response of external commercial borrowings (ECB) to one standard error shock to foreign exchange reserves to import ratio (rm)

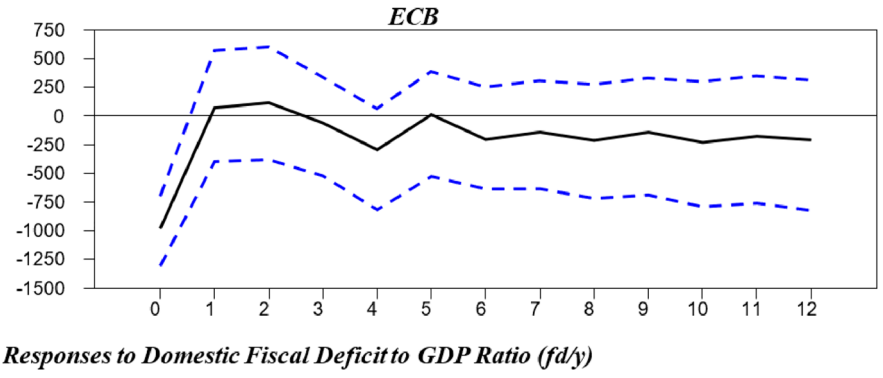


Fig. 7 Impulse response of external commercial borrowings (ECB) to one standard error shock to domestic fiscal deficit to GDP ratio (fd/y)

has easier access to international debt markets, allowing domestic corporates to raise more funds from external sources.

Figure 7 shows the impulse response of external commercial borrowings to one standard error shock to domestic fiscal deficit to GDP ratio. Note that the immediate response is highly negative and statistically significant, implying that a higher domestic fiscal deficit to GDP ratio discourages external commercial borrowings. A higher domestic fiscal deficit to GDP ratio indicates greater domestic macroeconomic instability, which increases investment risk and discourages domestic corporates from borrowing abroad to finance new investment ventures.

The impulse response of external commercial borrowings to a unit standard error shock to domestic output growth is shown in Fig. 8. We note that the response is positive and significant in initial quarters and continues to be positive even afterwards, which is in line with theory. This result indicates that an expanding domestic output provides a greater incentive for expansion and new investment ventures,

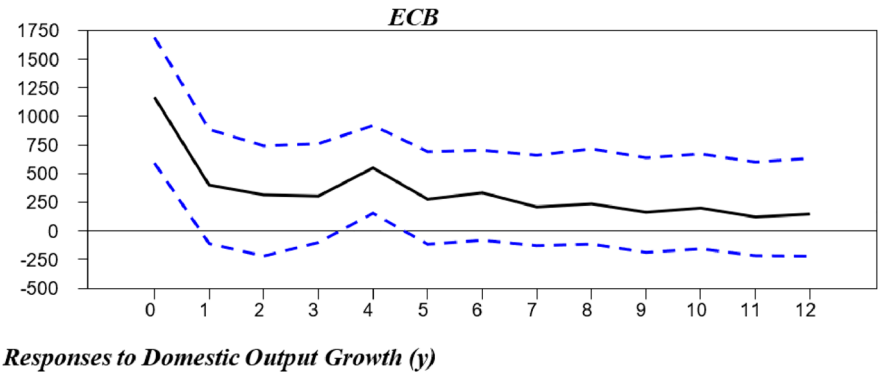


Fig. 8 Impulse response of external commercial borrowings (ECB) to one standard error shock to domestic output growth (y)

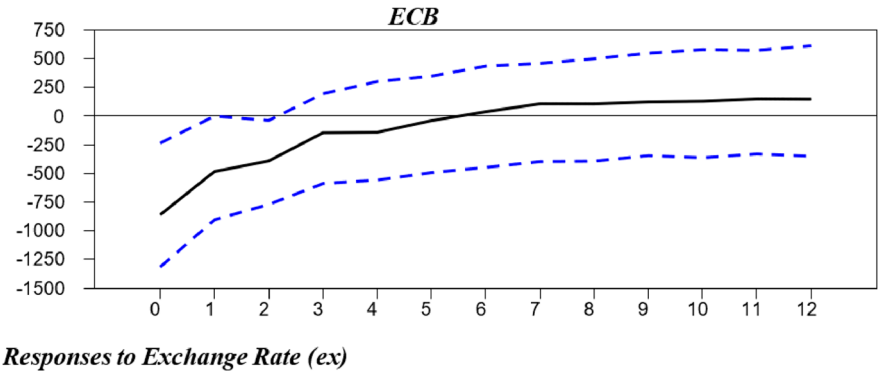


Fig. 9 Impulse response of external commercial borrowings (ECB) to one standard error shock to exchange rate (ex)

inducing domestic corporates to borrow more internationally. As expected, one standard error shock to the exchange rate, i.e., depreciation of the rupee**,** is found to have a negative and statistically significant impact on external commercial borrowings to India (Fig. 9). This result indicates that a rupee depreciation discourages external commercial borrowings by increasing the effective cost of overseas borrowings for Indian corporates.

5.2.2 Forecast error variance decomposition

The forecast error variance decomposition (FEVD) function indicates the proportion of overall variation in one variable that can be attributed to shocks from other variables in the model. In this study, it helps determine how much proportion of the total variation in ECBs to India is explained by the domestic monetary policy variables

Table 4 Forecast error variance decomposition

Period	*Mg*	*(i-i*)*	*y*	*ex*	*fd/y*	*rm*	*IIP*	*ECB*
1	1.306	15.961	10.937	6.804	5.084	0.930	3.726	55.252
2	1.955	15.036	10.823	7.556	5.258	6.412	5.912	47.048
3	2.210	15.267	10.942	8.351	4.545	6.211	5.948	46.526
4	2.215	15.137	10.297	8.269	4.144	7.064	6.345	46.529
5	2.401	15.299	10.139	8.021	4.359	7.181	6.172	46.428
6	2.386	13.367	10.520	7.866	4.128	7.621	6.228	47.884
7	2.048	13.628	10.855	7.744	4.035	7.823	6.112	47.755
8	2.046	13.858	11.014	7.683	3.942	8.198	6.033	47.226
9	2.081	14.162	11.150	7.685	3.907	8.342	5.949	46.724
10	2.075	14.474	11.118	7.747	3.859	8.501	5.886	46.340
11	2.087	14.737	10.048	7.859	3.878	8.555	5.847	46.989
12	2.076	14.948	10.948	8.146	3.907	8.592	5.817	45.566

and other variables in the SVAR model. The FEVD analysis (presented in Table 4) indicates that the interest rate differential explains about 14.6% and domestic money supply growth, on average, explains about 2% of the total variation in external commercial borrowings, implying that domestic monetary policy variables together explain about 16.6% of the total variation in ECBs to India. Other variables such as domestic industrial activity (represented by domestic IIP) explain about 5.8%, the exchange rate about 7.8%, domestic output growth about 10.7%, domestic macroeconomic instability (represented by domestic fiscal deficit to GDP ratio) about 4.2%, and domestic creditworthiness (represented by foreign exchange reserves to import ratio) about 7.1% of the total variation in external commercial borrowings to India.

5.2.3 Robustness Check

Interest Rate Differential We checked the robustness of the result using an alternative measure of interest rate differential, defined as the Call money rate minus 3-Month US dollar LIBOR. Figure 10 represents the impulse response of external commercial borrowings to one standard error shock to 'Call money rate minus LIBOR'. We note that the response is positive and significant, which verifies the results found in Fig. 3.

Exchange Rate We also checked the robustness of the result using an alternative measure of the exchange rate, the nominal effective exchange rate (NEER). Figure 11 represents the impulse response of external commercial borrowings to one standard error shock to the NEER. We note that the response is positive and significant for up to three quarters, suggesting that a shock to NEER (i.e. appreciation of the rupee) has positive impact on external commercial borrowings to India. This result is qualitatively similar to the result in Fig. 9.

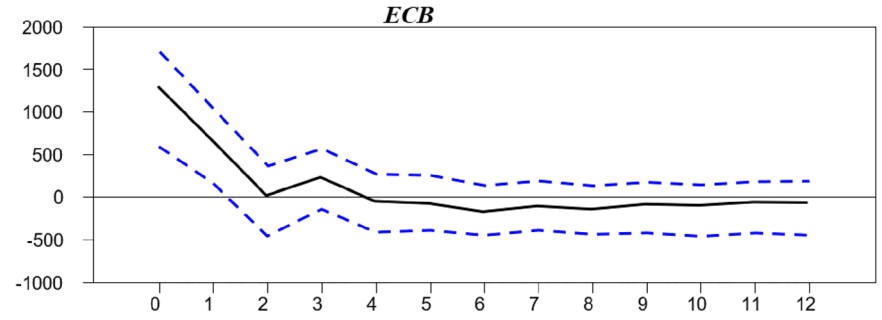


Fig. 10 Impulse response of external commercial borrowings (ECB) to one standard error shock to (call money rate minus LIBOR)

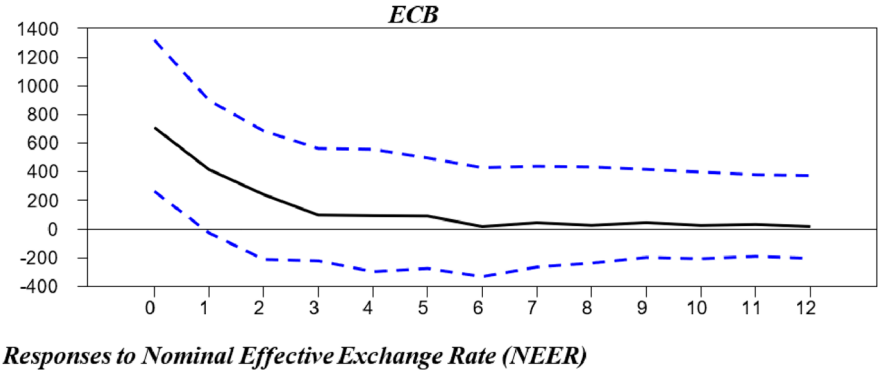


Fig. 11 Impulse response of external commercial borrowings (ECB) to one standard error shock to nominal effective exchange rate (NEER)

6 Conclusion and implication

This paper empirically examined the effect of domestic monetary policy shock on external commercial borrowings to India using the structural VAR methodology. Innovation accounting analysis under the structural VAR model (impulse response function and forecast error variance decomposition function) was utilized to derive empirical results.

The empirical results indicate that domestic monetary policy shock (measured through shocks in domestic money supply growth and interest rate differential) has a significant effect on external commercial borrowings to India, explaining about 16.6% of the total variation in ECBs to India. To be specific, the interest rate differential has a positive significant influence on external commercial borrowings (explaining about 14.6% of the total variation), and domestic money supply growth has a negative significant influence on external commercial borrowings (explaining

about 2% of the total variation). The results further indicate that the exchange rate, domestic industrial activity, domestic output growth, domestic' country creditworthiness, and domestic macroeconomic instability are the other significant determinants of external commercial borrowings to India, in the decreasing order of their importance.

The results suggest that the central bank, through monetary policy interventions, can influence and manage the flow of external commercial borrowings to India. Specifically, by implementing a contractionary monetary policy that results in a rise in interest rate differential, the central bank can persuade more ECBs, whereas by implementing an expansionary monetary policy that leads to a reduction in interest rate differential, the central bank can dissuade ECBs. The results further suggest that domestic policymakers should maintain adequate foreign exchange reserves to improve creditworthiness, follow policies to support economic and industrial growth, work on lowering macroeconomic instability, and control depreciation in the domestic currency to reduce the volatility associated with ECB flows.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40822-024-00275-x.

Funding The authors received no financial support for the research, authorship, and/or publication of this article.

Data availability The data can be made available upon request.

Declarations

Conflict of interest The authors have no conflicts of interest and have no relevant financial or non-financial interests to disclose.

References

Blanchard, O. J., & Watson, M. W. (1986). Are business cycles all alike? In R. Gordon (Ed.), *The American Business Cycle: Continuity and change* (pp. 123–156). University of Chicago Press.

Branson, W. H. (1972). *Macroeconomic equilibrium with portfolio balance in open economies*. Stockholm Institute for International Economic Studies, Seminar Paper no. 22.

Chinn, M. D., & Ito, H. (2008). A new measure of financial openness. *Journal of Comparative Policy Analysis, 10*(3), 309–322.

Dev, S. (2014). External commercial borrowings in India and its sensitivity to macroeconomic factors: An empirical analysis. *SSRN Electronic Journal,* 2406766.

Doan, T. A. (1990). *RATS user's manual. version 3.10. VAR econometrics*. Estima.

Han, Y., & Kim, M. H. (2023). Monetary shocks on the Korean stock index: Structural VAR analysis. *Eurasian Economic Review, 13*(1), 85–102.

Kim, S., & Roubini, N. (2000). Exchange rate anomalies in the industrial countries: A solution with a structural VAR approach. *Journal of Monetary Economics, 45*(3), 561–586.

Kouri, P. J., & Porter, M. G. (1974). International capital flows and portfolio equilibrium. *Journal of Political Economy, 82*(3), 443–467.

Maran, R. (2023). Impact of macroprudential policy on economic growth in Indonesia: A growth-at-risk approach. *Eurasian Economic Review, 13*(3), 575–613.

McKinnon, R. I., & Oates, W. E. (1966). *The implications of international economic integration for monetary, fiscal, and exchange-rate policy (No. 16)*. International Finance Section, Department of Economics, Princeton University.
Misra, P., Shettigar, J., & Oberoi, S. (2022). External commercial borrowings by Indian firms: A blessing or curse. *Indian Journal of Economics and Development, 18*(2), 488–494.
Patnaik, I., Shah, A. & Singh, N. (2016). Foreign Currency Borrowing by Indian Firms: Towards a New Policy Framework. National Institute of Public Finance and Policy, NIPFP Working paper series, working paper no. 167.
Pradhan, A. K., & Hiremath, G. S. (2019). External commercial borrowings by the corporate sector in India. *Journal of Public Affairs, 20*(1), e1987.
Pradhan, A. K., & Hiremath, G. S. (2020). Why do Indian firms borrow in foreign currency? *Margin: the Journal of Applied Economic Research, 14*(2), 191–211.
Sims, C. A. (1986). Are forecasting models usable for policy analysis? *Quarterly Review, 10*(Win), 2–16.
Sims, C. A., & Uhlig, H. (1991). Understanding unit rooters: A helicopter tour. *Econometrica, 59*(6), 1591–1599.
Singh, B. (2007). Corporate choice for overseas borrowings: The Indian evidence. Reserve bank of India Occasional Papers, *28*(3)
Singh, B. (2009). Changing contours of capital flows to India. *Economic Political and Weakly, XLIV*(43), 58–66.
Sur, A., Ray, P., & Nandy, A. (2019). India's external commercial borrowing: Pulled by domestic fundamentals or pushed by global conditions? *Journal of Asian Economics, 61*, 65–77.
Verma, R., & Prakash, A. (2011). Sensitivity of capital flows to interest rate differentials. *Department of Economic and Policy Research (DEPR)*, RBI Working Papers Series.

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.

An Improved Coot Optimization Algorithm for Optimal Power Flow in Small-Scale Power System

1st Udit Mittal
Electrical Engineering Department
Delhi Technological University
Delhi, India
udit.mittal27@gmail.com

2nd Uma Nangia
Electrical Engineering Department
Delhi Technological University
Delhi, India
umanangia@dce.ac.in

3rd Narender Kumar Jain
Electrical Engineering Department
Delhi Technological University
Delhi, India
vnarender84@yahoo.com

2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT) | 979-8-3503-8295-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICDCOT61034.2024.10515827

***Abstract*—The economic and reliable operation of modern electricity grids relies on insights from Optimal Power Flow (OPF) studies. Implementing a robust, swift, and reliable optimization algorithm is crucial for this purpose. This paper introduces the Improved Coot Optimization Algorithm (ICOA) to address OPF challenges in small-scale power systems. Integrating Lévy flight into population initialization strategically enhances the algorithm, improving the fitness of initial Coot individuals for superior OPF outcomes. The algorithm optimizes generation costs, considering the valve-point loading (VPL) effect and technical objectives related to voltage security. Simulations on the IEEE-30 bus system demonstrate ICOA's superior performance in minimizing fuel costs, improving the voltage profile, and enhancing stability compared to existing techniques. ICOA proves to be a promising tool for small-scale power grid management, offering economic and technical benefits with fast convergence. The findings highlight ICOA as an effective approach for addressing OPF challenges in small-scale grids for the considered set of objectives.**

Keywords—optimal power flow, coot algorithm, IEEE 30-bus, voltage deviation, L-index, valve-point loading

I. Introduction

In modern eco-friendly power practices, Optimal Power Flow (OPF) solutions are crucial for efficient network operation and planning. OPF optimizes control variables to achieve goals while adhering to essential network constraints. [1]. Contemporary power system management grapples with the intricate challenge of maximizing security and reliability within a complex framework of non-convex, large-scale, and non-linear constraints. Traditional approaches (linear programming, non-linear programming, interior point method, Newton's method etc.) are ill-suited for large-scale problems due to their complex and generalized calculations and heavy reliance on initial assumption values. Moreover, due to their non-linear characteristics, these methods cannot effectively address VPL effects, as they tend to converge towards local solutions [2]. In recent times, metaheuristic approaches have gained much popularity, especially when dealing with OPF problems. This is due to the fact that progressive metaheuristic techniques and evolutionary-based methods can avoid local minima by utilizing basic concepts that mimic nature and are versatile across a range of optimization problems. Researchers have developed innovative population-based optimization algorithms (OAs) and hybrids, as reviewed in [3] and [4]. Examples include well-established techniques like Genetic Algorithm (GA) [5], Particle Swarm Optimization (PSO) [6,7], alongside newer methods like Moth Flame Optimization (MFO) [8], Improved MFO [8], Teaching-Learning-Based Optimization (TLBO) [9], Moth Swarm Algorithm (MSA) [10], Jaya Optimizer [11], Modified Sine-Cosine Algorithm (MSCA) [12], Social-Spider Algorithm (SSO) [13], Equilibrium Optimizer (EO) [14], and Enhanced EO [14]. These methods have proven effective in solving OPF, but their application to different objectives results in a variety of solutions, which can affect the overall efficacy of the optimization process. Despite achieving satisfactory outcomes, the growing intricacies of optimization tasks present continuous challenges. Furthermore, the guarantee of achieving a global optimum is not consistent across all algorithms, as highlighted by the "No Free Lunch" theorem [15]. This limitation serves as an impetus for researchers to continually seek and develop new OAs to solve the OPF problem.

This paper introduces the Improved Coot Optimization Algorithm (ICOA) as a more efficient solution for OPF problems. ICOA incorporates a Lévy Flight (LF) component to enhance the basic Coot Optimization Algorithm detailed in reference [16]. This addition ensures comprehensive search coverage, enhances convergence towards optimal solutions, and effectively avoids local minima traps. Lévy flight is applied to randomly generated candidates during population initialization, enhancing their initial fitness for a robust algorithmic start. In the present work, ICOA has been applied to solve OPF with diverse objectives. The first objective is to minimize the total production cost, characterized as a non-convex function. Additionally, we aim to address crucial aspects of power system planning and operation, focused on improving the voltage profile and enhancing voltage stability, while considering the basic quadratic fuel cost (QFC). These constitute the second and third objectives, respectively.

The organization of the paper is as follows: Section II gives an overview of related OPF works for the IEEE 30-bus test system and offers the rationale behind the proposed ICOA. Section III outlines the mathematical model and objectives of the OPF problem, accompanied by a brief overview of the ICOA algorithm, incorporating the Lévy flight element. Section IV delves into simulation results, their detailed analysis, and initiates discussions on the obtained outcomes. Section V concludes the paper by summarizing key findings.

II. Related Works

Numerous detailed studies have been conducted in the field of OPF, examining a range of objectives across different test systems. In [17], the performance of the basic COA was investigated for the first time to address the OPF problem on a small-scale test system for a certain set of objectives. In [18], modifications and hybrid versions of the Flower Pollination Algorithm (FPA) were proposed to address its slow convergence rate and low precision. Specifically, the algorithm incorporated mutation and crossover processes inspired by Differential Evolution (DE), aiming to optimize the objective of minimizing fuel cost with VPL effect as one of the OPF objectives. In [19], an improved Salp Swarm Algorithm (ISSA) was introduced to address OPF with non-

979-8-3503-8295-2/24/$31.00 ©2024 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 06:02:07 UTC from IEEE Xplore. Restrictions apply.

smooth and non-convex generator fuel cost functions. The original SSA exhibits low search capability and may converge to the local optimal solution. To address this, ISSA incorporated random mutation for enhanced population diversity. While the obtained results were promising, there was a compromise in terms of voltage profile and voltage stability. The authors in [9] implemented TLBO approach for OPF. However, this approach may tend to converge to local optima due to its reliance on the current "best" solution during the teaching phase, potentially hindering exploration of untapped areas for globally better solutions. In [20], a multi-objective Manta Ray Foraging Optimizer (IMOMRFO) was introduced, incorporating a strategy based on crowding distance for Pareto archiving. Although this method was effective in lowering fuel costs, it necessitated a trade-off with the objective of minimizing voltage deviation. In [21], the authors introduced a hybrid approach combining crisscross search with the grey wolf optimizer, named CS-GWO, and enriched it with horizontal and vertical crossover techniques. While the outcomes for fuel cost, incorporating VPL effects, were favourable, they did not extend to improvements in voltage deviation performance when considering basic QFC. In [22], an Improved Archimedes Optimization Algorithm (IAOA) was proposed with an enhanced motion strategy featuring a dimension learning-based search approach. The authors considered fuel cost along with voltage deviation as objectives and achieved competitive results. The current study introduces an Improved Coot Optimization Algorithm (ICOA) for OPF. ICOA incorporates Lévy Flight for enhanced efficiency, faster optimal solution attainment, and avoidance of local minimum traps. The algorithm is applied effectively to diverse OPF objectives on the IEEE 30-bus test system.

III. Proposed Method: Improved Coot Optimization Algorithm (ICOA)

OPF poses a challenging task owing to its intricate, non-linear nature. It aims to determine optimal values for various control variables within the power system. This involves minimizing a specified objective function while adhering to operational constraints, including network limits.

A. Components of an OPF Model

The OPF tool solves power flow equations, assigns control variable values, and identifies optimal solutions with equality and inequality constraints. Mathematically, OPF can be stated as follows:

$$\text{Minimize: } J(x,u) \tag{1}$$

$$\text{Subject to: } g(x,u) = 0 \tag{2}$$

$$\text{and } h(x,u) \le 0 \tag{3}$$

Here, x represents a state vector, holding the system's dependent variables, while u denotes a vector of independent variables serving as the control vector. Equation (1) represents the objective function targeted for minimization, while (2) and (3) depict the equality and inequality constraints, respectively, that the solution must adhere to. The representation of state variables in OPF is provided in (4) as follows:

$$x = [P_{G_1}, \mathrm{V}_{L_1} \ldots \mathrm{V}_{L_{NLB}}, \mathrm{Q}_{G_1} \ldots \mathrm{Q}_{G_{NG}}, S_{BR_1} \ldots S_{BR_{NL}}] \tag{4}$$

At the slack bus (bus 1), the generated active power is referred to as PG1. The voltage at the load bus, the amount of reactive power produced by the generator, and the flow of apparent power in the branch are denoted as VL, QG and SBR respectively. NLB represents the total load buses, NG denotes the thermal generating units, and NL signifies the power transmission branches. The control vector is expressed as shown in (5):

$$u = [P_{G_2} \ldots P_{G_{NG}}, \mathrm{V}_{G_1} \ldots \mathrm{V}_{G_{NG}}, \mathrm{Q}_{Sh_1} \ldots \mathrm{Q}_{Sh_{NC}}, \mathrm{T}_1 \ldots \mathrm{T}_{NT}] \tag{5}$$

Where, VG signifies generator bus voltage, QSh represents shunt compensator injected reactive power, and T represents the tap-changing transformer setting. The count of shunt compensator units and tap-changing transformers is represented by NC and NT, respectively.

B. Objective Functions (OFs)

This study incorporates three single-objective OPF (SOOPF) functions, comprising an economy-based OF and two technical voltage-based OFs. The latter two functions focus on enhancing voltage stability (VSE) and improving the voltage profile (VPI).

1) Fuel cost minimization including valve-point loading effects (FCVPL): The cost-related objective is a fundamental aspect of OPF, extensively studied in the literature, constituting the primary and foremost OF. The fuel cost and the generated power exhibit an approximate quadratic relationship, characterized by the basic QFC objective function expressed in (6) [23]:

$$F(P_G) = \left(\sum_{i=1}^{NG} a_i P_{G_i}^2 + b_i P_{G_i} + c_i \right) (\$ / \mathrm{h}) \tag{6}$$

Here, F (PG) is the fuel cost function for ith generator in \$/h, dependent on its real power output P_{G_i} and influenced by coefficients ai, bi and ci. However, incorporating the VPL effect refines the quadratic cost function, enhancing its precision and realism. VPL effect arises when thermal generation unit valves open during steam admission, causing sudden losses and ripples in the cost function curve. Thus, the cost function with VPL effect is described as per (7) [24]:

$$F(P_G) = \left(\sum_{i=1}^{NG} a_i P_{G_i}^2 + b_i P_{G_i} + c_i \right) + \left| e_i \times \sin(g_i (P_{G_i}^{\min} - P_{G_i})) \right| \tag{7}$$

Where, e_i and g_i represent the VPL cost coefficients for the ith unit, with $P_{G_i}^{\min}$ representing the minimum active power generation from the same unit.

2) Voltage profile improvement and basic QFC minimization (VPIFC): The second essential objective in OPF involves optimizing the voltage profile throughout the power system. This is achieved by minimizing the deviation of voltage (VD) at each load bus from its ideal value of 1.0 p.u. In most instances, the consideration of the fuel cost function is typically linked with the voltage profile improvement (VPI), as outlined in (8) [25].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 06:02:07 UTC from IEEE Xplore. Restrictions apply.

$$F_{VPIFC} = \left(\sum_{i=1}^{NG} a_i P_{G_i}^2 + b_i P_{G_i} + c_i \right) + \mu_{vp} \sum_{i=1}^{NLB} |V_i - 1.0| \quad (8)$$

Here, the weighting factor (μvp) assigned to VPI was taken as 100.

3) Voltage stability enhancement and basic QFC minimization (VSEFC): Voltage stability is essential for the consistent performance of a power system, ensuring stable voltage levels under normal operating conditions. However, voltage instability can occur due to disruptions, changes in system configuration, and heightened load demand. Lowering the L-index (voltage stability index) serves as a method to enhance voltage stability [26]. Mathematically, (9) represents a single-objective function by incorporating both basic QFC minimization and L-index reduction.

$$\left.\begin{aligned} F_{VSEFC} &= \left(\sum_{i=1}^{NG} a_i P_{G_i}^2 + b_i P_{G_i} + c_i \right) + \mu_{si} \times L \\ where, \quad L &= \max \left\{ \left| 1 - \sum_{i=1}^{NG} F_{ji} \frac{V_i}{V_j} \right| \right\} = \max(L_j) \quad \forall\ j = 1,2,....NLB \end{aligned}\right\} \quad (9)$$

Here, L_j is the local voltage stability indicator for j^{th} load bus. F_{ji} is calculated using system bus admittance sub matrices. In this study, the weighting factor μ_{si} is set to 100.

C. Operating Constraints

Addressing the OPF problem demands considering two distinct constraint categories: equality constraints and inequality constraints [19]. The equality constraints pertain to non-linear power flow equations, while the inequality constraints encompass various operational limits of power system equipment. These limits include constraints on voltage magnitudes of generator and load buses, active and reactive powers of generators, tap changer ratios, and compensating reactive power limits.

D. Basic COOT Algorithm

The ICOA is an improved version of basic COA by Naruei and Keynia in 2021 [16]. The Coot algorithm draws inspiration from collective behaviours and movement movements of a swarm of birds, known as coots, reflecting their two distinct types of behavior on water. Initially, the birds disperse with unpredictable changes in direction. Subsequently, they synchronize their movements, swimming together with similar speeds, forming a denser group as shown in Fig. 1.

E. The Proposed ICOA

The proposed modification in COA incorporates Lévy flight (LF) features to escape local optima, allowing individuals to traverse a wider solution space and increasing the probability of identifying the global optimum. In [27], it is highlighted that Lévy flight effectively escapes local minima, enhancing efficiency in solving complex problems with challenging boundaries. The three fundamental mathematical steps of ICOA are as follows:

- Initialization with LF distribution.
- Updating fitness and designating the Coot leader.
- Updating Coot positions.



Fig. 1. Coots synchronized in elegant glide over water.

Step 1 Population initialization with LF distribution

In the ICOA algorithm, the Lévy Flight mechanism is used during the initialization phase to an initial population of coots (NPOP). The algorithm initializes the population with LF distribution, which ensures that the agents are spread out across the search space and move in a stochastic manner and explore different regions of the search space. The proposed modification to the basic COA, namely ICOA, revolves around utilizing LF distribution for population initialization, as per (10).

$$CP(i_initial) = LB + Levy(\lambda) \times (UB - LB) \quad (10)$$

Here, CP(i_initial) is the initial position of the of ith coot, Levy (λ) is a random search path formulated as follows in (11).

$$Levy(\lambda) = 0.01 \cdot \frac{u}{|v|^{1/\beta}} \quad (11)$$

Here, random numbers u and v follow a Gaussian distribution, and the distribution is controlled by the power-law index β. We have chosen a value of 1.5 for β.

Step 2 Updating fitness and designating the Coot leader

The fitness function is computed for each agent's position obtained from (10). Coot leaders (N_{Leader}) are randomly chosen from the total population, while the remaining coots are designated as coot followers (N_{COOT}, where $N_{COOT} = N_{POP} - N_{Leader}$).

Step 3 Updating Coot positions

In this step, the algorithm simulates four distinct movement behaviors of coots, mimicking their actions on water.

1) Random movement: The algorithm explores diverse areas of the solution space to avoid local optima, determining the new position randomly using (12).

$$CP(i) = CP(i) + T_1 \times r2 \times \left[\{ rand(1,D).*(UB - LB) + LB \} - CP(i) \right] \quad (12)$$

Here $r2$ is chosen arbitrarily between [0-1]. D denotes the problem dimension, determined by the count of involved decision variables and T_1 is determined using (13).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 06:02:07 UTC from IEEE Xplore. Restrictions apply.

$$T_1 = 1 - \left(\frac{It}{It_\max} \right) \tag{13}$$

Here, the current iteration number is *It*, with the maximum number of iterations denoted as *It*_max.

2) Chain movement: The chain movement for a pair of coots is described as per (14).

$$CP(i) = 0.5 \times \left[CP(i-1) + CP(i) \right] \tag{14}$$

Where, $CP(i)$ denotes second coot's position when it has moved half way towards the first.

3) Update through leaders: Coot leaders guide the group, and followers adjust positions accordingly. Each follower is associated with a leader, updating its location based on the average leader position as per (15). Every coot (i) in the group aligns its position with the leader (k) and adjusts its position accordingly. The next coot position concerning its associated leader is expressed in (16).

$$k = 1 + MOD\ (i,\ N_{\text{Leader}}) \tag{15}$$

$$CP(i) = LP(k) + 2 \times r1 \times \cos(2\pi R) \times \left[LP(k) - CP(i) \right] \tag{16}$$

Here, the position of the ith coot denoted as $CP(i)$, while the position of selected leader (k) is represented by $LP(k)$. Random numbers $r1$ and R assume values chosen from the intervals [0, 1] and [-1, 1], respectively.

4) Global best position update for leaders: The leader's position is updated using (17) to approach the optimal position.

$$LP(i) = \begin{cases} (G_{Best} - LP(i)) \times T_2 \times r3 \times \cos(2\pi R) + G_{Best} & ; r4 < \frac{1}{2} \\ (G_{Best} - LP(i)) \times T_2 \times r3 \times \cos(2\pi R) - G_{Best} & ; r4 \geq \frac{1}{2} \end{cases} \tag{17}$$

Here, G_{Best} represents the position of the global optimum, with $r3$ and $r4$ being randomly selected from the interval [0-1]. T_2 is determined using (18).

$$T_2 = 2 - (It\ /\ It_\max) \tag{18}$$

For additional clarity, the utilization of the ICOA in implementing an OPF solution is illustrated in Fig. 2 through a flowchart.

IV. Results and Discussion

The effectiveness of the proposed improved Coot algorithm (ICOA) in resolving the OPF issue was assessed using the IEEE-30 bus system as a test system. Performance evaluation involved solving SOOPF problems with fuel cost (with VPL effect), total voltage deviation, and voltage stability as individual OFs defined in cases 1–3, respectively. Numerical results were then presented and compared with outcomes from other metaheuristic methods. The ICOA configuration used a population size of 40 and ran for 100 iterations on a standard laptop equipped with an Intel Core i5 processor and 8GB RAM, running MATLAB 2018a.

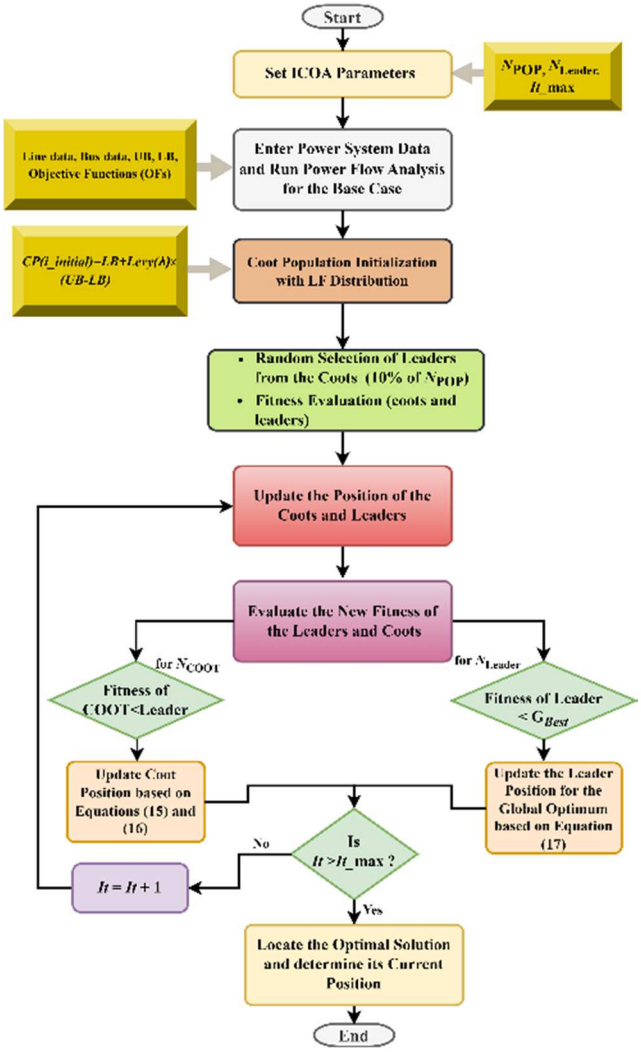


Fig. 2. Flowchart depicting the proposed ICOA for OPF.

A. IEEE 30-bus Test System

The IEEE 30-bus IEEE system represents a power grid with 6 generators, 24 load buses, and 41 connecting lines. In this configuration, bus 1 functions as the slack bus. The allowed voltage range for generator buses is defined as $0.95 \leq V_G \leq 1.1$ p.u., while load buses must maintain voltages within $0.95 \leq V_L \leq 1.05$ p.u. Tap changers operate between 0.9 and 1.1 p.u., while VAR compensators range from 0 to 0.05 p.u. Further details regarding the IEEE 30-bus test system can be found in [28].

B. Case Studies

1) Case 1- OPF for FCVPL: This case focuses on lowering fuel costs while accounting for VPL effects using ICOA, and the corresponding OF is formulated in (7). The fuel cost attributed to VPL effect decreases from 832.3648 \$/h for the basic COA to 832.1201 \$/h with ICOA. In this case, except for IMFO and PSO, the proposed method reduces fuel costs by 0.0074%, 0.0319%, 0.0384%, 0.0031%, and 10.62%, respectively, for EEO, MFO, GA, TLBO, and MSA. The proposed technique demonstrates better performance than other well-established metaheuristics in terms of accuracy and validation, as evidenced by the results.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 06:02:07 UTC from IEEE Xplore. Restrictions apply.

The obtained total cost with VPL is compared with other algorithms implemented to solve the given case under the same conditions, as shown in Table I.

TABLE I. PERFORMANCE COMPARISON OF VARIOUS METHODS FOR CASE 1

Method	Fuel cost with VPL effect ($/h)	Total voltage deviation (p.u.)	L-index
ICOA	832.1201	0.8235	0.1265
COA	832.3648	0.8311	0.1288
EEO [14]	832.1817	0.841403	0.129291
EO [14]	832.1969	0.841403	0.126421
PSO [8]	832.1095	NR*	NR
IMFO [8]	832.1023	0.8318	0.1281
MFO [8]	832.3871	0.5417	0.1320
GA [8]	832.4400	0.7282	0.1289
TLBO [8]	832.0944	0.8637	0.1280
MSA [10]	930.7441	0.4492	0.15676

*Not Reported

Additionally, the convergence traits of ICOA and basic COA are depicted in Fig. 3, revealing better convergence of ICOA within 100 iterations, especially for small-scale OPF problems.

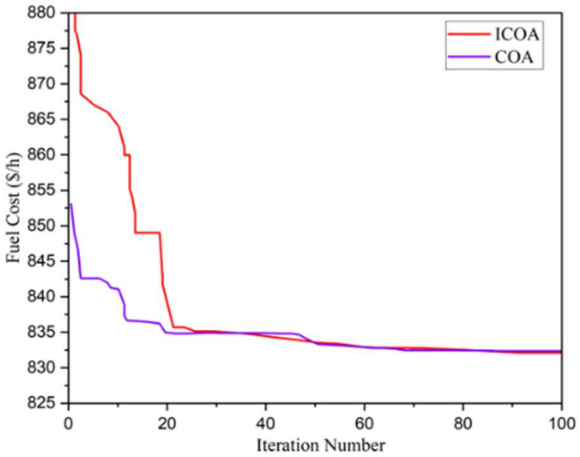


Fig. 3. Convergence plots for Case 1 by COA and proposed ICOA.

2) Case 2- OPF for VPIFC: In this case, OPF solutions are sought to simultaneously minimize the basic QFC and voltage deviation (i.e. improving voltage profile), forming a two-fold objective function as per (8). The reduction in load bus voltage deviation from 0.1126 (COA) to 0.0884 (ICOA) signifies a substantial improvement of around 21.43% compared to the basic COA. In relation to the initial case (1.1496 p.u.), the improvement reaches an impressive 92.33%. Results obtained from the proposed ICOA and well-established algorithms such as MGOA, GOA, EEO, GA, PSO, TLBO, Jaya, and MSCA for case 2 are presented in Table II, highlighting the superior solution achieved by the proposed algorithm over the others.

TABLE II. PERFORMANCE COMPARISON OF VARIOUS METHODS FOR CASE 2

Method	Total voltage deviation (p.u.)	L-index	Fuel cost (basic QFC) ($/h)
ICOA	0.0884	0.1365	803.4796
COA	0.1126	0.1369	804.3822
EEO [14]	0.089035	0.1363	864.2726
EO [14]	0.0905	0.1366	809.101
MGOA [29]	0.1107	0.1365	803.4176
GOA [29]	0.1709	0.1353	803.4488
GA [29]	0.1257	0.1379	803.9156
PSO [29]	0.1110	0.1367	803.6227
TLBO [29]	0.1087	0.1367	804.5827
Jaya [29]	0.1243	NR	840.7181

Fig. 4 depicts a bar chart illustrating load voltage deviation from 1.0 p.u., emphasizing ICOA's enhanced performance over basic COA and the base case.

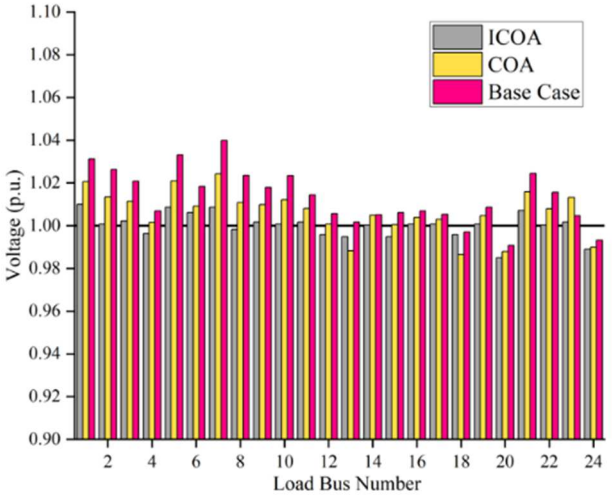


Fig. 4. Comparing voltage profiles from ICOA, COA, and base case.

3) Case 3- OPF for VSEFC: Maintaining bus voltages within operational limits is crucial for power system stability. The L-index, ranging from 0 to 1, signifies stability levels. A lower L-index signifies minimal or stable voltage loads, while an L-index near 1 suggests potential network collapse. This case aims to enhance voltage stability while simultaneously minimizing basic QFC through a two-fold objective approach. The OF related to this is defined in (9). The ICOA achieves an L-index value of 0.1243, indicating a 0.56% improvement over basic COA for the same case and a substantial 42.02% improvement over the base case value (i.e. 0.2144), as presented in Table III.

TABLE III. PERFORMANCE COMPARISON OF VARIOUS METHODS FOR CASE 3

Method	L-index	Total voltage deviation (p.u.)	Fuel cost (basic QFC) ($/h)
ICOA	0.1243	0.7532	801.912
COA	0.125	0.8162	802.661
EEO [14]	0.124236	0.980597	931.9871
EO [14]	0.124268	0.974147	887.8918
MGOA [29]	0.1255	0.8302	800.5595
GOA [29]	0.1281	0.6546	801.7874
IMFO [8]	0.1255	0.9086	800.4762
MFO [8]	0.1266	0.7789	800.9415
GA [29]	0.1266	0.7275	801.0195
PSO [29]	0.1278	0.7972	800.7512
TLBO [29]	0.1256	0.7450	800.7281
SSO [13]	0.1267	NR	NR
NISSO [13]	0.12547	NR	NR
ABC [30]	0.1379	NR	NR

Fig. 5 reveals that ICOA converges faster and achieves better solutions than the original COA.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 06:02:07 UTC from IEEE Xplore. Restrictions apply.

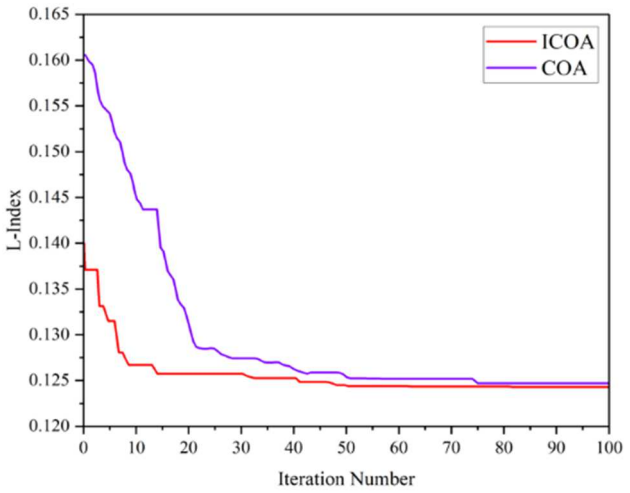


Fig. 5. Convergence plots for Case 3 by COA and proposed ICOA

V. Conclusion

This study effectively applies ICOA and basic COA to solve the SOOPF problem in small-scale power grids, prioritizing a well-defined set of objectives. The proposed ICOA optimizes generation costs by incorporating the VPL effect for thermal generators and includes technical objectives related to voltage stability and deviation, forming a two-fold objective function. The LF mechanism within the ICOA algorithm plays a vital role in diversifying the population by enabling particles to effectively explore the search space and prevent entrapment in local optima. Integrating LF into population initialization strategically improves the algorithm's starting point by enhancing the fitness of initial individuals, leading to better overall performance. The proposed ICOA method exhibits superior fuel cost reduction compared to existing techniques, as evidenced by results on the IEEE-30 bus system. This affirms its validity and accuracy in OPF problem-solving in small-scale power grids. Additionally, ICOA improves voltage profile and stability, ensuring feasible solutions with fast convergence, making it a promising tool for managing small-scale power grids with both economic and technical benefits.

While this paper focused on addressing SOOPF problems with constraints, there is potential for further development in future research. The proposed algorithm can be extended to handle multi-objective OPF (MOOPF) in real-world power systems with a diverse range of energy sources, including renewables and thermal generators. Additionally, there is an opportunity to explore incorporating extra objectives and constraints, addressing the complexities in present-day power systems.

References

[1] M. A. Abido, "Optimal power flow using particle swarm optimization", Int. J. Electr. Power Energy Syst., vol. 24, pp. 563–571, 2002.

[2] S. Frank, I. Steponavice and S. Rebennack, "Optimal power flow: A bibliographic survey I", Energy Syst., vol. 3, no. 3, pp. 221-258, Sep. 2012.

[3] S. Frank, I. Steponavice and S. Rebennack, "Optimal power flow: A bibliographic survey II", Energy Syst., vol. 3, no. 3, pp. 259-289, Sep. 2012.

[4] U. Mittal, U. Nangia and N. K. Jain, "Computational Intelligence-Based Optimal Power Flow Methods-A Review", 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, pp. 1-8, 2022.

[5] A.-F. Attia, Y. A. Al-Turki and A. M. Abusorrah, "Optimal power flow using adapted genetic algorithm with adjusting population size", Electr. Power Compon. Syst., vol. 40, no. 11, pp. 1285-1299, 2012.

[6] S. He, J. Wen, E. Prempain, Q. Wu, J. Fitch and S. Mann, "An improved particle swam optimization for optimal power flow", Proc. Int. Conf. Power Syst. Technol., pp. 1633-1637, Nov. 2004.

[7] J Hazra and AK Sinha, "A multi-objective optimal power flow using particle swarm optimization", Eur Trans Electr Power, vol. 21, no. 1, pp. 1028-1045, 2011.

[8] M. A. Taher, S. Kamel, F. Jurado and M. Ebeed, "An improved moth-flame optimization algorithm for solving optimal power flow problem", Int. Trans. Electr. Energy Syst., vol. 29, no. 3, pp. e2743, Mar. 2019.

[9] M. Ghasemi, S. Ghavidel, M. Gitizadeh and E. Akbari, "An improved teaching–learning-based optimization algorithm using Lévy mutation strategy for non-smooth optimal power flow", Int. J. Electr. Power Energy Syst., vol. 65, pp. 375-384, 2015.

[10] A.-A.-A. Mohamed, Y. S. Mohamed, A. A. M. El-Gaafary and A. M. Hemeida, "Optimal power flow using moth swarm algorithm", Electr. Power Syst. Res., vol. 142, pp. 190-206, Jan. 2017.

[11] W. Warid, H. Hizam, N. Mariun and N. Abdul-Wahab, "Optimal power flow using the Jaya algorithm", Energies, vol. 9, no. 9, pp. 678, Aug. 2016.

[12] A. F. Attia, R. A. El Sehiemy and H. M. Hasanien, "Optimal power flow solution in power systems using a novel sine-cosine algorithm", Int. J. Electr. Power Energy Syst., vol. 99, pp. 331-343, Jun. 2018.

[13] T. T. Nguyen, "A high performance social spider optimization algorithm for optimal power flow solution with single objective optimization", Energy, vol. 171, pp. 218-240, Mar. 2019.

[14] E. H. Houssein, M. H. Hassan, M. A. Mahdy and S. Kamel, "Development and application of equilibrium optimizer for optimal power flow calculation of power system", Appl. Intell., vol. 53, no. 6, pp. 7232-7253, Mar. 2023.

[15] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization", IEEE Trans. Evol. Comput., vol. 1, no. 1, pp. 67-82, Apr. 1997.

[16] I. Naruei and F. Keynia, "A new optimization method based on coot bird natural life model", Expert Systems with Applications, vol. 183, no. 30, pp. 115352, 2021.

[17] U. Mittal, U. Nangia and N. K. Jain, "Application of the Coot Optimization Algorithm for Optimal Power Flow Solution," 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, pp. 362-368, 2023.

[18] G. Chen, Q. Qin, Z. Ping, K. Peng, X. Zeng, H. Long and M. Zou, "A Novel Approach Based on Modified and Hybrid Flower Pollination Algorithm to Solve Multi-objective Optimal Power Flow". IAENG International Journal of Applied Mathematics, vol. 51, no. 4, pp.966-983, 2021.

[19] S. A. El-Sattar, S. Kamel, M. Ebeed and F. Jurado, "An improved version of salp swarm algorithm for solving optimal power flow problem", Soft Comput., vol. 25, pp. 1-26, Jan. 2021.

[20] H. T. Kahraman, M. Akbel and S. Duman, "Optimization of optimal power flow problem using multi-objective manta ray foraging optimizer", Appl. Soft Comput., vol. 116, Feb. 2022.

[21] A. Meng, C. Zeng, P. Wang, D. Chen, T. Zhou, X. Zheng, et al., "A high-performance crisscross search based grey wolf optimizer for solving optimal power flow problem", Energy, vol. 225, Jun. 2021.

[22] O. Akdag, "A improved Archimedes optimization algorithm for multi/single-objective optimal power flow", Electric Power Systems Research, vol. 206, pp. 107796, 2022.

[23] K. Abaci and V. Yamacli, "Differential search algorithm for solving multi-objective optimal power flow problem", Int. J. Electr. Power Energy Syst., vol. 79, pp. 1-10, Jul. 2016.

[24] T. Niknam, M. R. Narimani and R. Azizipanah-Abarghooee, "A new hybrid algorithm for optimal power flow considering prohibited zones and valve point effect", Energy Convers. Manage., vol. 58, pp. 197-206, Jun. 2012.

[25] O. Akdag, A. Ates and C. Yeroglu, "Modification of Harris hawks optimization algorithm with random distribution functions for optimum power flow problem", Neural Comput. Appl., vol. 33, no. 6, pp. 1959-1985, 2021.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 06:02:07 UTC from IEEE Xplore. Restrictions apply.

[26] K. P. Kessel and H. Glavitsch, "Estimating the Voltage Stability of a Power System", IEEE Transactions on Power Delivery, vol. PWRD-1, no. 3, pp. 346-354, July 1986.

[27] S. K. Dinkar and K. Deep, "An efficient opposition based Lévy flight antlion optimizer for optimization problems", J. Comput. Sci., vol. 29, pp. 119-141, Nov. 2018.

[28] E. Naderi, M. Pourakbari-Kasmaei and H. Abdi, "An efficient particle swarm optimization algorithm to solve optimal power flow problem integrated with FACTS devices", Appl. Soft Comput., vol. 80, pp. 243-262, Jul. 2019.

[29] M. A. Taher, S. Kamel, F. Jurado and M. Ebeed, "Modified grasshopper optimization framework for optimal power flow solution", Electr. Eng., vol. 101, no. 1, pp. 121-148, Apr. 2019.

[30] M. R. Adaryani and A. Karami, "Artificial bee colony algorithm for solving multi-objective optimal power flow problem", Int. J. Elect. Power Energy Syst., vol. 53, pp. 219-230, Dec. 2013.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 06:02:07 UTC from IEEE Xplore. Restrictions apply.

2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)
Amity University, Noida, India. Mar 14-15, 2024

2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) | 979-8-3503-5035-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICRITO61523.2024.10522184

Analysis of Area-Controlerror in Automatic Generation Control

Jitesh Wadhwa
Deptt. of Electrical Engineering
Delhi Technological University
Delhi, India - 110042
jitesh1941@gmail.com

Hayatunisha
Deptt. of Electrical Engineering
Delhi Technological University
Delhi, India - 110042
hayatunisha1@gmail.com

Narendra Kumar
Deptt. of Electrical Engineering
Delhi Technological University
Delhi, India - 110042
narendra.kumar@dtu.ac.in

***Abstract*—The Automatic Generation Control shares an objective to regulate the active power production according to the different power requirements, thus maintaining the expected energy flow and frequency in the network itself by connecting the desired control areas with power connection. A sudden change in energy can affect the magnetic currents in various motors and also cause instability in various electrical systems. In a modern and complex energy system, AGC plays an essential role and becomes an essential element in ensuring high quality energy supply to consumers, regardless of constant fluctuations in energy demand. When there are load fluctuations in the power grid, AGC plays a key role in maintaining optimal power exchange levels and grid frequencies while taking into account disturbance caused by load peaks. Therefore, to provide standard performance, the AGC system requires an intelligent efficient controller. The performance of different controllers depends on the choice of controller parameters used as well as the specific optimization algorithms.**

In literature research, the design and research of AGC has been further improved using advanced control strategies and intelligent metaheuristic strategies. Therefore, many researches have been conducted worldwide to provide many new control strategies to effectively solve the AGC problem. The centric goal of the proposed work is to study the concept of Automatic Generation Control while considering concept of area control with their respective simulations.

Keywords – Automatic Generation Control, Control Area, Area Control Error

I. INTRODUCTION

In a power system electrical energy plays a crucial role. The instability between load demand and generated power makes frequency of the system to diverge from its standard value. This will lead to unintended power transfer between various interconnected systems. The power system consists of many controlling areas and these controlling areas are being connected to each other through a transmission line. In such type of controlling areas, the generators swing (i.e., accelerate or deaccelerate) together. For maintaining the eminence of electrical power supply rather than storing it, is to maintain the frequency under different loaded conditions is defined as LFC (load frequency control). There will always be discrepancy between power demand and generation due to fluctuation in the load as shown in Fig. 1. In order to deliver reliable supply of electricity to buyers it is necessary to equalize between production of energy and demand energy. Due to environmental issues, it has become a challenging task to achieve the prerequisite levels of energy due to increase in requirement of electricity.

Modern power systems consist of several controlling areas that are interrelated to each other through a tie-line. During different time zones of a day, the change in load demand leads manual control of maintaining the precise balance to be ineffective. However, the balance among the generation and load power demand is achieved efficiently via AGC schemes. Whenever there is a sudden disturbance or the load change in any of the area it will lead to variation of frequency and if produced active power tends to become more than the required power, the generating unit's frequency starts increasing and vice versa. So, the system will always aim to maintain tie-line power and nominal frequency of a connected power system. To retain the systems tie-line power & frequency at their nominal values AGC plays a principal role in an interconnected power system. However, the speed governor alone cannot obtain the constant frequency.

II. THEORY

If there is a sudden change in the load and the load in the system starts increasing then the speed of the turbine starts dropping before the governor can alter the steam's input. The output power of electric generators within a defined area in respect to change in systems tie-line power and frequency change is regulated to maintain scheduled frequency of the system within the defined limits. For achieving the higher efficiency and to improve the quality as well stability of the electrical supply it is necessary to maintain the nominal tie-line power as well frequency deviation.

Whenever there is active power variation between load and generation this will lead to deviation in the frequency and deviation in voltage is due to imbalance reactive power in a power system. In a given power system by regulating the generation active power balance can be attained is termed as Automatic Generation Control (AGC). In a power system when there is a small load disruption, this will lead to a difference in real power and which will mainly affect the frequency deviation of the system but the level of the voltage remains unaffected as shown in Fig. 2. If there is a sudden change in reactive power this will affect the level of the bus voltage and the frequency of the system remains unaltered.

There are few objectives of an AGC system:

i. Frequency of the system should be maintained at a nominal value i.e., frequency deviation (Δf) should be reduced to zero without any delay.

ii. Change in tie-line power (ΔP_{tie}) must also be made zero as quickly as possible.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:14:48 UTC from IEEE Xplore. Restrictions apply.

iii. In each area the optimal scheduling of generators should be done.

The active power and reactive power flow in a transmission network are independent of each other and both demands are never balanced, they keep on changing depending upon the falling or rising movement. For determining the quality of power supply, voltage & frequency are the most important factors.

By managing the output of generators, LFC is required to maintain power interchange between areas. Irrespective of the conventional approach, new methods and controllers are being executed in the investigation and design of AGC recently. So, the main purpose of AGC is:

(1) to make sure that the bus voltages, frequencies and currents are sustained at a definite value.

(2) to observe the power supply frequency and tie-line power in load sharing areas. In Table 1.1 the overall operation and controlling of the power system on time scale.

III. Load Frequency Control

Load frequency control (LFC) is a simple governing mechanism in the operation of the power system. When the demand on the output changes, a sudden imbalance occurs between the input power and the output. This difference is due to the energy stored in the rotating part of the unit.

The kinetic-energy(KE) of any unit is given by

$$K.E. = 0.5\, J\omega^2$$

Here J is the moment-of-inertia of the rotatory part and ω is the angular speed of the rotatory part.

i. If the KE decreases, angular speed reduces which leads to decrease in speed and therefore the frequency will decrease.

ii. The difference in frequency Δf is detected and it is fed back using a speed governor system to adjust the positioning of the valve which is connected to the generation unit.

iii. Re-positioning of the valve result in the balance of real power input and output. Hence, the variation of frequency depends on the balance of real-power of the system.

iv. The LFCfurthergoverns the real-powertransmissionvia the interconnectedtransmission lines by detecting the variation in flow of power from the tie lines.

IV. Automatic Load - Frequency Control (ALFC)

The main functions of ALFC are: -

a) To regulate the required MW output of the generator with reference to continuous variation in load

b) It helps to control the frequency of vast interconnected systems.

c) To keep the absolute power interchange among pool members at the predetermined data.

The loop of Automatic Load Frequency Control maintains the operation only at slow and small differences i.e., the normal variations in load and frequency. Moreover,

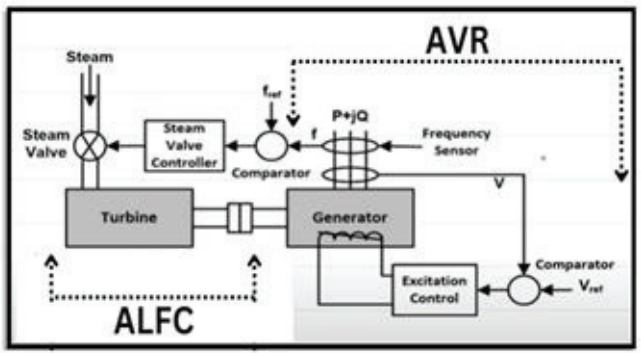


Fig. 1. Block diagram representation of load frequency and excitation control

this makes it unable to provide appropriate required balanced operation in emergency situations at the time of occurrence of large megawatts imbalance. Therefore, a more dynamic emergency control must be utilized.

V. Control Area Concept

The control area concept in power systems refers to a geographical or topological region where the electrical power generation, transmission, and consumption are coordinated and controlled as a single entity. A large power system can be divided into multiple areas of load-frequency control tied together using tie line. The issuefor the control of real power can be analysed by first interpreting the concept for a single area and then extending the solutionof a two or multi-area system.

Area Control Error - The integration control system comprises a frequency sensor and an integrator. This sensor senses the frequency error (Δf), and the resultant error signal is passed into the integrator as an input. The term used for the input to the integrator is "Area Control Error" (ACE).

VI. Area - Control Error (ACE)

A coherent area among which the generators oscillate together in reference to the variation in load or speed varying configurations in which the assumption of constant frequency is made in both static and dynamic environments can be considered as a Single Area. The isolated power system comprising a turbine, load, generator & speed governor can be a representation of single control area.

A. Area-Control-Error in single area case:

The input received by an integrator is known as ACE i.e. "Area Control Error".

The difference in area frequency is the ACE which pushes the error ofsteady state frequency to zero. Further, the production of the signal of real power command is done by an integrator which is represented as:

$$\Delta P_C = -K_I \int \Delta f\, dt = -K_I \int (ACE)\, dt$$

An APC signal is passed to the speed-changer and forces it to move where the integral gain constant is represented by KI operating the integration rate. The integral controller comprises of negative sign which reflects the production of a decrease command or negative command for attaining a positive frequency error. Similarly, the positive gain constant represented as K is responsible for controlling the integration rate and control loop's response speed. The integrator used is

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:14:48 UTC from IEEE Xplore. Restrictions apply.

the same type of the electronic integrator which is used in analog computers.

B. *Area Control Error in two area case:*

The Area Control Error, ACE maintains the linear combination of the difference in frequency and difference in line of power for two-area category. However, to achieve the zero power of steady state tie-line (P_{TL}=0), the other loop of integral-control needs to be introduced for each area with loop of integral frequency to further integrate the incremental power signal of tie-line power and passed back to speed changer.

The ACE of Area1 and 2 are:

$$ACE1\ (S) = \Delta PTL1\ (S) + b1\Delta F1\ (S)$$

$$ACE2\ (S) = \Delta PTL2\ (S) + b2\Delta F2\ (S)$$

VII. Problem Statement

A tie line connected with a two-area system at 1000 MVA common base has following parameters:

TABLE I. Values of Parameters, Constants & Coefficients

Area	1	2
Constant of Inertia	$H_1 = 5$	$H_2 = 4$
Regulation Speed	$R_1 = 5 \times 10^{-2}$	$R_2 = 6.25 \times 10^{-2}$
Time Constant of Governor	$T_{g1} = 2 \times 10^{-1}$sec	$T_{g2} = 3 \times 10^{-1}$sec
load coefficient of sensitive frequency	$D_1 = 6 \times 10^{-1}$	$D_2 = 9 \times 10^{-1}$
Base Power	10^3MVA	10^3MVA
Time Constant of Turbine	$T_1 = 5 \times 10^{-1}$sec	$T_2 = 6 \times 10^{-1}$sec

At nominal frequency of 60 Hz, the units operate in parallel. The coefficient of synchronizing power, Ps = 2.0 per unit which is derived from initial conditions of operation. A 187.5 MW of load variation occurs in Area 1.

(a) Determine the difference in tie line flow and the new frequency of steady state.

(b) Plot the block diagram of Simulink and determine the response of frequency deviation for the above condition in part (a)

VIII. Simulations

A. *Two Area part 1 Simulation*

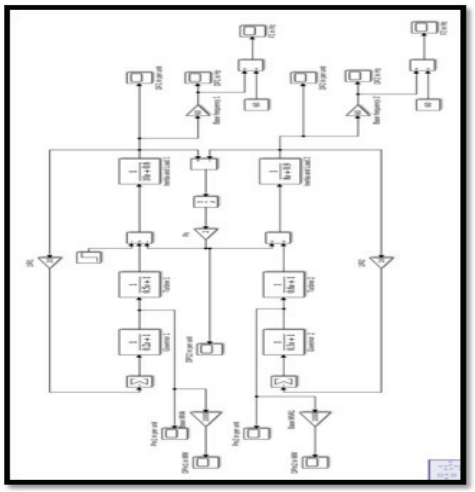


Fig. 2. MATLAB Model of Two Area Part 1

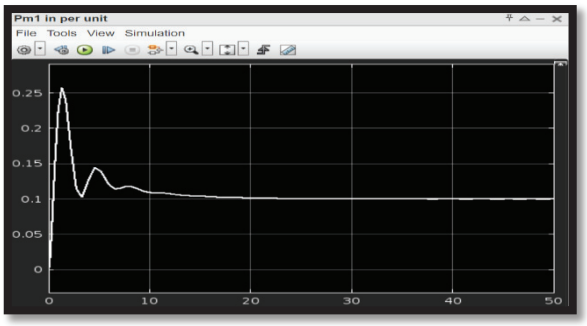


Fig. 3. (x,y): (t (s), Pm1 (pu))

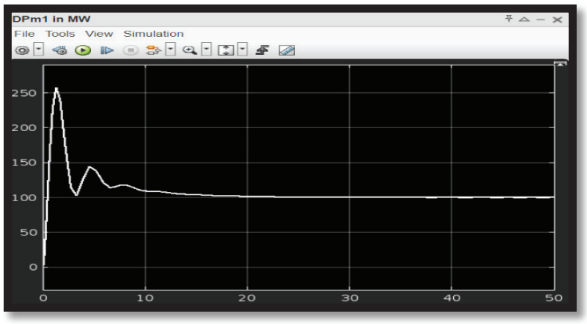


Fig. 4. (x, y): (t (s), DPm1 (MW))

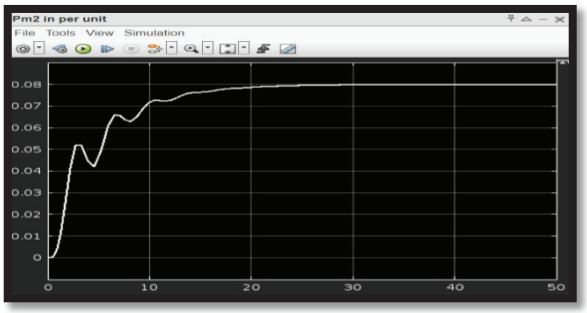


Fig. 5. (x, y): (t(s), Pm2 (pu))

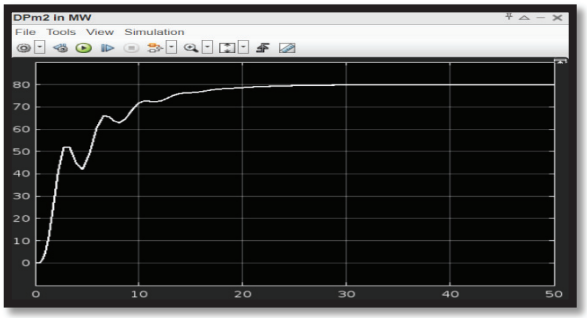


Fig. 6. (x, y): (t(s), DPm2 (MW))

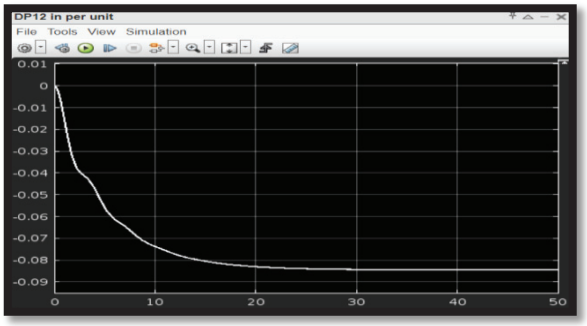


Fig. 7. (x, y): (t(s), DP12 (pu))

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:14:48 UTC from IEEE Xplore. Restrictions apply.

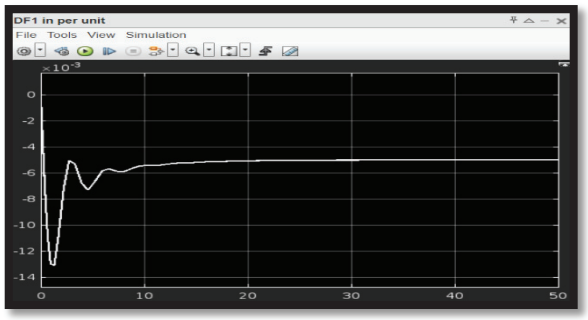


Fig. 8. (x, y): (t(s), DF1 (pu))

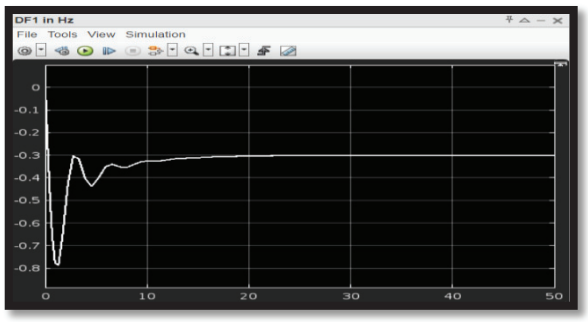


Fig. 9. (x, y): (t(s), DF1 (Hz))

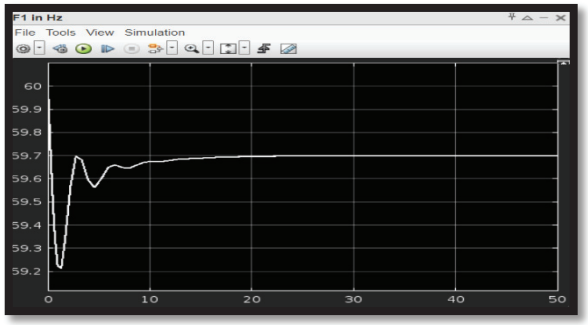


Fig. 10. (x, y): (t(s), F1 (Hz))

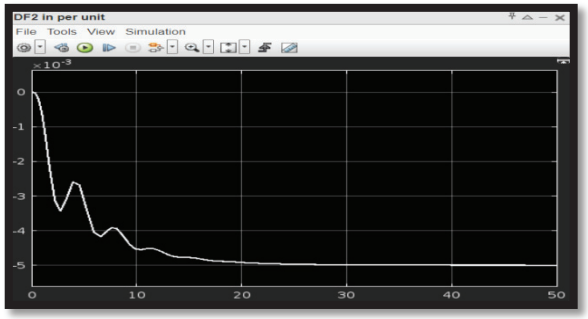


Fig. 11. (x, y): (t(s), DF2 (pu))

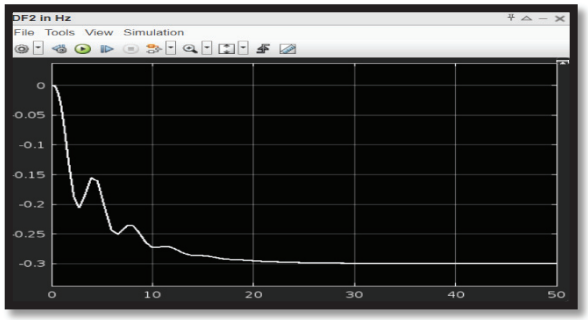


Fig. 12. (x, y): (t(s), DF2 (Hz))

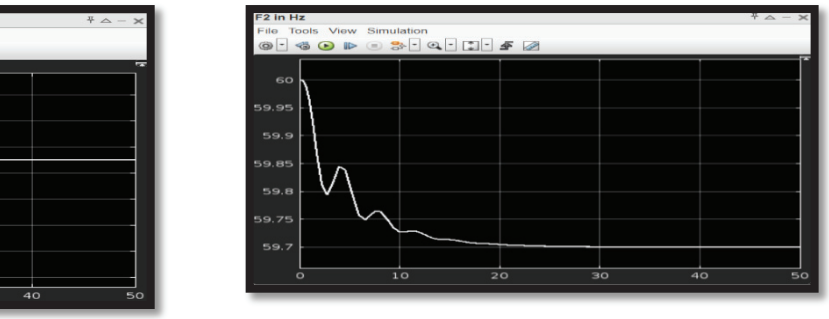


Fig. 13. (x, y): (t(s), F2 (Hz))

B. *Two Area part2 Simulation*

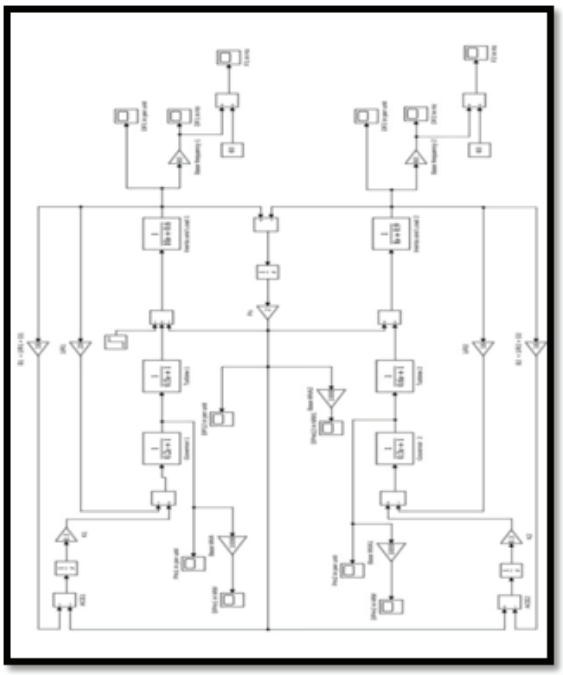


Fig. 14. MATLAB Model of Two Area Part 2

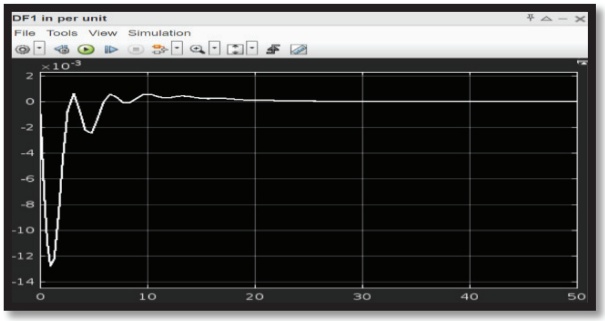


Fig. 15. (x, y): (t(s), DF1(pu))

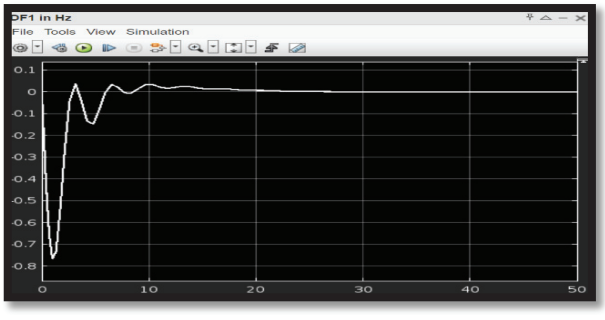


Fig. 16. (x, y): (t(s), DF1(Hz))

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:14:48 UTC from IEEE Xplore. Restrictions apply.

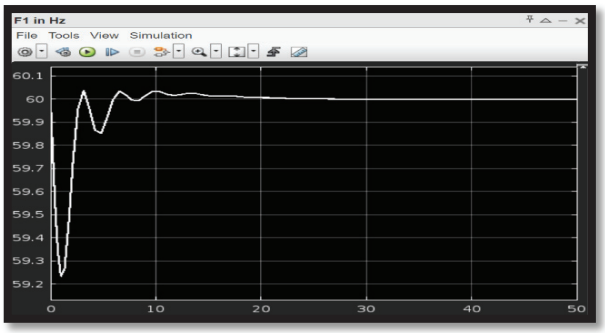


Fig. 17. (x, y): (t(s), F1(Hz))

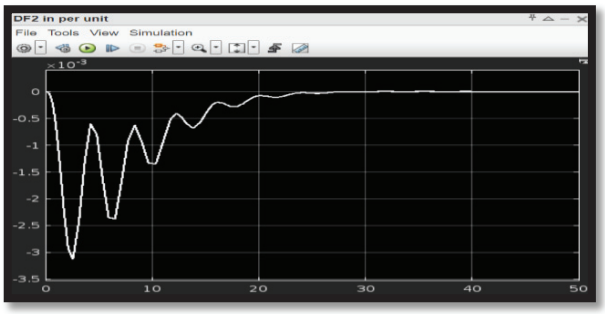


Fig. 18. (x, y): (t(s), DF2(pu))

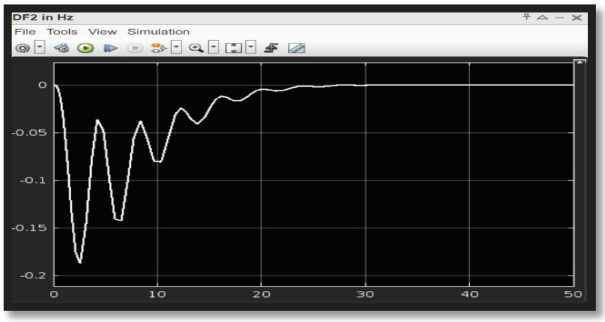


Fig. 19. (x, y): (t(s), DF2(Hz))

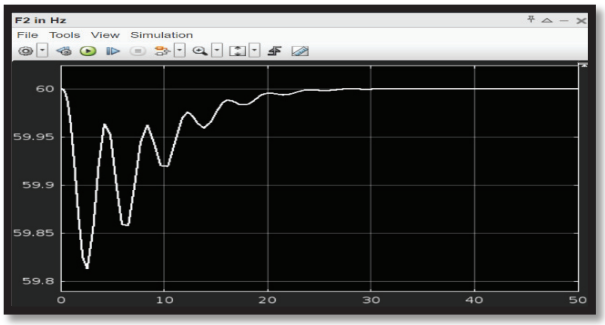


Fig. 20. (x, y): (t(s), F2(Hz))

In above shown simulations,

- Pm is Mechanical Power
- DPm is change in Mechanical Power
- F is Frequency
- DF is change in Frequency
- pu is per unit value
- t(s) is time in second

IX. ANALYSIS

Considering Steady State Frequency Deviation &difference in Tie Line Power

For Area 1 & Area 2 the speed regulation is calculated as:

$$\frac{1}{R_1} = \frac{1}{5 \times 10^{-2}} = 2 \times 10^{1}$$

$$\frac{1}{R_2} = \frac{1}{6.25 \times 10^{-2}} = 1.6 \times 10^{1}$$

For Area 1 and Area 2 the inertia and load iscalculated as:

$$\frac{1}{D_1 + 2H_1 s} = \frac{1}{0.6 + (2 \times 5s)} = \frac{1}{0.6 + 10s}$$

$$\frac{1}{D_2 + 2H_2 s} = \frac{1}{0.9 + (2 \times 4s)} = \frac{1}{0.9 + 8s}$$

For Area 1 the load change per unit is calculated as:

$$\Delta P_{\mathrm{LI}} = \frac{187.5}{10^3} = 18.75 \times 10^{-2}\ \text{p. u.}$$

The deviation in per unit frequency of steady state is calculated as:

$$\Delta\omega = \frac{-\Delta P_{ul}}{\left[\left(D_1 + \frac{1}{R_1}\right) + \left(D_2 + \frac{1}{R_2}\right)\right]}$$

$$= \frac{-18.75 \times 10^{-2}}{\left(\frac{1}{5 \times 10^{-2}} + 6 \times 10^{-1}\right) + \left(\frac{1}{6.25 \times 10^{-2}} + 9 \times 10^{-1}\right)}$$

$$\Delta\omega = \frac{-18.75 \times 10^{-2}}{(20 + 6 \times 10^{-1}) + (16 + 9 \times 10^{-1})}$$

$$= -5 \times 10^{-3}\text{pa}$$

Deviation in steady state frequency in Hertz is:

$$\Delta f_{\mathrm{Actual}} = \Delta f_{p.u.} \times f_{\mathrm{Base}}$$

$$\Delta f_{\mathrm{Actual}} = -5 \times 10^{-3} \times 60 = -3 \times 10^{-1}\text{Hz}$$

The steady state frequency in Hz is given by

$$f_{\mathrm{Actual}} = f_{\mathrm{original}} + \Delta f_{\mathrm{Actual}}$$

$$f_{\mathrm{Actinal}} = 60 - 3 \times 10^{-1} = 59.7\text{Hertz}$$

For each phase the mechanical power change is calculated as:

$$\Delta P_{\mathrm{m1}} = \frac{-\Delta\omega}{R_1} = \frac{-x - 5 \times 10^{-3}}{5 \times 10^{-2}} = 10^{-1}\text{pA}$$

$$\Delta P_{\mathrm{m1}} = 10^{-1} \times 10^{3} = 10^{2}\text{MW}$$

$$\Delta P_{\mathrm{m2}} = \frac{-\Delta\omega}{R_2} = \frac{-x - 5 \times 10^{-3}}{6.25 \times 10^{-2}} = 8 \times 10^{-2} p\Delta$$

$$\Delta P_{m2} = 8 \times 10^{-2} \times 10^{3} = 80\text{MW}$$

Thus, at new operating frequency of 59.7 Hz, 100MW of generation is increased by Area 1 and 80 MW by Area 2.

The total load difference in generation is observed as 180 MW, but load change is given as 187.5MW therefore, total load difference is 7.5 MW less because of the variation in area loads due to frequency drop.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:14:48 UTC from IEEE Xplore. Restrictions apply.

For Area 1, the load change is given as:

$$\Delta\omega \times D_1 = -5 \times 10^{-3} \times 6 \times 10^{-1} = -3 \times 10^{-3} pu$$

For Area 2, the load change is given as:

$$\Delta\omega \times D_2 = -5 \times 10^{-3} \times 9 \times 10^{-1} = -4.5 \times 10^{-3} pu$$

$$\Delta\omega \times D_1 = -4.5 \times 10^{-3} \times 10^{3} = -4.5\text{MW}$$

Thus, -7.5 MW of difference in total area load is observed.

The power flow in tie line is

$$\Delta P_{12} = \Delta\omega \times \left[\frac{1}{R_2} + D_2\right]$$

$$\Delta P_{12} = -5 \times 10^{-3} \times [16 + 9 \times 10^{-1}]$$

$$\Delta P_{12} = -8.45 \times 10^{-2} \times 10^{3} = -84.5\text{MW}$$

From above calculations, 84.5 MW of power flows from Area 2 to Area 1.

And 80 MW of power is received from the increase in generation of Area 2 whereas remaining 4.5 MW of power is received from the reduction in load of area 2 due to drop in frequency.

X. Conclusion

When the load in Area 1 of the power system is modified by 187.5 MW, a common assumption is that both generators will increase their output to meet the elevated demand. However, this assumption does not align with practical scenarios. In reality, when there is a sudden change in load in any area, each area is expected to handle its own variations. In other words, the load changes should be managed by the generators within that specific area. Specifically, for a load change of 187.5 MW, the increase in mechanical power in Area 1 should precisely match this value, while the variation in mechanical power for Area 2 must remain at zero. Additionally, the variation in tie line power should also be maintained at zero. To achieve this, adjustments were made to the model to ensure that alterations in the load of one area do not impact the generation of another area. It is worth noting that despite a 187.5 MW change in load, the system frequency only recovered to 59.7 Hz, not the nominal value of 60 Hz. As a result, only 180 MW of the load has been supplied by both generators, leaving a deficit of 7.5 MW. The full 187.5 MW will be supplied once the frequency recovers to 60 Hz. Therefore, there is a need to implement control techniques to bring the frequency back to its nominal value. Regarding the flow of powerfrom the tie line, although the change in mechanical power of generator 2 is 80 MW, the simulation results indicate a tie line power flow of 84.5 MW. This apparent discrepancy is explained by 80 MW being supplied from Area 2, while an additional 4.5 MW is redirected to Area 1 due to a reduction in load caused by the frequency drop in Area 2. In summary, the tie line carries a total power flow of 84.5 MW.

XI. Future Scope

- **In-depth Analysis and Simulation of AVR**: Undertake a comprehensive study of Automatic Voltage Regulator (AVR) systems, delving into the intricacies of their operation and control mechanisms.
- **Comparative Analysis of LFC Systems**: Conduct a comparative analysis of the single area and multi-area LFC systems. This will involve evaluating their performance, stability, and response characteristics under various operating conditions.
- **Single Area AVR System Simulation**: Develop simulations for a single area AVR system, this time integrating a Proportional-Integral-Derivative (PID) controller model to enhance system control.
- **Multi Area AVR System Simulation**: Similar to the LFC systems, expand the study to include multi-area AVR systems, with a PID controller model serving as the control mechanism.
- **Comparative Analysis of AVR Systems**: Carry out a comparative study of the single area & multi-area AVR systems. Assess their performance metrics, stability margins, and response times, among other parameters

References

[1] Ibraheem, P. Kumar, D.P. Kothari,"Recent philosophies of automaticgeneration control strategies inpower systems," IEEE Trans.Power Syst., vol. 20, no. 1, pp.346-357, 2005, doi:10.1109/TPWRS.2004.840438

[2] O. I. Elgerd, and C. E. Fosha, "Optimum megawatt-frequency control of multiarea electric energy systems," IEEE Trans. Power Apparat. Syst., vol. PAS-89, no. 4, pp. 556-563, 1970, doi: 10.1109/TPAS.1970.292602.

[3] Kothari DP, Nagrath IJ (2012) Modern power system analysis, 4th ed. McGraw Hill, New Delhi

[4] K. Prabha, "[Prabha Kundur"] Power System Stability and control.Pdf." p.1176,1994.

[5] "Hadi Saadat – Power Systems Analysis. 3rd Edition-PSA (2011).pdf."

[6] F. P. deMello, R. J. Mills and W. F. B'Rells, "Automatic Generation Control Part II-Digital Control Techniques," in IEEE Transactions on Power Apparatus and Systems, vol. PAS-92, no. 2, pp. 716-724, March 1973, doi: 10.1109/TPAS.1973.293777.keywords: {Automatic generation control;Error correction;Logic;Proportional control;Digital control;Control systems;Pulse measurements;Optimal control;Tuning;Frequency}

[7] Sahu RK, Panda S, Padhan S. Optimal gravitational search algorithm for automatic generation control of interconnected power systems. Ain shams engineering journal. 2014 Sep 1;5(3):721- 733.

[8] R. K. Sahu, S. Panda, and S. Padhan, "A hybrid firefly algorithm and pattern search technique for automatic generation control of multi area power systems," Int. J. Elect. Power Energy Syst., vol. 64, pp. 9–23, Jan. 2015.

[9] Abd-Elazim SM, Ali ES. Load frequency controller design of a two-area system composing of PV grid and thermal generator via firefly algorithm. Neural Computing and Applications. 2018 Jul 1;30(2):607-616.

[10] Khooban MH, Niknam T. A new intelligent online fuzzy tuning approach for multi-area load frequency control: Self Adaptive Modified Bat Algorithm. International Journal of Electrical Power & Energy Systems. 2015 Oct 1;71:254-261.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:14:48 UTC from IEEE Xplore. Restrictions apply.

Check for updates

Received: 24 October 2023 | Revised: 2 February 2024 | Accepted: 15 April 2024

DOI: 10.1002/acs.3823

RESEARCH ARTICLE

WILEY

Artificial neural network-based adaptive control for nonlinear dynamical systems

Kartik Saini[1] | **Narendra Kumar**[1,2] | **Bharat Bhushan**[1] | **Rajesh Kumar**[3]

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

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

[3]Department of Electrical Engineering, National Institute of Technology Kurukshetra, Kurukshetra, India

Correspondence
Kartik Saini, Department of Electrical Engineering, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi 110042, India.
Email: kartiksaini17@gmail.com

Summary

This research article presents an artificial neural network (ANN)-based indirect adaptive control method for nonlinear dynamical systems. In this article, a modified Elman recurrent neural network (MERNN) is proposed as an identifier and controller for controlling nonlinear systems. The architecture of the proposed controller is a modified form of the existing Elman recurrent neural network. The parameter training of ANN-based controllers is obtained by using the most popular optimization algorithm which is known as the back-propagation algorithm. A comparative study includes Elman, Diagonal, Jordan, feed-forward neural network (FFNN), and radial basis function network (RBFN)-based controllers to compare with the proposed MERNN controller. To determine the controller's robustness, parameter variations, and disturbance signals have been considered. The performance analysis of the proposed controller is illustrated by two simulation examples. The simulation results reveal that MERNN can not only identify the unknown dynamics of the plant but also adaptively control it compared to the others.

KEYWORDS

back-propagation, dynamic systems, indirect adaptive control, recurrent neural network

1 | INTRODUCTION

In recent years, neural networks have gained attention as an efficient global approximator. They are used to identify and control dynamical systems. Neural network technology is gaining popularity in data classification, adaptation, modeling, identification, and control.[1] The system identification procedure includes different methods. These include ANNs, fuzzy systems, and more. When system identification is achieved through the use of neural networks. They need the input–output dataset. This set can be used to approximate the unknown dynamics of the given system. They have self-learning capabilities. However, fuzzy logic systems lack self-learning. Another source must modify them. Also, fuzzy-based systems have another disadvantage. They require more mathematics to compute output than neural network models. Fuzzy-based systems generally have poor convergence speeds. Fuzzy models need more tuning and simulation than neural network models. Therefore, artificial neural network (ANN) technology is especially beneficial in the field of system identification and adaptive control. Due to the significant nonlinearity present in many systems, suitable controlling for dynamical systems by employing conventional control techniques is a challenging task.[2] However, it has been discovered that the conventionally constructed controller may not be able to meet some demanding requirements.[3,4] It is highly challenging to use conventional control methods when the mathematical model of the plant is unknown or only partially known.[5,6] The nonlinear systems have been controlled using various highly developed control techniques

 © 2024 John Wiley & Sons Ltd.

such as sliding mode control, back-stepping control, and feedback linearization control. There is a possibility that these advanced techniques may not work as expected due to their complexity.[7] The development of controllers for uncertain dynamical systems has become very effective through an adaptive control technique. Adaptive control is a method in which the response of a controller changes automatically based on changes in conditions within the process.[8] The primary benefit of the adaptive control method is that it can deliver superior performance to conventional control techniques.[9,10] There are different techniques for implementing neural network-based adaptive controllers that may be divided into two main categories: direct adaptive control and indirect adaptive control method. The direct adaptive control technique is suitable for linear systems. It uses a neural network as the controller whereas the indirect control method uses controllers designed based on a neural network model.[11,12] The indirect adaptive control technique is suitable for nonlinear dynamic systems and consists of two components, namely system identification, and control.[13,14] An ANN is one of the most widely used intelligent techniques which is capable of approximating any unknown nonlinear function to an arbitrary degree of accuracy and offering precise control instead of getting previous knowledge of the dynamics of the nonlinear systems.[15] These neural network-based adaptive controllers are capable of controlling highly nonlinear plants. The selection of the neural network to be used as a controller and the choice of the parameter learning technique to be utilized for adjusting its parameters are the fundamental problems in the field of ANN-based control.[16,17] There are several methods which are present in the literature that can be used to update parameters. The most extensive parameter optimization algorithm used to update the parameters of the identifier as well as the controller is the back-propagation optimization algorithm.[18,19] In the present study, a modification of the Elman recurrent neural network (MERNN) is constructed as an enhanced form of the original ERNN to determine the unknown dynamics of the plant. An indirect adaptive control scheme based on a MERNN is proposed to control the dynamical systems. In the suggested control structure, we employ two MERNNs as an identifier and a controller. The MERNN identifier approximates dynamical systems. It provides information on adaptability to the MERNN controller. The remaining paper is organized as follows. There are different sections in this research paper. The nonlinear dynamical system used in this article is described in detail in Section 2. The description of the proposed neural network model is discussed in Section 3. In Section 4 the identification process of the nonlinear system is explained. Section 5 describes the adaptive control scheme for dynamical systems using an ANN. The parameter update equations and control algorithm for dynamical systems are given in section 6. Results and discussion from the simulation and controllers are presented in Section 7. The robustness of controllers is tested and compared using disturbance signal and parameter variation. The conclusion is given in Section 8 and references are provided at the end.

1.1 | Literature review

In Reference 20 self-recurrent wavelet neural networks (SRWNNs) have been employed by the authors as an indirect adaptive control technique for dynamic systems. The SRWNN weights are trained using gradient descent with an adaptive learning rate (ALR). To ensure the convergence of the suggested control system, the ALR which is derived from the discrete Lyapunov stability theorem, is used. SRWNNs, which have both properties like attractor dynamics of RNN and the fast convergence of WNN, are used to identify and control nonlinear systems. But, they did not test the robustness of the proposed scheme. In Reference 21 the authors have presented diagonal recurrent neural network (DRNN)-based adaptive control for nonlinear dynamical systems. The Lyapunov stability criteria are used to build updated rules to assure stability. The robustness of the suggested control scheme is examined with parameter changes and disturbance signals. In Reference 22 this research develops an adaptive optimum tracking control technique for discrete-time nonlinear systems based on infinite horizon neural networks. As a result, the iterative adaptive dynamic programming (ADP) algorithm provides an easier way to obtain the iterative tracking control law, and neural networks are utilized to approximate the performance index function and compute the best control strategy. This proposed ADP algorithm gives optimal control and better convergence. They did not reveal the overall stability of the proposed algorithm. In Reference 23 the authors suggested a hybrid approach using feedforward neural networks to control a nonlinear dynamic system. A novel neural control approach has been created and presented. The approach is based on merging the Kohonen algorithm with the gradient descent method, the benefit of this hybrid learning algorithm is that it gives faster convergence. However, they did not examine the robustness of their suggested approach. In Reference 24 the authors have designed a novel adaptive multiple neural network controller (AMNNC) with a supervisor controller for a class of uncertain nonlinear dynamic systems. The AMNNC is a specific kind of adaptive feedback linearizing controller that approximates nonlinearity factors using multiple neural networks. The proposed control scheme (AMNNC) has better performance in comparison to the conventional adaptive neural network controller. However, they did not compare the proposed controller with different

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

controllers to determine the effectiveness of the scheme. In Reference 25 the dynamic surface control approach is used in radial-basis-function neural networks (NNs)-based control frameworks to solve the complexity explosion issue. By using the command filter, the command signals and derivatives have been generated without the need to perform analytical calculations. The stability of ultimate boundedness is established using the Lyapunov approach. The suggested approach delivers smoother parameter modification, more accuracy, and higher efficiency. However, the proposed scheme has not tested the effects of parameter variations and disturbance signals. In Reference 26 the authors proposed a NN-based direct adaptive control technique for nonlinear systems with unknown nonsymmetric input. A disturbance observer and NNs provide direct adaptive controllers. The proposed controlling technique has fast convergence for the uncertain nonaffine nonlinear system. However, they have not used a comparative study of different controllers. In Reference 27 the authors proposed an indirect adaptive control technique based on backstepping control to solve the stabilization issue, where an adaptive neural network approach approximates the unknown nonlinearity. In Reference 28 a novel adaptive controller with state feedback has been developed within the adaptive fuzzy back-stepping control approach to ensure that all signals in the resulting closed loop are bounded, and the state variables can be controlled to the origin. To deal with the network-induced delay, the Pade approximation has been introduced, and fuzzy logic systems have been used to estimate the unknown nonlinear characteristics of the system. The suggested controller can overcome the difficulty resulting from the delay. However, they did not test the robustness and used no comparative study. In Reference 29 Multilayer Perceptron (MLP), Elman, NARXSP, and Radial Basis Function (RBF) networks are the four neural networks employed by the authors. A benchmark nonlinear system and a DC servo motor are identified and controlled by using these neural networks. The controller is trained using the Levenberg–Marquardt technique and MATLAB's neural network toolbox. The proposed MLP network takes less time and minimum numbers of epochs to give better convergence. However, they have not tested robustness with parameter variation and disturbance signal. In Reference 30 fuzzy logic systems are used to discover unknown nonlinear dynamics before establishing an equivalent and an auxiliary system. Using the auxiliary system and the backstepping recursive design approach, an adaptive fuzzy inverse optimal scheme coupled with a meaningful objective function is developed. However, they did not use a comparative study. In Reference 31 a continuous-time recurrent neural network (CTRNN) is constructed to be utilized in the context of nonlinear model predictive control (NMPC) and using automated differentiation (AD) methods, an effective training procedure for the suggested network is created. The proposed model can reduce network training time and have better accuracy. However, they have not shown the robustness of the model. In Reference 32 a neuroadaptive stabilization scheme is proposed for the adaptive stabilization of continuous and discrete-time nonlinear systems. The neuroadaptive control is based on the universal function approximation property. The proposed technique guarantees the asymptotic stability of the closed-loop system by using the Lyapunov method. But, they haven't discussed the comparative results with different controllers.

1.2 | Motivation and contribution

The main motivation for this research is the development of an indirect adaptive controller based on a NN to control complex nonlinear dynamical systems. The suggested technique demonstrated robustness features, as it was able to deal with parameter change and disturbance signal impacts. This research study presents the following contribution:

1. A modified Elman recurrent neural network model is proposed for nonlinear system identification.
2. An indirect adaptive control scheme is used to design NN-based controllers.
3. An extensive evaluation of feed-forward, recurrent neural network-based identifiers, and controllers has been performed.
4. The back-propagation optimization method is used to optimize the parameters of the NN-based controllers.

2 | PROBLEM DESCRIPTION

Consider the following dynamical difference equation to explain the given plant.

$$y_p(k+1) = F[y_p(k), y_p(k-1), \ldots y_p(k-n+1)] + G[u_c(k), u_c(k-1), \ldots u_c(k-m+1)], \quad (1)$$

where $y_p(k+1)$ represents the plant's one step forward output value and $u_c(k)$ represents the plant's control input at kth time instant. The word n denotes the plant's order. Where F and G are accessible in terms of nonlinear or linear functions.

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

Let the following difference equation represent the dynamics of the stable reference model.

$$y_m(k+1) = P[y_m(k), y_m(k-1), ...y_m(k-a+1)] + Q[r(k), r(k-1), ...r(k-b+1)], \tag{2}$$

where $y_m(k+1)$ reveals the reference model's single step forward output signal and $r(k)$ indicates the external input. The terms a and b are both positive, although P and Q are both differentiable functions of linear or nonlinear types. The adaptive control aim is usually described as follows: the controller parameter update equations are designed to trigger the plant output to match the reference model output.

$$\lim_{k\to\infty} \left| y_m(k+1) - y_p(k+1) \right| \leq E_c(k), \tag{3}$$

where $E_c(k) \to 0$ and $E_c(k)$ represents the MSE (mean square error). For updating the parameters of ANN-based controllers, the error between reference model output and plant output will be used. It is also necessary for the control system to compensate for parameter variations and disturbance signals. Therefore, the proposed control scheme can compensate for uncertainties by adjusting the parameters of the controllers.

3 | OVERVIEW OF ERNN AND MERNN MODEL

In this article, Figure 1 shows the structure of the Elman recurrent neural network model. The proposed modified Elman recurrent neural network (MERNN) model is shown in Figure 2 and it is an enhanced version of the standard Elman recurrent neural network (ERNN) model. The ERNN model has four layers: the first is the input layer, the second is the hidden layer, the third is the context layer, and the fourth is the output layer. The context layer gets its input from the output of the hidden layer whereas the MERNN model consists of five layers: an input layer, a hidden layer, an output layer, and two context layers. The difference between the ERNN and MERNN models is that the MERNN model has an extra context layer (between the output layer and hidden layer) and an additional self-feedback loop with unit delayed values to the output node of the output layer. $X_j(k) = \{x_1(k), x_2(k)...x_m(k)\}$ represents the m dimensional external input vector. The weight vector between the input and hidden layer is denoted by $W^I(k)$. It consists of $w_1^I(k), w_2^I(k), ...w_j^I(k)$. The

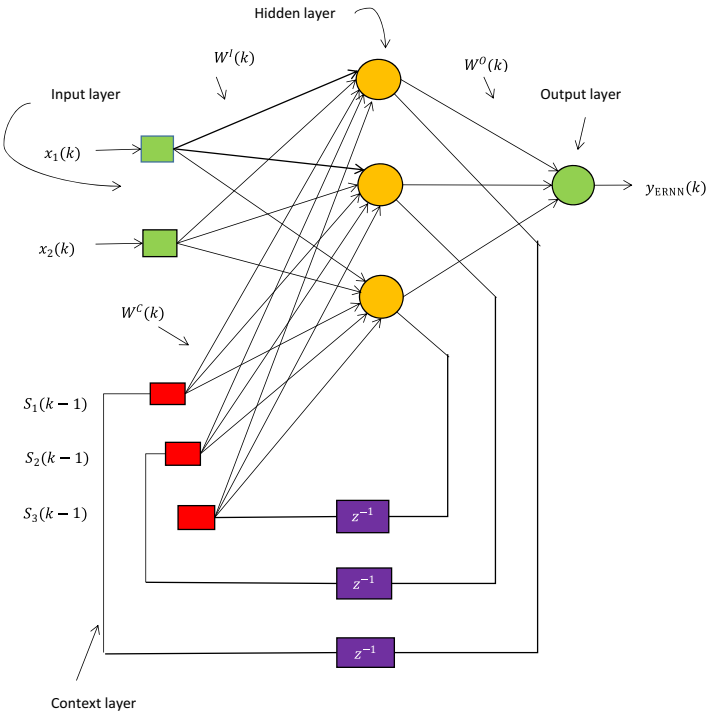


FIGURE 1 Elman recurrent neural network (ERNN) model.

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

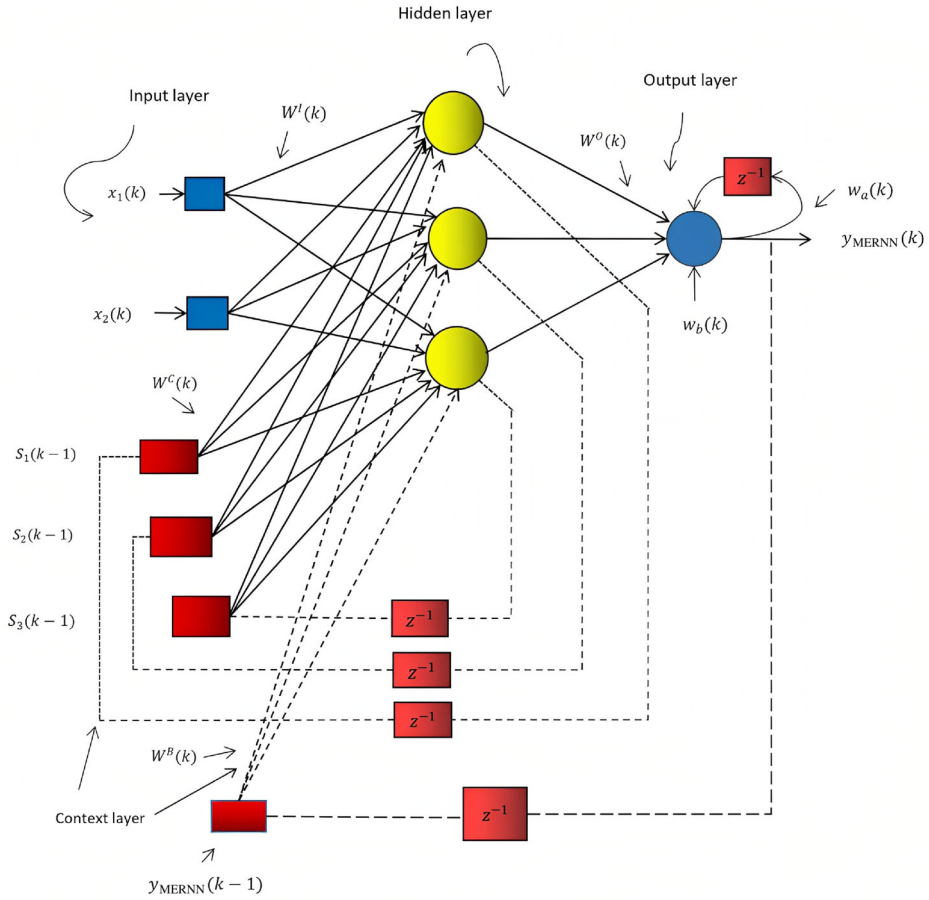


FIGURE 2 Modified Elman recurrent neural network (MERNN) model.

weight vector between the hidden layer and output layer is denoted as $W^O(k)$. This vector consists of elements $w_1^O(k)$, $w_2^O(k)$, and so on up to $w_q^O(k)$. The weight vector $W^C(k)$ contains three adjustable weights: $w_1^C(k)$, $w_2^C(k)$, and $w_3^C(k)$. It connects the first context layer to the hidden layer. In this context layer, the input is the output of the hidden layer with unit delay. In the second context layer, the input is the output of the MERNN model, which is delayed by one unit. This layer is connected to the hidden layer through the adjustable weight vector denoted as $W^B(k)$ which contains three adjustable weights: $w_1^B(k)$, $w_2^B(k)$, and $w_3^B(k)$. The MERNN model consists of an additional self-feedback loop with tunable weight $w_a(k)$ at the output node and $w_b(k)$ is the adjustable bias weight at the output node. The output of any *rth* recurrent neuron at any *kth* time instant is denoted by $S_r(k)$ and evaluated as

$$S_r(k) = V_r(k), \tag{4}$$

where $V_r(k)$ is the induced field produced by rth hidden neurons.

$$V_r(k) = \left\{ \sum_{j=1}^{m} x_j(k) w_{jr}^I(k) \right\} + \left\{ \sum_{i=1}^{h} S_i(k-1) w_{jr}^C(k) \right\} + \left\{ \sum_{j=1} w_{jr}^B(k) \times y_{\text{MERNN}}(k-1) \right\}, \tag{5}$$

where $w_{jr}^I(k)$ represents the adjustable weight between the input layer and hidden layer. w_{jr}^C and w_{jr}^B represent the adjustable weights connecting any jth input of the first context layer and second context layer to the rth hidden layer. $S_i(k-1)$ represents the previous output values of hidden neurons. $y_{\text{MERNN}}(k-1)$ represents the previous output value of the MERNN model. In this case, $T(.)$ represents the nonlinear tangent hyperbolic activation function that has both

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

positive as well as negative values, after nonlinear processing the output of the hidden layer is

$$T\{V_r(k)\} = \tanh\{V_r(k)\}, \tag{6}$$

$$T\{V_r(k)\} = \frac{e^{V_r(k)} - e^{-V_r(k)}}{e^{V_r(k)} + e^{-V_r(k)}}, \tag{7}$$

and the result of the output junction is

$$V_o(k) = \left\{ \sum_{q=1} w_{qr}^{O}(k) T\{V_r(k)\} \right\} + \{w_a(k) \times y_{\text{MERNN}}(k-1)\} - w_b(k), \tag{8}$$

where $w_{qr}^{O}(k)$ represents the adjustable weight parameter between rth hidden layer and qth output layer. The activation function of the output layer is linear, which is purelin $f(.)$. The outcome of the MERNN model with linear activation function is given as follows:

$$y_{\text{MERNN}}(k) = f\{V_o(k)\}. \tag{9}$$

4 | IDENTIFICATION OF NONLINEAR SYSTEM

System identification is a way of identifying or measuring the mathematical model of a system based on measurements of the system's inputs and outputs. Several controlling schemes are available if an accurate approximation mathematical model of the dynamical system is being controlled. In order to identify a system, we must first determine which model structure the given plant is related to and then approximate its order. Determining the model or category to which the presented plant belongs, as well as identifying the plant's order, is required for the identification. These nonlinear difference equations can be used to describe any of the four nonlinear dynamical plants introduced here:

$$\text{Model 1}: y_p(k+1) = F[y_p(k), y_p(k-1)....y_p(k-n+1), u_c(k), u_c(k-1)....u_c(k-m+1)]. \tag{10}$$

$$\text{Model 2}: y_p(k+1) = F[y_p(k), y_p(k-1)....y_p(k-n+1)] + G[u_c(k), u_c(k-1)....u_c(k-m+1)]. \tag{11}$$

$$\text{Model 3}: y_p(k+1) = \left[\sum_{i=0}^{n-1} a_i y_p(k-i) \right] + G[u_c(k), u_c(k-1)....u_c(k-m+1)]. \tag{12}$$

$$\text{Model 4}: y_p(k+1) = F[y_p(k), y_p(k-1)....y_p(k-n+1)] + \left[\sum_{i=0}^{m-1} b_i u_c(k-i) \right]. \tag{13}$$

All of the models described above illustrate a nonlinear system and the plant's output at $(k+1)$th is directly dependent on the present state of the plant's output and the prior state of the external inputs.

4.1 | Parallel and series-parallel identification model

Figure 3 shows the parallel identification scheme for the plant. Every dynamical plant utilized in this research is considered to be BIBO stable in the presence of a given class of input, assuming that all signals in the plants are equally BIBO stable. The series-parallel model involves feeding back the plant's output into the identification model, in contrast to the parallel model described above. The series-parallel identification scheme corresponding to a plant is shown in Figure 4.

5 | ADAPTIVE CONTROL FOR NONLINEAR DYNAMICAL SYSTEMS

Adaptive control is a technique that allows the controller's response to adjust automatically depending on changes in the process conditions. The main objective of control is to produce the expected results of the nonlinear plant, ensuring that

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

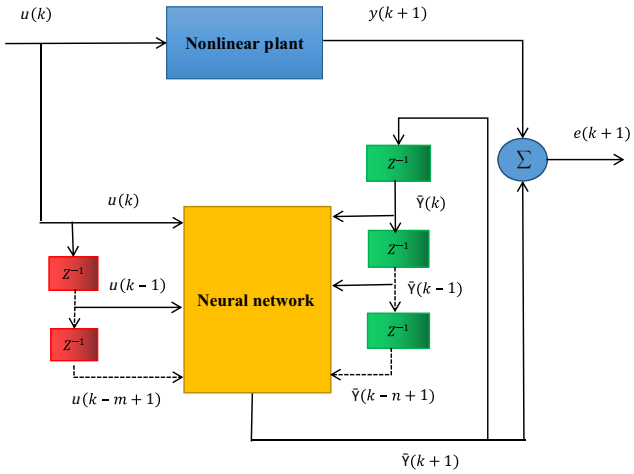


FIGURE 3 Parallel identification scheme.

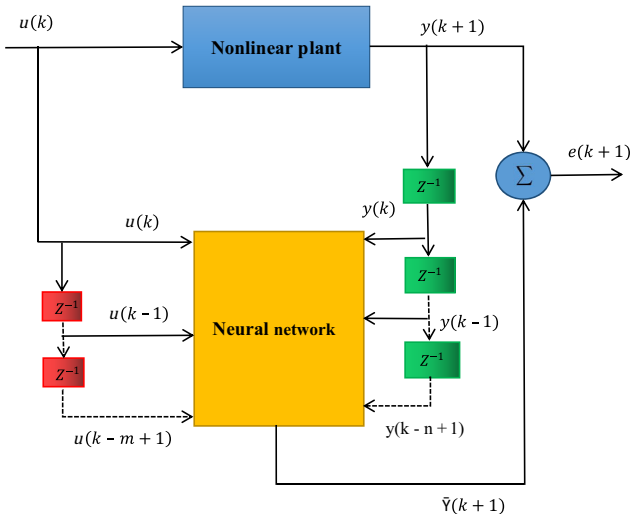


FIGURE 4 Series-parallel identification scheme.

it tracks the reference model output. There are various methods to implement adaptive control schemes, which fall into two categories.

1. Direct control
2. Indirect control

Direct control involves immediately adjusting the controller's settings at each time instant based on the difference between the plant's output and the desired reference model output. This method was limited to linear systems. Direct methods are not valid for nonlinear plants. Therefore, an indirect adaptive control approach is required to control nonlinear dynamical plants. The plant parameters are estimated, and the controller settings are computed in indirect adaptive control. In the case of indirect control, the parameter adjustment procedure demands knowledge of the mathematical model of the nonlinear plant with errors at each moment. In this article, the plant's dynamics have been considered to be unknown. Thus, in order to conduct control action in conjunction with the plant, an identification model will learn about the plant's dynamics. The indirect adaptive control technique is represented in Figure 5.

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

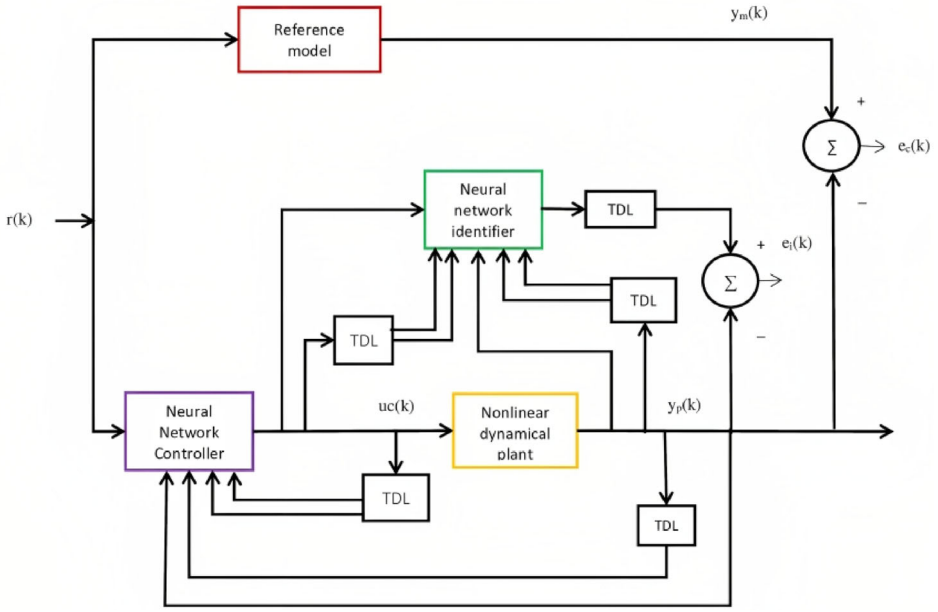


FIGURE 5 Indirect adaptive control scheme.

6 | PARAMETER OPTIMIZATION ALGORITHM

The backpropagation optimization approach is employed in this study to get the parameter update equations of the identifier and the controller.

$$E_c(k) = \frac{1}{2}e^2(k). \tag{14}$$

$$E_c(k) = \frac{1}{2}\left[y_p(k) - y_{\text{MERNN}}(k)\right]^2. \tag{15}$$

The weight update equation for each weight is obtained and described as follows:

$$w(k+1) = w(k) - \eta\frac{\partial E_c(k)}{\partial w(k)}. \tag{16}$$

Term η is known as the learning rate and lies between 0 and 1.

6.1 | Case 1: Weight update equations for identifier

The primary objective of the plant's training is to reduce identification error $e_i(k)$. The identifier's parameters between the input and hidden layer are defined by the input weight vector. The input weight vector is denoted as $W^I(k)$, consists of individual weights $w_1^I(k), w_2^I(k), \ldots w_j^I(k)$. The value of j represents the number of input nodes in the input layer. The input weight parameter update equation is described below by taking the partial derivative of an objective function $E_c(k)$ with respect to $w^I(k)$. At kth time instant, we have weight parameters: $w_1^I(k)$, $w_2^I(k)$, and $w_3^I(k)$. Each vector has three elements. $w_1^I(k)$ has elements $w_{11}^I(k)$, $w_{12}^I(k)$, and $w_{13}^I(k)$. $w_2^I(k)$ has elements $w_{21}^I(k)$, $w_{22}^I(k)$, and $w_{23}^I(k)$. $w_3^I(k)$ has elements $w_{31}^I(k)$, $w_{32}^I(k)$, and $w_{33}^I(k)$. We can write the weight update equations for all the elements of the input weight vector corresponding to input $x_j(k)$ by using Equation (16) as follows:

$$w_{11}^I(k+1) = w_{11}^I(k) - \eta\frac{\partial E_c(k)}{\partial w_{11}^I(k)}, \tag{17}$$

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

$$w_{12}^{I}(k+1) = w_{12}^{I}(k) - \eta \frac{\partial E_c(k)}{\partial w_{12}^{I}(k)}, \tag{18}$$

$$w_{13}^{I}(k+1) = w_{13}^{I}(k) - \eta \frac{\partial E_c(k)}{\partial w_{13}^{I}(k)}, \tag{19}$$

we can write the above equations as:

$$w_{jr}^{I}(k+1) = w_{jr}^{I}(k) - \eta \frac{\partial E_c(k)}{\partial w_{jr}^{I}(k)}. \tag{20}$$

The term $\frac{\partial E_c(k)}{\partial w_{jr}^{I}(k)}$ can be calculated as:

$$\frac{\partial E_c(k)}{\partial w_{jr}^{I}(k)} = \frac{\partial E_c(k)}{\partial y_{\mathrm{MERNN}}(k)} \times \frac{\partial y_{\mathrm{MERNN}}(k)}{\partial V_o(k)} \times \frac{\partial V_o(k)}{\partial S_r(k)} \times \frac{\partial S_r(k)}{\partial V_r(k)} \times \frac{\partial V_r(k)}{\partial w_{jr}^{I}(k)}, \tag{21}$$

$$where\ \frac{\partial E_c(k)}{\partial y_{\mathrm{MERNN}}(k)} = -e_i(k), \frac{\partial y_{\mathrm{MERNN}}(k)}{\partial V_o(k)} = 1, \frac{\partial V_o(k)}{\partial S_r(k)} = w_{qr}^{O}(k), \frac{\partial S_r(k)}{\partial V_r(k)} = \sum_{i=1}^{h}\left\{1 - S_i^2(k)\right\}\ and\ \frac{\partial V_r(k)}{\partial w_{jr}^{I}(k)} = x_j(k), \tag{22}$$

where $e_i(k)$ is the identification error. Now by using Equations (21) and (22)

$$\frac{\partial E_c(k)}{\partial w_{jr}^{I}(k)} = -e_i(k)\left\{w_{qr}^{O}(k)\right\}\left\{\sum_{i=1}^{h}\left\{1 - S_i^2(k)\right\}\right\}x_j(k), \tag{23}$$

$$w_{jr}^{I}(k+1) = w_{jr}^{I}(k) + \eta e_i(k)\left\{x_j(k)w_{qr}^{O}(k)\right\} \times \left\{\sum_{i=1}^{h}\left\{1 - S_i^2(k)\right\}\right\}, \tag{24}$$

similarly the weight update equations for weight parameter $w_{jr}^{C}(k)$ are obtained by using chain rule

$$\frac{\partial E_c(k)}{\partial w_{jr}^{C}(k)} = \frac{\partial E_c(k)}{\partial y_{\mathrm{MERNN}}(k)} \times \frac{\partial y_{\mathrm{MERNN}}(k)}{\partial V_o(k)} \times \frac{\partial V_o(k)}{\partial S_r(k)} \times \frac{\partial S_r(k)}{\partial V_r(k)} \times \frac{\partial V_r(k)}{\partial w_{jr}^{C}(k)}, \tag{25}$$

$$\frac{\partial E_c(k)}{\partial w_{jr}^{C}(k)} = -e_i(k)\left\{w_{qr}^{O}(k)\right\}\sum_{i=1}^{h}\left\{1 - S_i^2(k)\right\}S_i(k-1), \tag{26}$$

$$w_{jr}^{C}(k+1) = w_{jr}^{C}(k) + \eta e_i(k)w_{qr}^{O}(k)\left\{\sum_{i=1}^{h}\left\{1 - S_i^2(k)\right\}S_i(k-1)\right\}, \tag{27}$$

similarly the weight update equation for $w_{jr}^{B}(k)$ is:

$$\frac{\partial E_c(k)}{\partial w_{jr}^{B}(k)} = \frac{\partial E_c(k)}{\partial y_{\mathrm{MERNN}}(k)} \times \frac{\partial y_{\mathrm{MERNN}}(k)}{\partial V_o(k)} \times \frac{\partial V_o(k)}{\partial S_r(k)} \times \frac{\partial S_r(k)}{\partial V_r(k)} \times \frac{\partial V_r(k)}{\partial w_{jr}^{B}(k)}. \tag{28}$$

$$\frac{\partial E_c(k)}{\partial w_{jr}^{B}(k)} = -e_i(k)w_{qr}^{O}(k)\left\{\sum_{i=1}^{h}\left\{1 - S_i^2(k)\right\}\right\} \times y_{\mathrm{MERNN}}(k-1). \tag{29}$$

$$w_{jr}^{B}(k+1) = w_{jr}^{B}(k) + \eta e_i(k)w_{qr}^{O}(k)\left\{\sum_{i=1}^{h}\left\{1 - S_i^2(k)\right\}\right\} \times y_{\mathrm{MERNN}}(k-1). \tag{30}$$

The output parameter update equations are determined by using the chain rule

$$\frac{\partial E_c(k)}{\partial w_{qr}^{O}(k)} = \frac{\partial E_c(k)}{\partial y_{\mathrm{MERNN}}(k)} \times \frac{\partial y_{\mathrm{MERNN}}(k)}{\partial V_o(k)} \times \frac{\partial V_o(k)}{\partial w_{qr}^{O}(k)}, \tag{31}$$

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

$$\frac{\partial E_c(k)}{\partial w_{qr}^O(k)} = -e_i(k)\left\{\sum_{i=1}^{h} S_i(k)\right\}, \tag{32}$$

$$w_{qr}^O(k+1) = w_{qr}^O(k) + \eta e_i(k)\left\{\sum_{i=1}^{h} S_i(k)\right\}, \tag{33}$$

similarly, the weight update equation for $w_a(k)$ is given as:

$$\frac{\partial E_c(k)}{\partial w_a(k)} = \frac{\partial E_c(k)}{\partial y_{\text{MERNN}}(k)} \times \frac{\partial y_{\text{MERNN}}(k)}{\partial V_o(k)} \times \frac{\partial V_o(k)}{\partial w_a(k)}, \tag{34}$$

$$\frac{\partial E_c(k)}{\partial w_a(k)} = -e_i(k) y_{\text{MERNN}}(k), \tag{35}$$

$$w_a(k+1) = w_a(k) + \eta e_i(k)\{y_{\text{MERNN}}(k)\}, \tag{36}$$

similarly, the weight update equation for bias weight $w_b(k)$ can be written.

6.2 | Case 2: weight update equations for controller

Similarly, we can generalize the weight update equation for a controller. The parameter update equation for the input weight vector $W^I(k)$ can be calculated by using the partial derivative of an objective function $E_c(k)$ with respect to the weight vector $W^I(k) = \left\{w_1^I(k), w_2^I(k), \ldots w_j^I(k)\right\}$.

We can write weight update equations for the parameters $\left\{w_1^I(k), w_2^I(k), \ldots . w_j^I(k)\right\}$ corresponding to externally applied input $X_j(k) = \{x_1(k), x_2(k), \ldots x_m(k)\}$.

$$\frac{\partial E_c(k)}{\partial w_{jr}^I(k)} = \frac{\partial E_c(k)}{\partial y_p(k)} \times \frac{\partial y_p(k)}{\partial u_c(k)} \times \frac{\partial u_c(k)}{\partial S_r(k)} \times \frac{\partial S_r(k)}{\partial V_r(k)} \times \frac{\partial V_r(k)}{\partial w_{jr}^I(k)}. \tag{37}$$

The term $\frac{\partial y_p(k)}{\partial u_c(k)}$ is called sensitivity/jacobian of the plant and if the mathematical model of $y_p(k)$ is known, its value may be computed easily. If the plant's model is frequently unknown or just partially known then, we utilize neural network-based identifiers in parallel to the plant, and its configuration as well as the controller's parameters, are modified during online training. In that case as the training progresses, $\frac{\partial y_p(k)}{\partial u_c(k)} \approx \frac{\partial y_n(k)}{\partial u_c(k)}$, $y_n(k)$ is the identifier output. Then, the mathematical model of an identifier can approximate the dynamics of the nonlinear plant.

$$\text{where, } \frac{\partial E_c(k)}{\partial y_p(k)} = -e_c(k), \frac{\partial u_c(k)}{\partial S_r(k)} = w_{qr}^O(k), \frac{\partial S_r(k)}{\partial V_r(k)} = 1 - S_i^2(k), and \frac{\partial V_r(k)}{\partial w_{jr}^I(k)} = x_j(k), \tag{38}$$

where $e_c(k)$ is the control error. By using Equations (37) and (38) we can write

$$\frac{\partial E_c(k)}{\partial w_{jr}^I(k)} = -e_c(k)\frac{\partial y_p(k)}{\partial u_c(k)} w_{qr}^O(k) \times \left\{\sum_{i=1}^{h}\left\{1 - S_i^2(k)\right\}\right\} x_j(k), \tag{39}$$

by taking Equations (16) and (39)

$$w_{jr}^I(k+1) = w_{jr}^I(k) + \eta_c e_c(k)\frac{\partial y_p(k)}{\partial u_c(k)} w_{qr}^O(k)\left\{1 - S_i^2(k)\right\} x_j(k). \tag{40}$$

Similarly, the weight update equation for the output weight vector is determined as follows:

$$\frac{\partial E_c(k)}{\partial w_{qr}^O(k)} = \frac{\partial E_c(k)}{\partial y_p(k)} \times \frac{\partial y_p(k)}{\partial u_c(k)} \times \frac{\partial u_c(k)}{\partial w_{qr}^O(k)}. \tag{41}$$

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

$$\text{where,} \quad \frac{\partial E_c(k)}{\partial y_p(k)} = -e_c(k), \frac{\partial u_c(k)}{\partial w_{qr}^O(k)} = \left\{ \sum_{i=1}^{h} S_i(k) \right\}. \tag{42}$$

$$\frac{\partial E_c(k)}{\partial w_{qr}^O(k)} = -e_c(k) \frac{\partial y_p(k)}{\partial u_c(k)} \left\{ \sum_{i=1}^{h} S_i(k) \right\}. \tag{43}$$

Then,

$$w_{qr}^O(k+1) = w_{qr}^O(k) + \eta_c e_c(k) \frac{\partial y_p(k)}{\partial u_c(k)} \left\{ \sum_{i=1}^{h} S_i(k) \right\}. \tag{44}$$

7 | SIMULATION RESULTS

Two simulation examples are used to compare the performance of MERNN, ERNN, DRNN, JRNN, and feed-forward neural network (FFNN)-based controllers. Each controller contains only one hidden layer with three neurons. The initial learning rate in every example is assumed to be 0.00022.

7.1 | Example 1

Consider a nonlinear dynamical plant with dynamics defined by following the difference equation as given in Reference 17.

$$y_p(k+1) = \frac{y_p(k) y_p(k-1)[y_p(k)+2.5]}{1+y_p^2(k)+y_p^2(k-1)} + u_c(k), \tag{45}$$

where

$$F[y_p(k), y_p(k-1)] = \frac{y_p(k) y_p(k-1)[y_p(k)+2.5]}{1+y_p^2(k)+y_p^2(k-1)}, \tag{46}$$

$$r(k) = \sin\left(\frac{\pi k}{45}\right), \tag{47}$$

where $r(k)$ is the external input. The following difference equation describes the reference model.

$$y_m(k+1) = 0.6 y_m(k) + 0.2 y_m(k-1) + r(k). \tag{48}$$

The controller output error is as follows:

$$e_c(k+1) = y_m(k+1) - y_p(k+1). \tag{49}$$

The main aim of control is to determine the bounded input $u_c(k)$ by making $e_c(k+1) = 0$. Therefore,

$$y_m(k+1) = y_p(k+1). \tag{50}$$

$$y_m(k+1) = F[y_p(k), y_p(k-1)] + u_c(k). \tag{51}$$

$$u_c(k) = -F[y_p(k), y_p(k-1)] + 0.6 y_p(k) + 0.2 y_p(k-1) + r(k). \tag{52}$$

Using a neural network N, F can first be identified offline if it is unknown. Thus,

$$u_c(k) = -N[y_p(k), y_p(k-1)] + 0.6 y_p(k) + 0.2 y_p(k-1) + r(k). \tag{53}$$

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

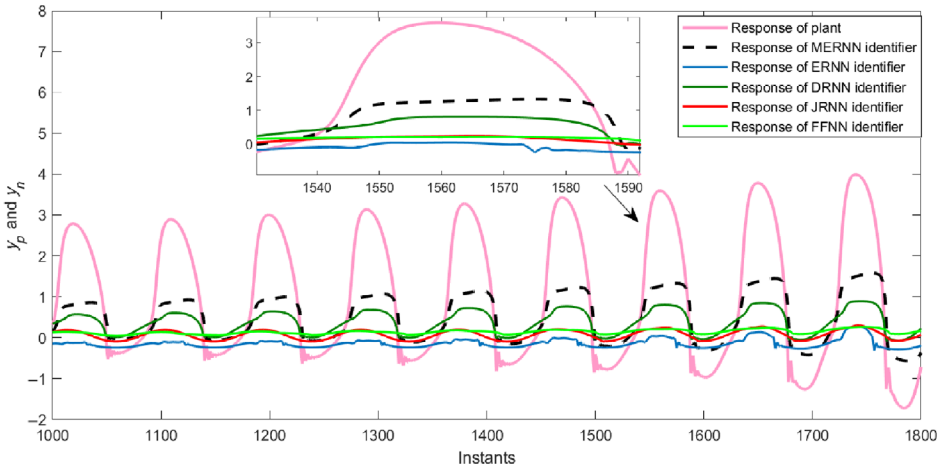


FIGURE 6 Identifier and plant output at the beginning of training [Example 1]

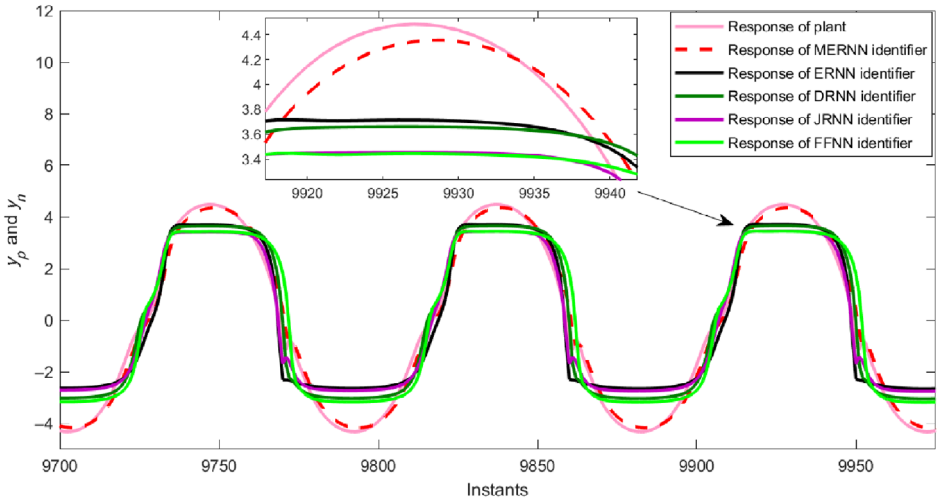


FIGURE 7 Identifier and plant output at the end of training [Example 1]

The nonlinear difference equation is

$$y_p(k+1) = F[y_p(k), y_p(k-1)] - N[[y_p(k), y_p(k-1)] + 0.6y_p(k) + 0.2y_p(k-1) + r(k). \tag{54}$$

In the given example the order of the plant is $n = 2$, which means there are $2n$ inputs to the controller. 0.00022 is the initial value of fixed learning rate η. The identifiers and plants' output at the beginning of training are shown in Figure 6. Their responses during the end phase of training are shown in Figure 7. Figure 8 shows both the controllers' response and the reference model's response at the beginning of training. In contrast, Figure 9 shows the reference model and controllers' responses at the final stage of training.

In example 1, Figures 10–14 show how parameter variation affects the output of MERNN, ERNN, DRNN, JRNN, and FFNN-based controllers. Table 1 shows the mean square error (MSE) comparison of identifiers and controllers. This comparison includes online training for the first nonlinear plant. The proposed MERNN model performs better than the ERNN, DRNN, JRNN, FFNN, and radial basis function network (RBFN) models. Table 2 compares the MSE of identifiers and controllers. It shows their MSE with online training for the second nonlinear plant. The proposed MERNN model performs better than the ERNN, DRNN, JRNN, FFNN, and RBFN models.

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

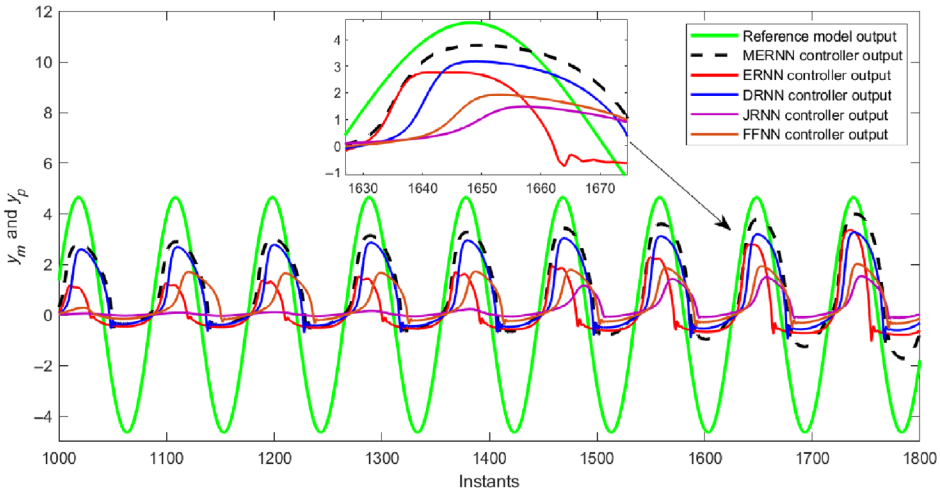


FIGURE 8 Controller and reference model output at the beginning of training [Example 1]

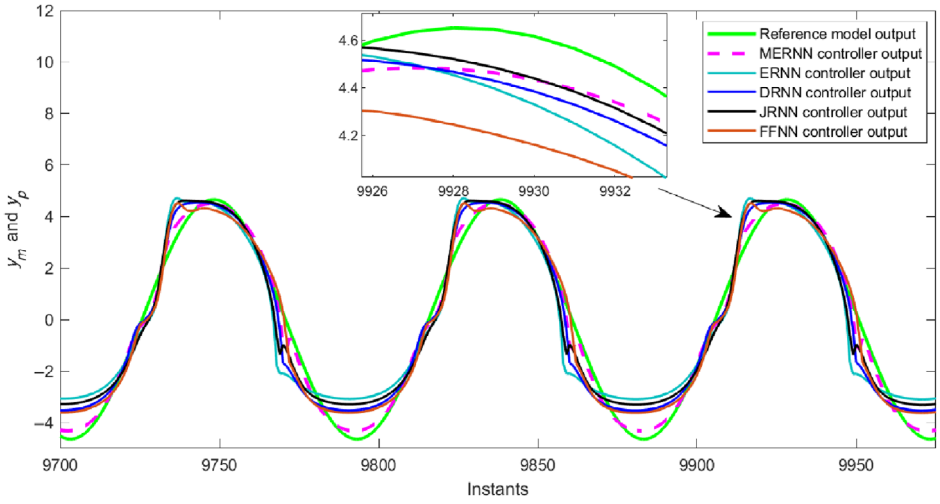


FIGURE 9 Controller and reference model output at the end of training [Example 1].

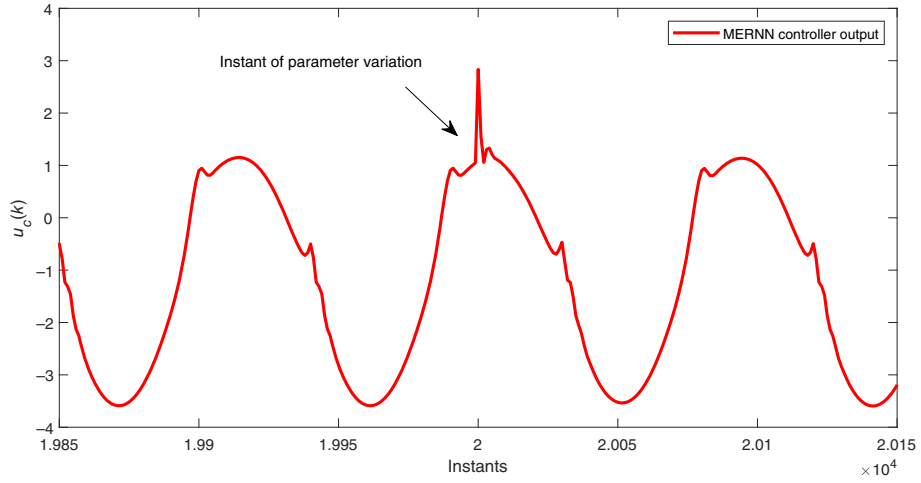


FIGURE 10 MERNN controller output with parameter variation [Example 1].

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

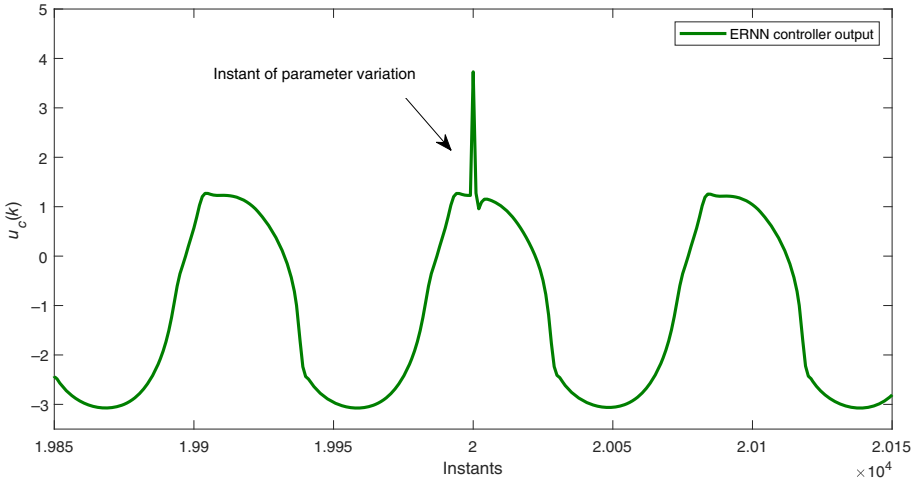


FIGURE 11 Elman recurrent neural network controller output with parameter variation [Example 1].

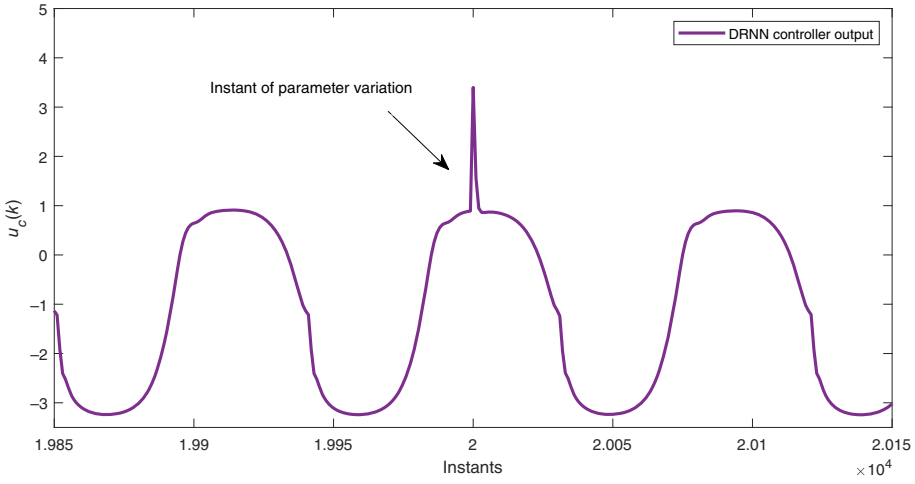


FIGURE 12 Diagonal recurrent neural network controller output with parameter variation [Example 1].

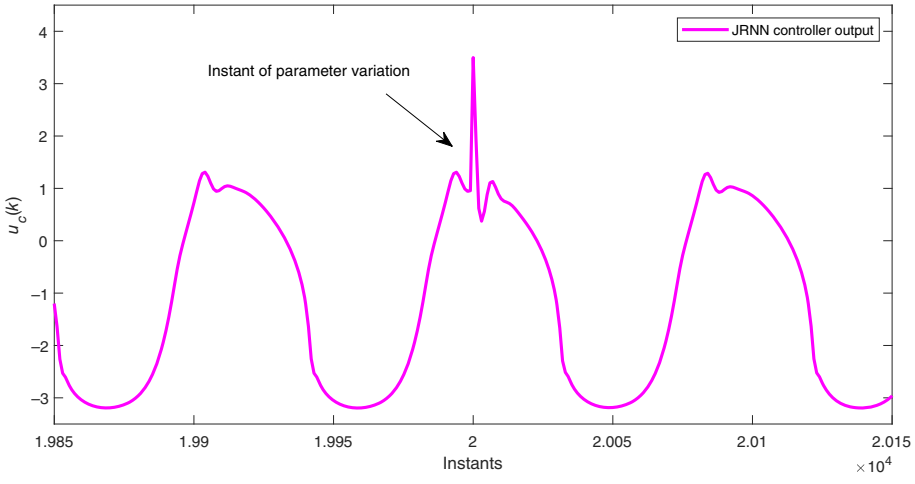


FIGURE 13 Jordan recurrent neural network controller output with parameter variation [Example 1].

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

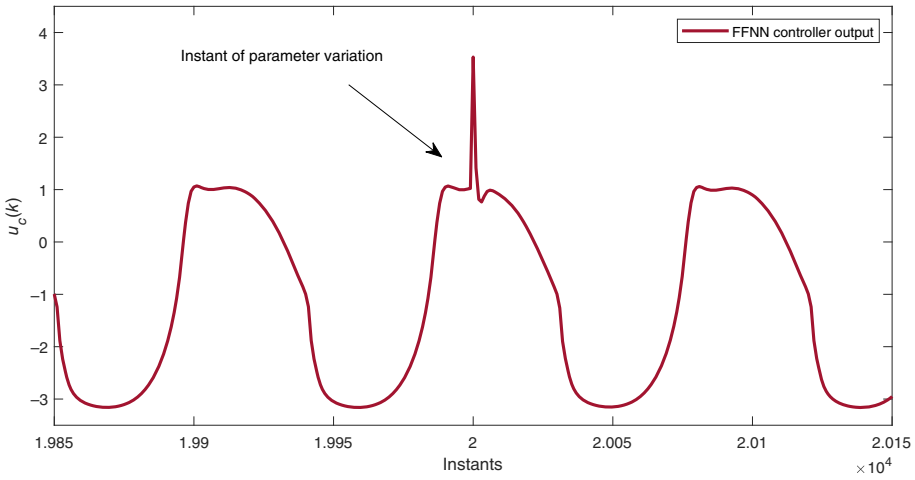


FIGURE 14 FFNN controller output with parameter variation [Example 1].

TABLE 1 Mean squared error (MSE) comparison of identifiers and controllers during online training [Example 1].

Sr. no.	Identifier			Controller		
	Model	MSE	RMSE	Model	MSE	RMSE
1.	**MERNN**	**0.00059**	**0.024**	**MERNN**	**0.0064**	**0.080**
2.	ERNN	0.0071	0.084	ERNN	0.0173	0.131
3.	DRNN	0.0136	0.116	DRNN	0.0236	0.153
4.	JRNN	0.0046	0.067	JRNN	0.0211	0.145
5.	FFNN	0.0175	0.132	FFNN	0.0150	0.122
6.	RBFN	0.0137	0.110	RBFN	0.0182	0.107

Abbreviations: DRNN, diagonal recurrent neural network; ERNN, Elman recurrent neural network; JRNN, Jordan recurrent neural network; RMSE, root mean square error.

TABLE 2 Mean square error (MSE) comparison of identifiers and controllers during online training [Example 2].

Sr. no.	Identifier			Controller		
	Model	MSE	RMSE	Model	MSE	RMSE
1.	**MERNN**	**0.00049**	**0.022**	**MERNN**	**0.00010**	**0.010**
2.	ERNN	0.00092	0.030	ERNN	0.0029	0.053
3.	DRNN	0.01160	0.107	DRNN	0.0020	0.044
4.	JRNN	0.03100	0.176	JRNN	0.02670	0.163
5.	FFNN	0.02680	0.163	FFNN	0.02520	0.158
6.	RBFN	0.0165	0.138	RBFN	0.0210	0.129

Abbreviations: DRNN, diagonal recurrent neural network; ERNN, Elman recurrent neural network; JRNN, Jordan recurrent neural network.

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

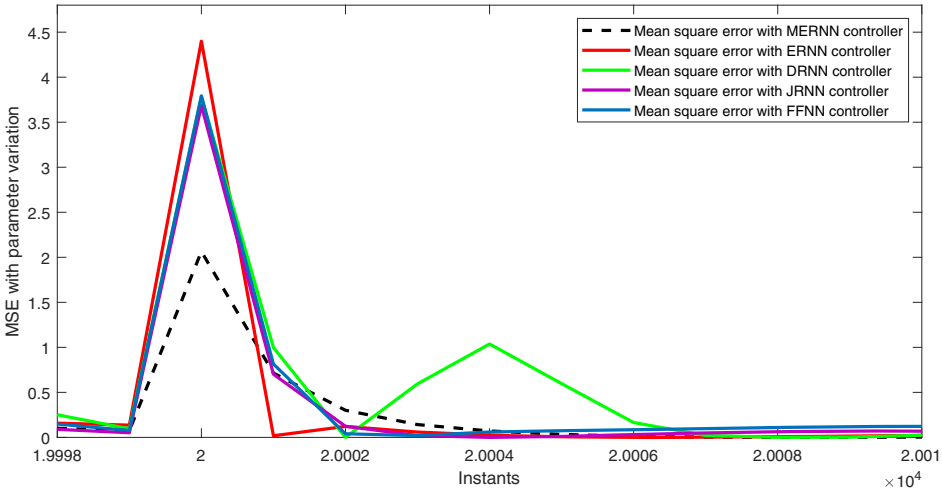


FIGURE 15 Mean squared error with parameter variation [Example 1].

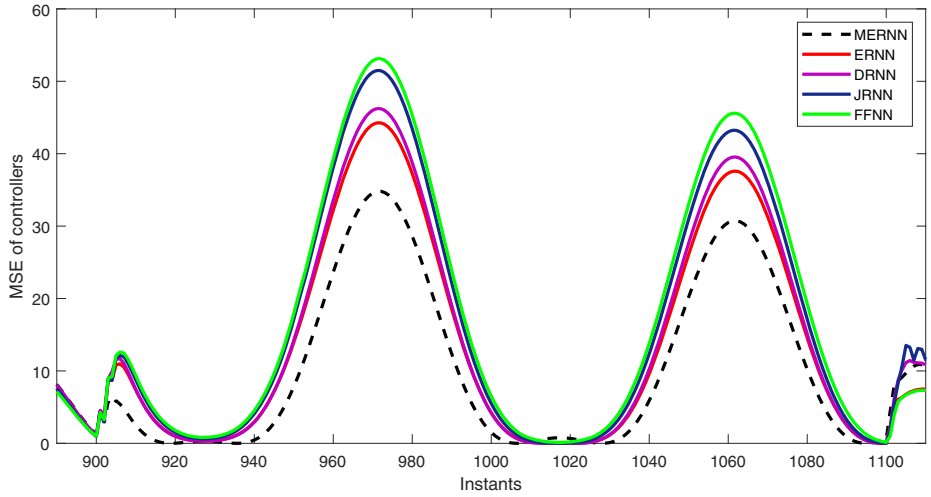


FIGURE 16 Mean squared error with disturbance signal [Example 1].

7.1.1 | Robustness testing

System robustness is defined as a system's ability to remain functioning under both uncertainties such as the disturbance signal and parameter variations. In this example, the robustness of all five controllers is analyzed and compared.

7.1.2 | Performance evaluation with parameter variation

In this case, at $k = 20{,}000$th time instant, a signal with a value of 2.5 units is applied to the controller's output to evaluate the effect of parameter variation. Figure 15 shows how parameter variation affects the MSE response of controllers. This figure shows that the increase in MSE in the proposed adaptive MERNN controller is minor at the time instant of parameter variation as compared to other controllers. MERNN-based controllers compensate for the growth in the MSE accordingly and restore it quickly to zero.

7.1.3 | Performance evaluation with disturbance signal

In the controller's output, a 1.8-unit disturbance signal is injected for values of k between 900 and 1100. The error would rise at this point; thus, if the controller is robust, this mistake must be compensated so that it can be reduced to zero

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

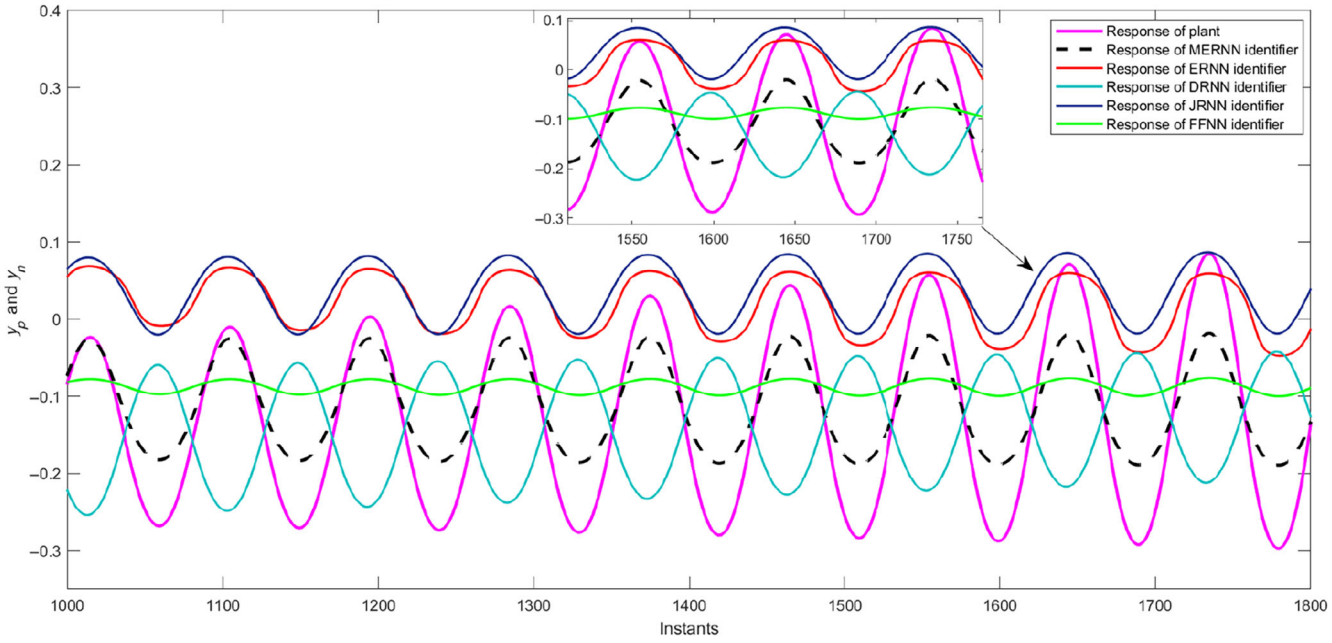


FIGURE 17 Identifier and plant output at the beginning of training [Example 2].

very quickly. Figure 16 demonstrates the instantaneous MSE plots of all the controllers affected by the disturbance signal. In comparison to the other controllers, this figure shows that the MERNN controller performs significantly better in adjusting for the impacts of the disturbance signal.

7.2 | Example 2

In this example, a nonlinear dynamical plant is considered whose dynamics are unknown, defined by the following difference equation as given in Reference 29.

$$y_p(k+1) = \frac{y_p(k)y_p(k-1)y_p(k-2)u_c(k-1)[y_p(k-2)-1]}{1+y_p^2(k-1)+y_p^2(k-2)} + \frac{u_c(k)}{1+y_p^2(k-1)+y_p^2(k-2)}, \tag{55}$$

$$r(k) = \sin\left(\frac{\pi k}{45}\right), \tag{56}$$

where $r(k)$ is the external input. The reference model equation is described as follows

$$y_m(k+1) = 0.08y_m(k) + 0.13y_m(k-1) + r(k). \tag{57}$$

If the output error $e_c(k+1)$ is defined as $e_c(k+1) = y_m(k+1) - y_p(k+1)$, the objective of control is to determine a bounded control input $u_c(k)$ by making $e_c(k+1) = 0$. In the case of a known function F, $u_c(k)$ is directly calculated based on $y_p(k)$, and its past values as $u_c(k)$.

$$u_c(k) = -F\left[y_p(k), y_p(k-1), y_p(k-2)\right] + 0.08y_p(k) + 0.13y_p(k-1) + r(k). \tag{58}$$

The function F is initially identified offline using neural network N if it is unknown. Therefore,

$$u_c(k) = -N\left[y_p(k), y_p(k-1), y_p(k-2)\right] + 0.08y_p(k) + 0.13y_p(k-1) + r(k). \tag{59}$$

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

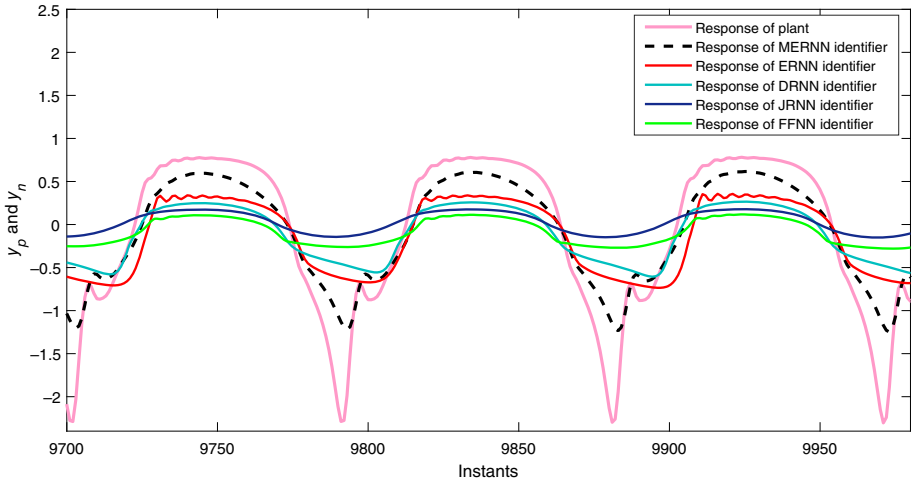


FIGURE 18 Identifier and plant output at the end of training [Example 2].

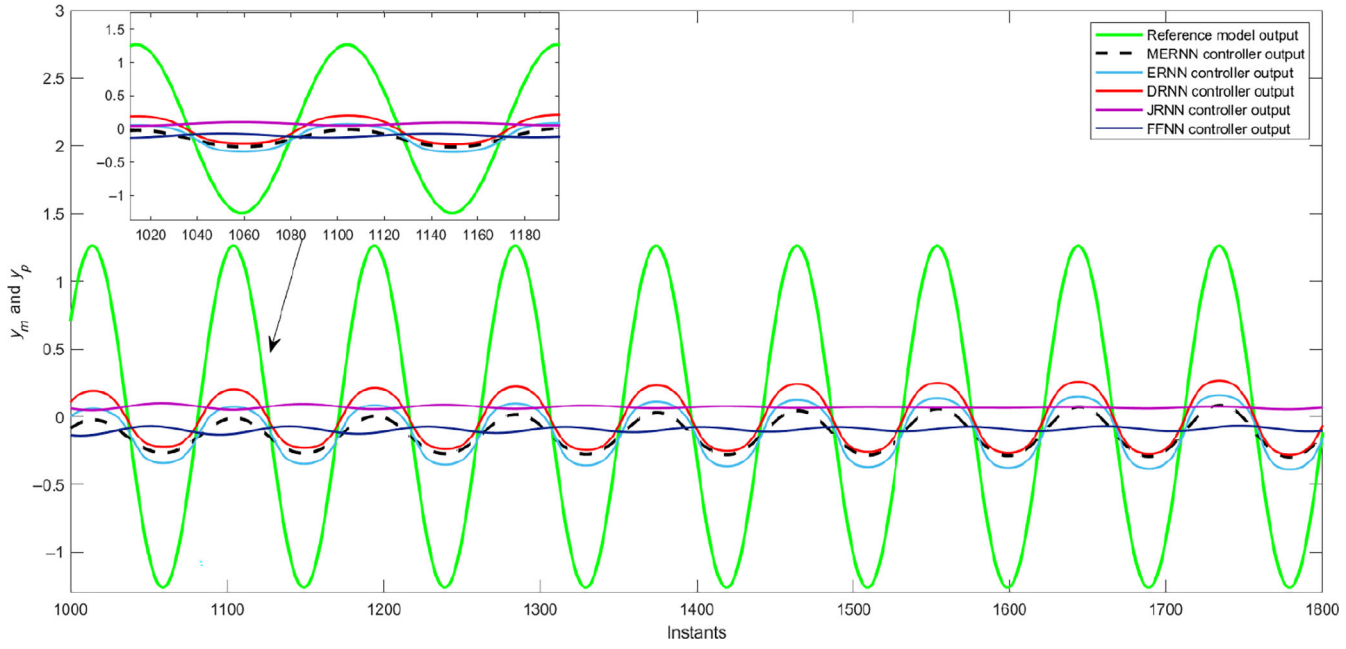


FIGURE 19 Controller and reference model output at the beginning of training [Example 2].

Then, the result of the nonlinear difference equation

$$y_p(k+1) = F\left[y_p(k), y_p(k-1), y_p(k-2)\right] - N[y_p(k), y_p(k-1), y_p(k-2)] + 0.08y_p(k) + 0.13y_p(k-1) + r(k). \quad (60)$$

Figure 17 shows the response of different identifiers. These include MERNN, ERNN, DRNN, JRNN, and FFNN with a nonlinear plant at the start of training. Whereas Figure 18 shows its response at the end of training. Figure 19 shows the NN-based controllers' response and the reference model's response at the start of controller training. Figure 20 shows the same at the end of training.

Figures 21–25 show how parameter variation affects the output of MERNN, ERNN, DRNN, JRNN, and FFNN-based controllers.

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

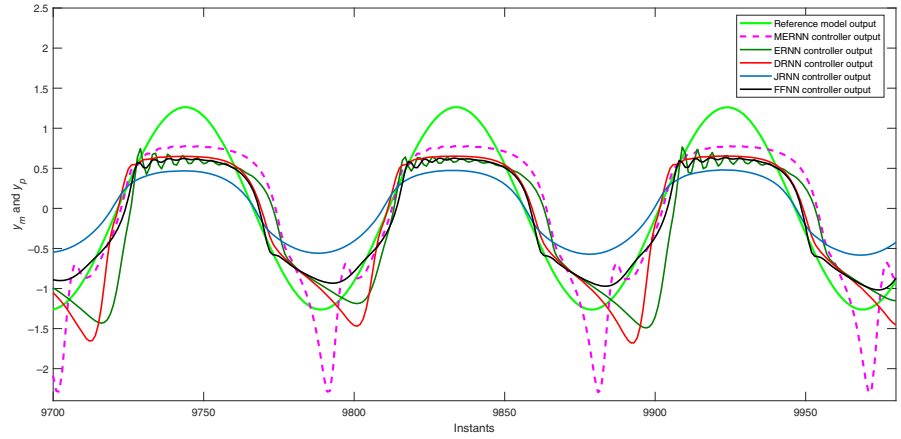


FIGURE 20 Controller and reference model output at the end of training [Example 2].

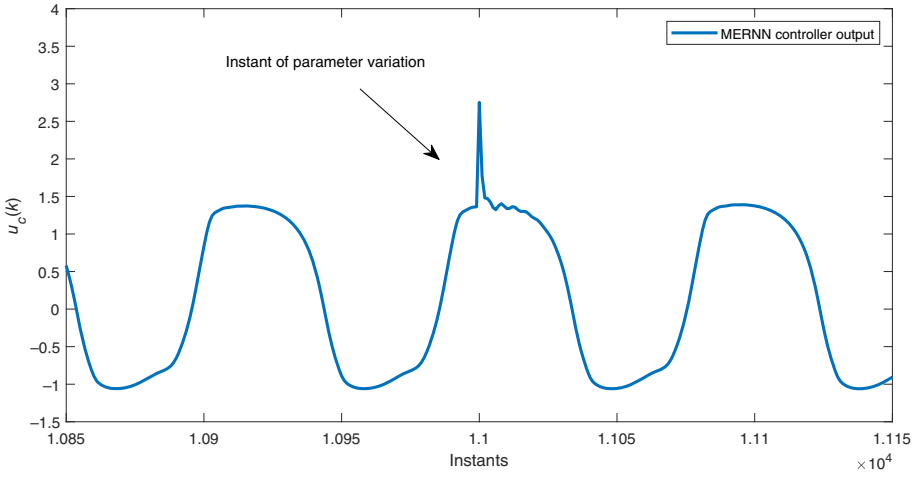


FIGURE 21 MERNN controller output with parameter variation [Example 2].

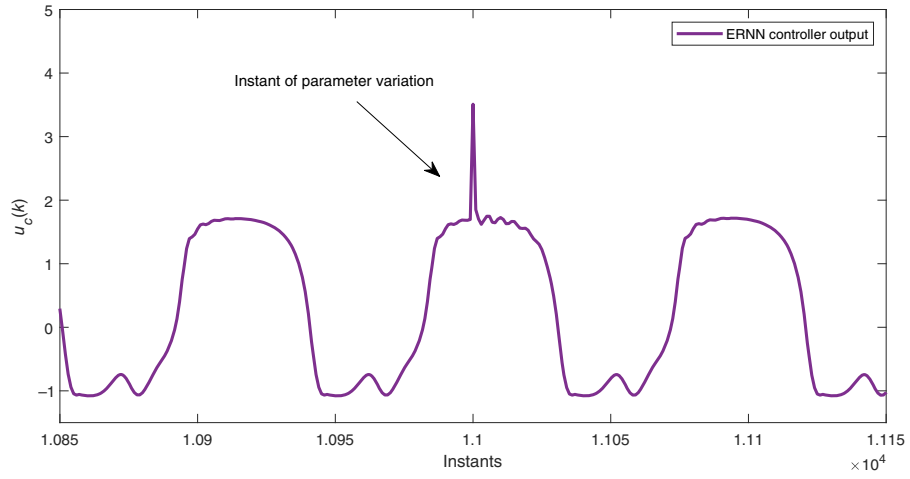


FIGURE 22 ERNN controller output with parameter variation [Example 2].

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

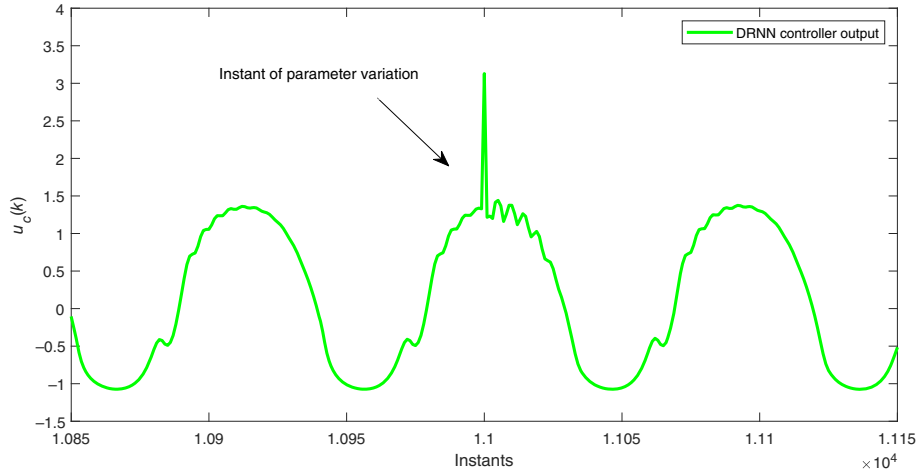


FIGURE 23 Diagonal recurrent neural network controller output with parameter variation [Example 2].

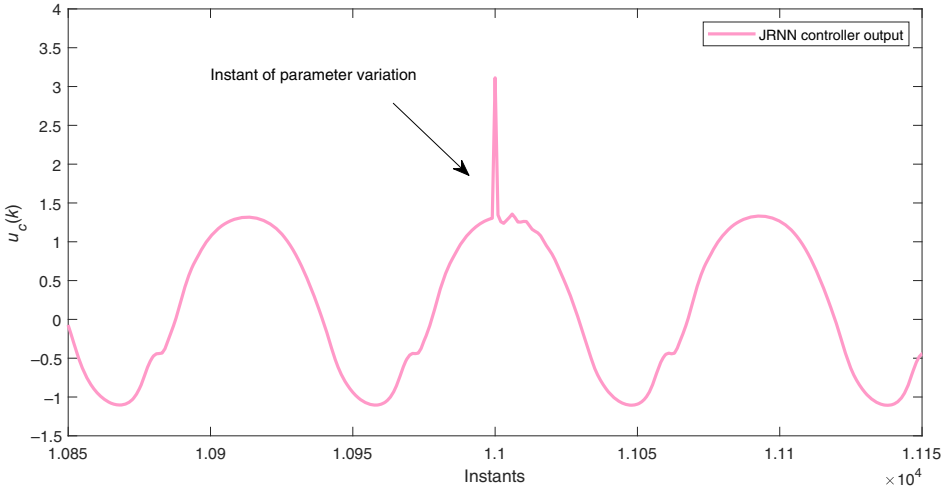


FIGURE 24 Jordan recurrent neural network controller output with parameter variation [Example 2].

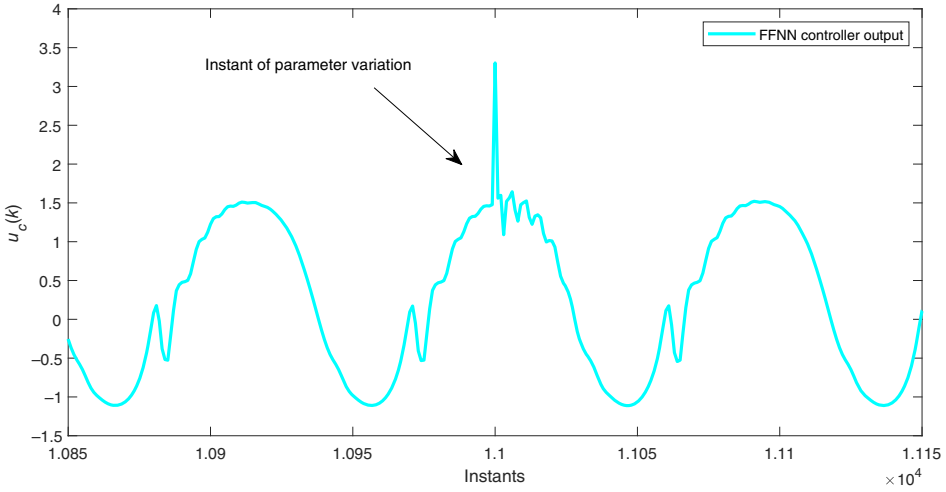


FIGURE 25 FFNN controller output with parameter variation [Example 2].

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

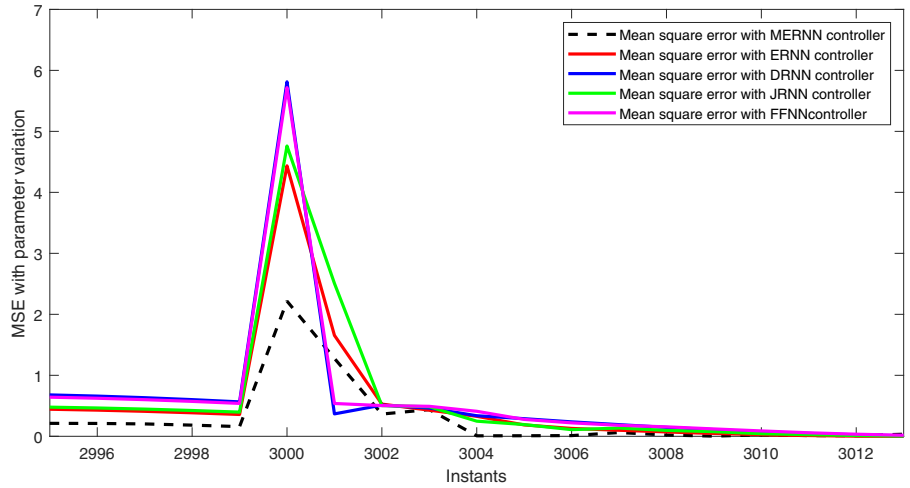


FIGURE 26 Mean square error with parameter variation [Example 2].

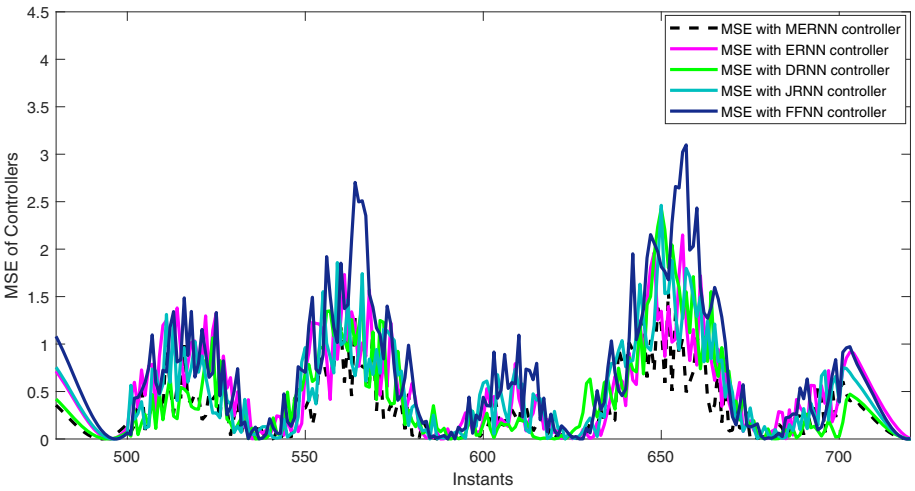


FIGURE 27 Mean square error with disturbance signal [Example 2].

7.2.1 | Performance evaluation with parameter variation

When a signal of magnitude 4.5 is applied to a controller, its performance is evaluated at the 3000th time instant. Figure 26 shows the rise in the MSE response of the controllers due to parameter variation. In this figure, the reaction of the MERNN controller shows a lesser amount of variation in comparison to the other controllers.

7.2.2 | Performance evaluation with disturbance signal

In the above example, a 1.5-unit disturbance signal is added to the controller output for k values ranging from 500 to 700. Figure 27 shows the response of instantaneous MSE of the controllers affected by the disturbance signal. From this figure, it can be noticed that the MSE response with the MERNN controller recovered more rapidly than the other controllers.

8 | CONCLUSION

This research presents a comparative study of different NN models for the identification and adaptive control of nonlinear dynamical systems. In this study, two simulation examples reveal that the identification and control of nonlinear systems

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

are very effective using the suggested methods. If the system dynamics are unknown, the control strategy suggested in this study is also effective. The back-propagation optimization method is applied for the training of adjustable parameters online. The performance analysis of MERNN as a controller is compared to that of ERNN, DRNN, JRNN, FNNN, and RBFN-based controllers for controlling dynamical systems and for compensating the effect of uncertainties like parameter variation and disturbance signal. According to the results, the MERNN model as an identifier and controller has the lowest mean square error and is superior to the others.

AUTHOR CONTRIBUTIONS

All the authors have contributed equally to the study, conception, design, and analysis. The whole draft of the manuscript was written jointly by all the authors.

FUNDING INFORMATION

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

ETHICS STATEMENT

This article does not contain any studies with human participants or animals performed by any of the authors.

ORCID

Kartik Saini https://orcid.org/0000-0002-6016-4993
Rajesh Kumar https://orcid.org/0000-0001-7172-1081

REFERENCES

1. Perrusquía A, Yu W. Identification and optimal control of nonlinear systems using recurrent neural networks and reinforcement learning: an overview. *Neurocomputing*. 2021;438:145-154.
2. Zheng D-D, Pan Y, Guo K, Yu H. Identification and control of nonlinear systems using neural networks: a singularity-free approach. *IEEE Trans Neural Netw Learn Syst*. 2019;30(9):2696-2706.
3. Liu Y-J, Zhao W, Liu L, Li D, Tong S, Chen CLP. Adaptive neural network control for a class of nonlinear systems with function constraints on states. *IEEE Trans Neural Netw Learn Syst*. 2023;34(6):2732-2741. doi:10.1109/TNNLS.2021.3107600
4. Jia C, Li X, Wang K, Ding D. Adaptive control of nonlinear system using online error minimum neural networks. *ISA Trans*. 2016;65:125-132.
5. Tiumentsev Y, Egorchev M. *Neural Network Modeling and Identification of Dynamical Systems*. Academic Press; 2019.
6. Haji SH, Abdulazeez AM. Comparison of optimization techniques based on gradient descent algorithm: a review. *PalArch's J Arch Egypt Egyptol*. 2021;18(4):2715-2743.
7. Meza-Aguilar M, Loukianov AG, Rivera J. Sliding mode adaptive control for a class of nonlinear time-varying systems. *Int J Robust Nonlinear Control*. 2019;29(3):766-778.
8. Chu Y, Fei J, Hou S. Adaptive global sliding-mode control for dynamic systems using double hidden layer recurrent neural network structure. *IEEE Trans Neural Netw Learn Syst*. 2019;31(4):1297-1309.
9. Takase T, Oyama S, Kurihara M. Effective neural network training with adaptive learning rate based on training loss. *Neural Netw*. 2018;101:68-78.
10. Daniel C, Taylor J, Nowozin S. Learning step size controllers for robust neural network training. Paper presented at: Proceedings of the AAAI Conference on Artificial Intelligence. 2016:30.
11. Minniti MV, Grandia R, Fäh K, Farshidian F, Hutter M. Model predictive robot-environment interaction control for mobile manipulation tasks. *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE; 2021:1651-1657.
12. Kumpati SN, Kannan P. Identification and control of dynamical systems using neural networks. *IEEE Trans Neural Netw*. 1990;1(1):4-27.
13. Ren G, Cao Y, Wen S, Huang T, Zeng Z. A modified elman neural network with a new learning rate scheme. *Neurocomputing*. 2018;286:11-18.
14. Kumar R. A lyapunov-stability-based context-layered recurrent pi-sigma neural network for the identification of nonlinear systems. *Appl Soft Comput*. 2022;122:108836.
15. Kumar R, Srivastava S, Gupta J, Mohindru A. Comparative study of neural networks for dynamic nonlinear systems identification. *Soft Comput*. 2019;23:101-114.

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

16. Liu P, Yu H, Cang S. Adaptive neural network tracking control for underactuated systems with matched and mismatched disturbances. *Nonlinear Dyn*. 2019;98(2):1447-1464.
17. Srivastava V, Sharma U, Schaub J, Pischinger S. Adaptive control concepts using radial basis functions and a kalman filter for embedded control applications. *Int J Engine Res*. 2024;25(1):125-139.
18. Huang W, Zhang W. Multi-objective optimization based on an adaptive competitive swarm optimizer. *Inform Sci*. 2022;583:266-287.
19. Ab Aziz MF, Mostafa SA, Foozy CFM, Mohammed MA, Elhoseny M, Abualkishik AZ. Integrating elman recurrent neural network with particle swarm optimization algorithms for an improved hybrid training of multidisciplinary datasets. *Expert Syst Appl*. 2021;183:115441.
20. Yoo SJ, Park JB, Choi YH. Indirect adaptive control of nonlinear dynamic systems using self recurrent wavelet neural networks via adaptive learning rates. *Inform Sci*. 2007;177(15):3074-3098.
21. Kumar R, Srivastava S, Gupta J. Diagonal recurrent neural network based adaptive control of nonlinear dynamical systems using lyapunov stability criterion. *ISA Trans*. 2017;67:407-427.
22. Wei Q, Liu D. Neural-network-based adaptive optimal tracking control scheme for discrete-time nonlinear systems with approximation errors. *Neurocomputing*. 2015;149:106-115.
23. Nasr MB, Chtourou M. Neural network control of nonlinear dynamic systems using hybrid algorithm. *Appl Soft Comput*. 2014;24:423-431.
24. Lee C-Y, Lee J-J. Adaptive control for uncertain nonlinear systems based on multiple neural networks. *IEEE Trans Syst Man Cybern Pt B Cybern*. 2004;34(1):325-333.
25. Xu B, Shi Z, Yang C, Sun F. Composite neural dynamic surface control of a class of uncertain nonlinear systems in strict-feedback form. *IEEE Trans Cybern*. 2014;44(12):2626-2634. doi:10.1109/TCYB.2014.2311824
26. Chen M, Ge SS. Direct adaptive neural control for a class of uncertain nonaffine nonlinear systems based on disturbance observer. *IEEE Trans Cybern*. 2013;43(4):1213-1225. doi:10.1109/TSMCB.2012.2226577
27. Cao X, Shi P, Li Z, Liu M. Neural-network-based adaptive backstepping control with application to spacecraft attitude regulation. *IEEE Trans Neural Netw Learn Syst*. 2018;29(9):4303-4313. doi:10.1109/TNNLS.2017.2756993
28. Wu C, Liu J, Jing X, Li H, Wu L. Adaptive fuzzy control for nonlinear networked control systems. *IEEE Trans Syst Man Cybern Syst*. 2017;47(8):2420-2430. doi:10.1109/TSMC.2017.2678760
29. Bhushan B, Singh M, Hage Y. Identification and control using mlp, elman, narxsp and radial basis function networks: a comparative analysis. *Artif Intell Rev*. 2012;37:133-156.
30. Li Y-M, Min X, Tong S. Adaptive fuzzy inverse optimal control for uncertain strict-feedback nonlinear systems. *IEEE Trans Fuzzy Syst*. 2020;28(10):2363-2374. doi:10.1109/TFUZZ.2019.2935693
31. Al Seyab R, Cao Y. Nonlinear system identification for predictive control using continuous time recurrent neural networks and automatic differentiation. *J Process Control*. 2008;18(6):568-581.
32. Hayakawa T, Haddad WM, Hovakimyan N. Neural network adaptive control for a class of nonlinear uncertain dynamical systems with asymptotic stability guarantees. *IEEE Trans Neural Netw*. 2008;19(1):80-89. doi:10.1109/TNN.2007.902704

How to cite this article: Saini K, Kumar N, Bhushan B, Kumar R. Artificial neural network-based adaptive control for nonlinear dynamical systems. *Int J Adapt Control Signal Process*. 2024;1-23. doi: 10.1002/acs.3823

10991115, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/acs.3823 by Delhi Technological University, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Check for updates

Received: 24 June 2023 | Revised: 19 March 2024 | Accepted: 19 April 2024

DOI: 10.1002/clen.202300198

RESEARCH ARTICLE

CLEAN Soil Air Water

Assessment of air quality benefits of vegetation in an urban-industrial region of India by integrating air monitoring with i-Tree Eco model

Mallika Vashist[1] | Thangamani Vijaya Kumar[2] | Santosh Kumar Singh[1,3]

[1]Department of Environmental Engineering, Delhi Technological University, Rohini, Delhi, India

[2]Department of Civil Engineering, Delhi Technological University, Rohini, Delhi, India

[3]Rajasthan Technical University, Kota, Rajasthan, India

Correspondence
Mallika Vashist, Department of Environmental Engineering, Delhi Technological University, Bawana Road, Shahbad Daulatpur, Rohini 110042, Delhi, India.
Email: m.vst979@gmail.com; mallikavashist_2k21phden01@dtu.ac.in

Abstract

In the last few years, urban trees have emerged as an effective nature-based solution to mitigate increasing air pollutant levels due to urbanization and industrialization. This study aims to assess the synergistic effect of urban trees on improving air quality by combining real-time $PM_{2.5}$ monitoring with the i-Tree Eco model. The monitoring was conducted during rush hours with high traffic volume and during non-rush hours, in both the tree alley and a non-tree road section within the industrial areas of the north-west region of the National Capital Territory of Delhi, India. The i-Tree Eco model was run using the diameter at breast height values of tree species present in the study area, and the $PM_{2.5}$ reduction ability of the trees was quantified. The results from both approaches indicated that urban trees can significantly reduce the traffic-fed $PM_{2.5}$ concentrations. Therefore, it is suggested that tree plantations be integrated into air pollution management strategies in urbanized regions with high traffic volumes. Although this study explores the initial link between trees and air quality in Delhi, further research incorporating local wind speed and direction measurements would provide a more comprehensive understanding of how trees influence air quality in any highly polluted urban setting.

KEYWORDS
nature-based solution, particulate matter, sustainable development goals, urban trees, vehicular emissions

1 | INTRODUCTION

Air pollution is a pressing global issue, with ≈4.2 million deaths attributed to it annually.[1] This hazard not only results in poor air quality but also elevates the risk of severe cardiovascular and respiratory disorders.[2–4] Moreover, it contributes to developing and worsening diabetes, mutagenicity, immunity, and other lifestyle-related disorders.[5–7] Beyond health concerns, poor air quality adversely affects urban wildlife, the environment, and the economy.[8] The air quality deterioration is primarily attributed to rapid urbanization, intensified industrial activities, high vehicular traffic volume, and increased population density. Elevating particulate matter (PM) emissions play a major role in the deterioration of air quality and represent a major concern in urban areas, with concentrations surging over the last decade, especially in Asia.[9] India, a South Asian country identified as one of the top five most polluted countries, particularly highlights the severity of this issue, with its megacity Delhi consistently ranking among the world's most polluted regions and recording high concentrations of $PM_{2.5}$ throughout the year.[4,10,11] Many studies have attributed these high concentrations of $PM_{2.5}$ in Delhi to vehicular emissions.[4,12] Winter conditions, including cool temperatures, shallow boundary layer heights, and frequent temperature inversions, contribute to a further escalation of already increased $PM_{2.5}$ levels by trapping pollutants within the boundary layer.[13] This

 © 2024 Wiley-VCH GmbH

necessitates potential solutions to mitigate air pollution, especially in rapidly urbanizing cities like Delhi.

Urban vegetation, comprising indoor plants, avenue trees, parks, green walls, green roofs, and other green systems, has emerged as an effective nature-based solution (NbS) to mitigate air pollution in urban areas.[14–16] It demonstrates enhanced efficiency in absorbing and catabolizing airborne pollutants compared to artificial surfaces while concurrently maintaining optimal urban environmental conditions.[17] It reduces $PM_{2.5}$ levels in the atmosphere through different mechanisms, such as deposition, dispersion, and modification. Deposition involves the 1D vertical settlement of pollutants on a uniform layer of vegetation,[18] dispersion leads to the alteration of the path and velocity of PM through the physical structure of urban trees,[19] and modification refers to the group of mechanisms which can alter of the actual properties of PM, like its size, which diminishes its adverse impacts.[20] The quantification of the reduction in $PM_{2.5}$ concentration through its dry deposition on urban vegetation is a widely studied mechanism.[20] The efficiency of trees in depositing $PM_{2.5}$ varies based on their morphology and meteorological factors such as wind speed.[21,22] A decrease in wind speed tends to increase the pollutant load owing to the poor ventilation ability and trapping effect at the leeward wall when the wind direction is perpendicular.[23,24] The intensity of this phenomenon depends on the wind direction, aspect ratio, number, distribution, and species of trees, as well as the density and porosity of vegetation. Moreover, this trapping effect mostly intensifies during winter.[25] Because of these highly varied micrometeorological conditions at the street level, it can be challenging to accurately quantify the air purification effect of urban trees. Thus, numerous studies have employed a combination of field investigations, wind tunnel studies, and numerical models to assess the influence of urban vegetation on wind flow patterns and pollution concentrations, demonstrating the potential of trees to filter out air pollutants at different locations.[26] For example, Riondato et al.[27] conducted air monitoring at vegetated and non-vegetated sites and found that the mean $PM_{2.5}$ concentration in the vegetative areas was almost half of that in the non-vegetated sites. Similarly, Zhu et al.[28] observed a 48% difference in $PM_{2.5}$ concentration in an urban park when compared to a nearby street with no vegetation in Wuhan.

The i-Tree Eco model, a user-friendly variant of the urban forest effects model, has been employed in various studies to quantify pollutants trapped by vegetation through dry deposition.[29] This model evaluates urban forest characteristics, such as species composition, leaf area, and tree health, to quantify the ecosystem services rendered by those trees. By utilizing the i-Tree Eco model's dry-deposition component, the pollution removal by trees during nonprecipitation periods and the hourly enhancements in air quality over a year can be estimated. The i-Tree Eco model can estimate pollutant removal by urban trees based on the wind tunnel parameters. It considers the urban forest as a leaf and estimates the $PM_{2.5}$ reduction across the boundary layer.[24] Several researchers worldwide have used this model to quantify the impact of trees on reducing air pollution.[30–34] However, the results obtained by the model require validation from field observations that employ high-accuracy $PM_{2.5}$ monitoring devices in multiple locations to provide high-resolution data on $PM_{2.5}$ mitigation through vegetation. For example, Su et al.[24] utilized an integrated model and mobile monitoring-based approach to quantify $PM_{2.5}$ mitigation at two different parks, one with dense vegetation and the other regarded as an open park, during the rush and non-rush hours, to calculate the PM removal. The results indicated that the park with a dense forest had a stable and stronger ability to improve air quality than open parks. The application of this integrated approach to assess the potential of urban vegetation in improving air quality remains unexplored in South Asian countries. Given the pressing environmental challenges faced by nations like India, particularly in cities like Delhi, with their poor air quality and high mortality rates, there is a critical need for NbS for air quality management, coupled with effective urban planning strategies. Therefore, the present study utilizes this approach of integrating mobile air monitoring with air quality modeling to assess the $PM_{2.5}$ deposition ability of urban vegetation in the industrial areas of the north-west district of Delhi, India, to: (a) assess the impact of urban vegetation on reducing PM levels in diverse urban settings: a tree alley (TA) and non-tree section (NTS) at different time periods, (b) evaluate the efficacy of urban trees in the reduction of air pollutants during both peak rush and non-rush hours, and (c) compare the results obtained from i-Tree Eco and those obtained from air quality monitoring. The results obtained from this study can help establish the crucial role of urban trees in mitigating both traffic-associated and background air pollution, thereby enhancing air quality within urban street canyons. Moreover, this research can serve as a baseline for emerging researchers and urban planners to conduct further research on air pollution mitigation through urban vegetation, particularly in highly polluted regions of South Asian countries.

2 | MATERIALS AND METHODS

2.1 | Site description

The urban-industrial areas investigated in this study are located in the north-west region of the National Capital Territory (NCT) of Delhi, which spans a geographical area of 1483 km^2 (28° 37′ N, 77° 12′ E). As the world's second most populous city, with 29 million population in 2018,[35] the NCT of Delhi experiences subtropical climatic conditions characterized by five well-defined seasons: winter (December–January), spring (February–March), summer (April–June), monsoon (July–September), and autumn (October–November). The weather conditions in this region feature hot summers and cold winters, with an average rainfall of 581 mm. Most of this rainfall (≈81%) occurs in July, August, and September, and the daily temperatures fluctuate between 5°C in winter and 45°C in summer.[36]

The air quality in Delhi undergoes notable fluctuations throughout the year, often surpassing the national ambient air quality standard value of 60 μg m^{-3} for 24-h mean $PM_{2.5}$ concentration. A notable spike is observed in the concentrations of $PM_{2.5}$, particularly during the winter season. This is primarily attributed to high pollutant build-up in winter due to weather conditions of a calm atmosphere, lower

18630669, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/clen.202300198 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

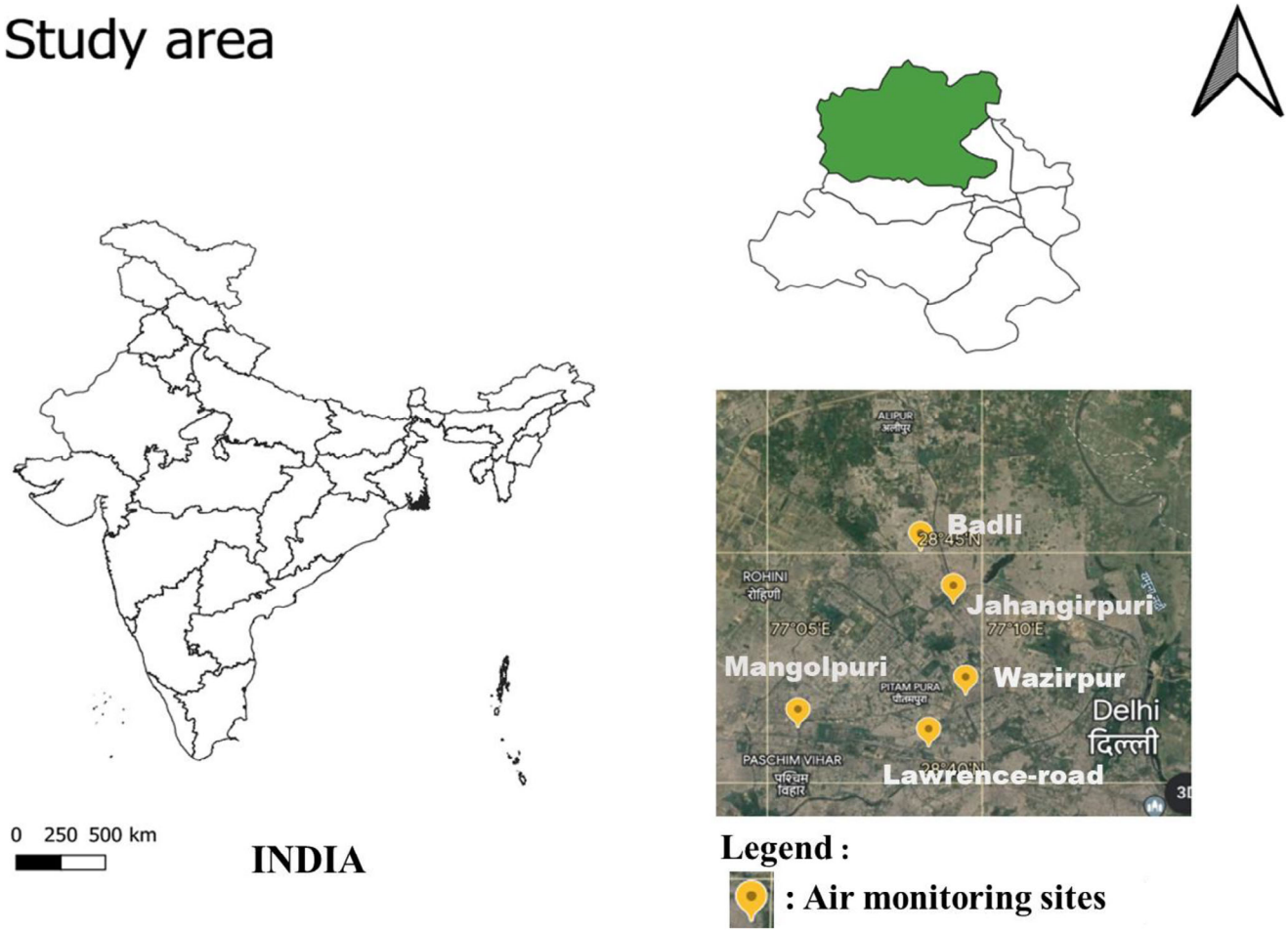


FIGURE 1 Reference map of the study areas used for the present analysis.

mixing heights due to colder temperatures, and temperature inversion that prevents the dispersion of pollutants.[37] The boundary layer is also the lowest in December, leading to the trapping of pollutants and a drastic increase in concentrations.[38] Additionally, annual Diwali celebrations involving fireworks further contribute to elevated $PM_{2.5}$ concentrations.[39] Consequently, the real-time monitoring study was conducted in December, when the most adverse air quality conditions prevailed.

The Central Pollution Control Board (CPCB) has categorized different areas of Delhi based on land use, including residential, commercial, and industrial areas. The primary sources of $PM_{2.5}$ in Delhi are vehicular emissions, construction progressions and suspended dust emissions, biomass burning, power plants, and industries.[40] The industrial areas of NCT of Delhi are hotspots of air pollutants owing to extensive industrial activities, heavy vehicular rush, commercial operations, petrol and diesel emissions, auto repair, natural gas combustion, combustion engines, and sewage sludge.[41] For this study, five different industrial areas were selected: Badli, Mangolpuri, Wazirpur, Jahangirpuri, and Lawrence-road industrial area (Figure 1). These areas were selected based on their vegetation, meteorological conditions, and anthropogenic activities. The TA and NTS chosen in each of the five industrial areas had similar characteristics. The road type chosen in each of the five areas was a street canyon, which comprised rows of buildings along both sides of the road. Two monitoring stations were set up at each site: one in the TA and the other in the NTS. Therefore, a total of 10 monitoring stations were set up for the present study. As these areas were established under proper strategic planning, vast vegetation was found with major patches of various plant species along the roadside and in various urban parks. Additionally, these five subsets of the Delhi industrial area mostly had common tree species with similar characteristics; therefore, their data sets, including the characteristics of trees, such as diameter at breast height (DBH), were pooled to obtain the final results.

These industrial areas were selected for air monitoring for several reasons:

a. Each of the streets selected for monitoring had a section with no planted trees. The other part had a TA with regularly distributed large and old trees along the road. Therefore, the presence of trees in one part of the road and their absence in another helped to compare the two sections.
b. Most trees in the five locations were planted on public property, ensuring that data for DBH for most of the local trees was measured and fed into the model.

18630669, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/clen.202300198 by Delhi Technological University, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

c. Each of the streets received reasonably heavy vehicular traffic, contributing to a higher concentration of $PM_{2.5}$, owing to combustion from their engines.[42] As the rate of pollutant deposition is proportional to their concentrations,[43] the impact of trees on air quality was more evident and significant in areas with poor air quality.

2.2 | Data collection

In the present study, pollution levels were recorded using air quality detector sensors (Xiaomi Smart mi $PM_{2.5}$). These sensors feature a three-color digital indicator with a high-precision laser-sensing system and accurately detects $PM_{2.5}$. Their large screen features enlarged numbers, which make it easy to read. Furthermore, its mini-body is convenient to carry at the monitoring site and can be powered by a USB cable. The air quality monitoring was conducted for 5 days in December 2021 at 10 monitoring sites (two sites comprising a TA and NTS in each industrial area) simultaneously. The monitors were run simultaneously for ≈60–120 min during rush and low-traffic hours at each site. A total of 1070 $PM_{2.5}$ values were recorded from each area, comprising values from both the TA and NTS. The values were recorded for 5 days from 8:00 to 10:00 a.m. and 2:00 to 7:00 p.m. Data from 5:00 to 7:00 p.m. was considered to be the time when traffic intensity was at its maximum on the road and thus has been referred to as the rush hour time period. Notably, Murthy et al.[44] found that $PM_{2.5}$ concentrations in the NCT of Delhi exhibit diurnal variation, with minimum levels occurring during mid-afternoon due to increased mixing caused by solar radiation. However, the concentrations elevate during the evening hours and remain high at night. Subsequently, the data from 8:00 to 11:00 a.m. and 1:00 to 5:00 p.m. was considered to be the hours when traffic volume was relatively low and thus have been referred to as a non-rush hour time period. The readings were recorded at 5-min intervals. It was observed that during rush hours, the predominant cause of air pollution stemmed from the increased traffic volume, whereas during non-rush hours, air pollution predominantly reflected the background concentration of the areas. The monitoring was thus carried out during these two periods to investigate urban vegetation's potential in mitigating traffic-fed air pollution and background pollution. Other studies focusing on $PM_{2.5}$ concentrations in Delhi have also considered the same time period of the day for rush and non-rush hours.[39,45] Moreover, during monitoring, the observations at all sites confirmed the duration of the peak traffic hours. Preparatory monitoring was also conducted to ensure that the concentration of air pollutants was low during non-rush hours; therefore, maximum monitoring was performed during rush hours only. The five $PM_{2.5}$ values recorded in the TA (one from each site) were later averaged to obtain a single value of $PM_{2.5}$ concentration in TA for that time of the day across the five monitoring sites. The same exercise was repeated to obtain a single numerical value for $PM_{2.5}$ in the NTS for that time of the day across the five monitoring sites. The concentrations of $PM_{2.5}$ were recorded at almost the midpoint of the TA and NTS at each monitoring site. Therefore, the monitors were placed to represent the $PM_{2.5}$ concentrations in both sections of the roads, TA and NTS.

For the i-Tree Eco model calculations, individual tree species were identified and marked at each monitoring site. The values of the circumference at the base height of the 38 most common tree species across all five sites in the TA were measured using a measuring tape, which was then converted to corresponding DBH values. Subsequently, the DBH values for the same tree species were averaged to a single value. This averaging process was undertaken because the DBH values for the same tree species were consistently similar, attributable to factors such as similar tree height and age, and exposure to similar anthropogenic activities such as industrial practices and heavy vehicular traffic. This procedure was followed for all the species. Finally, 38 distinct DBH values corresponding to 38 tree species were then used as input for the i-Tree Eco model, ensuring a representative assessment of urban vegetation's impact on air quality in the industrial areas of Delhi.

2.3 | i-Tree Eco model setup

The data collection process in the study area was conducted based on the guidelines provided in the i-Tree Eco user's manual.[46] When setting up the model, the user is required to specify whether the tree data is being randomly sampled or the inventory of all trees in the study area has been completed. In the present study, all tree species in the TA were identified and recorded, along with the number of individuals of each species. The minimum requirements for running the model are entering the name of the species and the values of their DBH. The model has the option to enter additional fields such as tree address, GPS coordinates, crown height, and crown health. For this study, only the name and DBH of each tree species were entered into the model. For Delhi, the model has pre-saved air quality and weather data from two meteorological stations, namely, IGI airport monitoring station, ID: 421810-99999 (latitude, longitude: 28.56, 77.10) and Safdarjung monitoring station, ID: 421820-99999 (latitude, longitude: 28.58, 77.20). The Safdarjung monitoring station was selected as it was closer to the industrial areas of Delhi. The model was run after selecting the monitoring station and entering the names of the tree species and their DBH.

2.4 | i-Tree calculations

The i-Tree Eco model uses the following formulas to calculate the air quality improvement:[47]

Total air pollutant mass per unit tree cover is calculated as

$$M_{total} = H\,C, \tag{1}$$

where M_{total} is the total air pollutant mass per unit tree cover ($g^{-2}\,h^{-1}$), H is the urban mixing height (m), and C is the air pollutant concentration ($g^{-3}\,h^{-1}$).

The hourly improvement in air quality per unit of tree cover resulting from the dry deposition of air pollutants is further calculated

18630669, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/clen.202300198 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

employing M_{total} in the subsequent steps:

$$I_{unit}\,(\%) = (F/F + M_{total}) \times 100, \quad (2)$$

where F is the pollutant flux ($g^{-2}\ h^{-1}$).

Now, if the net flux is negative for $PM_{2.5}$, then

$$I_{unit} = (F/M_{total}) \times 100. \quad (3)$$

Hourly augmentation of air quality resulting from total tree cover is computed as

$$I_{total} = \{[(F\,F_c)/100\,]/[\,(F\,F_c)/100 + M_{total}]\} \times 100, \quad (4)$$

where T_c is the total tree cover in the city (%).

If $F = M_{total}$ and $T_c = 100$ for $PM_{2.5}$, the denominator in the equation will become 0, and I_{total} will become infinity.

This is avoided by using the following equation:

$$I_{total} = (I_{unit}\,T_c)/100. \quad (5)$$

3 | RESULTS

3.1 | Variation of $PM_{2.5}$ between TA and NTS

The real-time air monitoring was carried out in the winter season in December because the characteristics of the winter season are different from other seasons due to the phenomenon of temperature inversion and low mixing height, which resulted in significantly higher pollutant accumulation in this month.[48,49] The findings of the present study demonstrated the reductive effect of urban vegetation on $PM_{2.5}$. The real-time measurement data from NTS and TA was collected from 10 monitoring sites to investigate the concentration of $PM_{2.5}$ particles. From each monitoring site, 532 sample points from the NTS and 538 from TA were recorded concurrently at 5 min intervals over 5 days in December 2021. There were a few occasions when $PM_{2.5}$ concentrations were not recorded due to instrument error. During the data analysis, it was assumed that minor differences in meteorological conditions near the monitoring sites would not significantly affect $PM_{2.5}$. The $PM_{2.5}$ values from each monitoring site were then averaged to obtain the final values analyzed in this study. The concentrations of $PM_{2.5}$ were more than 60 μg m^{-3} for 93.42% of the total sample points in NTS and 93% of the total sample points in TA. Furthermore, during rush hours with high traffic volume, $PM_{2.5}$ values were >60 μg m^{-3} for 97.86% of the sample points in NTS and 92% of the total sample points in TA. Additionally, the average daily mean concentrations of $PM_{2.5}$ in Delhi exceeded the threshold limit of 25 μg m^{-3} set by WHO for all the sample points at all times.

Table 1 presents the number of individual sample points where the $PM_{2.5}$ concentrations exceeded specific thresholds: 50, 60, 100, and 150 μg m^{-3} recorded across both rush and non-rush hours, whereas Table 2 reports the individual sample points where the $PM_{2.5}$ concentrations exceeded the same $PM_{2.5}$ values during rush hours (5:00–7:00 p.m.). The 50 μg m^{-3} threshold represents a moderately elevated level of $PM_{2.5}$ in a city like Delhi. The 60 μg m^{-3} threshold aligns with the ambient air quality standard for 24-h average concentration of $PM_{2.5}$ established by CPCB in India. This regulatory standard is crucial in air quality management and is used globally as a benchmark for evaluating air pollution. Additionally, thresholds of 100 and 150 μg m^{-3} signify hazardous air quality which can lead to increased respiratory and cardiovascular risks. These elevated $PM_{2.5}$ levels are mostly prevalent during winter in Delhi, primarily due to atmospheric inversions, high traffic activity, and industrial emissions occurring concurrently. The $PM_{2.5}$ concentrations for TA ranged between 42 and 221 μg m^{-3}, whereas that for NTS ranged between 37 and 267 μg m^{-3}. Air quality was comparatively better in TA than in NTS during all hours of monitoring. The majority of the values in the NTS and TA surpassed 100 μg m^{-3}. $PM_{2.5}$ values exceeding 150 μg m^{-3} were twice as frequent in the NTS than in the TA, providing evidence for the air-purifying capabilities of urban vegetation through the higher incidence of pollutant concentrations surpassing air quality thresholds. Similar results were shown by Wu et al.[50] who suggested that the fluctuation in the distribution of urban vegetation within a region can significantly influence the quantification of total pollution removal capacity and the removal rate using a dry deposition model. Another study by Jeong et al.[19] based on practical guidelines for managing urban vegetation suggested that multi-planting street trees and central reserves have more substantial PM reduction effects and provide health benefits.

3.2 | Statistical analysis

The $PM_{2.5}$ concentrations were analyzed using the Shapiro–Wilk test, which depicted that the values were not distributed normally ($p < 0.05$). The $PM_{2.5}$ concentrations from NTS and TA were visualized using histograms, indicating that both data sets were left-skewed (Figure 2). Consequently, both distribution sets were analyzed using a nonparametric Mann–Whitney U test. The test showed that both distribution sets differed significantly ($W = 105\,211$, $p < 0.05$). The average $PM_{2.5}$ concentration at NTS was 140.05 μg m^{-3} compared to 117.34 μg m^{-3} at TA. The $PM_{2.5}$ values were high between 9:00 and 10:00 a.m., then started to decrease around 1:00 pm, and then again started to increase to the maximum levels around 5:00–7:00 p.m.

The average $PM_{2.5}$ values during hours with high traffic intensity, that is, rush hours, were 176.23 μg m^{-3} in NTS and 118.79 μg m^{-3} in TA, whereas, during non-rush hours, the average values were 122.01 μg m^{-3} in NTS and 116.63 μg m^{-3} in TA. Figure 3 depicts the average $PM_{2.5}$ concentrations in NTS and TA during rush hours. Furthermore, during rush hours, the average difference between $PM_{2.5}$ values of TA and NTS was ≈52.89 and 75.46 μg m^{-3} at 5:00 and 6:00 p.m., respectively. This translated to an increase in the air quality of 59.52% at 5:00 p.m. and 73.23% at 6:00 p.m. in the TA as compared to the NTS. Additionally, the NTS had consistently higher $PM_{2.5}$ concentrations than TA. A trendline for each data set was plotted, which

18630669, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/clen.202300198 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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 Number of recorded $PM_{2.5}$ concentrations exceeding 50, 60, 100, and 150 μg m^{-3} during all monitoring hours.

			Number of cases exceeding thresholds of:							
	Sample points recorded		50 μg m^{-3}		60 μg m^{-3}		100 μg m^{-3}		150 μg m^{-3}	
Day	Non-tree	Tree	Non-tree	Tree	Non-tree	Tree	Non-tree	Tree	Non-tree	Tree
1	108	107	108	108	108	108	97	53	45	35
2	106	108	89	100	81	78	72	42	18	0
3	105	108	104	108	95	100	55	36	24	0
4	105	107	105	108	105	108	105	108	48	33
5	108	108	108	108	108	108	98	95	65	36
Total	532	538	514	532	497	502	427	334	200	104

depicted an upward increase in the concentrations over time. The average difference between the two locations during rush hours was 44.59, 52.89, and 75.46 μg m^{-3} at 10:00 a.m., 5:00 p.m., and 7:00 p.m., respectively, which rendered an improvement of 27.11%, 59.52%, and 73.23% of air quality in the TA as compared to the concentrations in NTS. The average $PM_{2.5}$ values, unit, and percent difference in concentrations between NTS and TA are recorded in Table 3. The cumulative distribution function was plotted for $PM_{2.5}$ values at all times and is represented in Figure 4.

3.3 | i- Tree Eco model findings

The vegetation growing in the TA of the selected industrial areas was highly diverse in terms of morphological and physiological characteristics. The DBH of 38 commonly grown tree species was calculated after measuring the circumference at base height and fed in the i-Tree Eco model. The values of DBH ranged from 12.95 to 114.04 cm, with an average diameter of 46.22 cm. Based on the input values of DBH, the model predicted the amount of average monthly removal of $PM_{2.5}$ by vegetation in the study area (Figure 5). The results indicated that the minimum $PM_{2.5}$ removal was observed in February, which coincided with the defoliation period and consequently explains that the dry deposition model calculates the PM removal rate by both pollution concentration and leaf area index values. The maximum $PM_{2.5}$ removal took place in May, which is the full-leaf period. By utilizing the meteorological and air quality database of the i-Tree Eco from 2018, the model calculated that the vegetation of the TA was responsible for an hourly air quality improvement of 2.5% for $PM_{2.5}$ in each site (Figure 6). Moreover, the results from the model also suggested that trees can influence air quality both in negative and positive ways, as during high wind speed, the resuspension of $PM_{2.5}$ into the atmosphere can worsen the air quality.[24] However, as the wind speed in the study areas was between 0.8 and 3 m s^{-1} during monitoring in December 2021,[51] there was negligible PM resuspension, leading to lower PM concentrations during rush hours in TA.

4 | DISCUSSION

The present study combined mobile air monitoring with a model-based approach to assess the impacts of urban trees on air quality. It aimed to investigate the air filtering efficiency of vegetation in varying temporal scales and provide a more holistic assessment of the relationship between vegetation and air quality dynamics. The first method involved monitoring $PM_{2.5}$ concentrations for 5 days in December at different industrial areas of Delhi to elucidate an immediate variation in air quality due to localized factors, such as vehicular emissions and high atmospheric inversion conditions. This approach provided granularity and detailed observations about potential fluctuations in

TABLE 2 Number of recorded $PM_{2.5}$ concentrations exceeding 50, 60, 100, and 150 μg m^{-3} during rush hours.

			Number of cases exceeding thresholds of:							
	Sample points recorded		50 μg m^{-3}		60 μg m^{-3}		100 μg m^{-3}		150 μg m^{-3}	
Day	Non-tree	Tree	Non-tree	Tree	Non-tree	Tree	Non-tree	Tree	Non-tree	Tree
1	38	38	38	38	38	38	38	22	22	13
2	38	38	38	37	37	29	37	13	18	0
3	37	38	36	38	34	32	33	13	22	0
4	36	38	36	38	36	38	36	38	32	15
5	38	38	38	38	38	38	38	25	38	13
Total	187	190	186	189	183	175	182	111	132	41

18630669, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/clen.202300198 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

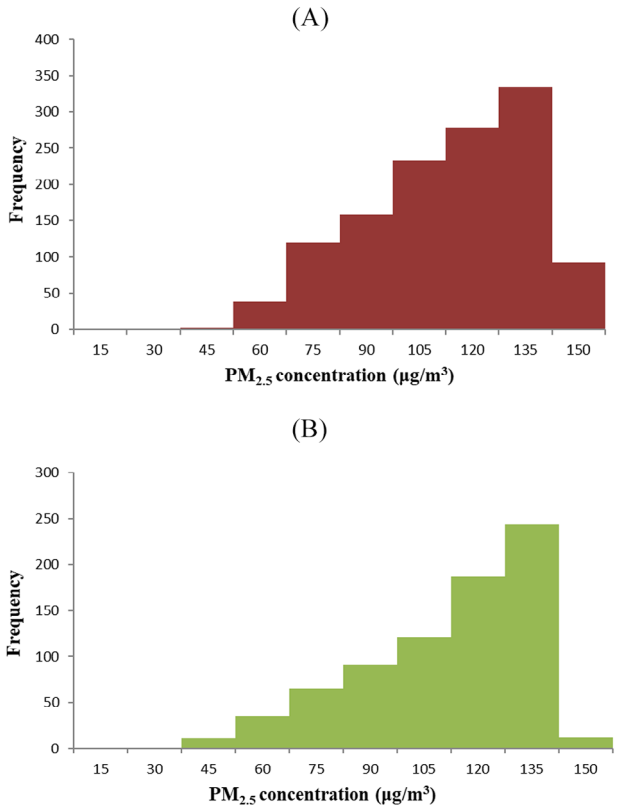


FIGURE 2 $PM_{2.5}$ concentrations in (A) non-tree road section and (B) tree alley.

the $PM_{2.5}$ levels influenced by the urban vegetation and the immediate effects of meteorological conditions. Conversely, the model-based approach provided the yearly trend of $PM_{2.5}$ mitigation by urban trees. This annual approach offered insights into the cumulative effects of vegetation on air quality, considering seasonal variations, changes in vegetation cover, and potential fluctuations in pollutant concentration over different seasons and environmental conditions. By employing both short- and long-term assessments, the study aimed to present a comprehensive understanding of the relationship between vegetation and air quality.

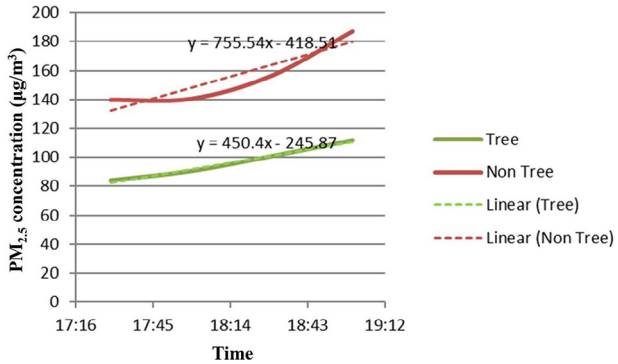


FIGURE 3 Average $PM_{2.5}$ concentrations in non-tree road section and tree alley during rush hours.

TABLE 3 Average concentrations of $PM_{2.5}$ in the non-tree section of the road and tree alley, along with unit and percent differences between them.

Time	Non-tree	Tree	Difference (µg m^{-3})	Difference (%)
8:00 a.m.	161.22	158.43	2.79	1.76
9:00 a.m.	173.58	161.85	11.73	7.25
10:00 a.m.	209.05	164.46	44.59	27.11
2:00 p.m.	114.23	106.96	7.27	6.8
3:00 p.m.	102.37	96.15	6.22	6.47
4:00 p.m.	87.38	89.65	−2.27	−2.53
5:00 p.m.	93.78	88	5.78	6.57
6:00 p.m.	141.75	88.86	52.89	59.52
7:00 p.m.	178.51	103.05	75.46	73.23

Based on the number of times the concentrations of pollutants exceeded certain air quality thresholds, the air-ameliorating properties of urban vegetation were apparent in the present study. The observed concentrations of $PM_{2.5}$ during non-rush hours were nearly the same between NTS and TA, fluctuating around 119 µg m^{-3}. These relatively similar values of $PM_{2.5}$ might represent the background pollution levels. Background concentrations occur due to contributions from all upwind source areas rather than any single source.[52] Knowledge about the concentration of background pollution can help determine a new source of pollution in an area.[53] An increase in $PM_{2.5}$ concentrations above the background levels may be due to high vehicular emissions in the monitoring locations. During rush hours, there was a sharp increase in the average $PM_{2.5}$ concentrations in both the monitoring locations, with the average concentration in TA being ≈18% less than that in NTS. Similar results have been obtained in several studies that compared the average $PM_{2.5}$ concentrations in the TA and non-tree road sections. For instance, Ozdemir et al.[54] showed that roadside $PM_{2.5}$ concentrations in Istanbul, Turkey, were significantly

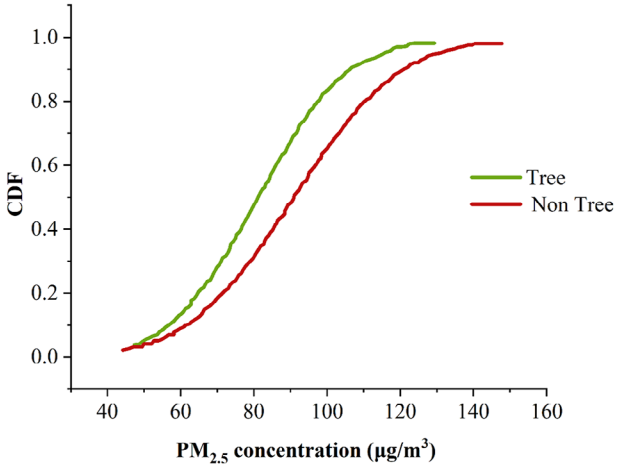


FIGURE 4 Cumulative distribution function (CDF) for $PM_{2.5}$ concentrations.

18630669, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/clen.202300198 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

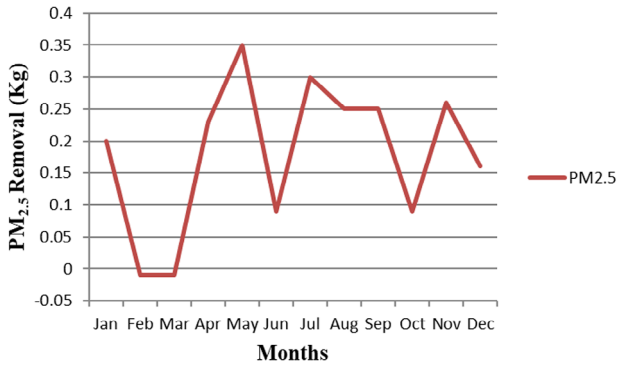


FIGURE 5 Average monthly removal of $PM_{2.5}$ by vegetation in industrial areas National Capital Territory (NCT) of Delhi.

lower in road sections with thick trees and no gaps than in road sections with no trees.

The results from i-Tree Eco revealed lower $PM_{2.5}$ removal during February and March, aligning with Delhi's defoliation period. This highlights the crucial role of leaves and their traits in pollutant removal. According to a study by Corada et al.,[55] leaf characteristics, mainly macrostructures (leaf area index and leaf area density) and microstructure (leaf shape, rigidity, and surface traits like wax and trichomes), impact PM removal. The epicuticular layer of leaves also plays a significant role by absorbing PM, with its effectiveness influenced by structure and composition. Leaves with rough surfaces retain more PM, and small leaves with complex structures and large, dense stomata efficiently accumulate PM as shown in studies by Song et al.,[56] Vashist and Singh,[57] and Chen et al.[58] The results from the model could have also been influenced by the wind speed and direction at the time of monitoring. Based on meteorological data from the DPCC website, the average daily wind speeds in the study areas ranged from 0.8 to 3 m s^{-1}, blowing in a west to north-west direction during December 2021.[48,51] A study by Nowak et al.[59] stated that around 4.5% of PM is resuspended at a wind speed of 3 m s^{-1}, which increases to ≈12% at a wind speed of 9 m s^{-1}. Although the mitigation ability of trees would have been altered due to the wind currents in the study area, there was still an evident decline in the concentration of $PM_{2.5}$ in the TA during rush hours. This observation suggests that the trees demonstrated a significant ability to filter the air in the study area during the elevated concentration levels, despite the prevailing wind direction or speed.

Additionally, as per the model findings, the urban trees present in the study area improved the air quality by 2% on an hourly basis. However, during field observations, the maximum recorded difference in air quality between the NTS and the TA was an improvement of 73.23%. This major difference between the results obtained from the model and real-time measurements can be due to the fact that the i-Tree Eco model used the air quality and meteorological data from the year 2018, and it is possible that the weather conditions and air quality during the monitoring period were significantly different, resulting in an inaccurate prediction from the model. Moreover, collecting additional data like tree height and crown size would have improved the data efficiency of the i-Tree Eco model and provided a better estimate of the role of trees in enhancing air quality. Additionally, there were further limitations linked to utilizing the i-Tree Eco model, such as the model operated on prescribed concentrations from the selected monitoring station without considering the feedback of deposition on $PM_{2.5}$ concentration and hence did not account for the effects of pollutant deposition on subsequent concentration levels. Furthermore, the model did not account for diverse interactions and processes responsive to $PM_{2.5}$ concentration changes, such as atmospheric chemistry and wet deposition. Despite these drawbacks, the model is still preferred because of its deposition-based approach, which helps gain a comprehensive depiction of vegetation through its detailed representation of trees or tree canopies. The environmental challenges faced by South Asian countries, particularly in Delhi, necessitate the

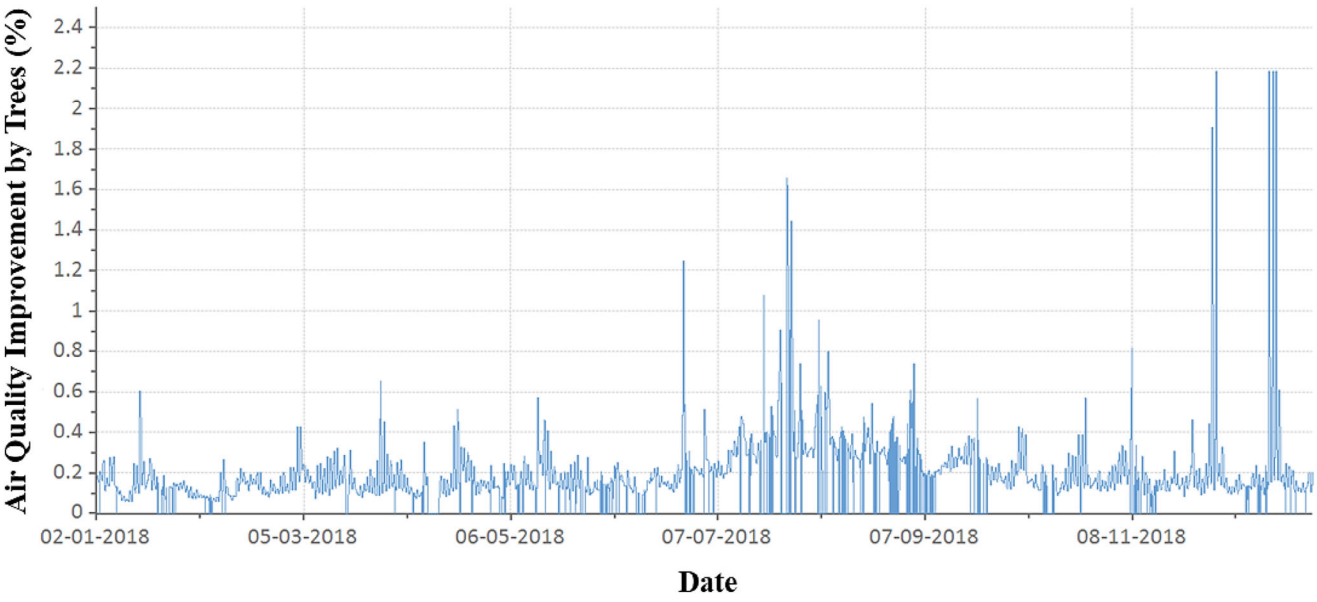


FIGURE 6 Improvement (%) of air quality by trees in industrial area National Capital Territory (NCT) of Delhi.

18630669, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/clen.202300198 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

implementation of NbS to address air pollution management, given their poor air quality and elevated mortality rates. The i-Tree Eco model offers quantifiable results on the potential of urban trees to mitigate air pollution, which could attract further research interest. Although this study offers a preliminary investigation into the correlation between trees and air quality in Delhi, conducting additional research that includes measurements of local wind speed and direction would provide a more comprehensive understanding of the impact of trees on air quality.

By 2030, $PM_{2.5}$ is predicted to cross an annual mean value of 74 μg m^{-3}, which is almost eight times the WHO standards/guidelines, and is expected to reduce life expectancy by more than 10 years in Delhi.[60] Several control strategies for improving the air quality have been implemented in Delhi, like Bharat stage IV vehicles, an odd-even scheme, a ban on open waste burning, and transferring entire public transportation to CNG fuel. A set of emergency measures termed a Graded Response Action Plan is also implemented in Delhi when $PM_{2.5}$ concentrations reach high levels, including reactive measures to control and reduce air pollution in the city.[61] Another project known as Clean Air for Delhi Through Interventions, Management, and Engagement has been implemented jointly by India and the United Kingdom with the aim of addressing air quality issues that affect public health in Delhi. The project has been initiated to identify a set of measures to improve present air quality by taking two future time horizons: 2030 and 2050. These options, however, have met with little success. Based on the findings of this study, it could be suggested that combining the strategy of developing vegetation strips along roads that serve as filters for air pollution from point sources along with other measures can help further reduce the concentration of $PM_{2.5}$. This is crucial because, along with improving air quality, urban green areas can help decrease stress, enhance moods, and improve social interaction, which can contribute to the prevention of cardiovascular disease, mental illnesses, and lower mortality rates.[62] Green areas can thus help address several urbanization challenges, as depicted in Figure 7, and can aid in making cities more resilient and sustainable.

Regarding the NCT of Delhi, evaluating the $PM_{2.5}$ removal capacity of urban vegetation is likely to provide a beneficial initial estimate for raising awareness in cities with elevated concentrations of $PM_{2.5}$. However, certain uncertainty exists in our study. First, the pollution removal capacity of the vegetation was only simulated during winter. As per research conducted by Abhijith et al.,[63] the dust-retaining capacity, deposition efficiency, and canopy morphology of plants would change to some extent in other seasons, which may differ from the results in winter. In another study conducted by Chaudhary and Rathore,[64] an analysis of the seasonal deposition efficiency of trees revealed that the highest dust deposition occurred during winter, followed by summer, with the lowest levels observed in the monsoon season. Second, the analysis in this study did not take into account the influence of meteorological factors such as wind speed and wind direction on the dispersion ability of trees. However, various studies have indicated that the alterations in wind direction and speed across different time intervals and heights influence the capacity of trees to capture $PM_{2.5}$.[63] Thirdly, the i-Tree Eco model ignored the resuspension process of $PM_{2.5}$. The movement of traffic can induce slight turbulence of pollutants close to the ground, facilitating the resuspension of $PM_{2.5}$; studies have identified that this process can reduce $PM_{2.5}$ levels, especially on two-way roads. Additionally, the calculation of $PM_{2.5}$ removal capacity was calculated based on the values of DBH of trees, overlooking other factors influencing the removal efficiency of tree species. This oversight could result in an underestimation of the actual deposition of pollutants. However, according to a study by Lin et al.,[65] the effect of small amounts of DBH measurement is negligible on the uncertainty of model output of i-Tree Eco at the landscape scale because the estimated degrees of input uncertainty arising from measurement errors are comparatively modest, leading to a limited influence on the uncertainty of the output.

There are several possible ways to reduce the uncertainty output of the model. One way of reducing uncertainty is by entering spatially distributed meteorological data into the model. However, that would require establishing more monitors, which can prove costly. Hence, overcoming the limitations poses a challenge, and model inputs will persistently depend on the most accurate local environmental data currently available. According to Lin et al.,[65] taking the inventory of entire street tree populations in individual tree management studies can help reduce the sampling error to zero. Furthermore, the sampling strategy can be modified from case to case depending on the area and/or the species being assessed. A significant hindrance to reducing uncertainty is the additional cost associated with reducing errors in the model inputs and sampling. Thus, although reducing uncertainty is crucial, it is essential to weigh the costs associated with it and evaluate whether the reduction is necessary. To summarize, for the reduction of uncertainty in subsequent studies, potential strategies involve the: (a) development of species-specific allometric relationships in cases where they are unavailable, (b) enhancement of the spatial accuracy of meteorological weather monitors, (c) subdivision of the study domain into subareas when multiple monitors are accessible for improved local meteorological estimates, and (d) optimization of sampling strategies to ensure representation of the urban forest diversity, balancing sampling intensities and data collection costs. The exploration of uncertainties in modeled attributes by i-Tree Eco remains a subject for further research.

To further enhance the role of urban vegetation as NbS for air quality management, certain areas need improvement. For example, introducing research briefs that extract findings and provide simple guidance to inform urban planners, multiple stakeholders, and policymakers will support the better implementation of urban planning. This can help stakeholders deliver NbS (in the form of urban vegetation) to improve air quality and public health with confidence. To ensure the selection and positioning of urban vegetation in the most efficient and effective way, the development of tools and services is necessary. Similar to the approach taken by Hewitt et al.,[14] a flowchart for decision-making in positive air quality interventions can aid management in making informed choices. Additionally, conducting a baseline study to consider all the benefits associated with urban vegetation in the early design stage can maximize its ability as an air purifier and health-promoting recreational resource. It is imperative

18630669, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/clen.202300198 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

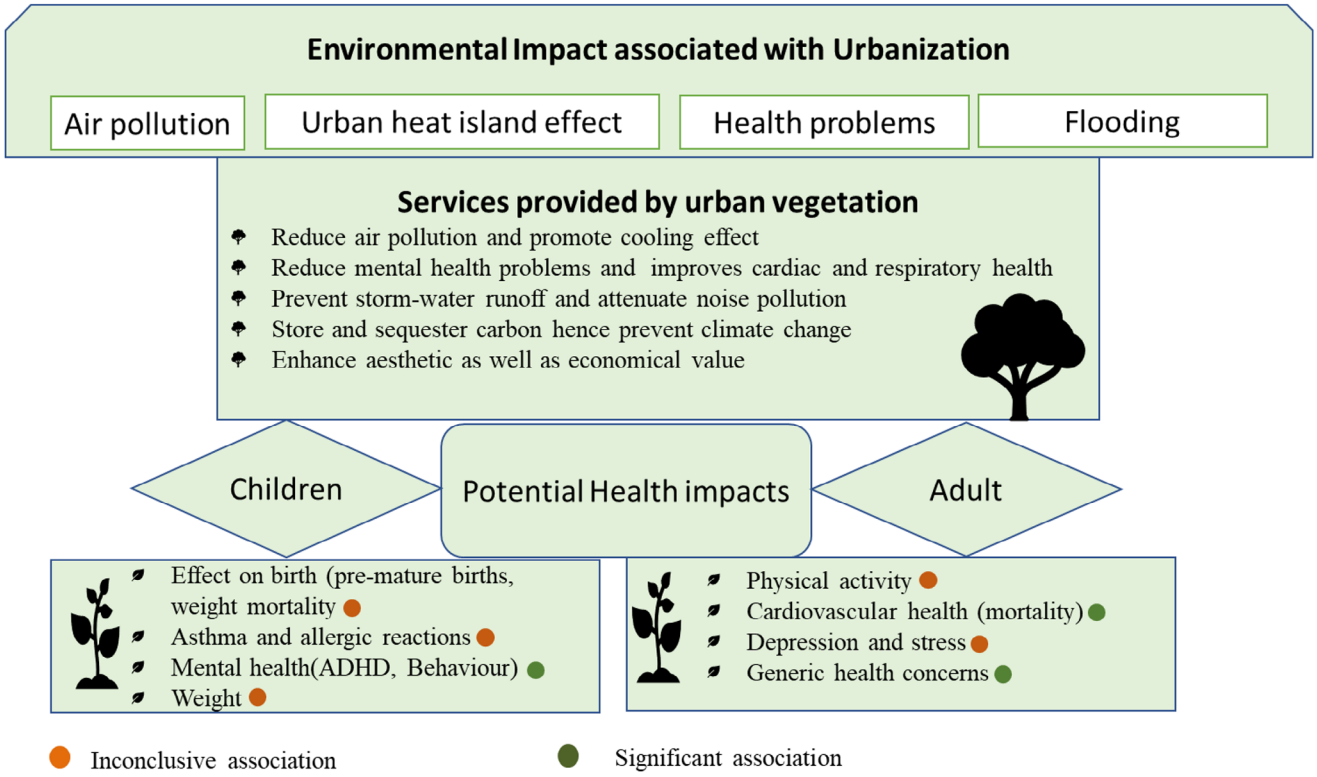


FIGURE 7 Role of urban vegetation in addressing urbanization challenge.

to continue researching the relationship between urban vegetation and air pollution to address existing scientific research gaps. Noteworthy studies by Deshmukh et al.[66] and Eisenman et al.[67] exemplify innovative research that enhances our understanding of how urban vegetation interacts with air pollutants. By documenting environmental factors that influence dispersion and plant-related variables that impact deposition, we can establish a more comprehensive evidence base to guide strategies for enhancing urban air quality. These elements, when combined, have the potential to reduce urban pollution. Furthermore, incorporating urban vegetation as an NbS into various disciplines such as economic development, educational programming, public health, and community building initiatives will enhance information exchange and facilitate the evaluation of direct and indirect health effects. Finally, it is essential to note that most studies incorporating NbS to urban problems come from high-income nations. Considering that future urban population growth is expected to occur primarily in low- and middle-income Asian and African nations, there is an urgent need to shift scientific attention to these regions.

5 | CONCLUSION

The present study focuses on quantifying the contribution of urban vegetation in reducing air pollution levels by measuring the dry deposition of PM on their leaves. This assessment has been conducted using a combination of real-time air quality monitoring and the i-Tree Eco model to evaluate the effectiveness of urban trees in mitigating air pollution in highly industrialized and polluted regions of NCT of Delhi. The findings of this study reveal that during rush hours when pollution was primarily attributed to traffic sources and vehicular emissions, the TA exhibited a significantly lower concentration of $PM_{2.5}$ compared to the NTS. However, no significant variation in $PM_{2.5}$ concentrations was observed outside of rush hours. Both the model-based approach and real-time measurements reveal that the presence of trees improves air quality. The preliminary results of this study provide valuable insights for informing national clean air strategies aimed at reducing pollutant emissions. However, conducting further research that compares additional pairs of TA and NTS across different seasons will facilitate a deeper understanding of how Delhi's trees influence air quality. The combination of real-time monitoring and modeling approaches demonstrated in this study serves the needs of various stakeholders, such as local governments, the general public, and nongovernmental organizations. Moreover, this research contributes to the paradigm shift from perceiving trees as costs and liabilities to recognizing their value and benefits, as observed through the i-Tree Eco model. This reinforces the rationale for investing in urban forests and supports decision-making processes. This study also provides evidence for integrating biodiversity in the sustainable planning of cities which can help achieve the goals outlined by the United Nations in its 2030 Agenda for Sustainable Development. Adopted in September 2015, the agenda comprises 17 Sustainable Development Goals (SDGs) aimed at enhancing global well-being through a balanced approach addressing social, environmental, and economic concerns, which strategically steer the world toward sustainability by 2030. The goal of reducing annual mean levels

18630669, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/clen.202300198 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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 PM in cities is addressed in Indicator 11.6.2 under SDG 11 (sustainable cities and communities), emphasizing the need to minimize PM concentration in cities due to its severe impact on public health. Additionally, the goal to integrate ecosystem and biodiversity values in government planning and policymaking of environmental issues is addressed in Target 15.9 of SDG 15 (Life on land). Therefore, urban vegetation can serve as a valuable resource for policymakers, researchers, and urban planners to develop effective strategies for managing air pollution and achieving SDG (11 and 15) aimed at urban resilience and sustainability.

AUTHOR CONTRIBUTIONS

Conceptualization; methodology; visualization; Writing—original draft; review and editing: Mallika Vashist. *Conceptualization; methodology; supervision and review*: Thangamani Vijaya Kumar. *Conceptualization; methodology; supervision; review and editing*: Santosh Kumar Singh.

ACKNOWLEDGMENTS

The authors express their gratitude to Delhi Technological University (DTU) for offering research facilities. The authors thank Mr. Harshit Chawla, a research scholar at DTU, for his valuable assistance throughout the study. The authors also thank the editor and anonymous reviewers whose valuable suggestions helped improve this manuscript.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest and financial or nonfinancial interests to disclose that are relevant to the content of this article.

FUNDING INFORMATION

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study is available from the corresponding author.

ORCID

Mallika Vashist https://orcid.org/0000-0003-3814-8150

REFERENCES

1. Gupta, A., Bherwani, H., Gautam, S., Anjum, S., Musugu, K., Kumar, N., Anshul, A., & Kumar, R. (2021). Air pollution aggravating COVID-19 lethality? Exploration in Asian cities using statistical models. *Environment, Development and Sustainability*, *23*, 6408–6417. https://doi.org/10.1007/s10668-020-00878-9A
2. Amsalu, E., Wang, T., Li, H., Liu, Y., Wang, A., Liu, X., Tao, L., Luo, Y., Zhang, F., & Yang, X. (2019). Acute effects of fine particulate matter (PM2.5) on hospital admissions for cardiovascular disease in Beijing, China: A time-series study. *Environmental Health*, *18*, 1–12. https://doi.org/10.1186/s12940-019-0506-2
3. Chawla, H., Singh, S. K., & Haritash, A. K. (2023). Reversing the damage: Ecological restoration of polluted water bodies affected by pollutants due to anthropogenic activities. *Environmental Science and Pollution Research*, *31*, 127–143. https://doi.org/10.1007/s11356-023-31295-w
4. Goyal, P., Gulia, S., & Goyal, S. K. (2021). Identification of air pollution hotspots in urban areas—An innovative approach using monitored concentrations data. *Science of The Total Environment*, *798*, 149143. https://doi.org/10.1016/j.scitotenv.2021.149143
5. Mainka, A., & Fantke, P. (2022). Preschool children health impacts from indoor exposure to PM2. 5 and metals. *Environment International*, *160*, 107062. https://doi.org/10.1016/j.envint.2021.107062
6. Neo, E. X., Hasikin, K., Mokhtar, M. I., Lai, K. W., Azizan, M. M., Razak, S. A., & Hizaddin, H. F. (2022). Towards integrated air pollution monitoring and health impact assessment using federated learning: A systematic review. *Frontiers in Public Health*, *10*, 851553. https://doi.org/10.3389/fpubh.2022.851553
7. Serafini, M. M., Maddalon, A., Iulini, M., & Galbiati, V. (2022). Air pollution: Possible interaction between the immune and nervous system? *International Journal of Environmental Research and Public Health*, *19*, 16037. https://doi.org/10.3390/ijerph192316037
8. Lu, J. G. (2020). Air pollution: A systematic review of its psychological, economic, and social effects. *Current Opinion in Psychology*, *32*, 52–65. https://doi.org/10.1016/j.copsyc.2019.06.024
9. Yadav, S. K., Mishra, R. K., & Gurjar, B. R. (2022). Assessment of the effect of the judicial prohibition on firecracker celebration at the Diwali festival on air quality in Delhi, India. *Environmental Science and Pollution Research*, *29*, 86247–86259. https://doi.org/10.1007/s11356-021-17695-w
10. Hama, S. M., Kumar, P., Harrison, R. M., Bloss, W. J., Khare, M., Mishra, S., Namdeo, A., Sokhi, R., Goodman, P., & Sharma, C. (2020). Four-year assessment of ambient particulate matter and trace gases in the Delhi-NCR region of India. *Sustainable Cities and Society*, *54*, 102003. https://doi.org/10.1016/j.scs.2019.102003
11. Sarma, R., & Singh, S. K. (2023). Assessment of groundwater quality and human health risks of nitrate and fluoride contamination in a rapidly urbanizing region of India. *Environmental Science and Pollution Research*, *30*, 55437–55454. https://doi.org/10.1007/s11356-023-26204-0
12. Singh, V., Singh, S., & Biswal, A. (2021). Exceedances and trends of particulate matter (PM2. 5) in five Indian megacities. *Science of the Total Environment*, *750*, 141461. https://doi.org/10.1016/j.scitotenv.2020.141461
13. Yadav, S. K., Mishra, R. K., & Gurjar, B. R. (2022). Ultrafine particle number concentration and its size distribution during Diwali festival in megacity Delhi, India: Are 'green crackers' safe? *Journal of Environmental Management*, *317*, 115459. https://doi.org/10.1016/j.jenvman.2022.115459
14. Hewitt, C. N., Ashworth, K., & MacKenzie, A. R. (2020). Using green infrastructure to improve urban air quality (GI4AQ). *Ambio*, *49*, 62–73. https://doi.org/10.1007/s13280-019-01164-3
15. Biswal, B. K., Bolan, N., Zhu, Y. G., & Balasubramanian, R. (2022). Nature-based Systems (NbS) for mitigation of stormwater and air pollution in urban areas: A review. *Resources, Conservation and Recycling*, *186*, 106578. https://doi.org/10.1016/j.resconrec.2022.106578
16. Kadaverugu, R., Dhyani, S., Dasgupta, R., Kumar, P., & Matli, C. (2022). Urban sustainability and resilience building: Blue-green infrastructure for air pollution abatement and realizing multiple co-benefits. In *Blue-green infrastructure across Asian countries: Improving urban resilience and sustainability*. Springer. https://doi.org/10.1007/978-981-16-7128-9_18
17. Kumar, R., Verma, V., Thakur, M., Singh, G., & Bhargava, B. (2023). A systematic review on mitigation of common indoor air pollutants using plant-based methods: A phytoremediation approach. *Air Quality, Atmosphere & Health*, *1*, 27. https://doi.org/10.1007/s11869-023-01326-z
18. Ysebaert, T., Koch, K., Samson, R., & Denys, S. (2021). Green walls for mitigating urban particulate matter pollution—A review. *Urban Forestry & Urban Greening*, *59*, 127014. https://doi.org/10.1016/j.ufug.2021.127014

18630669, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/clen.202300198 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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. Jeong, N. R., Han, S. W., & Ko, B. (2023). Effects of green network management of urban street trees on airborne particulate matter (PM2.5) concentration. *International Journal of Environmental Research and Public Health*, *20*, 2507. https://doi.org/10.3390/ijerph20032507
20. Vashist, M., Kumar, T. V., & Singh, S. K. (2024). A comprehensive review of urban vegetation as a nature-based solution for sustainable management of particulate matter in ambient air. *Environmental Science and Pollution Research*, *31*, 26480–26496. https://doi.org/10.1007/s11356-024-33089-0
21. Zhang, X., Lyu, J., Han, Y., Sun, N., Sun, W., Li, J., Liu, C., & Yin, S. (2020). Effects of the leaf functional traits of coniferous and broadleaved trees in subtropical monsoon regions on PM2.5 dry deposition velocities. *Environmental Pollution*, *265*, 114845. https://doi.org/10.1016/j.envpol.2020.114845
22. Zhang, X., Lyu, J., Zeng, Y., Sun, N., Liu, C., & Yin, S. (2021). Individual effects of trichomes and leaf morphology on PM2.5 dry deposition velocity: A variable-control approach using species from the same family or genus. *Environmental Pollution*, *272*, 116385. https://doi.org/10.1016/j.envpol.2020.116385
23. Grylls, T., & van Reeuwijk, M. (2022). How trees affect urban air quality: It depends on the source. *Atmospheric Environment*, *290*, 119275. https://doi.org/10.1016/j.atmosenv.2022.119275
24. Su, T. H., Lin, C. S., Lu, S. Y., Lin, J. C., Wang, H. H., & Liu, C. P. (2022). Effect of air quality improvement by urban parks on mitigating PM2. 5 and its associated heavy metals: A mobile-monitoring field study. *Journal of Environmental Management*, *323*, 116283. https://doi.org/10.1016/j.jenvman.2022.116283
25. Han, D., Shen, H., Duan, W., & Chen, L. (2020). A review on particulate matter removal capacity by urban forests at different scales. *Urban Forestry & Urban Greening*, *48*, 126565. https://doi.org/10.1016/j.ufug.2019.126565
26. Xue, F., & Li, X. (2017). The impact of roadside trees on traffic released PM10 in urban street canyon: Aerodynamic and deposition effects. *Sustainable Cities and Society*, *30*, 195–195. https://doi.org/10.1016/j.scs.2017.02.001
27. Riondato, E., Pilla, F., Basu, A. S., & Basu, B. (2020). Investigating the effect of trees on urban quality in Dublin by combining air monitoring with i-Tree Eco model. *Sustainable Cities and Society*, *61*, 102356. https://doi.org/10.1016/j.scs.2020.102356
28. Zhu, C., Przybysz, A., Chen, Y., Guo, H., Chen, Y., & Zeng, Y. (2019). Effect of spatial heterogeneity of plant communities on air PM10 and PM2.5 in an urban forest park in Wuhan, China. *Urban Forestry and Urban Greening*, *46*, 126487. https://doi.org/10.1016/j.ufug.2019.126487
29. Nowak, D. J., Hoehn, R. E., Bodine, A. R., Crane, D. E., Dwyer, J. F., Bonnewell, V., & Watson, G. (2013). *Urban trees and forests of the Chicago region*. United States Forest Service. https://doi.org/10.2737/NRS-RB-84
30. Amini Parsa, V., Salehi, E., & Yavari, A. (2020). Improving the provision of ecosystem services from urban forest by integrating the species' potential environmental functions in tree selecting process. *Landscape and Ecological Engineering*, *16*, 23–37. https://doi.org/10.1007/s11355-019-00401-x
31. Cimburova, Z., & Barton, D. N. (2020). The potential of geospatial analysis and Bayesian networks to enable i-Tree Eco assessment of existing tree inventories. *Urban Forestry & Urban Greening*, *55*, 126801. https://doi.org/10.1016/j.ufug.2020.126801
32. Song, P., Kim, G., Mayer, A., He, R., & Tian, G. (2020). Assessing the ecosystem services of various types of urban green spaces based on i-Tree Eco. *Sustainability*, *12*, 1630. https://doi.org/10.3390/su12041630
33. Ma, J., Li, X., Baoquan, J., Liu, X., Li, T., Zhang, W., & Liu, W. (2021). Spatial variation analysis of urban forest vegetation carbon storage and sequestration in built-up areas of Beijing based on i-Tree Eco and Kriging. *Urban Forestry & Urban Greening*, *66*, 127413. https://doi.org/10.1016/j.ufug.2021.127413
34. Xie, Y., Hirabayashi, S., Hashimoto, S., Shibata, S., & Kang, J. (2023). Exploring the spatial pattern of urban forest ecosystem services based on i-Tree eco and spatial interpolation: A case study of Kyoto City, Japan. *Environmental Management*, *72*, 991–1005. https://doi.org/10.1007/s00267-023-01847-4
35. United Nations. (2019). *World Urbanization Prospects 2018*. Department of Economic and Social Affairs, Population Division, United Nations. https://population.un.org/wup/Publications/Files/WUP2018-Highlights.pdf
36. Delhi Tourism and Transportation Development Corporation. Seasons of Delhi, Delhi tourism, Delhi, India. Delhi Tourism and Transportation Development Corporation. https://delhitourism.gov.in/delhitourism/aboutus/seasons_of_delhi.jsp
37. Garsa, K., Khan, A. A., Jindal, P., Middey, A., Luqman, N., Mohanty, H., & Tiwari, S. (2023). Assessment of meteorological parameters on air pollution variability over Delhi. *Environmental Monitoring and Assessment*, *195*, 1315. https://doi.org/10.1007/s10661-023-11922-2
38. Saxena, A., & Raj, S. (2021). Impact of lockdown during COVID-19 pandemic on the air quality of North Indian cities. *Urban Climate*, *35*, 100754. https://doi.org/10.1016/j.uclim.2020.100754
39. Chen, Y., Wild, O., Conibear, L., Ran, L., He, J., Wang, L., & Wang, Y. (2020). Local characteristics of and exposure to fine particulate matter (PM2. 5) in four Indian megacities. *Atmospheric Environment: X*, *5*, 100052. https://doi.org/10.1016/j.aeaoa.2019.100052
40. Jain, S., Sharma, S. K., Vijayan, N., & Mandal, T. K. (2020). Seasonal characteristics of aerosols (PM2.5 and PM10) and their source apportionment using PMF: A four year study over Delhi, India. *Environmental Pollution*, *262*, 114337. https://doi.org/10.1016/j.envpol.2020.114337
41. Guttikunda, S. K., & Calori, G. (2013). A GIS based emissions inventory at 1 km× 1 km spatial resolution for air pollution analysis in Delhi, India. *Atmospheric Environment*, *67101*, 101–111. https://doi.org/10.1016/j.atmosenv.2012.10.040
42. Squadrito, G. L., Cueto, R., Dellinger, B., & Pryor, W. A. (2001). Quinoid redox cycling as a mechanism for sustained free radical generation by inhaled airborne particulate matter. *Free Radical Biology and Medicine*, *31*, 1132. https://doi.org/10.1016/S0891-5849(01)00703-1
43. Nowak, D. J., & Crane, D. E. (2000). *The urban forest effects (UFORE) model: Quantifying urban forest structure and functions*. USDA Forest Service.
44. Murthy, B. S., Latha, R., Tiwari, A., Rathod, A., Singh, S., & Beig, G. (2020). Impact of mixing layer height on air quality in winter. *Journal of Atmospheric and Solar-Terrestrial Physics*, *197*, 105157. https://doi.org/10.1016/j.jastp.2019.105157
45. Madhwal, S., Tripathi, S. N., Bergin, M. H., Bhave, P., de Foy, B., Reddy, T. R., Chaudhry, S. K., Jain, V., Garg, N., & Lalwani, P. (2024). Evaluation of PM2. 5 spatio-temporal variability and hotspot formation using low-cost sensors across urban-rural landscape in Lucknow, India. *Atmospheric Environment*, *319*, 120302. https://doi.org/10.1016/j.atmosenv.2023.120302
46. US Department of Agriculture, Forest Service. (2021). i-Tree Eco user's manual. v.6. US Department of Agriculture, Forest Service. https://www.itreetools.org/documents/275/EcoV6_UsersManual.2021.09.22.pdf
47. Hirabayashi, S. (2015). *i-Tree Eco United States county-based hydrologic estimates*. US Department of Agriculture, Forest Service.
48. Mehta, D., Hazarika, N., & Srivastava, A. (2020). Diurnal variation of BTEX at road traffic intersection points in Delhi, India: source, ozone formation potential, and health risk assessment. *Environmental Science and Pollution Research*, *27*, 11093–117902. https://doi.org/10.1016/j.atmosenv.2020.117902
49. Sharma, D., Soni, V. K., & Pradhan, S. (2022). Assessment of variation in air quality over Delhi and neighbouring cities of Noida and

18630669, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/clen.202300198 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

Greater Noida. *Journal of Earth System Science*, *13*, 155. https://doi.org/10.1007/s12040-022-01902-47

50. Wu, J., Wang, Y., Qiu, S., & Peng, J. (2019). Using the modified i-Tree Eco model to quantify air pollution removal by urban vegetation. *Science of the Total Environment*, *688*, 673–683. https://doi.org/10.1016/j.scitotenv.2019.05.437
51. Delhi Pollution Control Committee (DPCC). Real time ambient air quality data. Delhi Pollution Control Committee. http://www.dpccairdata.com/dpccairdata/display/index.php
52. Gama, C., Monteiro, A., Pio, C., Miranda, A. I., Baldasano, J. M., & Tchepel, O. (2018). Temporal patterns and trends of particulate matter over Portugal: A long-term analysis of background concentrations. *Air Quality, Atmosphere & Health*, *11*, 397. https://doi.org/10.1007/s11869-018-0546-8
53. Ander, E. L., Johnson, C. C., Cave, M. R., Palumbo-Roe, B., Nathanail, C. P., & Lark, R. M. (2013). Methodology for the determination of normal background concentrations of contaminants in English soil. *Science of the Total Environment*, *454*, 604–618. https://doi.org/10.1016/j.scitotenv.2013.03.005
54. Ozdemir, H. (2019). Mitigation impact of roadside trees on fine particle pollution. *Science of the Total Environment*, *659*, 1176–1185. https://doi.org/10.1016/j.scitotenv.2018.12.262
55. Corada, K., Woodward, H., Alaraj, H., & Collins, C. M., & de Nazelle, A. (2021). A systematic review of the leaf traits considered to contribute to removal of airborne particulate matter pollution in urban areas. *Environmental Pollution*, *269*, 116104. https://doi.org/10.1016/j.envpol.2020.116104
56. Song, Y., Maher, B. A., Li, F., Wang, X., Sun, X., & Zhang, H. (2015). Particulate matter deposited on leaf of five evergreen species in Beijing, China: Source identification and size distribution. *Atmospheric Environment*, *105*, 53–60. https://doi.org/10.1016/j.atmosenv.2015.01.032
57. Vashist, M., & Singh, S. K. (2023). Potential of individual leaf traits and leaf-associated microorganisms in the removal of particulate matter. In *Recent developments in energy and environmental engineering: Select proceedings of TRACE 2022*. Springer Nature. https://doi.org/10.1007/978-981-99-1388-6_16
58. Chen, L., Liu, C., Zhang, L., Zou, R., & Zhang, Z. (2017). Variation in tree species ability to capture and retain airborne fine particulate matter (PM2. 5). *Scientific Reports*, *7*, 3206. https://doi.org/10.1038/s41598-017-03360-1
59. Nowak, D. J., Hirabayashi, S., Bodine, A., & Hoehn, R. (2013). Modeled PM2. 5 removal by trees in ten US cities and associated health effects. *Environmental Pollution*, *178*, 395–402. https://doi.org/10.1016/j.envpol.2013.03.050
60. Lee, K., & Greenstone, M. (2021). *Annual update 2021*, *Air quality life index*. The Energy Policy Institute at the University of Chicago. https://aqli.epic.uchicago.edu/wp-content/uploads/2021/08/AQLI_2021-Report.EnglishGlobal.pdf
61. Central Pollution Control Committee (CPCB). (2017). Graded response action plan for Delhi &NCR. Central Pollution Control Board, Ministry of Environment, Forest & Climate Change (Government of India). https://cpcb.nic.in/graded-response-action-plan-for-delhi-ncr/
62. Triguero-Mas, M., Dadvand, P., Cirach, M., Martínez, D., Medina, A., Mompart, A., Basagaña, X., Gražulevičienė, R., & Nieuwenhuijsen, M. J. (2015). Natural outdoor environments and mental and physical health: Relationships and mechanisms. *Environment International*, *77*, 35–41. https://doi.org/10.1016/j.envint.2015.01.012
63. Abhijith, K. V., Kumar, P., Gallagher, J., McNabola, A., Baldauf, R., Pilla, F., Broderick, B., Di Sabatino, S., & Pulvirenti, B. (2017). Air pollution abatement performances of green infrastructure in open road and built-up street canyon environments—A review. *Atmospheric Environment*, *162*, 71–86. https://doi.org/10.1016/j.atmosenv.2017.05.014
64. Chaudhary, I. J., & Rathore, D. (2018). Suspended particulate matter deposition and its impact on urban trees. *Atmospheric Pollution Research*, *9*, 1072. https://doi.org/10.1016/j.apr.2018.04.006
65. Lin, X., Chamecki, M., & Yu, X. (2020). Aerodynamic and deposition effects of street trees on PM2.5 concentration: From street to neighborhood scale. *Building and Environment*, *185*, 107291. https://doi.org/10.1016/j.buildenv.2020.107291
66. Deshmukh, P., Isakov, V., Venkatram, A., Yang, B., Zhang, K. M., Logan, R., & Baldauf, R. (2019). The effects of roadside vegetation characteristics on local, near-road air quality. *Air Quality, Atmosphere & Health*, *12*, 259–270. https://doi.org/10.1007/s11869-018-0651-8
67. Eisenman, T. S., Churkina, G., Jariwala, S. P., Kumar, P., Lovasi, G. S., Pataki, D. E., Weinberger, K. R., & Whitlow, T. H. (2019). Urban trees, air quality, and asthma: An interdisciplinary review. *Landscape and Urban Planning*, *187*, 47–59. https://doi.org/10.1016/j.landurbplan.2019.02.010

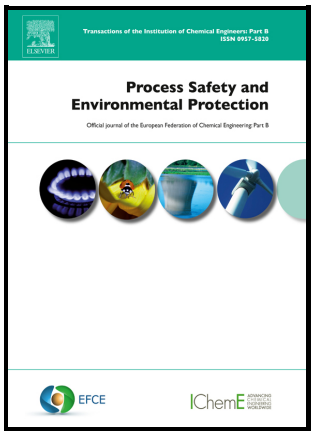
How to cite this article: Vashist, M., Kumar, T. V., & Singh, S. K. (2024). Assessment of air quality benefits of vegetation in an urban-industrial region of India by integrating air monitoring with i-Tree Eco model. *CLEAN – Soil, Air, Water*, e2300198. https://doi.org/10.1002/clen.202300198

18630669, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/clen.202300198 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

Journal Pre-proof

Assessment of Solar-Biomass Hybrid Power System for Decarbonizing and Sustainable Energy Transition for Academic Building

Neelam Baghel, K. Manjunath, Anil Kumar



PII: S0957-5820(24)00516-0

DOI: https://doi.org/10.1016/j.psep.2024.05.004

Reference: PSEP5679

To appear in: *Process Safety and Environmental Protection*

Received date: 15 October 2023
Revised date: 4 December 2023
Accepted date: 2 May 2024

Please cite this article as: Neelam Baghel, K. Manjunath and Anil Kumar, Assessment of Solar-Biomass Hybrid Power System for Decarbonizing and Sustainable Energy Transition for Academic Building, *Process Safety and Environmental Protection,* (2024) doi:https://doi.org/10.1016/j.psep.2024.05.004

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier.

Assessment of Solar-Biomass Hybrid Power System for Decarbonizing and Sustainable Energy Transition for Academic Building

Neelam Baghel[1], Manjunath K.[1] and Anil Kumar[1, 2,*]

[1]Department of Mechanical Engineering, Delhi Technological University, Delhi-110042 (India)

[2]Centre for Energy and Environment, Delhi Technological University, Delhi-110042 (India)

ORCID ID: 0000-0002-2846-4347 (Anil Kumar)

*Corresponding Author: anilkumar76@dtu.ac.in (Anil Kumar)

Contact No: +91-9425680448

ORCID ID:

Neelam Baghel: 0000-0002-8076-6772

Manjunath K.: 0000-0001-7732-954X

Anil Kumar: 0000-0002-2846-4347

Journal Pre-proof

Abstract

A solar-biomass hybrid power system may be a feasible approach for decarbonizing and achieving a sustainable energy transition in an academic building. This cutting-edge technology generates clean, sustainable power by fusing the advantages of biomass and solar energy. The present research analyses and optimises a Hybrid Renewable Energy System (HRES) for the Department of Mechanical Engineering, Delhi Technological University (located at 28°44' N and 77°06' E). Hybrid optimization model for electric renewable (HOMER) software is employed for simulation, and analysis primarily focuses on the techno-environomic assessment of the system. HRES is designed to fulfil daily energy consumption of 588 kWh at a peak load of 65 kW. System generates 3,76,780 kWh of power every year. The cost of energy (COE) is ₹16.96 (US$ 0.207)/kWh and the gross net present cost (NPC) is ₹41,603,969.8 (US$ 5,07,737). Throughout lifetime (25 years),

biomass gasifier and solar PV contribute 23.4% and 76.6% of the total energy production. Around 161 metric tons of CO_2 is prevented from entering the atmosphere. Therefore, proposed HRES is technically consistent, dependable, economically viable, and environmentally sustainable.

Keywords:

Hybrid; Renewable energy; Feasibility analysis; Environmental sustainability; HOMER

Journal Pre-proof

1. Introduction

Energy consumption is directly proportional to the advent of manpower with the rising population, improving humanity's standard of living and the economic prosperity of emerging countries. With the rise in demand for energy, there is a subsequent need to increase and diversify energy production and energy sources (Ahmet Bozgeyik, Lutfiye Altay, 2023). Global warming is increasing greenhouse gas (GHG) emissions, and the depletion of non-renewable energy sources has raised interest in protecting the world's energy supplies (Razaghi et al., 2024). However, according to World Economic Outlook 2020, 940 million people, or 13% of the global population face shortage of grid power or other energy sources ("Access to Energy," n.d.). However, as a result of urbanization and population growth, electricity demands are increasing. Therefore, more energy is needed when the population grows, which comes from fossil fuels but also generates emissions. Utilizing cutting-edge, suitable machinery and sustainable energy sources in rural regions is a very efficient approach to reducing fossil fuel use (Mobin Korpeh, Poorya Asadbagi et al., 2023). The main issue with non-grid systems is the unpredictability of the electrical supply, which may be addressed by combining renewable and conventional technologies. Over the past ten years, research into HRESs has drawn significant attention in the Middle East, Asia, and Africa as a result of rising electricity costs, rural electrification, and rapid global climate change. Ghenai et al. investigated renewable resources in Sharjah city to find the best arrangement for meeting the city's intended power loads. In Sharjah city, a hybrid system may meet up to 14% of the entire yearly electrical demand, with photovoltaic (PV) panels contributed for 74% and a biogas generator for 26% of the total electrical demand (Ghenai and Janajreh, 2016). Kumar et al.

analyzed that in Tamil Nadu, India, there are six separate temperature zones. The best climate zone to implement this system is the arid interior (Kanyakumari) and findings of the sensitivity analysis indicate that ₹1854 (₹49/lit) was saved as NPC for the arid interior climate zone (Suresh Kumar and Manoharan, 2014). Shafiqur et al. examined actual hybrid systems worldwide in various climatic situations to discover the optimal hybrid system configurations for Saudi Arabia to enhance its current energy spectrum and found that Wind-PV HPS with battery storage is the best option (Rehman, 2021). Al-Najideen et al. analysed that 96 MWh/year of energy is consumed by Engineering Faculty at Mu'tah University. The system produces 97.02 MWh of electricity to the grid yearly through installing an on-grid solar PV system of 56.7 kW and the payback period of the system is 5.5 years (Al-Najideen and Alrwashdeh, 2017). Kavuma et al. examined the practicality of grid-connected solar PV systems in Uganda by analyzing the system performance that was already in operation. An estimated 69.52 GWh of energy was produced annually. The capacity factor varies from 13.1% to 17.5% and the performance ratio is 0.76, both within the intended ranges (Kavuma et al., 2022). Kamali et al. analysed the environmental and economic parameters of solar PV systems connected to the Malaysian electrical grid system (Kamali and Mekhilef, 2009). Goyal et al. evaluated the economic feasibility of concentrated solar power plants in Jodhpur using three distinct techniques: Solar Power Tower (SPT), Parabolic Trough Collector (PTC), and Linear Fresnel Reflector (LFR) of 50 MW and 100 MW capacity. A 50 MW Parabolic Trough Collector plant with storage has the highest NPV of ₹6940.4 million whereas the 50 MW Linear Fresnel Reflector plant without thermal storage has the lowest calculated Levelized Cost of Energy (LCOE) ₹9.27/kWh (Goyal et al., 2022). Kumar et al. analysed a 95 kW on-grid solar PV system. The maximum estimated annual specific energy generation from PVWatts is 1612.968 kWh/kWp compared to 1406.632 kWh/kWp for PVGIS. PVWatts and PVGIS both estimate capacity factors were 16.05% and 18.41%, correspondingly. An insignificant variation was perceived in the performance among PVGIS, PVWatts, and investigational data (Kumar et al., 2018). Sharma et al. did parametric analysis as well as solar photovoltaic system modelling based on varied load situations in a building to obtain performance ratio, maximum power, and improved system efficiency (Sharma et al., 2018). Wibowo et al. demonstrated that the stationary panel produces more energy, with an average final output of 3.3% and a capacity factor of 13.6% compared to 2.3% and 9.6% for PV panels with solar trackers. Due to the daily average maximum power produced between 11:00 AM and 2:00 PM, Palembang does not require solar trackers for

Journal Pre-proof

PV panels (Wibowo et al., 2021). Saha et al. evaluated the efficacy of the proposed HRES system in contrast to the conventional utility grid expansion. It was affordable, cost-effective, and environmentally safe and made for a demand of 189.26 kWh/d and a peak load of 61.78 kW, and it generates power using a solar photovoltaic and a wind turbine system. The micro-grid concept produces 1,70,872 kWh of extra energy each year (Saha et al., 2022). Al-Ghussain et al. determined the ideal size of a HRES with and without storage. The hybrid system saved the university US$ 1.37×10^6 in annual electricity costs and has a RES component of 99.59%, Demand Supply Fraction (DSF) of 98.86%, and an electricity cost of 0.1626 US$/kWh based on increasing the DSF and renewable energy percentage with a net value larger than or equal to zero (Al-Ghussain et al., 2021). Wang et al. recommended a variety of energy storage solutions used for renewable energy to address the fluctuation and unpredictability issue (Wang et al., 2022). Sen et al. highlighted the need for residential, institutional, commercial, agricultural, and small-scale industrial development. The best off-grid option is identified and evaluated against conventional grid expansion (Sen and Bhattacharyya, 2014). Koussa et al. discussed the principal advantage of photovoltaic/wind/diesel hybrid with battery storage, it enhanced the system's reliability (Saheb-Koussa et al., 2009). Zhou et al. analyzed stand-alone hybrid solar-wind power generation systems which were acknowledged as a feasible alternative to grid supply. Such hybrid power generation systems are more appropriate than systems with only one renewable energy source for electricity supply to off-grid applications (Sayed et al., 2021). Sheikh et al. explored the thermal control of a solar cell that was being cooled by a phase-change material. When phase change material is used for cooling, electrical energy efficiency rises by 4.82, 8.1, and 7.17%, respectively, while power output rises by 5.12, 8.4, and 7.29% (Sheik et al., 2022). Singh et al. estimated COE of a biomass gasifier generator set, solar PV, and fuel cell crossover energy system is 15.064 ₹/kWh, with a total net current cost of ₹51,89,003 and the suggested framework's plentiful power is 36 kWh/year with zero tariffs unmet electrical load (Singh and Baredar, 2016). Turkay et al. discussed the economics of standalone and grid-connected HRESs. Results reveal that grid-connected HRES are more likely to adapt than standalone configurations (Türkay and Telli, 2011). Decentralized green energy is becoming more feasible to supply electricity to one billion people where no admittance to the power grid (Sawle et al., 2016). While government regulations cut back on emissions of agricultural waste, solar energy prices are declining. In India, more than 70% of the population resides in isolated areas deprived of access to grid-connected power (Krishnamoorthy et al., 2020).

Journal Pre-proof

A major limitation in HRESs is a lack of a fixed optimal structure of the HRES. Simulation and modelling are based on the ease of access to traditional and non-traditional sources, existing base structures, cost of building and manufacturing, etc. As a result, proper designing of HRES needs a comprehensive survey of the planned project location and resources in addition to the renewable energy sources.

Academic institutions play an important role in demonstrating sustainable practices. Installing a solar-biomass hybrid power system demonstrates a building's dedication to sustainability and promotes similar initiatives. Reducing dependency on fossil fuels and GHG emissions can be achieved by integrating solar and biomass power. Students and researchers can analyze the integration of various renewable sources, comprehend their efficiency, and investigate breakthroughs in sustainable energy technology when an academic building has a hybrid system. Institutions can directly support decarbonization and sustainable energy transitions while also promoting research, education, and a community-wide sustainability culture by installing a solar-biomass hybrid power system in an academic building.

The literature review reveals that some techno-environomic analysis has been carried out, however, meteorological data based on an academic institute, for optimizing the HRES in Delhi, India is not present. The literature review, meanwhile, also does not mention how HRESs are prioritized in terms of energy output, NPC, COE, operational cost, and carbon emissions for an academic institute. An abundant amount of solar energy and adequate biomass is present at the site location. Hence, the main focus is to use more than one renewable energy source to provide a reliable source of power instead of the traditional form of electricity. The present work aims to tackle the proposed system's environmental assessment to ensure sustainability. Generally, research on HRESs involves a technical and economic assessment to determine the system's feasibility. The viability and sustainability of the proposed system will be thoroughly determined through technical, economic, and environmental studies.

Along with the system's technical and economic feasibility, it is also crucial to understand the extent of the positive environmental impact it can provide, which can be seen very significantly in this system. The system uses a combination of renewable energy sources like solar energy and biogas energy to provide a consistent source of power instead of the traditional form of electricity. The analysis is done using HOMER, to assess the proposed system i.e. technically consistent,

Journal Pre-proof

dependable, economically viable and environmentally sustainable. To achieve this, the system is analysed first, assessing the available resources; second choosing system parameters, third, assessing the load profile, and fourth analyzing the techno-environomic aspects of the system.

2. Methodology

In this HRES, solar PV along with a biomass gasifier generator and batteries has been used for energy storage. Figure 1 represents a schematic diagram of power generation using a PV array, biogas genset, battery and power converter. Biomass is burned or utilized after transforming it into various forms of solid, liquid, and gaseous fuels, CO_2 is emitted into the atmosphere that the biomass had recently taken from the atmosphere during its photosynthesis process. As a result, there is no net CO_2 increase (Abbasi and Abbasi, 2010). The system should have low or zero carbon emissions and be easily integrated with current technologies or grid and in the end, the cost of the system should be optimal (Azar et al., 2006).

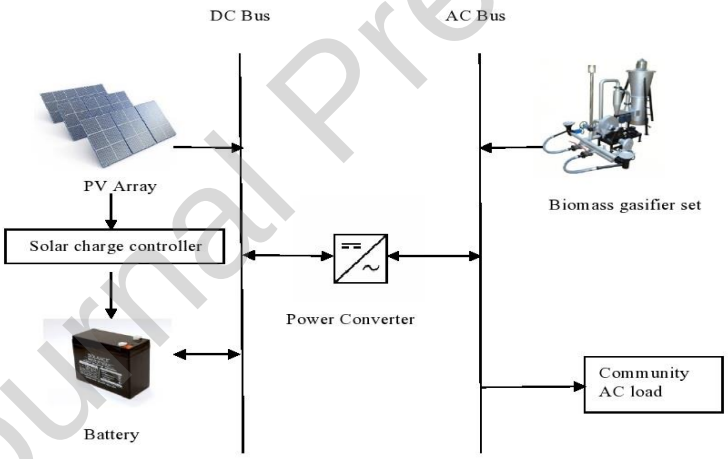


Figure 1: Hybrid renewable energy system

Proposed system is an amalgam of solar PV power sources with other power sources. The system is set to be a standalone system. It is usually advantageous in remote or inaccessible areas. Components of this system include a biomass generator set, solar PV energy generator, battery and a converter. Figure 2 shows a brief procedure to develop, simulate, and optimize the HRES using HOMER.

Journal Pre-proof

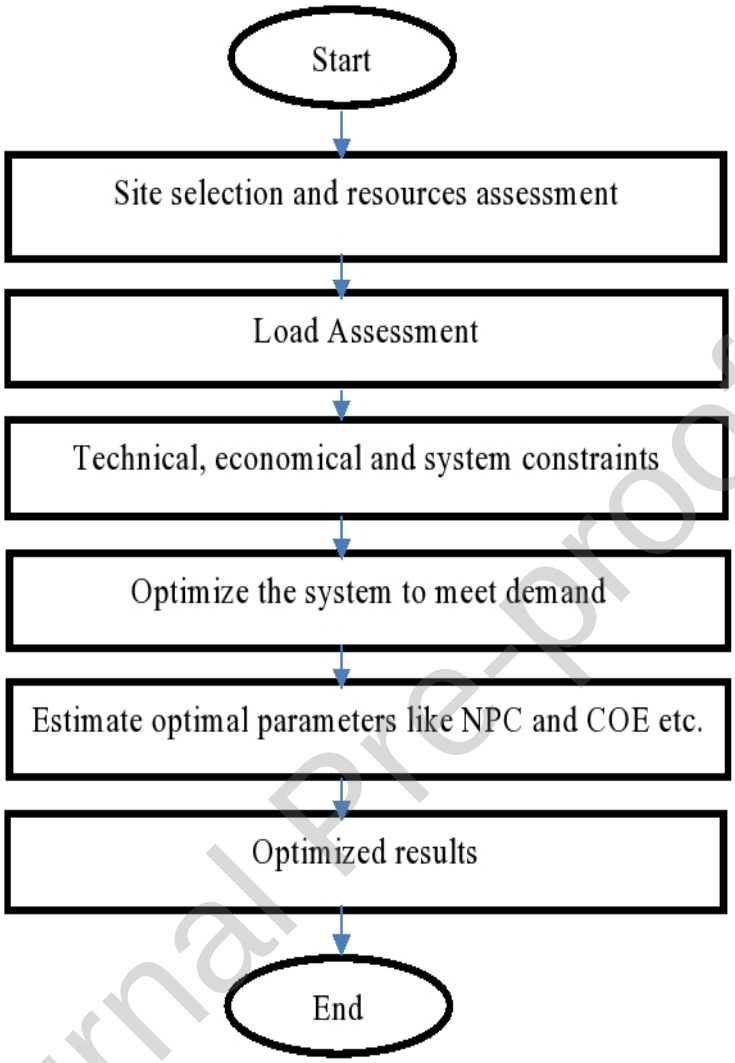


Figure 2: Flowchart to design and optimise HRES

Journal Pre-proof

2.1 Components detail

HRES comprises solar PV, biomass energy sources, a converter and a battery to store the generated power. Optimization is of utmost importance in determining the optimal system's size and cost. It can be an indication of a good ratio between the system's cost and its performance.

Solar photovoltaic system

Sunlight is converted into direct current power by solar PV cells. The cells consist of several varieties of semiconductor materials. Solar cell technology, which includes positive and negative charges, is designed with a low cost and a high conversion efficiency rate (Shaikh, 2017).

Photovoltaic system output is mainly dependent on incident solar radiation. Furthermore, the output is in terms of voltage. Hence the power generated is determined by the product of current and voltage. A 1 kW solar photovoltaic system and maintenance costs are nearly ₹1,50,000. The lifetime of this system is approximately twenty years and does not include any tracking system (Singh and Baredar, 2016).

The output of the photovoltaics in HOMER is given by Eq. (1):

$$P_{PV} = A_{PV}\, f_{PV}\, (R_T / R_{T,STC})\, (1 + a_P(T_{cell} - T_{cell,STC})) \quad (1)$$

If the option chosen is not to model the coefficient of temperature, then the simplified equation is given by Eq. (2):

$$P_{PV} = A_{PV}\, f_{PV}\, (R_T / R_{T,STC}) \quad (2)$$

PV system Output is obtained from the product of the current and voltage. Eq. (3) represents the energy distributed by the panel in the system (Kazem and Khatib, 2013).

$$P_{PV}Output = P_{NPV} \times R / R_{ref} \times (1 + K_T(T_{cell} - T_{ref}) \quad (3)$$

Biomass gasifier generator set

When an organic resource is burnt at an extremely high temperature and combined with a limited flow of air, the structure of the source breaks down into a simpler form. This organic material is heated with a limited quantity of air in a biomass gasifier to propagate incomplete combustion, producing a mix of highly combustible gases. Complete combustion of biomass resources will produce N_2, H_2O, CO_2, and excess O_2. Since incomplete combustion occurs in the gasifier, resulting gases are CO, H_2 and minimal methane gas (Farret and Simões, 2006). Collected food waste in the campus is used as biomass to produce biogas and the LHV of biogas is assumed 5.50 MJ/kg. Biogas/fossil ratio is 8.50%. The density of syngas is about 0.72 kg/m^3, then the gasification ratio is calculated at 0.70 kg/kg biomass.

The electrical efficiency of the system is defined by Eq. (4) and Eq. (5) (Singh and Baredar, 2016):

$$\eta_{elec} = (P_o - P_n) / (Input\ Biomass)_{LHV} \quad (4)$$

$$\eta_{elec} = (P_{tot}) / (Input\ Biomass)_{LHV} \quad (5)$$

Here, $P_{tot} = P_o - P_n$

Journal Pre-proof

Power converter

In this HRES, the power converter is of utmost importance. An electronic converter aims to maintain a continuous stream of power between the component's DC and AC power generated. HOMER assumes that the rectifier and inverter capacity do not exceed what the device can handle for an extended period. Instead, they are considered continuous capacities, the apparatus can resist for the required period (Farret and Simões, 2006). With research about present market rates, we anticipate the approximate capital cost to be ₹15,000, with maintenance and upkeep costs at ₹10,000. The approximate lifespan of a converter is 20 years and it maintains nearly a 90% efficiency level (Givaki et al., 2014).

Battery

A battery storage bank is used to reinforce the system. Similarly, a battery is used to keep the voltage constant over the electrical component. Common usage of the battery bank is for storing the electrical power generated in a highly effective. Its discharging capacity is unable to exceed a base breaking point. Battery bank's storage capacity (C_{Wh}) is determined by Eq. (6) (Malheiro et al., 2015):

$$C_{Wh} = (E_{tot} \times \mathrm{AD})\eta_{in} \times \eta_{batt} \times DOD_{batt} \quad (6)$$

The software performs calculations to determine the maximum amount of electricity that can be stored in the storage bank for each time step. The maximum quantity of electricity stored is used to figure out whether the storage bank can take in extra renewable energy or the extra power needed to be generated by a generator. This maximum charge power fluctuates in between steps due to the varying state of charge and previous discharge ("Homer," n.d.). The maximum amount of power that the two-tank system can withstand is provided by Eq. (7):

$$P_{batt} = (rE_1\, exp^{-r\Delta t} + Qrs\,(1 - exp^{-r\Delta t})) / (1 - exp^{-r\Delta t} + s\,(r\Delta t - 1 + exp^{-r\Delta t}) \quad (7)$$

Another limitation imposed by HOMER is the maximum charge rate. Battery storage bank power related to it is given by Eq. (8):

$$P_{batt} = (1 - exp^{-\alpha\Delta t})\,(E_{max} - E)/\Delta t \quad (8)$$

The last limitation imposed by HOMER is the maximum charge of the storage component. The

Journal Pre-proof

formula for maximum storage bank charge power for this charge current is given by Eq. (9):

$$P_{batt} = N_{batt} I_m V_n / 10 \quad (9)$$

In HOMER, the least of the previous three values is used as the maximum storage charge power after taking into account losses due to charging, leading to Eq. (10):

$$P_{batt,cmax} = MIN\,(P_{batt,cmax,kbm}, P_{batt,cmax,mcr}, P_{batt,cmax,mcc}) / \eta_{batt,c} \quad (10)$$

Capital cost ₹5,110,000 substitution cost ₹5,036,270 after 15 years and battery life is estimated to be 15 years in this project (da Silva Lima et al., 2021).

2.2 Site selection and resource assessment

Figure 3 shows the site of the study area i.e., the Department of Mechanical Engineering, Delhi Technological University, Delhi. DTU is situated in the North-West of Delhi with a 165-acre campus. It is a large campus comprising many hostels, residential buildings, educational buildings, biogas, waste plant, etc. Hence sufficient raw material is available daily to produce the required biomass. NASA Surface Meteorology used to perform irradiation at site location 28°44'N latitude and 77°06'E longitude.

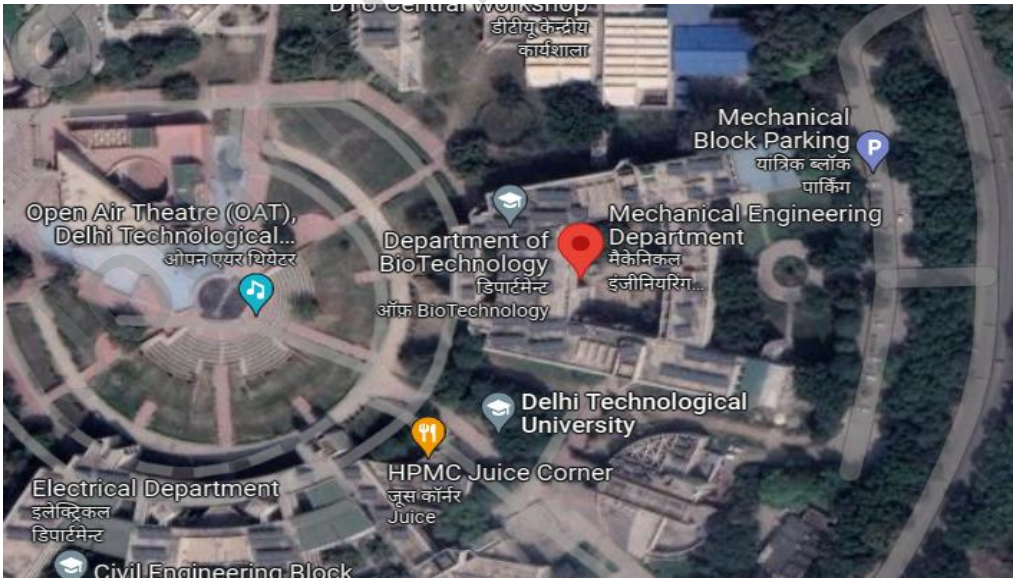
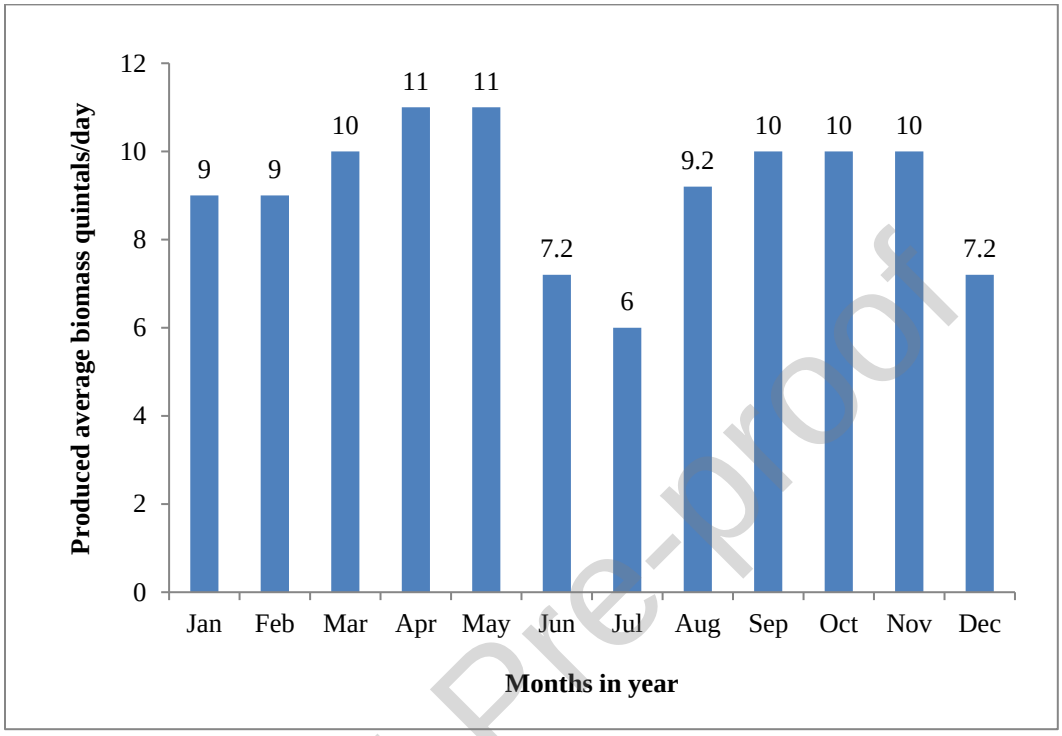


Figure 3: Location of the study area, Mechanical Engineering Department, DTU, Delhi (India)

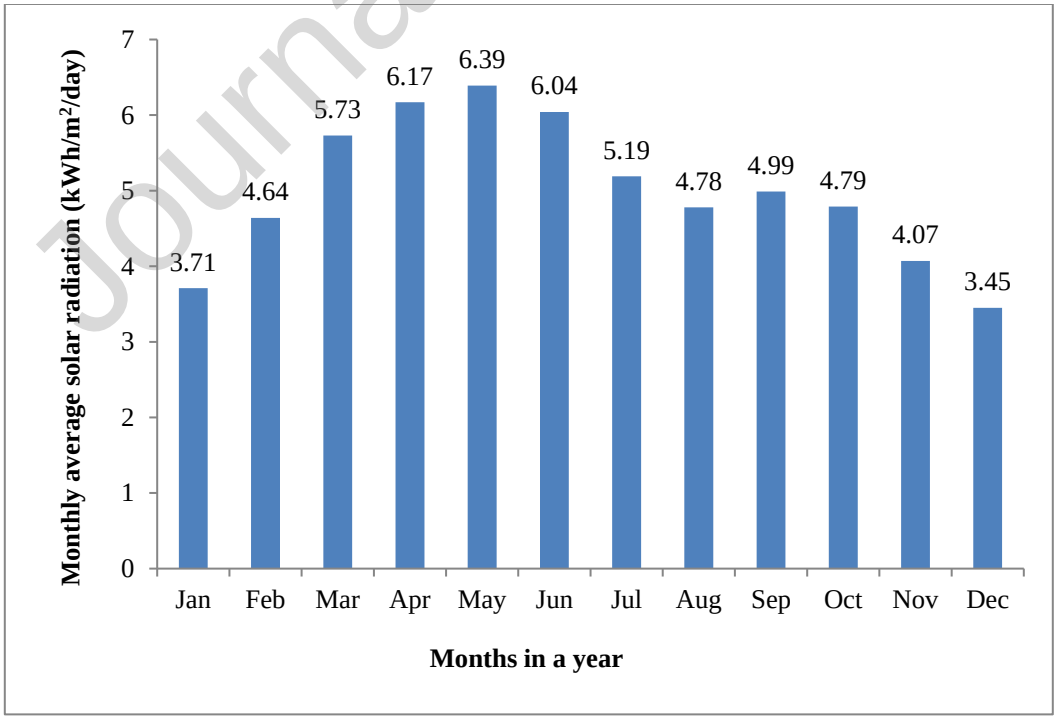
Figure 4(A) shows biomass resources available for power generation of the proposed system. The average year's biomass production amount is 9.08 quintals/day. Figure 4(B) illustrates the monthly

Pre-proof

average solar radiation. Yearly average solar power can be scaled up to 4.99 kWh/m^2 per day, which is considerable.



(A) Monthly average biomass resources

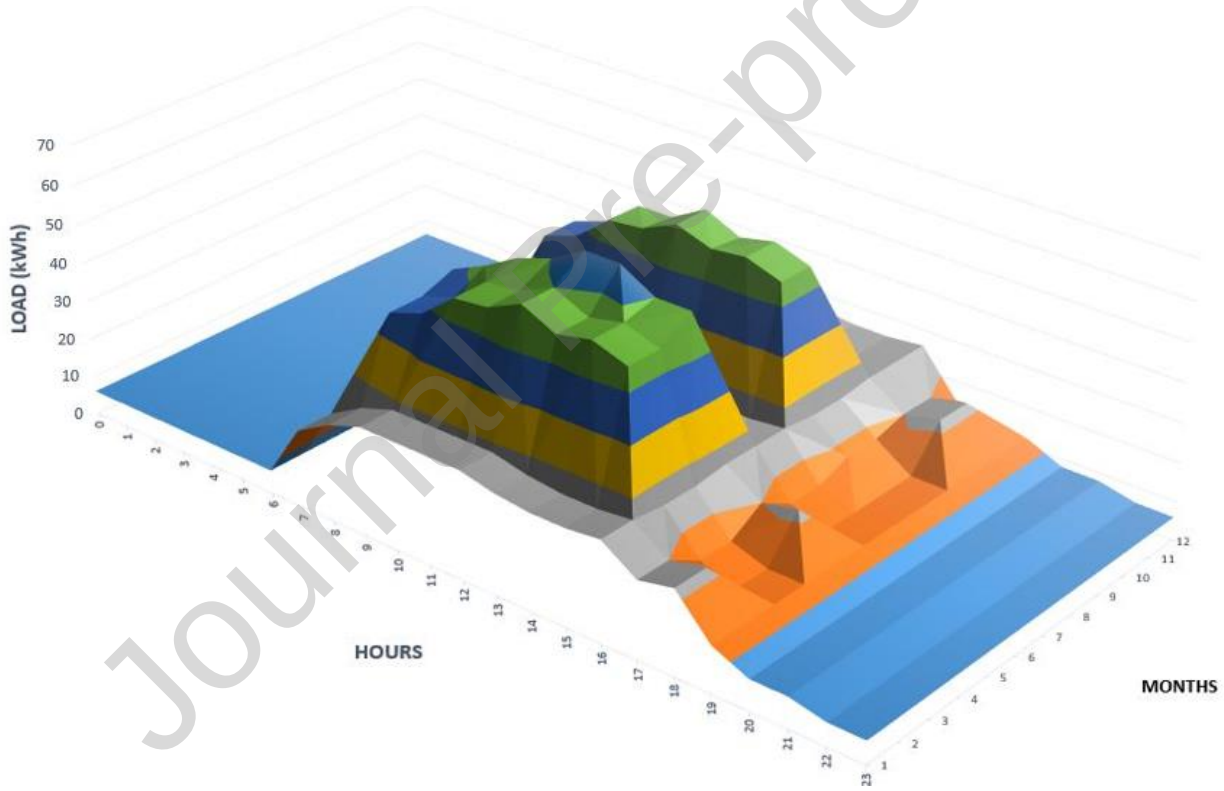


Journal Pre-proof

(B) Monthly average solar radiation

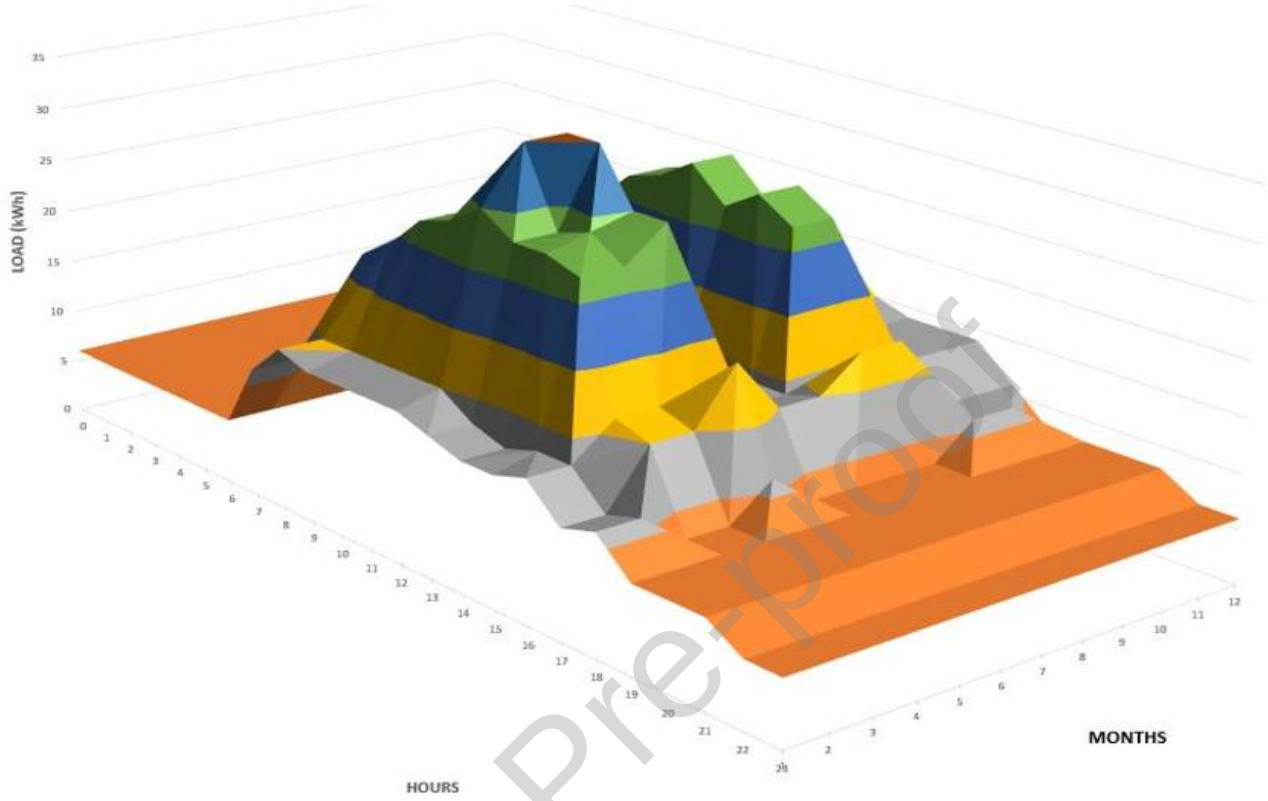
Figure 4: Renewable energy resources assessment

Figure 5(A) shows a 3D surface plot of each month of the year for the weekday load profile for the Mechanical Department, DTU, India. Figure 5(B) shows the exact data for weekends. It translates to a 65 kW peak load on weekdays and 35 kW on weekends. The average load calculated over the year amounts to 588 kWh per day. Load profiles vary based on the users, for example, homes versus commercial sites. There is the consideration that specific equipment runs continuously like refrigerators. Academic building comprises laboratory equipment and electrical appliances that contribute to their load requirement.



(A) Weekdays

Journal Pre-proof



(B) Weekends

Figure 5: Load profile for Mechanical Department, DTU, Delhi (India)

2.3 Cost analysis of the HRES

Net present cost (NPC): The net present cost of the hybrid system is the cost difference between all present costs incurred by the system over its lifetime and all present revenue collected during the project's lifetime. It is given by the Eq. (11) (Sahu et al., 2021).

$$C_{NPC} = \frac{C_{AT}}{CRF(i_r P_L)} \tag{11}$$

Levelized cost of energy (LCOE): HOMER calculates the LCOE by dividing the annualized cost of producing electricity by the total useful electric energy production, which is computed by Eq. (12) (Li et al., 2020).

$$LCOE = \frac{C_{ann,tot}}{E_{prim,AC}+E_{prim,DC}+E_{grid,sales}} \tag{12}$$

Salvage value: Salvage value is the cost of a power system component that is still operational at

Journal Pre-proof

the end of the project's lifespan. To assess the worth of each component at the conclusion of the project's life cycle, HOMER employs Eq. (13) (Bonah et al., 2022).

$$Salvage = C_{rep} \frac{R_{rem}}{R_{comp}} \quad (13)$$

Total annualized cost: As shown in Eq. (14), the annualized cost is calculated by multiplying the net present cost by the capital recovery factor (Riayatsyah et al., 2022).

$$C_{ann,tot} = CRF\ (i, R_{proj}) \times C_{NPC,tot} \quad (14)$$

The first step in this research is to decide on the site location. Decision of the site location is based on two main factors. Firstly, the given site has a significant load and is operational and relatively consistent throughout the year. Secondly, the location is easily accessible for the required data collection. Once the site location is finalized, analyze the available resources to design the system. It is done by analyzing all the naturally available resources in that particular area of Delhi, such as rain, wind, biomass, and sunlight. This particular data is sourced from the NASA meteorological website.

After that, data collection about how much energy is consumed daily in the decided location is required. This secondary data is sourced from the technicians responsible for the location. After breaking down the load profile according to weekdays and weekends and then further to summers and winters, the yearly energy required is calculated. This is done by monitoring and controlling the necessary variables. A significant review of past work in the field is done to understand the limitations faced and where the gap can be overcome.

This material is sourced from various sites, including published articles in reputed journals, reports, case studies, and government websites. These materials are chosen based on their data range, where preference and higher weightage are given to the more recent materials. After the system design, data is prepared and analysed using HOMER. Data is used to conduct three technical, economic, and environmental analyses. Results and findings are summarised, and relevant conclusions are drawn by concluding the assessment.

Tool used for Simulation: HOMER is utilized to simulate and optimize the HRES. HOMER is developed by the National Renewable Energy Laboratory (NREL) (“Homer,” n.d.). To meet thermal and electrical demands, a variety of energy sources, including photovoltaic (PV) modules,

Journal Pre-proof

wind turbines, small hydro, micro-turbines, biomass, batteries, hydrogen storage, and fuel cells, can be combined and simulated using HOMER. Studying and building these systems can be quite difficult since there are a lot of design possibilities available and uncertainty about key factors like fuel prices and load sizes. The inconsistency, seasonality, non-dispatchability, and unclear availability of renewable energy sources increase the complexity of issues. HOMER was created to deal with these problems. The main activities that HOMER performs are simulation, optimization, and sensitivity analysis and then it produces the best-optimized outcomes.

3. Results and Discussion

HOMER software is used along with the relevant parameters, inputs, and constraints to simulate the system. HOMER simulates multiple combinations and arrangements of the system and arranges the solutions based on the aggregate net present cost, and the system matches the current energy sources. The ideal arrangement of the HRES components for our proposed system is a solar photovoltaic power source, which can vary from 0 to 150 kW. The biomass gasifier generator set produces 160 kWh daily and also an idealized 600V battery and a 75 kW converter. The cost of various elements in the system is as follows: Capital Cost (CC), COE, and NPC_{total} for the HRES are ₹16.96 (US$ 0.207)/kWh, ₹32,028,543.3 (US$ 3,90,878) and ₹41,603,969.8 (US$ 5,07,737) respectively.

The yearly average load utilized is 190,116 kWh and the average energy production of the biomass gasifier and solar PV system are as follows: Solar PV system = 288,620 kWh/year and biomass gasifier generator set = 88,160 kWh/year as shown in Figure 6. The annual energy production of HRES is 376,780 kWh. Biomass gasifier supports approximately 23.4% of and solar PV approximately 76.6% of the total energy production.

Journal Pre-proof

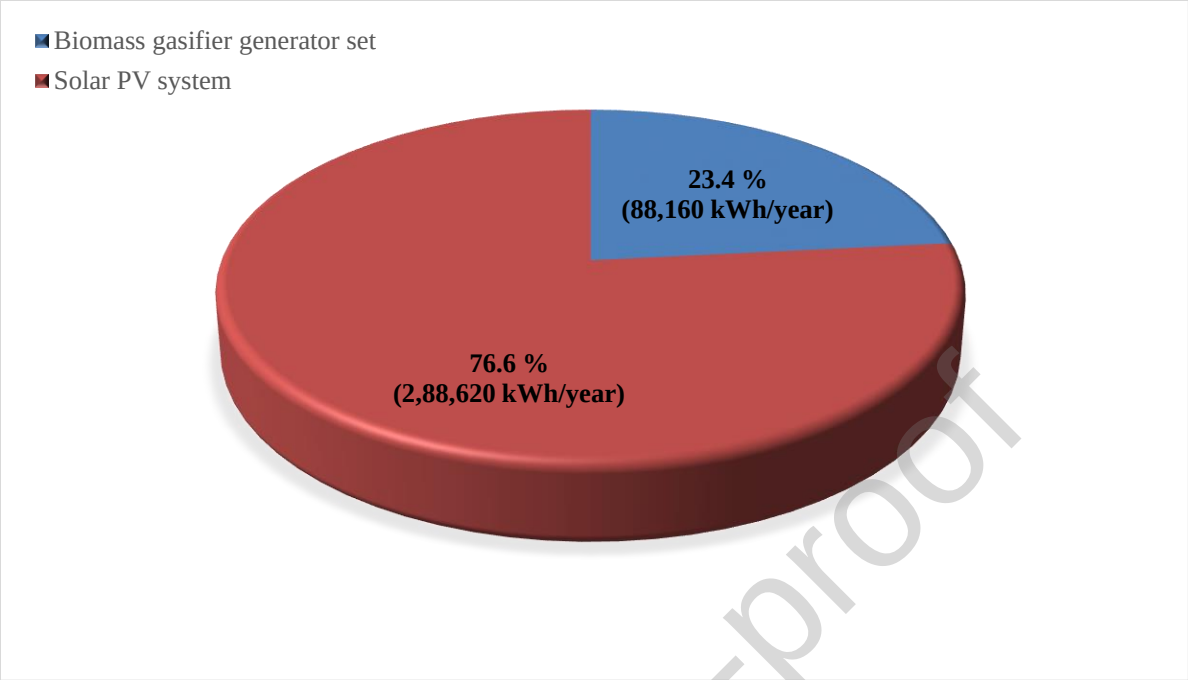
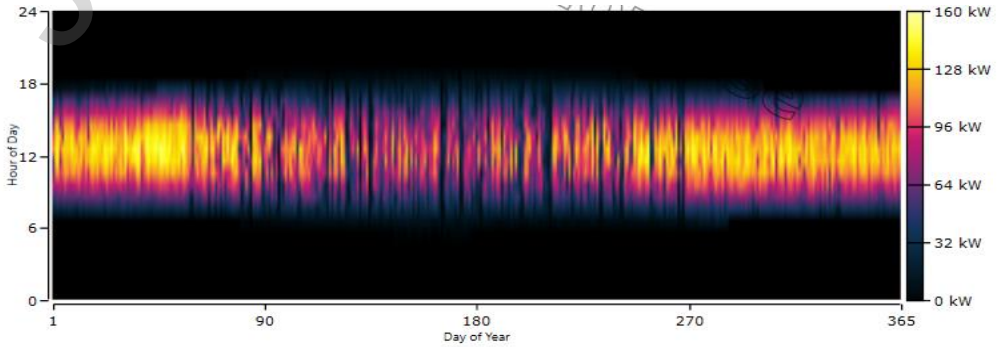


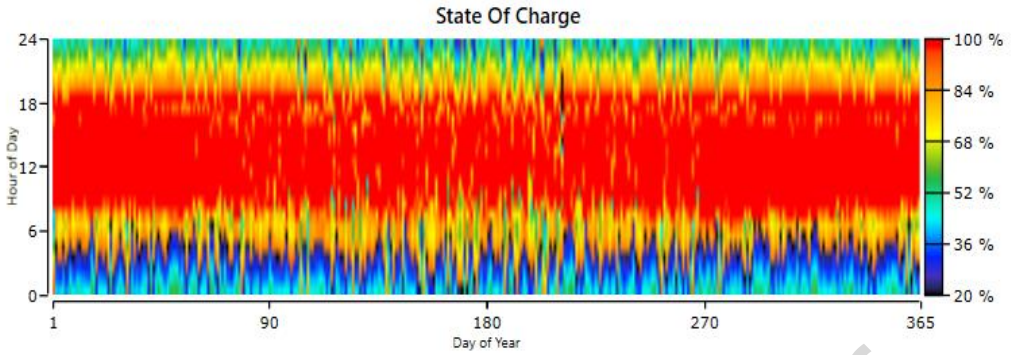
Figure 6: Average energy production from different resources

During the day, when ample solar power is accessible, there is a direct supply to the load through the solar PV system. Additionally, the biomass gasifier generator is accessible during the daytime. However, biomass gasifier works to provide a consistent supply at night when load demand is high since solar energy is not accessible. When the energy generation surpasses the energy demand, the spare energy is utilized in charging the battery. Load demand presented for the HRES of solar PV, and biomass for the Mechanical Engineering Department. In this HRES, an unmet electrical load of 48.8 kW per year is present and surplus energy of 173,147 kWh per year energy. A shortage in the capacity of 0.0375% is observed.

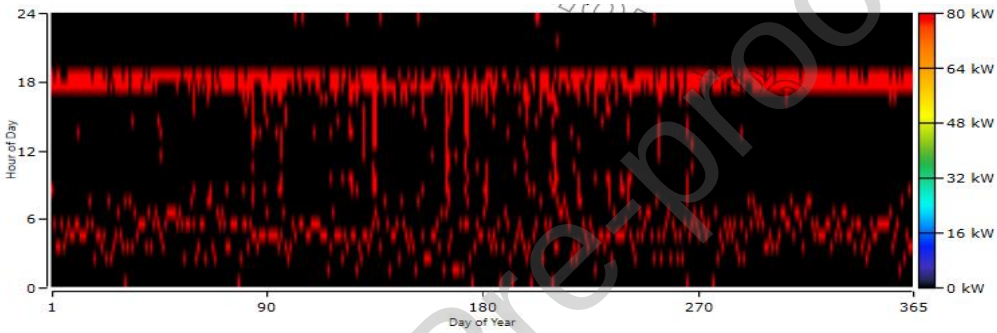


(A) Output of Solar photovoltaic system

Journal Pre-proof



(B) Battery state of charge (SOC)



(C) Biomass gasifier power generation

Figure 7: Output of different components of hybrid power generation system

Figure 7(A) shows the output of the solar PV system. PV module can produce energy in the morning even when there is not enough sunlight. However, most energy is produced between 10:00 AM to 04:00 PM i.e., 6 hours when PV modules work at full capacity. Figure 7(B) depicts the amount of charge present in the battery at a particular time. It can be deduced that the charge is always full during the daytime but constantly depletes once daylight is gone and the load is met using batteries only. The charge reduces to 15%-20% until it gets replenished in the morning. For a healthy battery life, trickle charging is suggested, and the minimum charge should never be lower than 15%. Biomass generator charges batteries and meets the load demands when PV output is unavailable. Figure 7(C) shows that the generator works at its peak during twilight and just before dawn. Figure 8 shows energy produced from PV and biogas generators every month. Throughout the system's lifetime, i.e., 25 years, biomass gasifier and solar PV contribute 23.4% and 76.6% of the total energy produced. Excess energy produced per annum is about 173,147 kWh, mainly due to the dynamic load profile, which considers the holidays and vacations in June, July and

Journal Pre-proof

December (2022) when the classes are suspended, and the halls and rooms in the academic buildings are not in that much use. It reduces the load requirements significantly. In June and July, the Northern Hemisphere is warm and sunny in summer; therefore, solar irradiation is at its fullest. Therefore, the PV at its total capacity will produce more energy than the whole Mechanical Engineering Department needs. To utilize this excess energy, the authors suggest that other departments can use this energy to power themselves, particularly during the months of vacations, rather than drawing power from the grid.

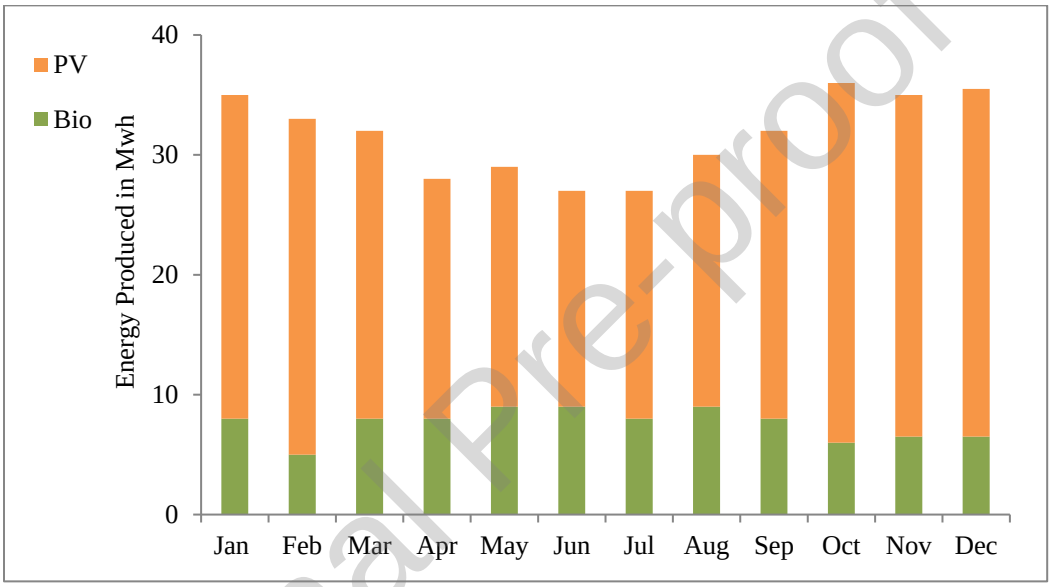


Figure 8: Monthly electricity production of HRES

The benefit of this hybrid plant does not limit itself to a lower carbon footprint compared to conventional energy generation methods. Biogas used for the gasifier comes from food waste collected on campus. Digestate is a by-product of the anaerobic digestion of food waste and can be used as fertilizer for gardening. At the same time, the biogas is extracted and utilized in power generation. In Table 1, the feedstock supplied on an average timely basis to the gasifier has been shown. In the gasifier, it has been estimated that from 1 ton of biomass, 400 kg of fertilizer is produced. Hence, with an annual food waste of 277 tons collected, around 110.8 tons of fertilizer is produced as a digest.

Journal Pre-proof

Table 1: Biomass consumption

Feedstock	Quantity	Units
Gross feedstock consumed	277	Tons
Daily average feedstock	0.759	Tons/day
Hourly average feedstock	0.0316	Tons/hour

Table 2: Yearly emissions

Pollutant	Yearly Emissions (kg/yr)
CO_2	49.90
CO	0.554
UHC	0.00
PM	0.00
SO_2	0.00
NO_X	0.346

Journal Pre-proof

HOMER calculates the analysis and consequently the results of emissions based on the following six pollutants: Carbon dioxide (CO_2), Carbon monoxide (CO), Particulate matter (PM), Unburnt hydrocarbons (UHC), Sulphur dioxide (SO_2), Nitrous oxide (NO_X). These pollutants are emitted due to electricity production by the generators, thermal energy released by the boiler and electricity consumption. HOMER simulates the gaseous emissions of both the boiler and the generator. HOMER first establishes the emissions factor, which is the weight of each pollutant in kg produced per unit of the fuel used, after which it simulates the power system. Next, it evaluates the yearly gaseous emissions of the particular pollutant from the emissions factor product and total yearly fuel consumption. The user inserts the emissions factor of 4 pollutants: CO, UHC, PM and NOx. Using this input as well as the contents of the fuel (carbon and sulphur), the software analyses to

evaluate emissions factors for the two other air pollutants, CO_2 and SO_2. HOMER also makes use of some assumptions in these calculations, which are they are as follows:

Carbon contents in the fuel which are not released as CO or UHC are treated as CO_2. The carbon ratio of the UHC produced is equivalent to that of the fuel. Sulphur contents in burnt fuel that are not released as PM are treated as SO_2 ("Homer Pro 3.15," 2022). Table 2 shows the yearly emissions that have been considered in the software for the feasibility analysis and output of the modelled HRES. According to research 1 kWh of electricity produces 0.85 kilograms of carbon dioxide annually ("Central Electricity Authority," 2022). According to the load profile, the total output of CO_2 generated by the Mechanical Engineering Department is estimated from Eq. (11).

$$190116\,\frac{\text{kWh}}{\text{year}}\,(\text{yearly average load}) \times 0.85 = 161598.6\ \text{kg}\ (\text{Output amount of } CO_2) \quad (11)$$

The total amount of CO_2 emission is calculated as 161,598.6 kg/year using Eq. (11). This shows a significant improvement of nearly 99% which is roughly 161 carbon credits per year. One carbon credit means that 1 metric tonne of CO_2 is prevented from entering the atmosphere. Similarly, 1 kWh of electricity produces approximately 7 grams/kilowatt-hour of SO_2 and 4.3 grams/kilowatt-hour of NO_x (Mittal, 2012). On the contrary, the HRES designed and analysed produces no SO_2 and only 0.346 kg/year of NOx. Hence it is a significant improvement in environmental impact and successful environmental sustainability analysis. All the feasible configurations with cost analysis for the proposed system are given in Table 3. The cost analysis of each component for a lifespan of 25 years is given in Table 4 and the Gross cost of the system is ₹41,603,935.4 (5,07,736.58US$).

Journal Pre-proof

Table 3: Feasible configurations with cost for HRES

Photovoltaics (kW)	Bioenergy (kW)	Number of batteries	Converter (kW)	Net Present Cost (US$)	Cost of energy (US$)	Operating cost (US$/yr)
144	160	1	73.7	5,07,737	0.207	9,040

133	160	3	88.0	6,59,382	0.268	10,570
261	160	2	115	7,06,777	0.287	7,054
145	160	2	79.2	5,60,841	0.228	7,506

(US$ 1= ₹81.94 accessed on April 15, 2023) ("USD to INR," n.d.)

Table 4: Cost analysis data of HRES components for 25 years

Component	**Capital (US$)**	**Replacement (US$)**	**Operation& Maintenance (US$)**	**Fuel (US$)**	**Salvage (US$)**	**Gross (US$)**
Generic 100kWh Li-ion	70,000.00	68,990.23	12,927.52	0.00	-1,577.08	150,340.67
Generic Biogas Genset	122,000.00	0.00	28,492.25	0.00	-1,173.83	149,318.41
Generic flat plate PV	173,069.37	0.00	10,949.84	0.00	0.00	173,380.12
System Converter	25,808.42	10,949.84	0.00	0.00	-2,060.87	34,697.38
System	390,877.79	79,940.07	41,730.51	0.00	-4,811.78	507,736.58

(US$ 1= ₹81.94 accessed on April 15, 2023) ("USD to INR," n.d.)

Journal Pre-proof

Validation of the proposed system results with previous published research

In contrast to earlier research, this study has created a novel HRES configuration with different component sizes, load requirements, and renewable resource usage. Therefore, it is impossible to compare those systems precisely. However, a comparison that is established based on the systems' economic designs can be considered acceptable and appropriate. This proposed system's NPC and COE have been compared to published research findings from previous studies (Table 5).

Table 5: Various hybrid power configurations using HOMER

S.No.	Author	Year	Components	Total NPC (US$)	LCOE (US$/kWh)
1	Hassan et al. (Hassan et al., 2022)	2022	PV/FC	10,166	0.23
2	Sawle et al. (Sawle et al., 2021)	2021	PV/WT/BG/DG	8,31,217	0.19
3	Basser et al. (Baseer et al., 2019)	2019	PV/WT/BG/DG	5,55,492	0.25
4	Sahu et al. (Sahu et al., 2021)	2021	WT/PV/BS	4,54,242	0.27
5	**Proposed system**	**2023**	**PV/BG**	**5,07,737**	**0.20**

Journal Pre-proof

Employment Opportunity

The design, setting up, maintenance and operation of plants lead to green employment for individuals in society. Skills required for the jobs vary from very little knowledge to high-level expertise, aiming for the community to position a post.

Improved sanitation: Adopting a hybrid project including biogas and biomass will positively

impact the waste management of society as a significant portion of the biological waste collected, that is initially discarded in landfills is now being sent to produce electricity.

Site assessment and location: Location and choice of setting up a biogas plant to be judicious and as least convenient to the area residents. Biogas plants make use of biological refuse that discards a pungent odour. The location of the biogas plant should be such that citizens feel minimal discomfort by its presence so that the plant can keep operating.

By replacing the traditional role of electricity, in academic institutions and other buildings, with green energy sources, there will be several positive impacts, most notably in the number of harmful substance emissions. Furthermore, renewable energy systems and plants can offer direct and indirect work to several people and businesses along the line of processes. HRES encourages good hygiene and sanitation since there is a significant focus on waste collection and management. It consequently reduces the waste going into dumpsites and uses it to convert it into usable energy. Therefore, all buildings and institutions can obtain their energy needs from renewable energy sources and fully optimized HRESs.

4. Conclusion

The techno-economic and environmental sustainability analysis to power the Mechanical Engineering Department at Delhi Technological University, Delhi has been done. The following remarks have been made on the basis of assessment, evaluation, and Output of HRES:

i. The cost of energy of the proposed HRES is ₹16.96 (US$0.207)/kWh, and the gross NPC is determined ₹41,603,969.8 (US$ 5,07,737).

ii. During a lifetime (25 years), biomass gasifier and solar PV contribute 23.4% and 76.6% of the total energy produced.

iii. System is optimized to meet the mean energy constraint of 588 kWh per day with a peak of 65 kW and 35 kW on weekdays and weekends, respectively.

iv. Total power generated by the system is 3,76,780 kWh per year to match the load demand of 1,90,116 kWh per year.

v. HRES has an unmet electrical load of 48.8 kW per year. A surplus of 1,73,147 kWh per year of energy is produced and a shortage of 0.0375% is observed in the capacity.

Journal Pre-proof

vi. HRES produces no SO_2 and only 0.346 kg per year of NOx, demonstrating a significant improvement in environmental effect and a successful environmental sustainability analysis and around 161 metric tonnes of CO_2 is prevented from entering the atmosphere.

The outcome of this research work highlights the future role of HRESs in meeting the load demands of educational institutions instead of relying on fossil fuel sources for power. The cost of energy is also reduced by using two renewable energy sources. It acts as a stepping-stone where energy demands can be met sufficiently by combining two or more non-conventional/renewable energy resources for power generation rather than depending on a single renewable energy source.

Recommendation for Future Scope:

i. Reduce the hazards of global warming by successfully exploiting HRES using AI-based multi-objective optimisation techniques like machine learning, artificial neural network, and deep learning models.

ii. Technological advancements in fuel cells make them economical, more dependable, and safe. Taking into account research topics such as connecting fuel cells with grid, safety concerns of storing H_2 in fuel cells, and maximum power utilisation from fuel-utilising AI-based optimisation algorithms.

Therefore there is a significant opportunity for intellectuals to further investigate proposed study areas in depth.

Ethics approval (include appropriate approvals or waivers):

Not Applicable

Journal Pre-proof

Funding:

There is no funding.

Acknowledgment

The authors are extremely thankful to Prof. D. Yogi Goswami, University of South Florida for

supporting the creation of Centre for Energy and Environment, Delhi Technological University, Delhi (India). The present research is conducted at the Centre for Energy and Environment.

Declaration

Availability of data and materials: All data given in either tables or figures.

Code availability (software application or custom code): Not Applicable

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

Additional declarations for articles in life science journals that report the results of studies involving humans and/or animals

Consent to participate (include appropriate statements): Not Applicable

Consent for publication (include appropriate statements): Not Applicable

Authors' contributions:

Neelam Baghel: Visualization Methodology, Writing - original draft and Investigation. **Manjunath K.:** Writing - review & editing, and Supervision.

Anil Kumar: Conceptualization, Writing - review & editing, and Supervision.

Journal Pre-proof

Nomenclature

A_{PV} = Rated capacity of the photovoltaic array

a_P = Temperature coefficient of power

α= Maximum charge rate of storage bank

DOD_{batt} = Depth of discharge of the battery

$C_{ann,tot}$ = Total annualised cost in $/year

C_{AT} = Total annualized cost

C_{NPC} = Net Present Cost

C_{rep} = Component replacement cost in $

C_{Wh} = Storage capacity

$E_{prim,AC}$ = AC primary load served in kWh/year

$E_{prim,DC}$ = DC primary load served in kWh/year

$E_{grid,sales}$ = Total grid sales in kWh/year

EL = Total energy demand

E_{tot} = Gross energy demand

E_{max}= Net capacity of the storage bank

E_1 = Available energy in storage bank-start of the time step

E_{tot}= Net amount of power in storage bank at the start of the time step

f_{PV} = Derating factor

I_m= Maximum charge current of the storage bank

N_{batt}= No. of batteries in the storage bank

P_{PV} = Energy output from photovoltaic cell

P_{NPV} = Rated power-reference

P_o = Electrical power output of the system

P_n = Power needed by system components

P_{tot} = Effective electrical energy system generated, (Input biomass) $_{LHV}$ = input biomass- lower heating value

Q = the total amount of energy in the storage at the beginning of the time step

r = The storage rate constant

R_{comp} = Component life in year

Journal Pre-proof

R_{rem} = Component remaining life in year

R_T = Incident solar radiation on the photovoltaic array in the current time step

$R_{T,STC}$ = Incident solar radiation, solar radiation under standard reference conditions

R = Solar radiation

R_{ref} = Solar radiation-reference (G_{ref} = 1000 W/m)

r = Constant (depending on storage rate)

s= Storage capacity ratio

T_{ref} = Temperature-reference conditions (Tref = 25 °C)

T_{cell} = Photovoltaic cell temperature

$T_{cell,STC}$ = Photovoltaic cell temperature under standard test conditions

Δt= Duration of the time step

V_n= Nominal voltage of the storage bank

Greek Symbol

η_{in} = Efficiency of inverter

η_{batt} = Efficiency of battery

$\eta_{batt,c}$ = Charge efficiency of the storage battery bank

Abbreviations

AD = Daily Autonomy

CC = Capital cost

COE = Cost of energy

CRF = Capital recovery factor

CUF = Capacity Utilization Factor

DSF = Demand supply fraction

Journal Pre-proof

GHG = Greenhouse gas

HRES = Hybrid renewable energy system

HOMER = Hybrid Optimization of Multiple Energy Resources

IRENA = International Renewable Energy Agency

LCOE = Levelized cost of energy

LFR = Linear Fresnel Reflector

NPC = Net present cost

NPV= Net present value

NREL = National Renewable Energy Laboratory

PTC = Parabolic Trough Collector

PV = Photovoltaic

UHC = Unburnt hydro carbon

Journal Pre-proof

References

Abbasi, T., Abbasi, S.A., 2010. Biomass energy and the environmental impacts associated with its production and utilization. Renewable and Sustainable Energy Reviews 14, 919–937. https://doi.org/10.1016/j.rser.2009.11.006

Access to Energy [WWW Document], n.d. URL https://ourworldindata.org/energy-access (accessed on April 15, 2023)

Ahmet Bozgeyik, Lutfiye Altay, A.H., 2023. Energetic, exergetic, exergoeconomic, environmental and sustainability analyses of a solar, geothermal and biomass based novel multi-generation system for production of power, hydrogen, heating, cooling and fresh water. Process Safety

and Environmental Protection 177, 400–415. https://doi.org/https://doi.org/10.1016/j.psep.2023.07.018

Al-Ghussain, L., Darwish Ahmad, A., Abubaker, A.M., Mohamed, M.A., 2021. An integrated photovoltaic/wind/biomass and hybrid energy storage systems towards 100% renewable energy microgrids in university campuses. Sustainable Energy Technologies and Assessments 46, 101273. https://doi.org/10.1016/j.seta.2021.101273

Al-Najideen, M.I., Alrwashdeh, S.S., 2017. Design of a solar photovoltaic system to cover the electricity demand for the faculty of Engineering- Mu'tah University in Jordan. Resource-Efficient Technologies 3, 440–445. https://doi.org/10.1016/j.reffit.2017.04.005

Azar, C., Lindgren, K., Larson, E., Möllersten, K., 2006. Carbon capture and storage from fossil fuels and biomass - Costs and potential role in stabilizing the atmosphere. Climatic Change 74, 47–79. https://doi.org/10.1007/s10584-005-3484-7

Baseer, M.A., Alqahtani, A., Rehman, S., 2019. Techno-economic design and evaluation of hybrid energy systems for residential communities: Case study of Jubail industrial city. Journal of Cleaner Production 237, 117806. https://doi.org/10.1016/j.jclepro.2019.117806

Bonah, E., Dankwa, J., Afrane, S., Sunday, T., Agbozo, E., 2022. ScienceDirect employment potential assessment of a hybrid energy system for tropical weather conditions e Combination of HOMER software, shannon entropy, and TOPSIS. International Journal of Hydrogen Energy 47, 31073–31097. https://doi.org/10.1016/j.ijhydene.2022.07.049

Central Electricity Authority [WWW Document], 2022. URL https://cea.nic.in/cdm-co2-baseline-database/?lang=en (accessed on April 15, 2023).

da Silva Lima, L., Quartier, M., Buchmayr, A., Sanjuan-Delmás, D., Laget, H., Corbisier, D., Mertens, J., Dewulf, J., 2021. Life cycle assessment of lithium-ion batteries and vanadium redox flow batteries-based renewable energy storage systems. Sustainable Energy Technologies and Assessments 46. https://doi.org/10.1016/j.seta.2021.101286

Farret, F.A., Simões, M.G., 2006. Integration of Alternative Sources of Energy, Integration of Alternative Sources of Energy,February 2006. https://doi.org/10.1002/0471755621

Ghenai, C., Janajreh, I., 2016. Design of Solar-Biomass Hybrid Microgrid System in Sharjah.

Journal Pre-proof

Energy Procedia 103, 357–362. https://doi.org/10.1016/j.egypro.2016.11.299

Givaki, K., Parker, M., Jamieson, P., 2014. Estimation of the power electronic converter lifetime in fully rated converter wind turbine for onshore and offshore wind farms. 7th IET International Conference on Power Electronics, Machines and Drives, PEMD 08-10 April 2014, DOI: 10.1049/cp.2014.0301.

Goyal, N., Aggarwal, A., Kumar, A., 2022. Financial feasibility of concentrated solar power with and without sensible heat storage in hot and dry Indian climate. Journal of Energy Storage 52, 105002. https://doi.org/10.1016/j.est.2022.105002

Hassan, Q., Jaszczur, M., Hafedh, S.A., Abbas, M.K., Abdulateef, A.M., Hasan, A., Abdulateef, J., Mohamad, A., 2022. Optimizing a microgrid photovoltaic-fuel cell energy system at the highest renewable fraction. International Journal of Hydrogen Energy 47, 13710–13731. https://doi.org/10.1016/j.ijhydene.2022.02.108

Homer [WWW Document], n.d. URL https://www.homerenergy.com/products/pro/docs/3.13/index.html (accessed on April 15, 2023)

Homer Pro 3.15 [WWW Document], 2022. URL https://www.homerenergy.com/products/pro/docs/latest/how_homer_calculates_emissions.html (accessed on April 15, 2023).

Kamali, S.K., Mekhilef, S., 2009. Evaluation study on grid-connected PV system at university of Malaya. International Conference for Technical Postgraduates 2009, TECHPOS 2009. https://doi.org/10.1109/TECHPOS.2009.5412043

Kavuma, C., Sandoval, D., de Dieu, H.K.J., 2022. Analysis of solar photo-voltaic for grid integration viability in Uganda. Energy Science and Engineering 10, 694–706. https://doi.org/10.1002/ese3.1078

Kazem, H.A., Khatib, T., 2013. A novel numerical algorithm for optimal sizing of a photovoltaic/wind/ diesel generator/battery microgrid using loss of load probability index. International Journal of Photoenergy 2013. https://doi.org/10.1155/2013/718596

Krishnamoorthy, M., Periyanayagam, A.D.V.R., Santhan Kumar, C., Praveen Kumar, B.,

Journal Pre-proof

Srinivasan, S., Kathiravan, P., 2020. Optimal Sizing, Selection, and Techno-Economic Analysis of Battery Storage for PV/BG-based Hybrid Rural Electrification System. IETE Journal of Research 0, 1–16. https://doi.org/10.1080/03772063.2020.1787239

Kumar, N.M., Subathra, M.S.P., Moses, J.E., 2018. On-Grid Solar Photovoltaic System: Components, Design Considerations, and Case Study. Proceedings of the 4th International Conference on Electrical Energy Systems, ICEES 2018 616–619. https://doi.org/10.1109/ICEES.2018.8442403

Li, J., Liu, P., Li, Z., 2020. Optimal design and techno-economic analysis of a solar-wind-biomass off-grid hybrid power system for remote rural electrification: A case study of west China. Energy 208, 118387. https://doi.org/10.1016/j.energy.2020.118387

Malheiro, A., Castro, P.M., Lima, R.M., Estanqueiro, A., 2015. Integrated sizing and scheduling of wind/PV/diesel/battery isolated systems. Renewable Energy 83, 646–657. https://doi.org/10.1016/j.renene.2015.04.066

Mittal, M.L., 2012. Estimates of Emissions from Coal Fired Thermal Power Plants in India 39, 1–22. Available: www.eia.doe.gov.

Mobin Korpeh, Poorya Asadbagi, A., Lotfollahi, Sina Ghaemi, A.A.-, Moghaddam, 2023. 4E Analysis and Optimization of a Novel Hybrid Biomass-Solar System: Focusing on Peak Load Management and Environmental Emissions. Process Safety and Environmental Protection. https://doi.org/https://doi.org/10.1016/j.psep.2023.11.035

Pyke, T., n.d. The energy debate: Renewable energy cannot replace fossil fuels [WWW Document]. URL https://developmenteducation.ie/feature/the-energy-debate-renewable-energy-cannot-replace-fossil-fuels/ (accessed on April 15, 2023).

Razaghi, M., Rahdar, M.H., Aghamohamadi, N., Mohammadi, M., Ahmadi, P., 2024. Biofuel, solar-biofuel or geothermal energy; which resource can better contribute to an integrated energy system for residential energy demands. Process Safety and Environmental Protection 181, 103–130. https://doi.org/10.1016/j.psep.2023.10.063

Rehman, S., 2021. Hybrid power systems – Sizes, efficiencies, and economics. Energy Exploration and Exploitation 39, 3–43. https://doi.org/10.1177/0144598720965022

Journal Pre-proof

Riayatsyah, T.M.I., Geumpana, T.A., Rizwanul Fattah, I.M., Rizal, S., Indra Mahlia, T.M., 2022. Techno-Economic Analysis and Optimisation of Campus Grid-Connected Hybrid Renewable Energy System Using HOMER Grid. Sustainability (Switzerland) 14. https://doi.org/10.3390/su14137735

Saha, P., Kar, A., Behera, R.R., Pandey, A., Chandrasekhar, P., Kumar, A., 2022. Performance optimization of hybrid renewable energy system for small scale micro-grid. Materials Today: Proceedings 63, 527–534. https://doi.org/10.1016/j.matpr.2022.03.656

Saheb-Koussa, D., Haddadi, M., Belhamel, M., 2009. Economic and technical study of a hybrid system (wind-photovoltaic-diesel) for rural electrification in Algeria. Applied Energy 86, 1024–1030. https://doi.org/10.1016/j.apenergy.2008.10.015

Sahu, P.K., Jena, S., Sahoo, U., 2021. Techno-Economic Analysis of Hybrid Renewable Energy System with Energy Storage for Rural Electrification. Hybrid Renewable Energy Systems 63–96. https://doi.org/10.1002/9781119555667.ch3

Sawle, Y., Gupta, S.C., Bohre, A.K., 2016. PV-wind hybrid system: A review with case study. Cogent Engineering 3. https://doi.org/10.1080/23311916.2016.1189305

Sawle, Y., Jain, S., Babu, S., Nair, A.R., Khan, B., 2021. Prefeasibility Economic and Sensitivity Assessment of Hybrid Renewable Energy System. IEEE Access 9, 28260–28271. https://doi.org/10.1109/ACCESS.2021.3058517

Sayed, E.T., Wilberforce, T., Elsaid, K., Rabaia, M.K.H., Abdelkareem, M.A., Chae, K.J., Olabi, A.G., 2021. A critical review on environmental impacts of renewable energy systems and mitigation strategies: Wind, hydro, biomass and geothermal. Science of the Total Environment 766, 144505. https://doi.org/10.1016/j.scitotenv.2020.144505

Sen, R., Bhattacharyya, S.C., 2014. Off-grid electricity generation with renewable energy technologies in India: An application of HOMER. Renewable Energy 62, 388–398. https://doi.org/10.1016/j.renene.2013.07.028

Shaikh, M.R.S., 2017. A Review Paper on Electricity Generation from Solar Energy. International Journal for Research in Applied Science and Engineering Technology V, 1884–1889. https://doi.org/10.22214/ijraset.2017.9272

Journal Pre-proof

Sharma, S., Kurian, C.P., Paragond, L.S., 2018. Solar PV System Design Using PVsyst: A Case Study of an Academic Institute. 2018 International Conference on Control, Power, Communication and Computing Technologies, ICCPCCT 2018 123–128. https://doi.org/10.1109/ICCPCCT.2018.8574334

Sheik, M.A., Aravindan, M.K., Beemkumar, N., Chaurasiya, P.K., Jilte, R., Shaik, S., Afzal, A., 2022. Investigation on the thermal management of solar photo voltaic cells cooled by phase change material. Journal of Energy Storage 52, 104914. https://doi.org/10.1016/j.est.2022.104914

Singh, A., Baredar, P., 2016. Techno-economic assessment of a solar PV, fuel cell, and biomass gasifier hybrid energy system. Energy Reports 2, 254–260. https://doi.org/10.1016/j.egyr.2016.10.001

Suresh Kumar, U., Manoharan, P.S., 2014. Economic analysis of hybrid power systems (PV/diesel) in different climatic zones of Tamil Nadu. Energy Conversion and Management 80, 469–476. https://doi.org/10.1016/j.enconman.2014.01.046

Türkay, B.E., Telli, A.Y., 2011. Economic analysis of standalone and grid connected hybrid energy systems. Renewable Energy 36, 1931–1943. https://doi.org/10.1016/j.renene.2010.12.007

USD to INR [WWW Document], n.d. URL https://www.google.com/finance/quote/USD-INR?sa=X&ved=2ahUKEwjQsu-1_pL-AhXByTgGHdufDaEQmY0JegQIBhAd (accessed on April 15, 2023)

Wang, W., Yuan, B., Sun, Q., Wennersten, R., 2022. Application of energy storage in integrated energy systems — A solution to fluctuation and uncertainty of renewable energy. Journal of Energy Storage 52, 104812. https://doi.org/10.1016/j.est.2022.104812

Wibowo, H., Bow, Y., Sitompul, C.R., 2021. Performance Comparison Analysis of Fixed and Solar-Tracker Installed Panel at PV System. IOP Conference Series: Earth and Environmental Science 709. https://doi.org/10.1088/1755-1315/709/1/012003

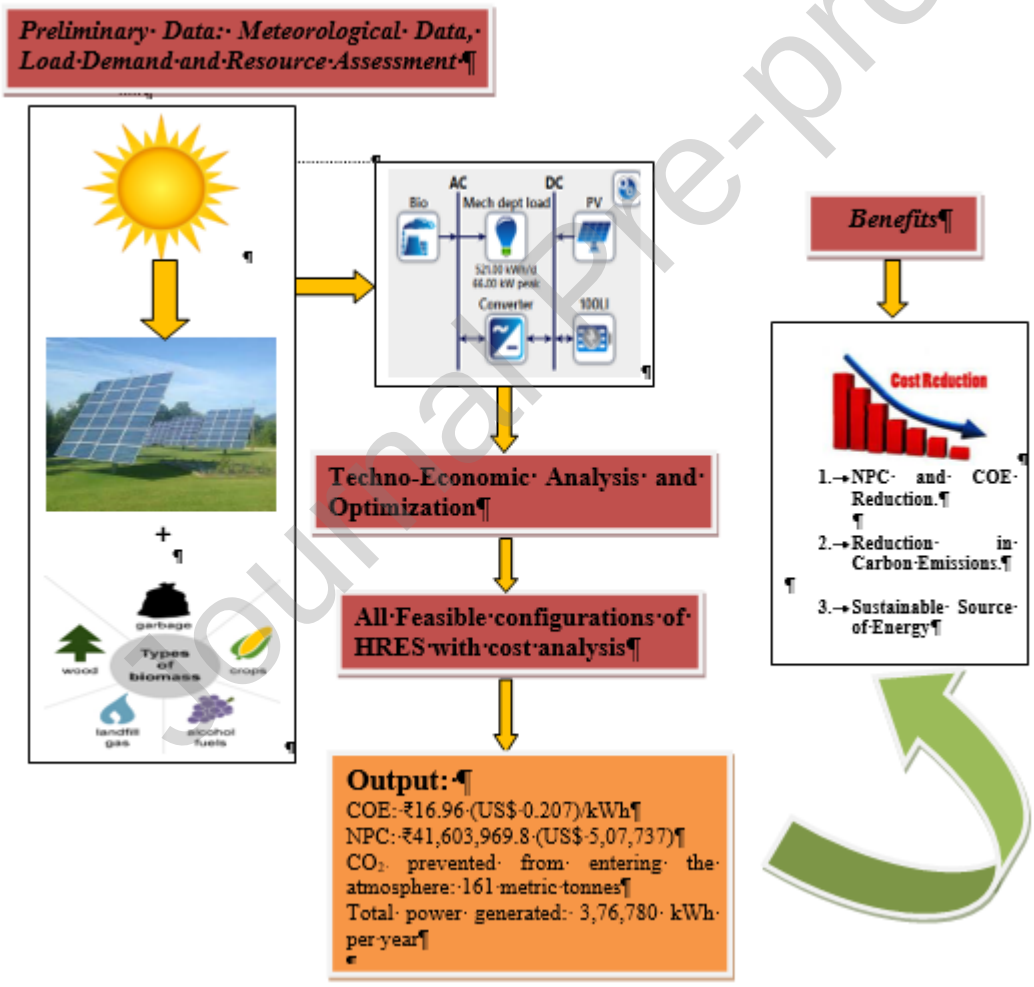
Journal Pre-proof

Declaration of interests

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

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Graphical Abstract



Journal Pre-proof

Highlights

- Calculated net present cost and energy cost for the hybrid system.
- Biomass gasifier and PV contribute 23.4% and 76.6% to total energy.
- Designed to meet the average energy need of 588 kWh.
- Generates 3,76,780 kWh/year to fulfill the energy requirement.
- Prevents 161 metric tonnes of CO_2 from entering the atmosphere.

Journal Pre-proof

2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE) | 979-8-3503-8328-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICWITE59797.2024.10502721

Attribute-Based Encryption Scheme for Secure and Efficient Access in Blockchain

Anita Thakur
Department of Information Technology
Delhi Technological University
Delhi, INDIA
anitathakur_2k21phdit504@dtu.ac.in

Virender Ranga
Department of Information Technology
Delhi Technological University
Delhi, INDIA
drvirender.ranga@gmail.com

Ritu Agarwal
Department of Information Technology
Delhi Technological University
Delhi, INDIA
ritu.jeea@gmail.com

***Abstract*— Due to the large volume of sharing and storage of sensitive information on centralized or distributed storage, there is a growing necessity for employing encryption techniques that provide promising security and access control to the sensitive data on these platforms. Access control is a fundamental aspect of information security, encompassing the mechanisms and policies governing who can access, modify, or use the data within a system. Attribute-based encryption (ABE) is a cryptographic technique that offers a fine-grained access control mechanism by associating attributes with encrypted data and secret keys. This paper suggests the proposed attribute-based encryption as a means to establish secure access control in Blockchain, thereby attaining the secrecy and freshness aspects of cryptographic protocols. The proposed scheme is simulated on SPAN AVISPA. The Security Protocol Animator SPAN functions by simulating an active intruder, enabling interactive identification and construction of potential attacks against various protocols.**

Keywords—Blockchain, Attribute-Based Encryption, Security, Access Control, AVISPA

I. INTRODUCTION

Security attacks in the Blockchain ecosystem can have profound repercussions. They can alter data, enable double spending, lead to fund theft, disrupt the network's operation, and, most significantly, erode trust in the technology. These attacks compromise the core principles of Blockchain, such as immutability, transparency, and security, making it essential for Blockchain networks to implement robust security measures to mitigate these threats and maintain their integrity. Cryptography protocols serve as a crucial defense mechanism in protecting Blockchain networks against security attacks. They play a pivotal role by encrypting data stored on the Blockchain, ensuring the confidentiality of sensitive information, and rendering it inaccessible to unauthorized parties. Additionally, cryptographic techniques are applied to secure transactions, making it extremely challenging for attackers to alter or intercept data during the transfer of assets.

In the context of cryptographic protocols, an identity-based encryption (IBE) system distinguishes individual users through distinct identity strings. Conversely, an ABE system operates by characterizing users through a collection of attributes. It utilizes a specified function based on these attributes to determine their decryption capabilities for a given ciphertext. Identity-based encryption (IBE), as pioneered by Adi Shamir in his seminal work [1], represents a cryptographic paradigm that affords individuals the capability to employ arbitrary character strings as their public keys, such as email addresses. This novel approach obviates the necessity for a distinct infrastructure for the dissemination of public keys, thereby streamlining the process of message transmission. In practical terms, it enables message senders to securely communicate with recipients solely based on the latter's identity, which can be, for instance, an email address. Shamir's motivation for developing Identity-Based Encryption (IBE) was to streamline and simplify the management of certificates in email systems.

Boneh and Franklin [2] first introduce fully functional identity-based encryption. Prior to their proposal, several IBE schemes had been put forth, but they exhibited various limitations. Some of these earlier solutions necessitated that users not collaborate or collude, while others imposed significant computational overhead on the Private Key Generator (PKG) during private key generation requests. Additionally, certain solutions called for tamper-resistant hardware, as identified by Boneh and Franklin. Nonetheless, in many cases, individuals are typically distinguished by their distinctive attributes rather than a singular unique identity. Individuals may not possess precise knowledge of the specific identities of all other individuals granted access to the data; instead, they may only possess the means to describe potential access recipients based on their descriptive credentials or attributes [3]. ABE is a form of access control in which the data owner defines access policies based on attributes, and only users with the necessary attributes can access the encrypted data. Sahai and Waters [4] introduced the notion of ABE. In an ABE framework, both a user's cryptographic keys and ciphertexts are associated with sets of descriptive attributes, and the decryption of a specific ciphertext is contingent upon the alignment of attributes between the ciphertext and the user's key.

The primary objective of this research work is to evaluate the efficacy of ABE in safeguarding sensitive data while ensuring robust security measures such as confidentiality and freshness.

A. Our Contribution

In essence, this paper contributes by proposing a novel approach to fortify access control in Blockchain through the integration of Attribute-Based Encryption. By leveraging ABE, the paper aims to elevate the security standards and responsiveness of access control mechanisms in the ever-evolving landscape of information sharing and storage. We utilized the AVISPA SPAN tool to simulate the functioning of the proposed scheme under various conditions. This simulation allows

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:53:45 UTC from IEEE Xplore. Restrictions apply.

us to thoroughly analyze the ABE's capability to protect sensitive information from intruders or attackers.

II. Related work

The advent of identity-based encryption has brought forth a distinctive approach introduced by Sahai and Waters, known as ABE [4]. Subsequently, various iterations such as KP-ABE [5] and CP-ABE [6] schemes have been introduced to advance and optimize the efficacy of ABE. A Blockchain is an ever-expanding series of blocks connected by hash values containing the ledger of validated transactions with timestamps. A Blockchain is a peer-to-peer network with each node in the network replicating and storing its copy of the data. Consequently, this distributed, decentralized approach excludes the need for truth in third parties. ABE in Blockchain can be used to enhance privacy, security, and access control in various ways such as fine-grained access control, privacy preservation, flexible access policy, key management, multi-party access, and data sharing.

Authors [7] introduced a Blockchain-centered framework integrating ABE. This scheme, beyond upholding data integrity and non-repudiation, notably safeguards the privacy of transactional data. The proposed scheme for a privacy-preserving Blockchain incorporates access control mechanisms to specifically tackle the confidentiality of shared data within the Blockchain ecosystem. Consequently, it endeavors to establish an end-to-end privacy-preserving system tailored for the Internet of Things (IoT). Authors [8] proposed an efficient and privacy-preserving traceable attribute-based encryption scheme that leverages Blockchain technology to ensure data integrity and non-repudiation. Employing pre-encryption methods, it enables swift ciphertext generation while preserving user anonymity through attribute hiding within anonymous access control structures using attribute bloom filters. This scheme allows auditing of abused secret keys, distinguishing between user and attribute authority origins, thus enhancing security. Comprehensive security and performance analyses affirm the scheme's robustness and efficiency, addressing critical issues in attribute-based encryption within distributed networks.

The primary limitation associated with ABE pertains to its computational overhead, which escalates proportionally with the intricacy of the access policy employed. Authors [9] introduced a solution that tackles the escalating computational burden in ABE by securely outsourcing decryption computation. The proposed ABE scheme introduces fairness through Blockchain and sampling techniques. Leveraging smart contracts within Blockchain ensures proxies receive rightful rewards for valid outsourced decryption results, establishing a system that upholds fairness while maintaining the confidentiality of the underlying plaintext. In the present investigation, the authors [10] have introduced a keyword-searchable ABE scheme. This innovative approach aims to enhance the security of digital certificate transfers while addressing concerns pertaining to data privacy. The findings of the study indicate that the proposed scheme demonstrates notable cost-effectiveness, characterized by a low computational cost. Furthermore, the system exhibits proficiency in attribute revocation, thereby bolstering its overall efficacy.

In the study [11], the authors introduced a novel multi-authorization attribute-based verifiable encryption scheme, leveraging Blockchain technology. The proposed scheme facilitates the validation of key accuracy distributed by numerous authorization centers, effectively resolving the challenge associated with the necessity for all authorization centers to retransmit the private key in the event of an incorrect key discovery. The study [12] introduced an innovative collaborative framework designed for secure cloud file sharing, employing a combination of Blockchain technology and ABE. The utilization of Blockchain facilitates the implementation of access control through smart contracts established between data owners and users. This decentralized scheme exhibits fault tolerance and resilience against Denial of Service (DoS) attacks, ensuring robust security measures

III. Preliminaries

Within this section, we provide a formal definition along with essential background information about cryptographic assumptions, bilinear maps, access structure, and attribute-based encryption.

A. Bilinear Maps

Considering two multiplicative cyclic groups denoted as $\mathbb{G}_0$ and $\mathbb{G}_T$, both having a prime order p. Let e be a bilinear map, e: $\mathbb{G}_0 \times \mathbb{G}_0 \rightarrow \mathbb{G}_T$ with g be a generator of $\mathbb{G}_0$ has the properties below [13]:

- Bilinearity: $\forall\ a,\ b \in \mathbb{Z}p$ and $g_1,\ g_2 \in \mathbb{G}_0$, $e(g_1^a, g_2^b) = e(g_1, g_2)^{ab}$.
- Computability: $e(g_1, g_1)$ is efficiently computable for all $g_1, g_1 \in \mathbb{G}_0$.
- Non-degeneracy: There exists $g_1, g_1 \in \mathbb{G}_0$, $e(g_1, g_1) \neq 1$.

Bilinear maps in ABE provide the mathematical framework to implement fine-grained and flexible access control policies and efficient key generation while maintaining security.

B. Security Assumption

The security of the Boneh and Franklin (in their fully functional IBE) is based on the Computational Diffie Hellman (CDH) assumption on the elliptic curve. This system can be constructed using any bilinear maps, provided that the variant of the CDH problem in $\mathbb{G}_0$ is NP-Hard.

Computational Diffie Hellman assumption states that for a probabilistic polynomial time algorithm $\mathcal{B}$, the computational challenge lies in computing $\mathbb{Z} = g^{ab}$ when provided with known tuple g^a, g^b, where $(a,b) \in \mathbb{Z}_p$ is randomly selected. The advantage ε in addressing the CDH assumption is described as:

$$|Pr[\mathcal{B}(g^a, g^b) = g^{ab}]| \geq \varepsilon$$

Another popular security assumption, Decisional Bilinear Diffie Hellman (DBDH) [4] [14] [15] describes that g is a generator of $\mathbb{G}_0$. a, b, c, z $\in \mathbb{G}_0$ are four randomly chosen elements, and let Y_1, Y_2, Y_3 are $Y_1 = g^a$, $Y_2 = g^b$, $Y_3 = g^c$, no probabilistic polynomial time algorithm $\mathcal{B}$ can distinguish from $e(g,g)^{abc}$ and $e(g,g)^z$ with advantage ε in solving:

$$Adv_{\mathcal{B}} = |Pr[\mathcal{B}(Y_1, Y_2, Y_3, e(g,g)^{abc}) = 0] - Pr[\mathcal{B}(Y_1, Y_2, Y_3, e(g,g)^z)] = 0| \geq \varepsilon$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:53:45 UTC from IEEE Xplore. Restrictions apply.

C. Access Structure

Access structures [16] define individuals who possess the requisite qualifications or those who lack permission to access particular data, contingent upon the authorization of specific attribute sets.

Definition: Let $\{P_{r1}, P_{r2}, \ldots, P_{rn}\}$ in access structure be a set of parties. A collection $\mathbb{A} \subseteq 2^{\{Pr1, Pr2, \ldots, Prn\}}$ is monotone $\forall$ B and C, if $B \in \mathbb{A}$, $B \subseteq C$, then $C \in \mathbb{A}$. An access structure is a collection $\mathbb{A}$ of nonempty subsets of $\{P_{r1}, P_{r2}, \ldots, P_{rn}\}$, i.e. $\mathbb{A} \subseteq 2^{\{Pr1, Pr2, \ldots, Prn\}} \setminus \{\emptyset\}$. The attributes have taken the role of the parties here.

Access Tree: Consider a tree, denoted as $\top$, representing an access structure. Within this structure, every non-leaf node within the tree signifies a threshold gate. Each gate is defined by its children node and a specific threshold value. In ABE (Figure 1), authorized users must satisfy the access policy formulated using a set of attributes to gain access to sensitive data.

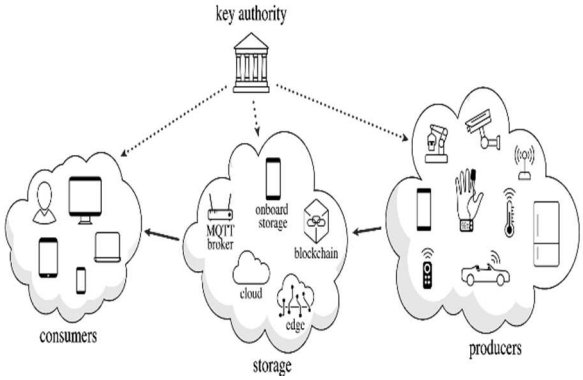


Figure 1: ABE Scheme [17]

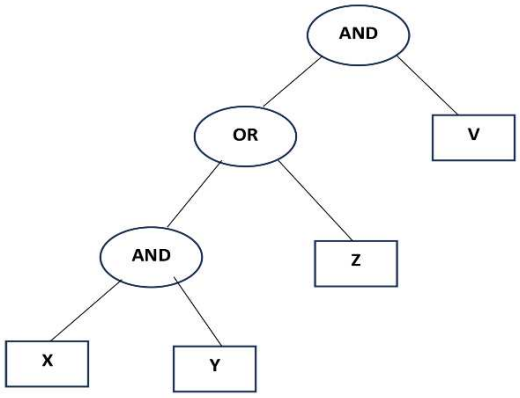


Figure 2: Access Tree

Figure 2 illustrates an access structure that gives access to those users who satisfy the access policy (((X∧Y)∨Z)∧V).

IV. System Model

Within our model (as shown in Figure 3) the encryption of the message is achieved through the implementation of the CP-ABE scheme. The proposed model consists of four key entities, namely the Data Owner, Data Consumer, Attribute Authority, and Blockchain.

Blockchain: Blockchain serves as the repository for data storage, ensuring non-repudiation, consistency, and immutability of the stored information. The data owner uploads the ciphertext on the Blockchain, also the access approach furnished by the data owner is archived, while the data user undergoes authentication in alignment with the access strategy delineated by the data owner.

Data Owner: The Data Owner assumes the role of a data sharer, storing information within the Blockchain. Utilizing an access structure, the Data Owner encrypts the message to generate ciphertext before storage.

Data Consumer: Access to the data is granted to users whose attributes align with the ciphertext access policy. Upon meeting this criterion, the user employs the decryption algorithm alongside the secret key to decrypt the file and acquire the plaintext.

Attribute Authority: The Attribute Authority releases the public parameter and formulates the master key for the system. Subsequently, the AA issues a secret key to the user, considering the user's specific attributes.

There are four major algorithms in attribute-based encryption briefly described as:

- Setup: *Setup{ λ }→PK, MSK,* In this algorithm security parameter λ is taken as input which then generates the public key *(PK),* known to everyone, and the master secret key *(MSK),* kept as secret.
- Key Generation: *KeyGen{ PK, MSK, S }→SK,* It is sometimes referred to as the attribute key generation algorithm, which takes the public key PK, set of attributes S, and master secret key MSK as input and generates the Secret Key SK.
- Encryption: *Encrypt{ PK, M, γ }→CT,* This algorithm takes the public key parameters, a message, and a set of attributes γ as the input and returns the ciphertext *CT.*
- Decryption: *Decrypt{PK, CT, SK} →M*, As stated earlier, the user acquires the decryption key that corresponds to their attribute set *S.* The user then decrypts the ciphertext using their secret key.

Security Goal

Access Control and Confidentiality: The preservation of confidentiality within the system necessitates the continuous safeguarding of message secrecy during transmission. Simultaneously, access control mechanisms guarantee that message decryption and access are exclusively granted to users whose attributes align with the access policies defined by the data owner.

Data Security: The scheme should ensure that the data is accessed by the authenticated users only and upon a user's departure from the system, their access privileges must be promptly revoked, barring any further access to the data or messages.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:53:45 UTC from IEEE Xplore. Restrictions apply.

Collision Resistance: Collusion can occur when multiple users pool their keys together to collectively decrypt a ciphertext, even if individually they might not possess enough attributes to satisfy the access policy. Collusion tolerance aims to prevent or mitigate such attacks. Central authority can prevent collusion by managing key distribution and maintaining control over the keys.

Forward and Backward Security: Newly joined users are restricted from accessing ciphertexts shared before their enrolment to prevent access to past data. Similarly, former users are prohibited from accessing current ciphertexts and any forthcoming encrypted information. This measure ensures that only the current authorized user base can access historical and ongoing information, safeguarding the confidentiality of past, present, and future data exchanges.

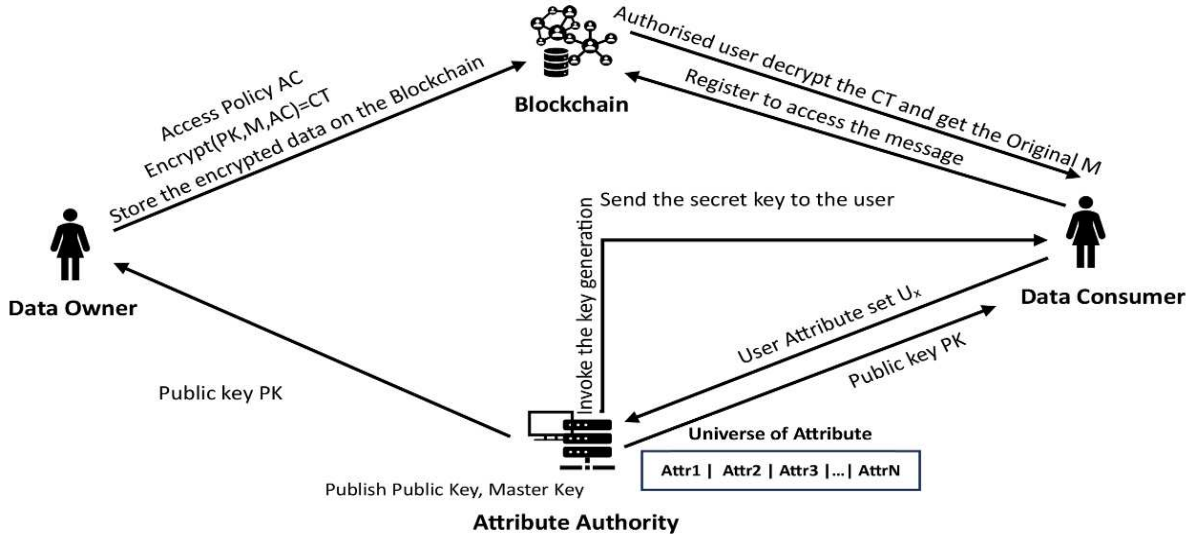


Figure 3: Work Flow of System Model

V. SIMULATION

AVISPA conducts analyses on network security protocols and software applications scripted in HLPSL.

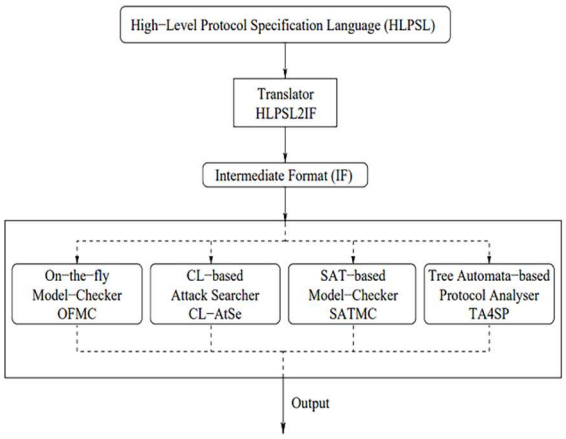


Figure 4: AVISPA Tool Architecture [18]

HLPSL encompasses fundamental roles delineating diverse entities and configurations characterizing pivotal role scenarios. The roles operate independently, gathering initial data through parameters and employing AVISPA backend parameters like CL-AtSe to verify the authenticity of the communication method with distinct characteristics.

```
SUMMARY
  SAFE
DETAILS
  BOUNDED_NUMBER_OF_SESSIONS
  TYPED_MODEL
PROTOCOL
  /home/span/span/testsuite/results/ABE.if
GOAL
  As Specified
BACKEND
  CL-AtSe
STATISTICS
  Analysed   : 39 states
  Reachable  : 34 states
  Translation: 0.00 seconds
  Computation: 0.00 seconds
```

(a)

```
% OFMC
% Version of 2006/02/13
SUMMARY
  SAFE
DETAILS
  BOUNDED_NUMBER_OF_SESSIONS
PROTOCOL
  /home/span/span/testsuite/results/ABE.if
GOAL
  as_specified
BACKEND
  OFMC
COMMENTS
STATISTICS
  parseTime: 0.00s
  searchTime: 0.23s
  visitedNodes: 1102 nodes
  depth: 10 plies
```

(b)

Figure 5: Simulation results of ABE scheme using CL-AtSe and OFMC backend

AVISPA serves as a singular interface solution facilitating mutual authentication of communication protocols. Throughout transmission, it furnishes a role-specific and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:53:45 UTC from IEEE Xplore. Restrictions apply.

comprehensive validation mechanism. The output format of AVISPA is accessible via one of its four back-end systems, namely TA4SP, SATMC, CL-AtSe, and OFMC as shown in Figure 4. In AVISPA the resultant report provides information about the tested protocol in terms of summary, details, protocol, backend, and statistics Figure 5.

SUMMARY: Evaluate the safety status of the protocol, determining whether it is safe, unsafe, or if the analysis outcome remains inconclusive.

DETAILS: Describe the reason behind affirming the protocol's safety, describe the criteria exposing it to potential vulnerabilities, or illustrate the reasons contributing to an inconclusive analysis.

PROTOCOL: encapsulates the HLPSL specification of the protocol in an intermediary format.

GOAL: involves the utilization of HLPSL specifications by AVISPA in the execution of its assessments.

BACKEND: denotes the specific backend employed for conducting the analysis.

VI. Results and Analysis

The proposed approach is replicated within the High-Level Protocol Specification Language (HLPSL), specifying the distinct roles for the data owner, data consumer, Blockchain, and attribute authority, alongside essential roles such as session, environment, and goals. This segment will offer a concise overview of the procedural sequence adopted in our implementation, along with a summary of the obtained results. Within this model, the data owner initiates attribute requests and formulates the access policy (AC).

In the context of the Attribute-Based Encryption (ABE) scheme discussed earlier, access to data or messages is restricted solely to users possessing a set of attributes aligning with the key, thereby satisfying the designated access policy. After acquiring the public key of the system(PKs) from the attribute authority, the data owner proceeds to encrypt the message, transmitting *SND(exp(exp(M,AC'),PKs'))* to the Blockchain. The operator *exp* is a special operator if a file/message has the properties *exp(exp(A,B),C)= exp(exp(A,C),B)* and *exp(exp(A,B),inv(B))=A*. The data consumer here named Alice, transmits her user identity *UID,* alongside specific attributes *Attr*, to the attribute authority. Subsequently, the attribute authority furnishes the user with a Global Identity *GID*, aimed at mitigating collusion concerns within ABE when operating beyond simulated environments.

```
role environment()
def=
        const
                bob,attribute_authority,blockchain,alice:agent,
                kaa,kdc:public_key,m:text,attrrq:text,ac:text,
                attr:text,sec_1:protocol_id
        intruder_knowledge={bob,attribute_authority,blockchain,alice}
composition
        session(bob,attribute_authority,blockchain,alice,kaa,kdc,m,attr)
        /\session(bob,attribute_authority,blockchain,alice,kaa,kdc,m,attr)
end role
goal
        secrecy_of sec_1
end goal
environment()
```

(a)

```
SUMMARY
  SAFE
DETAILS
  BOUNDED_NUMBER_OF_SESSIONS
  TYPED_MODEL
PROTOCOL
  /home/span/span/testsuite/results/ABE.if
GOAL
  As Specified
BACKEND
  CL-AtSe
STATISTICS
  Analysed   : 54962 states
  Reachable  : 54832 states
  Translation: 0.00 seconds
  Computation: 1.98 seconds
```

(b)

Figure 6: Simulation results of ABE scheme for two sessions

Assigning a distinct Global Identity *GID* to each user in correspondence with their specific attributes effectively eradicates the potential for collusion between two users. The data user sends *SND(GID',PWD'),* that is Global Identity and associated password to the Blockchain. Following verification, the Blockchain securely transmits encrypted data to the authenticated data user. To realize the objective of maintaining secrecy, the declaration of *secrecy_of* includes the parameter *sec_1* allocated to the entities of both the Blockchain and the data consumer Figure 6.

HLPSL applies the Dolev-Yao threat model, and subsequently, AVISPA assesses the security protocol's viability, determining its safety or susceptibility. Within AVISPA, an active role is assumed by an intruder *i* during communication Figure 7. For formal security verification under the SPAN in AVISPA Figure 6, In this study, the widely used OFMC and CL-AtSe back-ends are employed. Results from TA4SP and SATMC backends are disregarded due to the current limitation of the HLPSL implementation, which does not recognize the *exp* operation. The protocol additionally guarantees the freshness property within the tested protocol. In the course of the protocol session, a message qualifies as fresh when it has been exclusively constructed for that particular protocol session. The study provides a secure data exchange between the owner and the consumer and has been compared with [19], showing that our scheme is safe and satisfies the secrecy, and freshness aspects of cryptographic protocol.

VII. Conclusion

In this study, our model, formulated using the HLPSL language, has been simulated within the AVISPA SPAN tool. The primary objective was to ensure the security of sensitive information exchanged between entities, Bob as the data owner and Alice as the data consumer. We analyzed the data's secrecy between the Blockchain and the data consumer in our model. Additionally, we focused on preserving data freshness across multiple sessions during model execution. In the future, our main goal will be to conduct a comprehensive examination of prospective instances of replay attacks and man-in-the-middle (MITM) attacks and also analyze the attack traces comprehensively and thoroughly with precise attention to detail.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:53:45 UTC from IEEE Xplore. Restrictions apply.

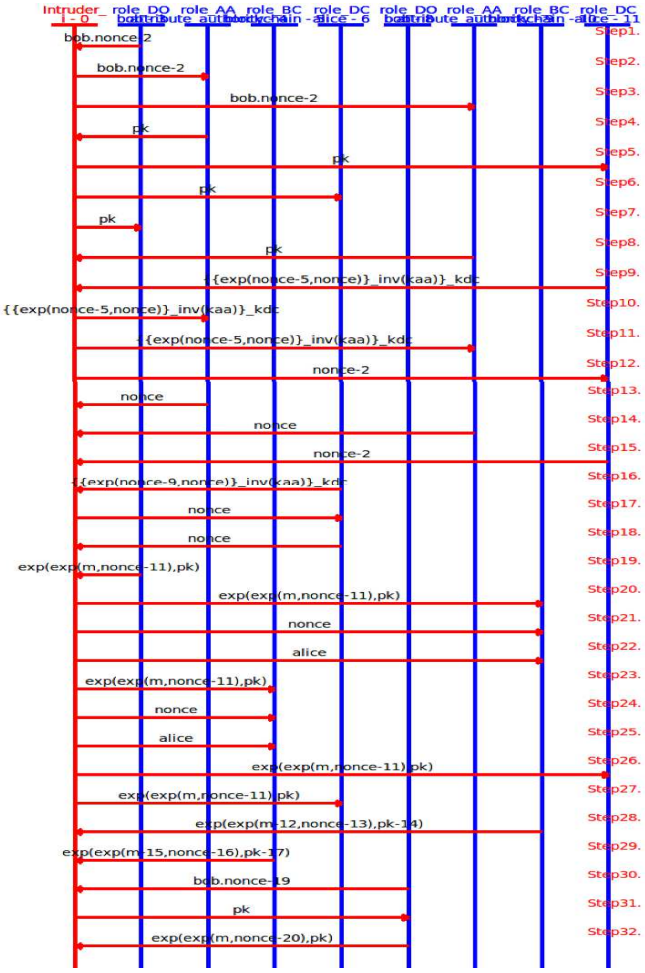


Figure 7: Intruder Simulation result of our scheme in the AVISPA using CL-AtSe code backend

REFERENCES

[1] A. Shamir, "Identity-based cryptosystems and signature schemes," Springer, 1985, pp. 47–53.

[2] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in Annual international cryptology conference, Springer, 2001, pp. 213–229.

[3] A. Lewko and B. Waters, "Decentralizing attribute-based encryption," in Annual international conference on the theory and applications of cryptographic techniques, Springer, 2011, pp. 568–588.

[4] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in Advances in Cryptology–EUROCRYPT 2005: 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005. Proceedings 24, Springer, 2005, pp. 457–473.

[5] X. Si, P. Wang, and L. Zhang, "Kp-abe based verifiable cloud access control scheme," in 2013 12th IEEE International Conference on Trust, Security, and Privacy in Computing and Communications, IEEE, 2013, pp. 34–41.

[6] J. Lai, R. H. Deng, and Y. Li, "Fully secure ciphertext policy hiding cp-abe," in Information Security Practice and Experience: 7th International Conference, ISPEC 2011, Guangzhou, China, May 30–June 1, 2011. Proceedings 7, Springer, 2011, pp. 24–39.

[7] Y. Rahulamathavan, R. C.-W. Phan, M. Rajarajan, S.Misra, and A. Kondoz, "Privacy-preserving blockchain-based iot ecosystem using attribute-based encryption," in 2017 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), IEEE, 2017, pp.1–6.

[8] A. Wu, Y. Zhang, X. Zheng, R. Guo, Q. Zhao, and D. Zheng, "Efficient and privacy-preserving traceable attribute-based encryption in blockchain," Annals of Telecommunications, vol. 74, pp. 401–411,2019.

[9] H. Zheng, J. Shao, and G. Wei, "Attribute-based encryption with outsourced decryption in blockchain," Peer-to-peer networking and applications, vol. 13, pp. 1643–1655, 2020.

[10] G. Verma and S. Kanrar, "Secure document sharing model based on blockchain technology and attribute-based encryption," Multimedia Tools and Applications, pp.1–18,2023.

[11] T. Feng, F. Kong, C. Liu, and Y. Lu, "Multi-authorization attribute-based verifiable encryption scheme based on blockchain," Mobile Networks and Applications, pp. 1–8, 2023.

[12] M. Almasian and A. Shafieinejad, "Secure cloud file sharing scheme using blockchain and attribute-based encryption," Computer Standards & Interfaces, vol. 87,p. 103 745, 2024.

[13] R. Ostrovsky, A. Sahai, and B. Waters, "Attribute-based encryption with non-monotonic access structures," in Proceedings of the 14th ACM conference on Computer and Communications Security, 2007, pp. 195–203.

[14] D. Boneh and X. Boyen, "Efficient selective-id secure identity-based encryption without random oracles," in Advances in Cryptology-EUROCRYPT 2004: International Conference on the Theory and Applications of Cryptographic Techniques, Interlaken, Switzerland, May 2-6, 2004. Proceedings 23, Springer, 2004, p. 223–238.

[15] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in Proceedings of the 13th ACM conference on Computer and Communications Security, 2006, pp. 89–98.

[16] A. Beimel et al., "Secure schemes for secret sharing and key distribution," 1996.

[17] M. Rasori, M. La Manna, P. Perazzo, and G. Dini, "A survey on attribute-based encryption schemes suitable for the Internet of Things," IEEE Internet of Things Journal, vol. 9, no. 11, pp. 8269–8290, 2022.

[18] The AVISPA project. [Online]. Available: https://www.avispa-project

[19] H. Garg and M. Dave, "Securing user access at iot middleware using attribute-based access control," in 2019 10th International Conference on Computing, communication and Networking Technologies (ICCCNT), IEEE,2019, pp. 1–6

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:53:45 UTC from IEEE Xplore. Restrictions apply.

Journal of Electronic Materials
https://doi.org/10.1007/s11664-024-11144-3

ORIGINAL RESEARCH ARTICLE

Characterization of Luminescent $Li_3Ba_2Gd_3Mo_8O_{32}:Tb^{3+}$ Phosphors Prepared via Sol–Gel Method

Vijay Singh[1] · B. R. V. Rao[2] · Aadil Ahmad Bhat[1] · Ch. B. Annapurna Devi[3] · A. S. Rao[4] · Ji Bong Joo[1]

Received: 17 February 2024 / Accepted: 23 April 2024
© The Minerals, Metals & Materials Society 2024

Abstract

Luminescent $Li_3Ba_2Gd_3Mo_8O_{32}:Tb^{3+}$ phosphors were synthesized using the sol–gel method, and extensively characterized through various spectroscopic techniques, including x-ray diffraction (XRD), scanning electron microscopy (SEM), Fourier-transform infrared (FT-IR) spectroscopy, photoluminescence (PL) excitation, and emission. SEM micrographs revealed the irregular and non-uniform morphology and size of the prepared phosphors. The FT-IR spectra offered insights into the functional groups and bonding within the phosphor. The PL excitation spectra at a 545 nm emission wavelength displayed a slight shift in peak positions with a prominent peak at 283 nm, corresponding to the characteristic $4f^8 \rightarrow 4f^75d^1$ transitions of Tb^{3+} ions, while under excitation at $\lambda_{exc} = 283$ nm, PL emission spectra exhibited distinct peaks associated with $^5D_4 \rightarrow {}^7F_j$ ($j = 6, 5, 4, 3$) transitions. Notably, the transition $^5D_4 \rightarrow {}^7F_5$ at 545 nm demonstrated strong and particularly high intensity in the $Li_3Ba_2Gd_3Mo_8O_{32}:0.09Tb^{3+}$ phosphor. Photometric analysis showed the potential of this phosphor to serve as a vivid green light source for various display applications. The $Li_3Ba_2Gd_3Mo_8O_{32}:Tb^{3+}$ phosphor may be used for creating sensors for environmental monitoring. Future research could explore its biocompatibility for biomedical applications and improve its targeting and photostability.

Keywords Sol–gel · Tb^{3+} ions · $Li_3Ba_2Gd_3Mo_8O_{32}$ · Luminescence

Introduction

White light-emitting diodes (LEDs) are recognized as a fourth-generation lighting source due to their numerous advantages, which include high efficiency, extended lifespan, rapid response, safety, and environment-friendly nature.[1–4] Notably, phosphor-converted white LEDs have in recent times become commercially viable and are extensively deployed as white LEDs. Consequently, the pursuit of novel phosphor materials, specifically phosphors doped with Ln^{3+} ions and tailored for use in white LEDs has become a dynamic and vibrant area of research.[5–8]

Trivalent terbium ions (Tb^{3+}) have garnered significant attention due to their primary emission bands falling within the visible region. This emission results from transitions $^5D_4 \rightarrow {}^7F_j$ ($j = 3, 4, 5, 6$) when subjected to violet or near-ultraviolet illumination.[9] Researchers have found that the unique luminous properties and wide range of applications of Tb^{3+} -doped phosphor materials have generated significant interest across various industries.[10–12] Due to their strong green emission upon excitation, phosphors doped with $Tb^{3+}$are useful for biomedical imaging, lighting, and display technologies. To customize the emission properties and improve the functionality of Tb^{3+} -doped phosphors, researchers have investigated a variety of host matrices, such as oxides, sulfides, and fluorides. Baig et al.[13] reported that, by integrating Tb^{3+} ions into various host materials, these phosphors can be used for fluorescent labeling in biological assays, solid-state lighting, and LED color conversion. They also offer tunable emission wavelengths, improved quantum efficiency, and increased stability.[13] Garcia-Cortes et al.[14] reported Yb^{3+} -doped $Li_{0.75}Gd_{0.75}Ba_{0.5}(MoO_4)_2$ crystals

✉ Vijay Singh
vijayjiin2006@yahoo.com

1 Department of Chemical Engineering, Konkuk University, Seoul 05029, Republic of Korea

2 Department of Physics, V.R. Siddhartha Engineering College, Vijayawada 52007, India

3 Department of Physics, Vignan's Lara Institute of Technology and Science, Vadlamudi, Guntur, Andhra Pradesh 522 213, India

4 Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi 110042, India

Published online: 28 May 2024

with small quantum defects, while Song et al.[15] reported the photoluminescence characterization of Nd^{3+} -doped $Li_3Ba_2Gd_3(MoO_4)_8$.

The crystal of $Li_3Ba_2Gd_3(MoO_4)_8$, which is part of the $Li_3Ba_2Ln_3(MoO_4)_8$ family (where Ln = La-Lu) is characterized by a monoclinic crystal system with the space group C2/c.[16] In this study, $Li_3Ba_2Gd_3Mo_8O_{32}$ was selected as the host material due to its outstanding luminescent properties. Lithium possesses unique characteristics that set it apart from other elements within its group, and molybdenum exhibits a range of valence states, encompassing Mo^{3+}, Mo^{4+}, Mo^{5+}, and Mo^{6+}. When molybdenum forms compounds with other elements, it gives rise to intriguing electrical and optical phenomena. Rare-earth-doped phosphors offer exceptional chemical, optical, and electronic properties that make them ideal for a wide range of applications, including luminosity and display areas.[16,17]

The uniform distribution of dopants in the phosphor matrix and precise control over their chemical composition are facilitated by the sol–gel method. Verma et al.[18] demonstrated that phosphors produced through sol–gel exhibited superior homogeneity compared to phosphors manufactured using other methods. This low-temperature processing technique allows for the preservation of the luminescence of sensitive dopants and enables the incorporation of organic additives. The sol–gel method is highly recommended for phosphor synthesis, particularly in applications that require precise control over composition, morphology, and optical properties.[19,20]

The primary objectives of this study include the examination of structural phenomena, assurance of phase purity, analysis of surface morphology, characterization of bond properties, and study of luminescent properties. Additionally, the research aims to determine the CIE chromaticity coordinates and color purity, and to provide microscopic insights into the $Li_3Ba_2Gd_3Mo_8O_{32}$:xTb^{3+} phosphor network. To achieve these goals, the investigation involves spectroscopic techniques, including x-ray diffraction (XRD; Rigaku), Fourier-transform infrared (FT-IR; Nicolet 6700; Thermo Fisher) spectroscopy, scanning electron microscopy (SEM; S-3400; Hitachi), photoluminescence (PL; RF-5301PC; Shimadzu) excitation, and emission spectroscopy, as well as photometric analysis. Future research can improve the properties of $Li_3Ba_2Gd_3Mo_8O_3$:Tb^{3+} phosphor for different sensor applications. Additional studies are needed to enhance its performance.

Materials preparation and analysis

Phosphors with the chemical formula $Li_3Ba_2Gd_3Mo_8O_{32}$:xTb^{3+} (x = 0.005–0.13 mol) were synthesized using a sol–gel procedure. Stoichiometric quantities of the following compounds were employed: $LiNO_3$ for lithium, $Ba(NO_3)_2$ for barium, $Gd(NO_3)_3{\cdot}6H_2O$ for gadolinium, $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$ for molybdenum, SiO_2 for silicon, $Tb(NO_3)_3{\cdot}xH_2O$ for terbium, and citric acid. Citric acid was added to the mixture at a molar ratio of 3.5:1 to the metal ions. The weighed compounds were homogeneously mixed with 10 mL of distilled water in a 150-mL beaker, using a magnetic stirrer. After 1 h of stirring, the transparent solution was dried in an oven at 112°C overnight. The resulting dried sol was subsequently crushed and transferred to a 50-mL porcelain crucible, which was then placed in a 400°C furnace for 2 h. The black precursors obtained were subjected to further heat treatment at 880°C for 3 h in an air atmosphere to complete the synthesis of the samples. Figure 1 shows the synthesis process.

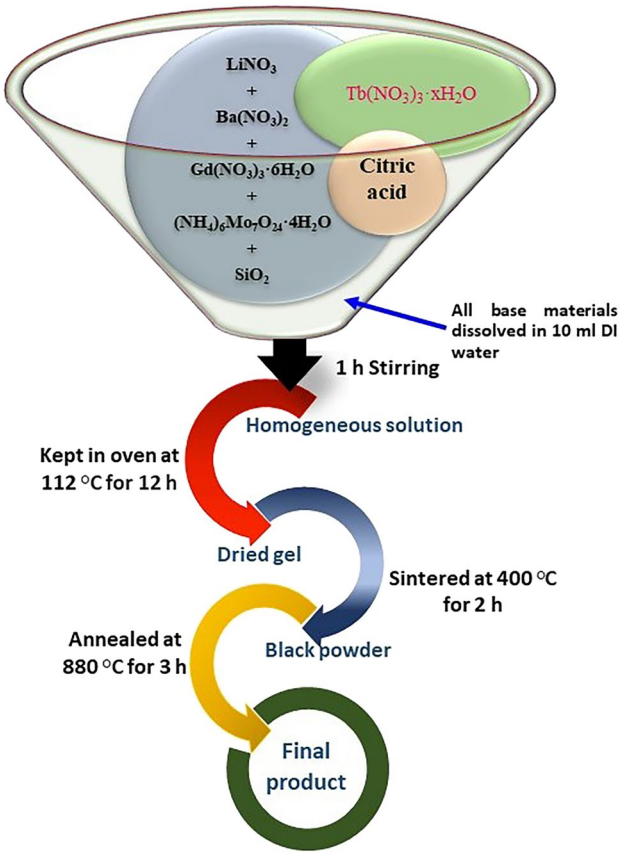


Fig. 1 The entire preparation scheme of $Li_3Ba_2Gd_3Mo_8O_{32}$:xTb^{3+}.

The phase of the prepared samples was characterized by XRD, the morphological details were investigated through SEM, the FT-IR spectra of the samples were analyzed, and the PL excitation and emission spectra were recorded.

Results and Discussion

Crystal structures and XRD examination

Figure 2a–d illustrates the standard orientation of the crystal structure of $Li_3Ba_2Gd_3Mo_8O_{32}$ and its orientation within the (011), (010), and (100) planes, which were generated using VESTA software. These observations confirm the monoclinic symmetry of $Li_3Ba_2Gd_3Mo_8O_{32}$. The structure of the $Li_3Ba_2Gd_3Mo_8O_{32}$ phosphor comprises honeycomb layers of Gd polyhedra with 8 vertices, with each layer being connected to two Mo tetrahedra featuring oxygen corners on both sides. The Li atoms are found at two distinct crystallographic positions: one-third of them occupy rare-earth sites with a coordination number of 8, while the remaining two-thirds occupy a specific position on a twofold axis with octahedral-coordinated oxygen.[21] Barium atoms possess a coordination number of 10 and are situated between the Mo and Gd polyhedral layers.

XRD pattern of the $Li_3Ba_2Gd_3Mo_8O_{32}:xTb^{3+}$ phosphors were compared with the Joint Committee on Powder Diffraction Standards (JCPDS) card 77-0830, as depicted in Fig. 3. These XRD patterns exhibit excellent agreement with all the peaks present in the reference pattern. Notably, the XRD patterns of all the prepared phosphors remain consistent across all Tb^{3+} concentrations, thereby confirming the presence of a monoclinic deformed scheelite structure in the samples with the

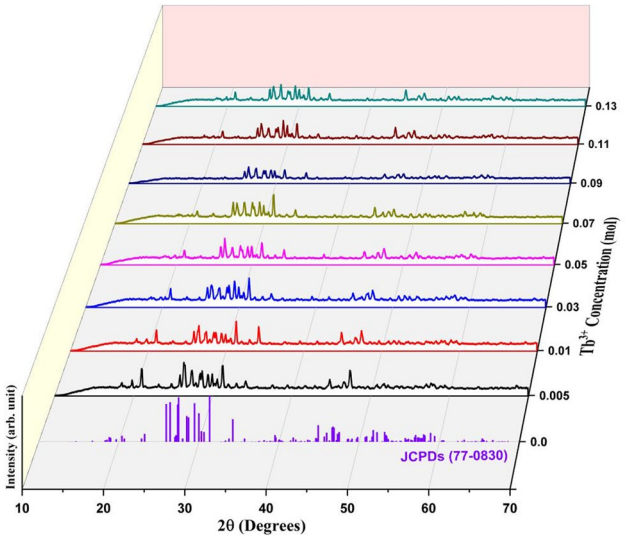


Fig. 3 XRD patterns of $Li_3Ba_2Gd_3Mo_8O_{32}:xTb^{3+}$ phosphors.

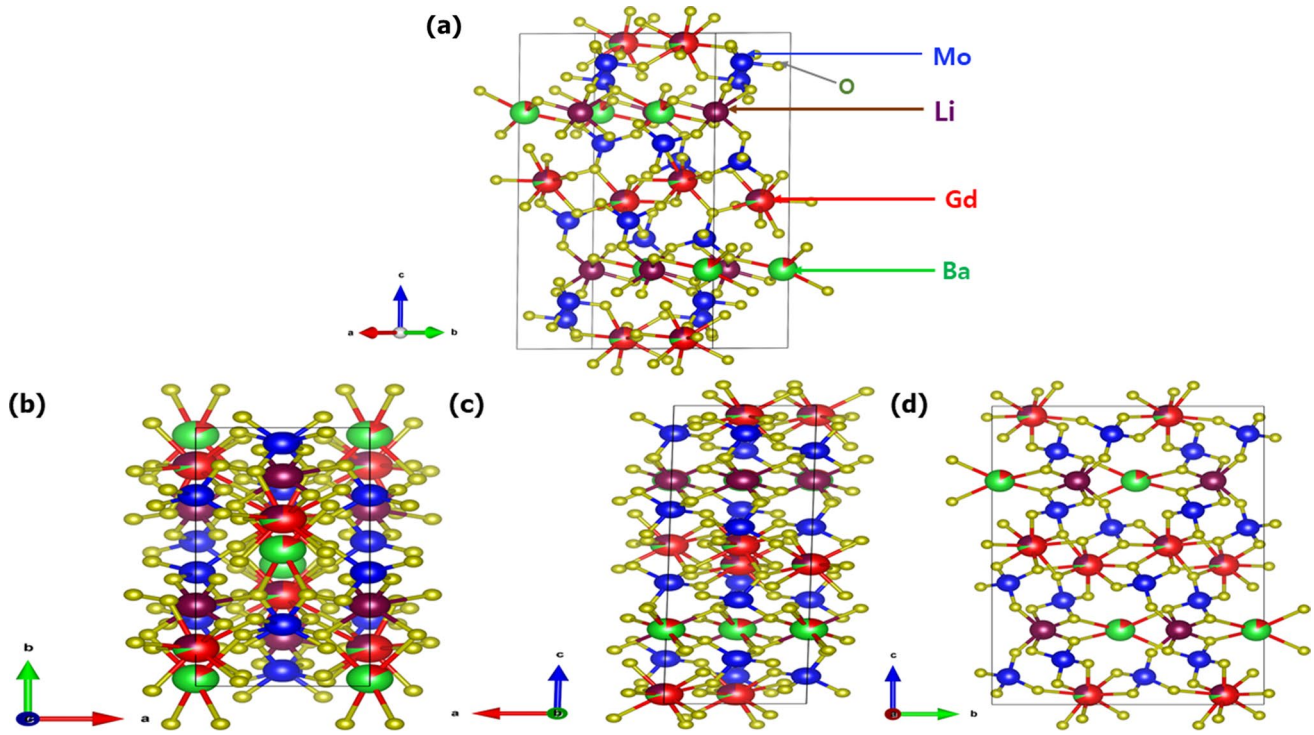


Fig. 2 $Li_3Ba_2Gd_3Mo_8O_{32}$ crystal structure: (a) standard orientation, (b–d) three different orientations of 001, 010, and 100.

C2/c space group.[22] The unit-cell parameters for the $Li_3Ba_2Gd_3Mo_8O_{32}$ single crystal were determined as follows: a = 5.341 Å, b = 12.953 Å, c = 19.367 Å, and β = 91.27 Å. By analyzing the XRD spectra, the full width at half-maximum (FWHM) values were calculated corresponding to the diffraction angles for the most intense peaks observed in all the prepared samples. Utilizing these FWHM values in the Debye–Scherrer equation,[23] the average crystallite size of the prepared samples was determined to be less than 50 nm. Table I presents the lattice strain and crystallite size.

Table I FWHM, crystallite size, and lattice strain of $Li_3Ba_2Gd_3Mo_8O_{32}$:xTb^{3+} phosphor

S. no.	Composition	FWHM	Crystallite size	Lattice strain
1.	0.005	0.22	39.18	0.0034
2.	0.01	0.18	47.89	0.0028
3.	0.03	0.20	43.10	0.0031
4.	0.05	0.24	35.92	0.0037
5.	0.07	0.20	43.10	0.0031
6.	0.09	0.22	39.17	0.0034
7.	0.11	0.19	45.36	0.0030
8.	0.13	0.20	43.10	0.0031

Morphological analysis using SEM

Figure 4a–d presents the SEM micrographs of the $Li_3Ba_2Gd_3Mo_8O_{32}$:0.09Tb^{3+} phosphors at various magnifications (×1000, ×4000, ×8000, and ×15,000). The particle size and surface morphology of phosphors are significant features that play a vital role in determining luminescence efficiency. In the present study, the particles exhibited irregular and non-uniform shapes, and their diameters range was 2–8 μm.[24] The irregular morphology observed in the SEM micrographs of the $Li_3Ba_2Gd_3Mo_8O_{32}$:0.09Tb^{3+} phosphors is likely due to the agglomeration of particles with non-uniform shapes. The agglomeration process can be influenced by various factors, such as the synthesis method, reaction conditions, and the nature of the materials involved. These conditions may lead to the formation of particles with irregular shapes and uneven surfaces. The irregular morphology observed in the $Li_3Ba_2Gd_3Mo_8O_{32}$:0.09Tb^{3+} phosphors can be attributed to the sol–gel synthetic process, in which the formation of the material occurs through the evolution of a sol, which is a colloidal suspension of small particles in a liquid, into a gel, a three-dimensional network of interconnected particles. Several factors during the sol–gel synthesis can contribute to the irregular morphology, like

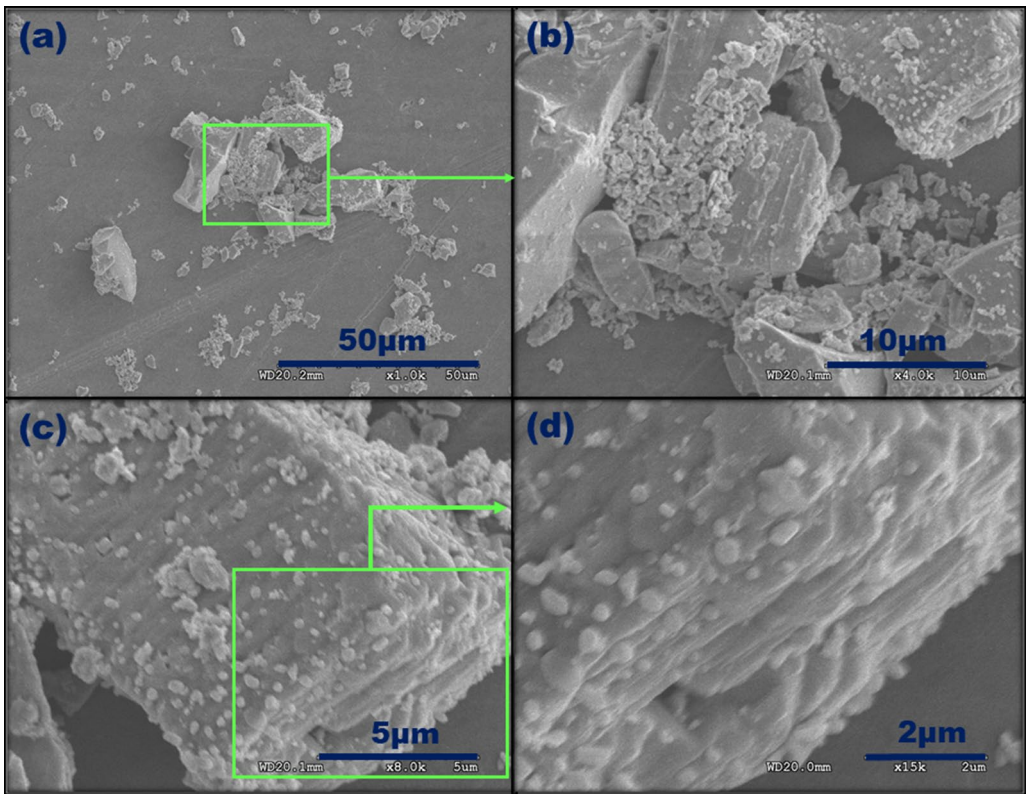


Fig. 4 Scanning electron microscopy (SEM) images, at different magnifications: (a) 1 K, (b) 4 K, (c) 8 K, and (d) 15 K of $Li_3Ba_2Gd_3Mo_8O_{32}$:0.09Tb^{3+} phosphors; (b) is the magnified view of the green zone marked in (a), and (d) is the magnified view of the green zone marked in (c) (Color figure online).

agglomeration during gelation, non-uniform drying, particle size distribution, chemical inhomogeneity, and reaction kinetics. This characteristic makes these phosphors well-suited to commercial LED fabrication, aligning with the desired particle size range of 5–10 μm.[25]

FT-IR spectral analysis

Figure 5 shows the FT-IR spectra of the $Li_3Ba_2Gd_3Mo_8O_{32}$:0.005Tb^{3+} and $Li_3Ba_2Gd_3Mo_8O_{32}$:0.09Tb^{3+} phosphors. The bands observed at 3400–3450 cm^{-1}, 1600–1650 cm^{-1}, and 1520–1500 cm^{-1} are attributed to the O-H stretching vibrations and H-O-H bending vibrations, respectively.[26–28] The presence of an asymmetric stretching band around 1400 cm^{-1} is indicative of C=O vibrations.[29] The Mo-O stretching frequency of MoO_6 octahedra is evident in the spectra, appearing in the 1000–970 cm^{-1} range.[30] Vibrations from isolated MoO_4 units may contribute to the shoulder around 890 cm^{-1}.[31] The intense absorption peaks at 840 cm^{-1} and in the 790–740 cm^{-1} range correspond to the stretching vibration of O-Mo-O in the MoO_4^{2-} tetrahedron.[30] The existence of M-O (where M is a metallic element) is confirmed by the appearance of absorption bands at 700 cm^{-1} and 630 cm^{-1}.[32,33] Vibrations involving the modification of cations concerning barium atoms are associated with absorption bands at 600 cm^{-1} or to lower bands in the range 570–530 cm^{-1}, indicative of typical metal–oxygen (M-O) lattice vibrations.[34,35] The vibration observed at 460 cm^{-1} is linked to the bending mode vibration of M-O-M, while the bending modes of Mo-O are associated with a faint absorption band around 420 cm^{-1}.[36] A comprehensive FT-IR spectral investigation confirms the formation of the prepared phosphor.

PL excitation and emission spectral analysis

The luminescent characteristics of the synthesized phosphors were investigated through PL spectra analysis. Figure 6 shows the PL excitation spectra of the $Li_3Ba_2Gd_3Mo_8O_{32}:xTb^{3+}$ phosphors that were examined within the 220 to 350-nm range (UV region), with a green light source of 545-nm wavelength ($^5D_4 \rightarrow {}^7F_5$) employed for emission. The obtained spectra reveal a single broad hump, with distinct peak positions corresponding to different doping concentrations. This observation suggests that the PL excitation transitions originate from the ground state $^7F_6$ to various excited states within the charge transfer band (CTB). A broad excitation peak is observed around 280 nm, which corresponds to the CTB.[37] The PL excitation peaks exhibit a notable shift from 279.99 to 293.48 nm with an increase in Tb^{3+} concentration from 0.005 to 0.13 mol, as depicted in Fig. 6. This shift toward longer wavelengths indicates a red shift of 13.49 nm in the CTB. Excitation intensities in the PL spectra show a linear correlation with the quantity of Tb^{3+} ions, suggesting an unaffected crystal field strength. Notably, the $Li_3Ba_2Gd_3Mo_8O_{32}$:0.09Tb^{3+} phosphor exhibits its maximum excitation intensity at 283 nm, corresponding to the spin-allowed inter-configurational $4f^8 \rightarrow 4f^75d^1$ transitions of Tb^{3+} ions within the host lattice.

The PL emission spectra of the $Li_3Ba_2Gd_3Mo_8O_{32}:xTb^{3+}$ phosphors were recorded in the range 400–650 nm under UV excitation (283 nm), as shown in Fig. 7. The PL emission spectra exhibit four distinct emission bands with

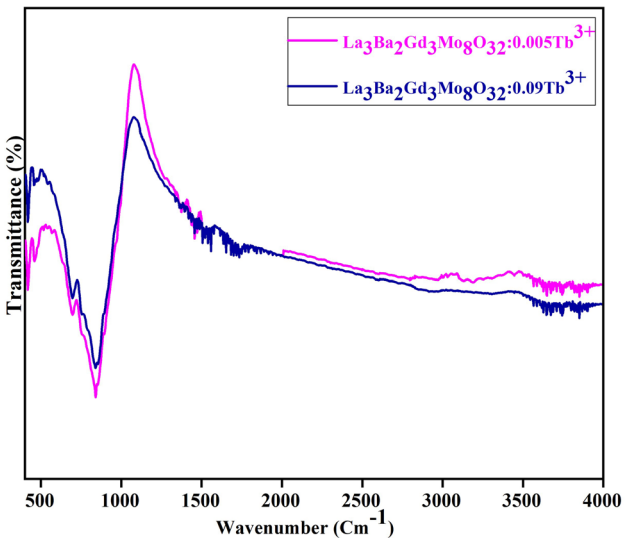


Fig. 5 FT-IR spectra of $Li_3Ba_2Gd_3Mo_8O_{32}$:0.005Tb^{3+} and $Li_3Ba_2Gd_3Mo_8O_{32}$:0.09Tb^{3+} phosphors.

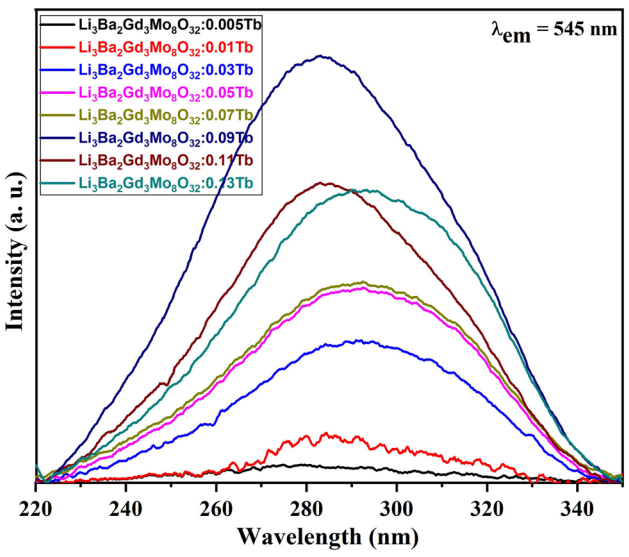


Fig. 6 Photoluminescence (PL) excitation spectra of $Li_3Ba_2Gd_3Mo_8O_{32}:xTb^{3+}$ (λ_{em} = 545 nm).

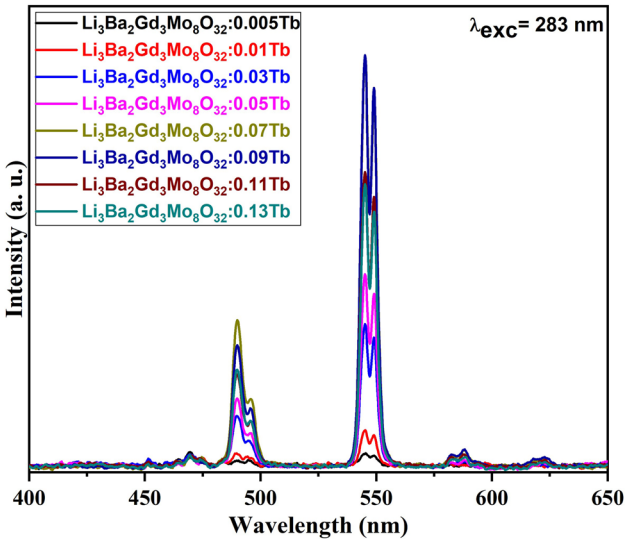


Fig. 7 Photoluminescence emission spectra of $Li_3Ba_2Gd_3Mo_8O_{32}$:xTb^{3+} (λ_{exc} = 283 nm).

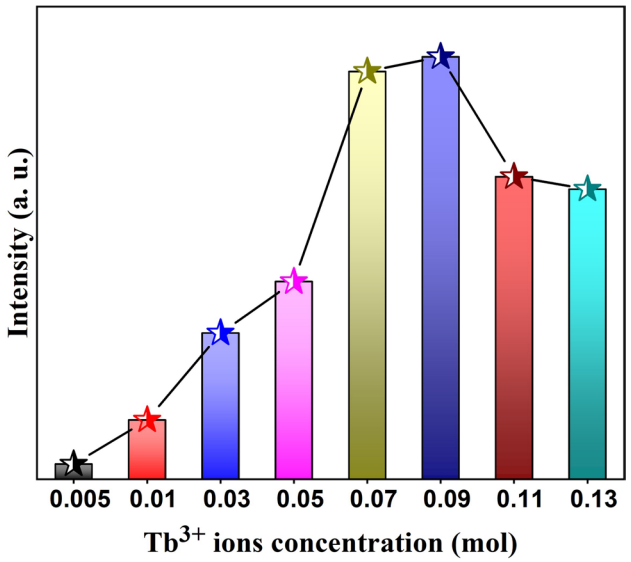


Fig. 8 Tb^{3+} concentration variation on emission intensity and increment of intensity with concentration (at 545-nm strong emission).

peaks located at 490, 545, 588, and 622 nm, corresponding to the transitions $^5D_4 \rightarrow$ ($^7F_6$, $^7F_5$, $^7F_4$, and $^7F_3$), respectively.[38] These transitions were assigned following the framework established by Carnall et al.[39]. Notably, the $^5D_4 \rightarrow {}^7F_5$ transition at 545 nm (green) is prominently intensified across all phosphors. The relationship between the emission intensity of the strong $^5D_4 \rightarrow {}^7F_5$ transition at 545 nm and the concentration of Tb^{3+} ions in the host is depicted in Fig. 8. The analysis reveals a gradual increase in emission intensity with the rising concentration of Tb^{3+} ions up to 0.09 mol, followed by a significant reduction. This observed behavior is attributed to concentration-quenching, a phenomenon arising from non-radiative energy transfer between adjacent luminescent centers (Tb^{3+} ions). In line with Blasse's theory, this quenching effect is contingent on the critical distance (R_c) between the nearest activators (Tb^{3+} ions). For the optimum phosphor composition of the $Li_3Ba_2Gd_3Mo_8O_{32}$:0.09Tb^{3+}, the critical separation R_c can be evaluated with the help of the Blasse formula[40]:

$$R_c = 2\left[\frac{3V}{4\pi x_c Z}\right]^{\frac{1}{3}} \quad (1)$$

Using $\frac{V}{Z}$= 629.1 Å^3, V = cell volume, Z = number of cations per cell, and x_c = 0.09 (optimum concentration) for the $Li_3Ba_2Gd_3Mo_8O_{32}$:0.09Tb^{3+} phosphor, R_c was calculated to be 23.73 Å.

The non-radiative relaxations between activator ions in the $Li_3Ba_2Gd_3Mo_8O_{32}$:0.09Tb^{3+} phosphor can be categorized based on the critical distance (R_c) value. In this context, the R_c for Tb^{3+} ions significantly exceed 5 Å.

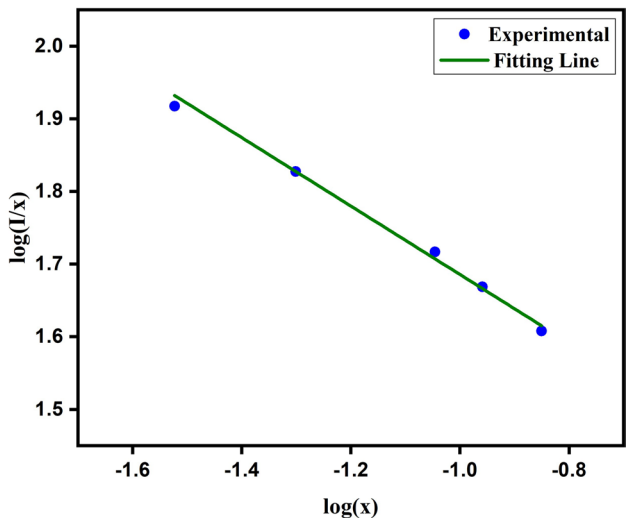


Fig. 9 log (I/x) versus log (x) plotting for $Li_3Ba_2Gd_3Mo_8O_{32}$:xTb^{3+} phosphor.

This observation supports the conclusion that electric multipole–multipole interactions are the dominant mechanism in the concentration-quenching process, as indicated by the literature.[41] The nature of multipolar interaction can be elucidated through the application of Dexter's theory.[42] According to this theory, the luminescence intensity (I) per activator concentration (x) can be expressed as:

$$\frac{I(x)}{x} = \frac{C}{[1+\beta.x^{\frac{P}{3}}]} \quad (2)$$

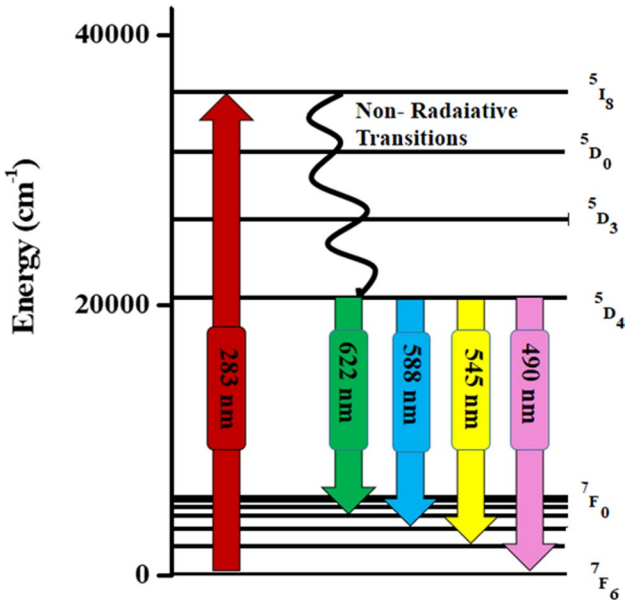


Fig. 10 Possible transition of trivalent terbium ions.

where C and β are constants for a particular host, and P is the multipolar interaction constant ($P = 6$ for dipole–dipole, $P = 8$ for dipole–quadrupole, and $P = 10$ for quadrupole–quadrupole interactions). The determination of P value involves the organization of a graph between $\log(I/x)$ and $\log(x)$, as illustrated in Fig. 9. This graph manifests as a straight line with a slope of $(-P/3)$, leading to a calculated P value close to 6 (specifically 5.9232). This outcome strongly suggests that dipole–dipole interactions are accountable for the non-radiative relaxations, thereby confirming the occurrence of concentration quenching when the Tb^{3+} ions concentration exceeds 0.09 mol in the $Li_3Ba_2Gd_3Mo_8O_{32}$ phosphors.[43] In the typical energy level diagram, depicted in Fig. 10, the excitation and emission transitions of Tb^{3+} ions in the $Li_3Ba_2Gd_3Mo_8O_{32}$ host are represented. All the samples exhibit a singular excitation transition at 283 nm, originating from the $^7F_6$ ground state to a higher energy band with a $4f^75d^1$ configuration. Additionally, four prominent non-radiative transitions occur from the $^5D_4$ level to lower energy levels $^7F_6$, $^7F_5$, $^7F_4$, and $^7F_3$ at 490 nm, 545 nm, 588 nm, and 622 nm, respectively. Figure 10 shows these crucial transitions, along with their corresponding wavelengths, in the partial energy level diagram of Tb^{3+} ions.

Photometric analysis

Photometric analysis has been utilized to discern the color emitted by the phosphors when subjected to specific lighting conditions. Figure 11 presents typical images of two phosphors: (1) $Li_3Ba_2Gd_3Mo_8O_{32}:0.005Tb^{3+}$, and (2) $Li_3Ba_2Gd_3Mo_8O_{32}:0.09Tb^{3+}$, captured under room light (Fig. 11a, a') and 254-nm UV light (Fig. 11b, b'). Both phosphor powders exhibit a white appearance under daylight. However, when exposed to UV illumination at 254 nm, the $Li_3Ba_2Gd_3Mo_8O_{32}:0.005Tb^{3+}$ phosphor appears dull-green, while the $Li_3Ba_2Gd_3Mo_8O_{32}:0.09Tb^{3+}$ phosphor emits a green hue. Photometric analysis conducted under UV excitation substantiates the green color emission of the $Li_3Ba_2Gd_3Mo_8O_{32}:0.09Tb^{3+}$ phosphor.

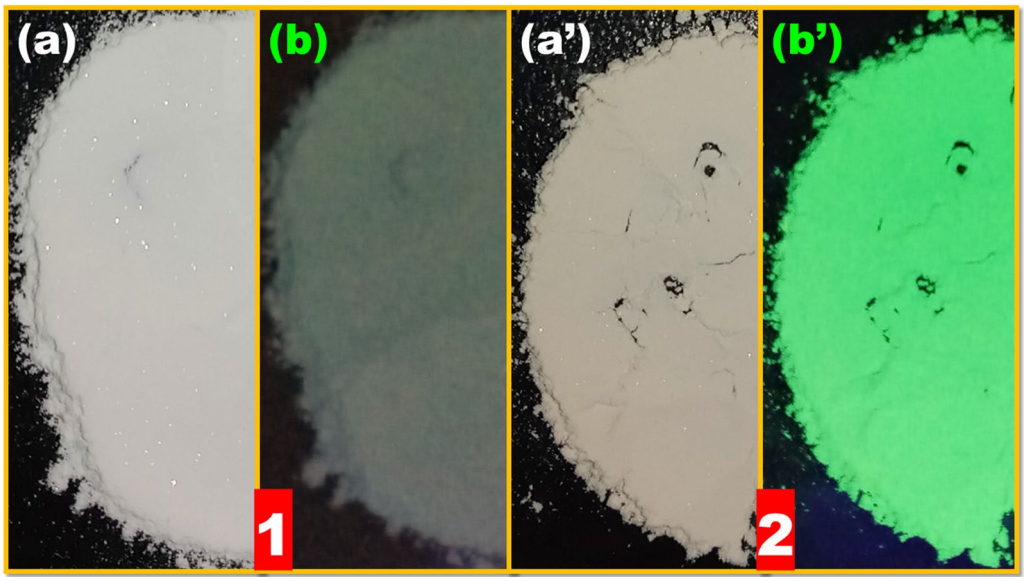


Fig. 11 Typical images of (1) $Li_3Ba_2Gd_3Mo_8O_{32}:0.005Tb^{3+}$ and (2) $Li_3Ba_2Gd_3Mo_8O_{32}:0.09Tb^{3+}$: (a, a') white powder image in direct room light and (b, b') green visible image under 254-nm ultraviolet lamp (Color figure online).

Table II CIE color coordinates for $Li_3Ba_2Gd_3Mo_8O_{32}$:xTb^{3+} phosphors utilizing PL emission results (λ_{exc}= 283 nm)

Phosphors $Li_3Ba_2Gd_3Mo_8O_{32}$:xTb^{3+}	λ_{exc}= 283 nm		Color purity %: (Sign)
	x	*y*	
x = 0.005	0.2163	0.3756	27.36 (●)
x = 0.01	0.2409	0.4869	45.39 (●)
x = 0.03	0.2385	0.5690	64.75 (●)
x = 0.05	0.2435	0.6177	76.08 (●)
x = 0.07	0.2543	0.6281	78.08 (●)
x = 0.09	0.2618	0.6358	79.66 (●)
x = 0.11	0.2551	0.6165	75.24 (●)
x = 0.13	0.2502	0.6277	78.16 (●)

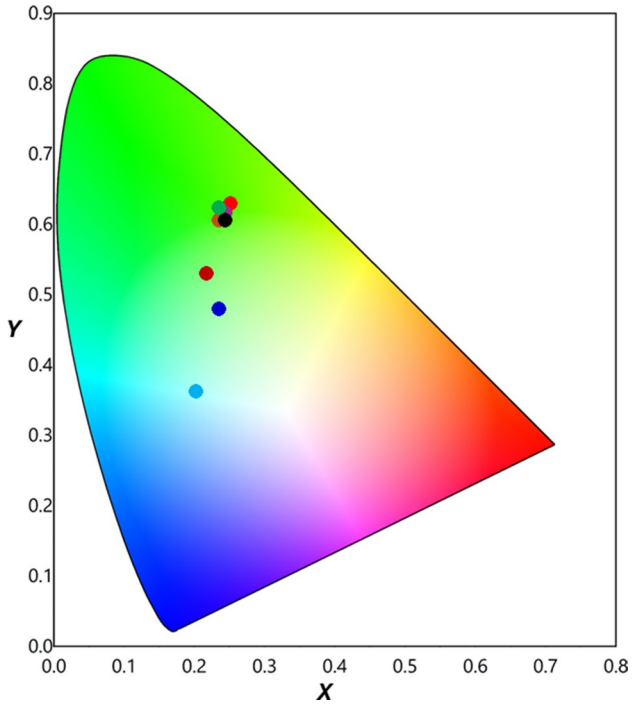


Fig. 12. CIE-chromaticity diagram for $Li_3Ba_2Gd_3Mo_8O_{32}$:xTb^{3+} phosphors (λ_{exc} = 283 nm).

The color emitted by the phosphors was predicted utilizing the Commission Internationale de l'Eclairage (CIE) chromaticity diagram, with chromaticity coordinates (*x*, *y*) derived from PL emission spectral data at an excitation wavelength of 283 nm. Table II presents these coordinates, along with the corresponding color purity percentages calculated using the established formulae.[44,45] The CIE coordinates of all prepared samples have been plotted on the CIE chromaticity diagram (Fig. 12), revealing a trend from pale blue (0.2163, 0.3756) to bright green (0.2618, 0.6358). A comparison of the estimated CIE coordinates highlights that the $Li_3Ba_2Gd_3Mo_8O_{32}$:0.09Tb^{3+} phosphor, with coordinates (0.2618, 0.6358), closely aligns with the ideal green color coordinates (0.29, 0.60), and is positioned within the bright green region. Notably, this phosphor demonstrates the highest color purity percentage (79.66%) among all the samples. Consequently, the $Li_3Ba_2Gd_3Mo_8O_{32}$:0.09Tb^{3+} phosphor emerges as a promising candidate for applications involving green light emission under UV excitation.

Conclusions

A series of $Li_3Ba_2Gd_3Mo_8O_{32}$:xTb^{3+} phosphors were successfully synthesized through the sol–gel method, resulting in a monoclinic crystal structure with the space group C2/c, as confirmed by XRD profiles. The SEM images indicated good crystallinity, with irregular and non-uniform particle shapes. The formation of the phosphors was verified through FT-IR spectra. The PL excitation spectra displayed a red shift in the $4f^8 \rightarrow 4f^75d^1$ transitions of Tb^{3+} ions, while the PL emission profiles exhibited strong emissions in the green region (545 nm) of the visible spectrum under UV excitation, specifically the $^5D_4 \rightarrow {}^7F_5$ transition. The optimal concentration of Tb^{3+} ions within the host material was determined to be 0.09 mol, beyond which concentration-quenching occurred due to dipole–dipole interactions. Photometric analyses collectively confirmed the bright green color emission of the $Li_3Ba_2Gd_3Mo_8O_{32}$:0.09Tb^{3+} phosphor. This study supports its potential for applications in green light emission technologies, including but not limited to white LEDs and solid-state lighting.

Acknowledgments This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C1092509). This paper was also supported by the KU Research Professor Program of Konkuk University.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Y. Wang, Y. Gong, X. Xu, and Y. Li, Recent progress in multicolor long persistent phosphors. *J. Lumin.* 133, 25 (2013).
2. A.K. Parchur and R.S. Ningthoujam, Behaviour of electric and magnetic dipole transitions of Eu^{3+}, $^5D_0 \rightarrow {}^7F_0$ and Eu-O charge transfer band in Li^+ co-doped YPO_4: Eu^{3+}. *RSC Adv.* 2, 10859 (2012).
3. K.P. Mani, P. Sreekanth, G. Vimal, P.R. Biju, N.V. Unnikrishnan, M.A. Ittyachen, R. Philip, and C. Joseph, Tunable emission and excited state absorption induced optical limiting in $Tb_2(MoO_4)_3$:Sm^{3+}/Eu^{3+} nanophosphors. *Opt. Mater.* 62, 110 (2016).

4. L. Zhou, W. Wang, S. Yu, B. Nan, Y. Zhu, Y. Shi, H. Shi, X. Zhao, and Z. Lu, Single-phase $LiY(MoO_4)_{2-x}(WO_4)x:Dy^{3+}$, Eu^{3+} phosphors with white luminescence for white LEDs. *Mater. Res. Bull.* 84, 429 (2016).
5. V. Singh, C.B.A. Devi, A.S. Rao, and J.B. Joo, Ultraviolet-B radiation from Gd (III) doped hardystonite. *Int. J. Mater. Res.* 114, 1122 (2023).
6. J. Grigorjevaite and A. Katelnikovas, Luminescence and luminescence quenching of $K_2Bi(PO_4)(MoO_4):Eu^{3+}$ phosphors with efficiencies close to unity. *ACS Appl. Mater. Interfaces* 8, 31772 (2016).
7. S. Cheng, X. Xu, D. Li, D. Zhou, F. Wu, Z. Zhao, and J. Xu, Spectral properties of Yb^{3+} ions in $Lu_3ScAl_4O_{12}$ single crystal. *J. Alloys Compd.* 506, 513 (2010).
8. A. Wojciechowski, M. Swirkowicz, A. Karas, L.R. Jaroszewicz, A. Majchrowski, E. Gondek, K. Ozga, and I.V. Kityk, Optical features of $Ca_4NdO(BO_3)_3$:Yb single crystals under CW laser illumination. *J. Alloys Compd.* 504, 197 (2010).
9. V. Singh, C.B.A. Devi, B.R.V. Rao, and A.S. Rao, Green emission from calcium zirconium silicate doped with trivalent terbium phosphor. *J. Electron. Mater.* 52, 50 (2023).
10. R.S. Yadav, Energy transfer induced color tunable photoluminescence in Tb^{3+}/Sm^{3+} co-doped Y_2O_3 nano-phosphor for warm white LEDs. *J. Alloys Compd.* 931, 167579 (2023).
11. R.S. Yadav and S.B. Rai, Concentration and wavelength dependent frequency downshifting photoluminescence from a Tb^{3+} doped yttria nano-phosphor: A photochromic phosphor. *J. Phys. Chem. Solids* 114, 179 (2018).
12. R.S. Yadav and S.B. Rai, Surface analysis and enhanced photoluminescence via Bi^{3+} doping in a Tb^{3+} doped Y_2O_3 nano-phosphor under UV excitation. *J. Alloys Compd.* 700, 228 (2017).
13. N. Baig, R.S. Yadav, N.S. Dhoble, V.L. Barai, and S.J. Dhoble, Near UV excited multi-color photoluminescence in RE^{3+} (RE=Tb, Sm, Dy and Eu) doped $Ca_2Pb_3(PO_4)_3Cl$ phosphors. *J. Lumin.* 215, 116645 (2019).
14. A. Garcia-Cortes, C. Zaldo, C. Cascales, X. Mateos, and V. Petrov, Laser operation of Yb^{3+} in disordered $Li_{0.75}Gd_{0.75}Ba_{0.5}(MoO_4)_2$ crystal with small quantum defect. *Opt. Express.* 15, 18162 (2007).
15. M. Song, W. Zhao, G. Wang, M. Zhao, and L. Wang, Growth, thermal and polarized spectral properties of Nd^{3+}-doped $Li_3Ba_2Gd_3(MoO_4)_8$ crystal. *J. Alloys Compd.* 509, 2164 (2011).
16. V. Singh, C.B.A. Devi, A.S. Rao, and J.L. Rao, Narrowband ultraviolet B emitting Gd^{3+} activated $NaSrPO_4$ phosphor prepared by co-precipitation method. *Optik* 208, 163632 (2020).
17. V. Singh, C.B.A. Devi, B.R.V. Rao, A.S. Rao, N. Singh, and B.M. Mistry, Narrow-band ultraviolet B (UVB) emitting $CaZr_4(PO_4)_6$ doped with Gd^{3+} phosphor. *Optik* 226, 165932 (2021).
18. A. Verma, R. Sharma, D.P. Bisen, N. Brahme, K. Thakkar, S. Tigga, and P. Chandrakar, Luminescence investigation of $CaY_2Al_4SiO_{12}:Dy^{3+}$ phosphor synthesized by sol–gel method. *Luminescence* 38, 576 (2023).
19. N.M. Kozhevnikova, Synthesis and luminescence properties of a $Li_3BaCaY_3(MoO_4)_8:Er^{3+}$ phosphor with a layered scheelite-like structure. *Inorg. Mater.* 53, 419 (2017).
20. V. Singh, S. Kaur, C.B.A. Devi, A.S. Rao, and J.B. Joo, Electron paramagnetic resonance and optical properties of green emitting Mn(II)-doped hardystonite phosphor prepared by sol-gel method. *Optik* 266, 169553 (2022).
21. V. Singh, C.B.A. Devi, S. Kaur, A.S. Rao, and N. Singh, Green emission of Er^{3+}-doped $Sr_2La_8(SiO_4)_6O_2$ obtained by a sol–gel method. *Optik* 243, 167322 (2021).
22. N.M. Kozhevnikova and S.Y. Batueva, Erbium-doped upconversion phosphor in the Li_2MoO_4-$BaMoO_4$-$Y_2(MoO_4)_3$ system. *Inorg. Mater.* 56, 286 (2020).
23. U. Holzwarth and N. Gibson, The scherrer equation versus the Debye–Scherrer equation. *Nat. Nanotechnol.* 6, 534 (2011).
24. Z.Q. Peng, R. Chen, and W.L. Feng, Photoluminescence properties of $Ca_3Si_2O_7:Pr^{3+}$ orange-red phosphors prepared by high-temperature solid-state method. *Z. Naturforsch. A* 73, 555 (2018).
25. J.B. Talbot and J. McKittrick, Electrophoretic deposition of phosphors for solid-state lighting. *ECS J. Solid State Sci. Technol.* 5, 3107 (2015).
26. L.J. Burcham and I.E. Wachs, Vibrational analysis of the two non-equivalent, tetrahedral tungstate (WO_4) units in $Ce_2(WO_4)_3$ and $La_2(WO_4)_3$. *Spectrochim. Acta A: Mol. Biomol. Spectrosc.* 54, 1355 (1998).
27. A. Kato, S. Oishi, T. Shishido, M. Yamazaki, and S. Iida, Evaluation of stoichiometric rare-earth molybdate and tungstate compounds as laser materials. *J. Phys. Chem. Solids* 66, 2079 (2005).
28. A.A. El-Rehim, K.S. Shaaban, H.Y. Zahran, I.S. Yahia, A.M. Ali, M.A. Halaka, S.A. Makhlouf, E.A. Wahab, and E.R. Shaaban, Structural and mechanical properties of lithium bismuth borate glasses containing molybdenum (LBBM) together with their glass–ceramics. *J. Inorg. Organomet. Polym. Mater.* 31, 1057 (2021).
29. J.S. Revathy, N.C. Priya, K. Sandhya, and D.N. Rajendran, Structural and optical studies of cerium doped gadolinium oxide phosphor. *Bull. Mater. Sci.* 44, 1 (2021).
30. S. Marzouk, S.M. Abo-Naf, M. Hammam, Y.A. El-Gendy, and N.S. Hassan, FTIR spectra and optical properties of molybdenum phosphate glasses. *J. Appl. Sci. Res.* 7, 935 (2011).
31. S.K. Jaganathan, J.P. Anthuvan, and I.S. Banu, Synthesis, luminescent properties and energy transfer in Tb^{3+} and Eu^{3+} co-doped $Li_3Ba_2Gd_3(MoO_4)_8$ phosphors for W-LED's. *J. Mater. Sci.: Mater. Electron.* 29, 2363 (2018).
32. Q. Chen, J. Li, and W. Wang, Synthesis and luminescence properties of Tb^{3+}/Eu^{3+} co-doped $GdAlO_3$ phosphors with enhanced red emission. *J. Rare Earths* 36, 924 (2018).
33. A. Yadav, P. Narwal, M.S. Dahiya, T. Dahiya, A. Agarwal, and S. Khasa, XRD and FTIR analysis heat treated lithium bismo-borate glasses doped with 1.0 mol.% copper ferrite. *AIP Conf. Proc.* 1953, 1 (2018).
34. H.A. Silim, Composition effect on some physical properties and FTIR spectra of alumino-borate glasses containing lithium, sodium, potassium and barium oxides. *Egypt. J. Sol.* 29, 293 (2006).
35. H.H.S. Oliveira, M.A. Cebim, A.A. Da Silva, and M.R. Davolos, Structural and optical properties of $GdAlO_3:RE^{3+}$ (RE= Eu or Tb) prepared by the pechini method for application as x-ray phosphors. *J. Alloys Compd.* 488, 619 (2009).
36. W.S. Kim, K.H. Yoon, and E.S. Kim, Far-infrared reflectivity spectra of $CaTiO_3$-$Li_{1/2}Sm_{1/2}TiO_3$ microwave dielectrics. *Mater. Res. Bull.* 34, 2309 (1999).
37. A. Raja, G. Annadurai, D.J. Daniel, and P. Ramasamy, Synthesis, optical and thermal properties of novel Tb^{3+} doped $RbCaF_3$ fluoroperovskite phosphors. *J. Alloys Compd.* 683, 654 (2016).
38. D.T. Lien, D.T.M. Huong, and N.N. Long, Structure and photoluminescence characterization of Tb^{3+}-doped $LaPO_4$ nanorods prepared via the microwave-assisted method. *J. Lumin.* 161, 389 (2015).
39. W.T. Carnall, P.R. Fields, and K. Rajnak, Electronic energy levels of the trivalent lanthanide aquo ions. IV. Eu^{3+}. *J. Chem. Phys.* 49, 4450 (1968).
40. G.P.R.L. Blasse, Energy transfer in oxidic phosphors. *Phys. Lett. A* 28, 444 (1968).
41. C. Zeng, H. Huang, Y. Hu, S. Miao, and J. Zhou, A novel blue-greenish emitting phosphor $Ba_3LaK(PO_4)_3F:Tb^{3+}$ with high thermal stability. *Mater. Res. Bull.* 76, 62 (2016).
42. D.L. Dexter, A theory of sensitized luminescence in solids. *J. Chem. Phys.* 21, 836 (1953).

43. S. Chahar, R. Devi, M. Dalal, M. Bala, J. Dalal, P. Boora, V.B. Taxak, R. Lather, and S.P. Khatkar, Color tunable nanocrystalline $SrGd_2Al_2O_7:Tb^{3+}$ phosphor for solid state lighting. *Ceram. Int.* 45, 606 (2019).
44. S.K. Mahamuda, F. Syed, C.B.A. Devi, K. Swapna, M.V.V.K.S. Prasad, M. Venkateswarlu, and A.S. Rao, Spectral characterization of Dy^{3+} ions doped phosphate glasses for yellow laser applications. *J. Noncryst Solid* 555, 120538 (2021).
45. V. Singh, C.B.A. Devi, S. Kaur, A.S. Rao, and N. Singh, Optical properties of $Sr_2La_8(SiO_4)_6O_2$ doped with Ho^{3+} phosphor. *Optik* 242, 167268 (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.

Check for updates

Received: 6 January 2024 | Revised: 22 March 2024 | Accepted: 19 April 2024

DOI: 10.1111/ijac.14805

International Journal of Applied Ceramic Technology

REPLY

Comparative study of luminescence in alkali-metal-based yttrium fluoride nanophosphor for biophotonic applications

Sumandeep Kaur[1] | **Simran Arora**[2] | **A. Srinivasa Rao**[1]

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

[2]Department of Physics, Indian Institute of Technology Bombay, Mumbai, India

Correspondence
A. Srinivasa Rao, Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi 110 042, India.
Email: drsrallam@gmail.com

Funding information
Council of Scientific and Industrial Research, India, Grant/Award Number: 08/133(0053)/2020-EMR-I

Abstract

Alkali-metal-based yttrium fluoride host ($AY_{1-x}F_4:xEu^{3+}$; A: Li/Na/K) nanophosphors were developed using co-precipitation method. Phase analysis and emission spectral behavior were established for the said nanophosphors. The highest emission observed for $NaYF_4$ doped with Eu^{3+} was then optimized to achieve maximum photoluminescence (PL) emission for varied dopant amount. The downconversion and upconversion PL studies revealed pure red emission under n-ultraviolet and infrared excitations. The CIE coordinates are observed to be lie in the red hue of the CIE diagram. The PL decay curves reported to study lifetime of the emitting state and to identify quantum efficiency using Auzel's model for the said nanophosphor. Thermal stability was also evaluated from temperature-sensitive luminescent spectra. The studies conducted on $AYF_4:.05Eu^{3+}$ (A: Li/Na/K) nanophosphors contemplates that $NaYF_4$ nanophosphors doped with Eu^{3+} ions are best suited for biophotonic applications as the later one show relatively high emission.

KEYWORDS
activation energy, downconversion, quantum efficiency, temperature-dependent PL, upconversion

1 | INTRODUCTION

There is a process to generate photons of high energy by employing photons of low energy, the process of conversion is known as upconversion (UC). The researchers are greatly attracted toward UC, as there are multifarious applications contingent on the process. The applications include solar cell converters, photocatalysis, drug delivery, in vitro/in vivo imaging, and other biophotonic applications.[1–5] Biophotonic indicates the interaction of photons with biological matter, where in appropriate organic/inorganic material comes in contact with biological matter to regulate biological activities, ultrasensitive detection, and many more.[6] Thus, biophotonic is an important tool for bioimaging, biosensors, phototherapy, laser surgery, and optogenetics. The escalating advancement in the size and structure of phosphors from micro to nanoscale has led to perspicacious encouragement for researchers to develop nanophosphors for biophotonic applications.[7,8] Nanophosphors (NP) are desirable for biophotonic applications due to their altered optical properties owing to the reduction in the size of particles from micro to nano.[9] When the size of particles encroaches in the nano range, surface to volume ratio increases, which enables more atoms at the surface to interact with biocells. Thus, the efficiency of these nanophosphors gets significantly enhanced and they become quite useful in biophotonic applications.

Generally, upconverting nanophosphors serve as excellent probe in comparison with fluorescent dyes and quantum dots, since later have certain shortcomings, such as fast photobleaching, low luminescent intensity,

 © 2024 The American Ceramic Society.

autofluorescence, loss of light due to scattering, and limited imaging length.[8,10] These pitfalls can be overcome by employing rare-earth (RE) doped NPs, since narrow and sharp f–f transitions of REs help in reducing spectral overlap of excitation and emission transitions.[11,12] In this context, alkali yttrium fluorides are identified as good near-infrared (NIR) to visible upconverter for its utilization in biophotonic applications.[13,14] Fluoride-based hosts are found suitable for nanophosphor materials as these materials have low phonon energy, which gives better luminescent efficiency and low toxicity of these materials favor their utility in biophotonic applications.[15] Typically, 980 nm laser excitation is considered as an NIR source, which makes it difficult for in vivo imaging, since water absorbs this incident radiation and makes it difficult for deep tissue imaging using 980 nm excitation. This can be over shooted if 800 nm excitation can be used instead of 980 nm excitation. Therefore, Eu^{3+} ions are chosen among various REs, as these can be excited by 800 nm and can serve as NIR to visible upconverter.[16]

In this article, AYF_4 (A: Li/Na/K) doped with Eu^{3+} ions nanophosphors ($A(Y_{1-x}Eu_x)F_4$) with low phonon energy in the range of 300–600 cm^{-1} were developed by co-precipitation route.[17,18] The structural tunability of the aforementioned nanophosphors by varying the alkali metal exhibits different luminescent properties. The structural, downconversion (DC) and UC luminescent characteristics of the said nanophosphors have been discussed in detail in the perspective of biophotonic applications.

2 | EXPERIMENT AND CHARACTERIZATION

To prepare the desired nanophosphors, the chemical precursors AF (LiF/NaF/KF), from Fisher Scientific, and yttrium acetylacetonate, from Sigma–Aldrich, were dissolved in 5 mL methanol, individually, and the solution of AF was added dropwise into the solution of yttrium acetylacetonate. The mixture was set on stirring for 10 min under room temperature and $EuCl_3{\cdot}6H_2O$ dissolved in 5 mL methanol was added dropwise into the above mixture. All the chemicals taken for the synthesis of nanophosphor were 99.9% pure. Now, the obtained solution was kept on stirring at 60°C for 5 h without any disturbance. After that, the solution gets precipitated and stirring and heating were ceased, so that it attains room temperature naturally. Then, the precipitate was centrifuged and washed five times with methanol. The obtained precipitate was kept in vacuum oven at 50°C overnight to achieve dried sample, which was grinded to carry out characterizations. The synthesis route is explained in Figure 1A.

To investigate the structure of the prepared samples, X-ray diffraction (XRD) patterns were recorded using Bruker diffractometer. Micrograph images were recorded by using scanning electron microscopy (SEM) instrument of Carl Zeiss with model no. EVO-50. Photoluminescence (PL) DC and UC studies were conducted on JASCO 8300 spectrofluorometer provided with Xenon lamp serving as irradiation source having excitation power 150 W. Decay curves were recorded using Edinburg instrument FL920 fluorescence spectrometer. The temperature-dependent photoluminescence (TDPL) measurements were made on nanophosphor by using an ocean optics spectrometer (model no. FLAME-S-XR1-ES). The powder nanophosphor was dissolved in dimethyl sulfoxide (DMSO) and polyvinylpyrrolidone (PVP) in ethanol solution.

3 | RESULTS AND DISCUSSION

3.1 | Structural analysis

Figure 1B presents XRD patterns of $AYF_4{:}.05Eu^{3+}$ (A: Li/Na/K) nanophosphor. The XRD peaks show similarity with the peaks indexed using standard JCPDS data bearing card no. 77-2042 for $NaYF_4$ cubic phase. The XRD peaks in the pattern for $LiYF_4{:}.05Eu^{3+}$ and $KYF_4{:}.05Eu^{3+}$ show some deviation along with some extra peaks, which show that mixed phase exists for these samples. The XRD peaks for $NaYF_4{:}.05Eu^{3+}$ show more resemblance with reference to the standard data suggesting the $NaYF_4$ phase has been achieved in this host. The cubic structure with space group Fm3m is formed in the synthesized nanophosphor. The XRD patterns for $NaY_{1-x}F_4{:}xEu^{3+}$ are shown in Figure S1. The average crystallite size was computed using Debye–Scherrer equation: $d = 0.9\lambda/\beta\cos\theta$, where d is crystallite size, λ is wavelength of X-rays used ($\lambda = 1.5406$ Å), and β is full width at half maximum (FWHM) corresponding to diffraction angle θ. The computed values are 2.80 and 5.75 nm for $LiYF_4$ and KYF_4 and 5.07, 4.80, 4.77, 4.67, and 4.57 nm for $NaY_{1-x}F_4{:}xEu^{3+}$ ($x = 5, 10, 15, 20$, and 25 mol%), respectively.

3.2 | Morphological studies

The as-prepared nanophosphors were taken to record SEM images in order to identify the morphology of the prepared particles. As observed from Figure 2A–C, the morphology of the prepared alkali-metal-based yttrium fluorides changes from nanospheres to nanoflakes to nanorods when alkali metal varies from lithium to potassium in the periodic table in AYF_4 host. The difference in the morphology of the prepared samples is observed to be

17447402, 0, Downloaded from https://ceramics.onlinelibrary.wiley.com/doi/10.1111/ijac.14805 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

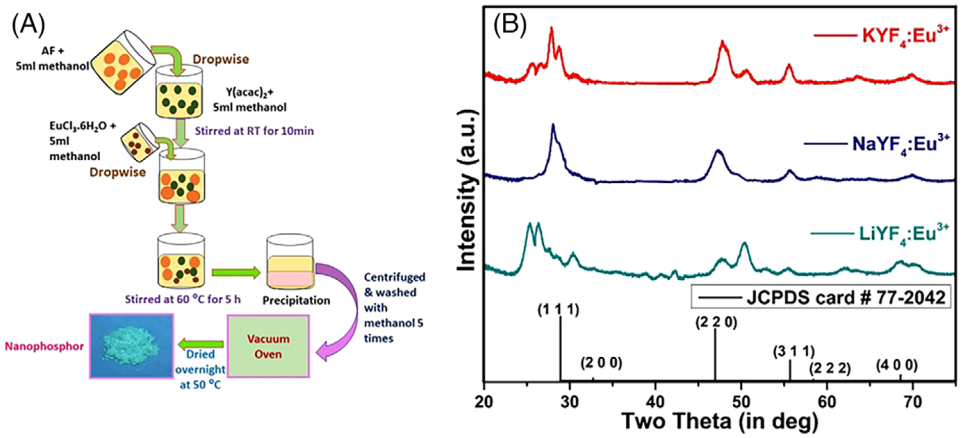


FIGURE 1 (A) Schematic illustration for synthesis of $AYF_4:Eu^{3+}$ (A: Li/Na/K) nanophosphor. (B) X-ray diffraction (XRD) patterns of $AYF_4:.05Eu^{3+}$ (A: Li/Na/K) nanophosphor.

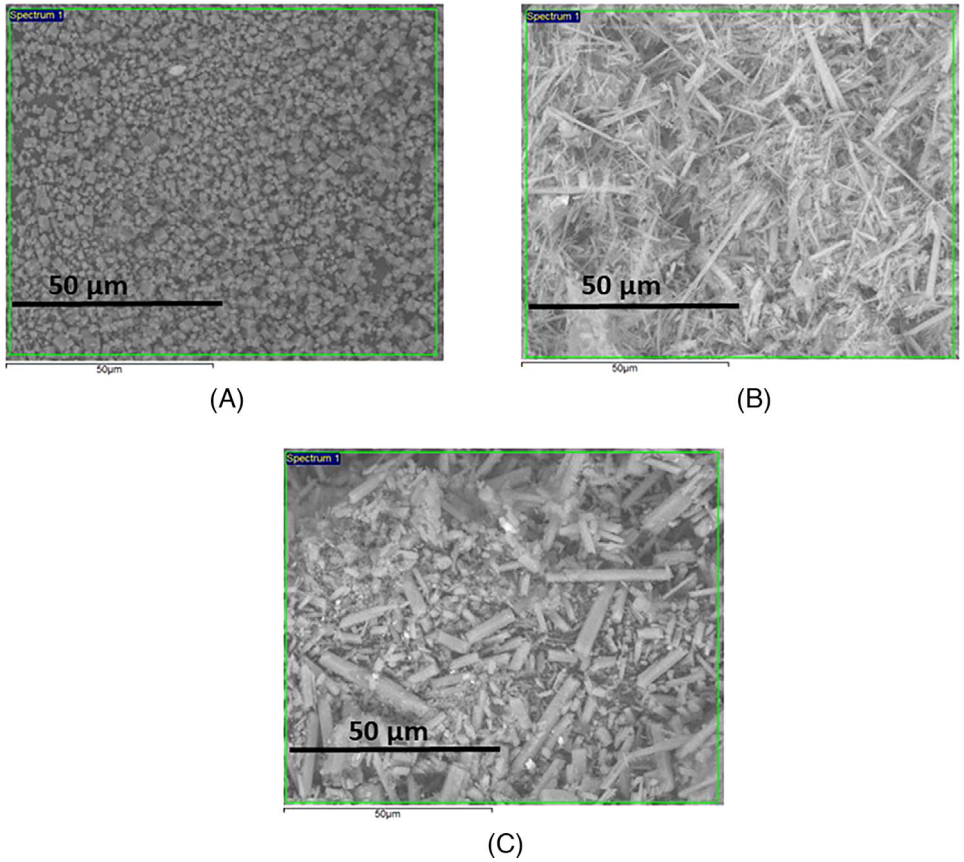


FIGURE 2 (A–C) Scanning electron microscopy (SEM) images of $AYF_4:0.05Eu^{3+}$ (A: Li/Na/K nanophosphor, respectively).

different, which might be due to different host precursor in these nanophosphors. However, the morphology observed suggested that particles are not uniformly distributed in $NaYF_4:0.05Eu^{3+}$ and $KYF_4:0.05Eu^{3+}$. Moreover, the size of the particles varies in the aforementioned host matrices, as some agglomeration is also observed, whereas the particles are uniformly distributed in $LiYF_4:0.05Eu^{3+}$ with agglomerated morphology. The size distribution of the as-prepared nanophosphor is from 100 nm to few microns as agglomeration can be seen in the prepared samples.

3.3 | X-ray photoelectron spectral analysis

X-ray photoelectron spectroscopy (XPS) is an analytical way to study properties of the sample, which are surface

17447402, 0, Downloaded from https://ceramics.onlinelibrary.wiley.com/doi/10.1111/ijac.14805 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

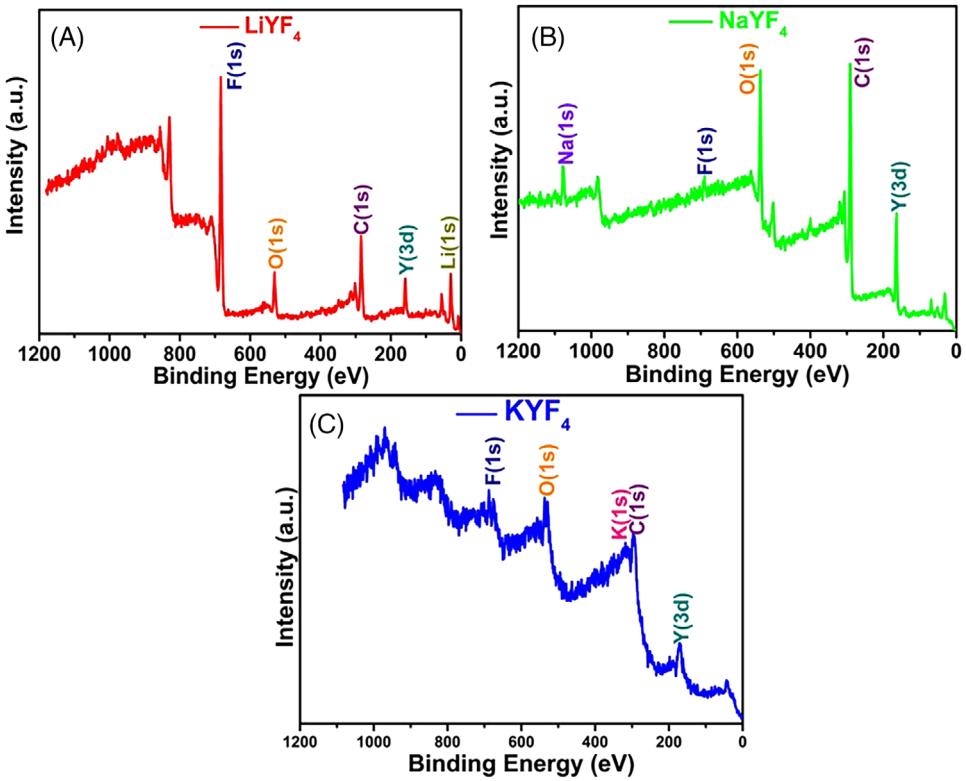


FIGURE 3 (A–C) X-ray photoelectron survey scan of AYF_4 (A: Li, Na, and K, respectively).

sensitive. It helps to identify elements, oxidation state, and binding energy of the constituent elements of the samples.[19,20] Figure 3A–C represents X-ray photoelectron survey scan of AYF_4 (A: Li, Na, K) in the range 0–1200 eV to identify elemental composition of the prepared samples, respectively. In general, fortuitous presence of carbon in 1s state is taken as reference to determine binding energy of the elements present in the host.[21] In Figure 3A, binding energy at 283 eV indicates carbon 1s peak and considered as reference for charge shift correction to identify other elements. Peaks at 55, 156, 533, and 683 eV suggest the binding energy of Li, Y, O, and F in 1s, 3d, 1s, and 1s states, respectively. The presence of oxygen in 1s state is due to the exposure of sample to air. Similarly, Figure 3B shows peak at 164, 286, 536, 686, and 1076, which indicates Y(3d), C(1s), O(1s), F(1s), and Na(1s), respectively. Figure 3C shows the presence of all the elements of KYF_4 for 166, 287, 395, 534, and 685 eV binding energy values corresponding to Y(3d), C(1s), K (1s), O(1s), and F(1s), respectively. Thus, it is well founded that the prepared samples are prepared successfully with the corresponding composition as described above.

3.4 | DC PL studies

The steady-state PL investigation of $AY_{1-x}F_4{:}xEu^{3+}$ (A: Li/Na/K) nanophosphors were conducted by observing excitation and emission spectra. Figure 4A represents the PL excitation spectrum of 5 mol% of Eu^{3+}-doped $NaYF_4$ nanophosphor from the fixed emission wavelength of Eu^{3+} ions (i.e., 613 nm). The spectrum possesses the sharp peaks related to 4f transitions of Eu^{3+} ions in the nanophosphor. The sharp excitation peaks at 380, 391, 412, and 463 nm are indexed to $^7F_0$–$^5L_7$, $^7F_0$–$^5L_6$, $^7F_0$–$^5D_3$, and $^7F_0$–$^5D_2$ transitions, respectively, as observed in 5 mol% of Eu^{3+}-doped $NaYF_4$ nanophosphor.[22] The strongest excitation at 463 nm validates the potential of the nanophosphor to get significantly excited by blue LED chips.

In order to study the intensity variation in the luminescence profile of the prepared samples, approximately the same amount of $AY_{1-x}F_4{:}xEu^{3+}$ (A: Li/Na/K) nanophosphors was taken to relate the quality of luminescence of the aforementioned nanophosphors with the replacement of alkali metals of higher order in the fluoride-based AYF_4 host. The PL emission spectra of the said nanophosphors under blue excitation are seen in Figure 4B. Usually, the transitions of RE^{3+} ions remain the same, if the host in which RE ion is doped is varied with the slight variation in their peak position. Here, in Figure 4B, almost all the transitions are common in the three emission spectral profiles with slight shift in their position. It is observed from the figure that there are four emission peaks due to transition from $^5D_0$ to $^7F_1$, $^7F_2$, $^7F_3$, and $^7F_4$ energy states.[23,24] Among these four emission lines, the one corresponding to $^5D_0$ to $^7F_1$ is magnetic dipole transition that does not

17447402, 0, Downloaded from https://ceramics.onlinelibrary.wiley.com/doi/10.1111/ijac.14805 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

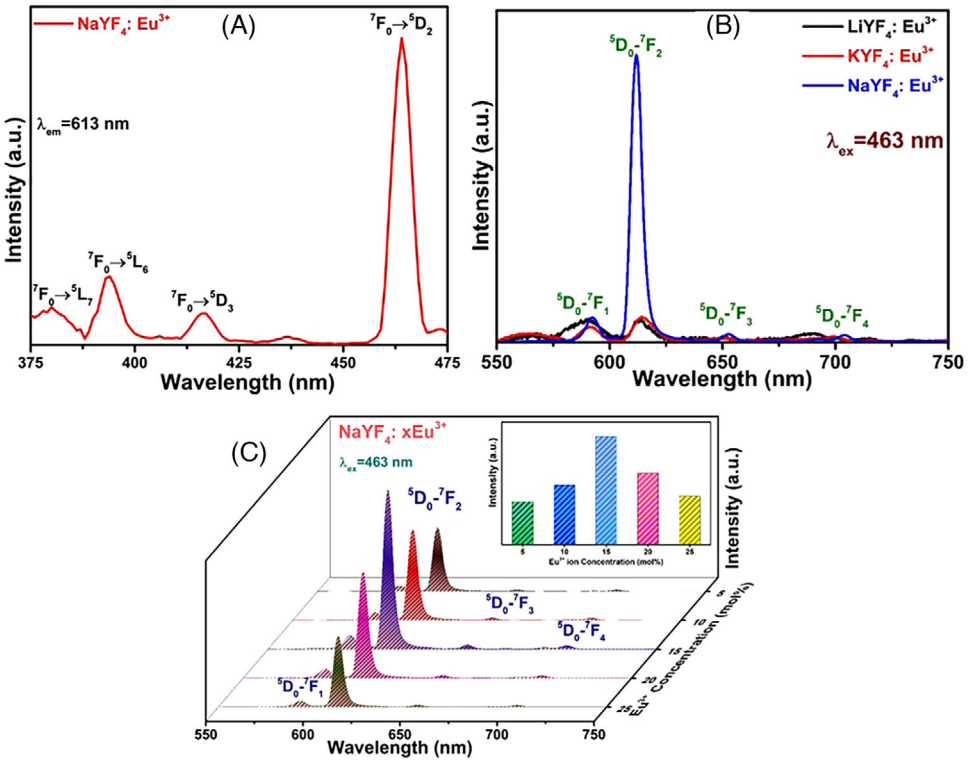


FIGURE 4 (A) Excitation spectrum of $NaYF_4$:0.05Eu^{3+} nanophosphor under 613 nm. (B) Emission spectra of AYF_4:0.05Eu^{3+} (A: Li/Na/K) nanophosphor under 463 nm. (C) Emission spectra of $NaYF_4$:xEu^{3+} (x = 0.05, 0.10, 0.15, 0.20, and 0.25 mol) nanophosphor under 463 nm. (Inset) Variation of intensity of peak at 613 nm for different concentration of Eu^{3+} ion.

get influenced by the site symmetry and the other one, the most intense one, corresponding to $^5D_0$ to $^7F_2$ is electric dipole transition that gets regulated by the site symmetry of the host crystal lattice. The f–f intra-configurational electric dipole transitions are forbidden by Laporte selection rule ($\Delta S = 0$, $\Delta L = 0$, and $\Delta J = 0, \pm 1$) and become parity allowed by mixing with opposite parity orbitals.[25–27] This induced electric dipole transition is also known as hypersensitive transition due to the dependency of this transition on the structural environment around Eu^{3+} ions. The relative dominance of transitions originating from magnetic/electric dipole dictates the site occupancy of Eu^{3+} ions in the host material. In the present analysis, all the three fluoride-based AYF_4 hosts show the dominant characteristic of electric dipole transition at 613 nm, which suggests that Eu^{3+} ions have occupied the non-inversion symmetric sites in the host crystal structure.[23,28] Furthermore, Figure 4B shows the highest emission profiles observed for sodium alkali metal in AYF_4 host, which suggested relatively more potential of Eu^{3+} ion-doped $NaYF_4$ nanophosphor in biophotonic applications. Also, the luminescence properties were studied in $NaYF_4$ nanophosphor for the different amounts of Eu^{3+} ions as emission spectral profiles recorded by keeping excitation wavelength fixed at 463 nm are presented in Figure 4C. Figure 4C shows that more doping of Eu^{3+} ions facilitates increase in emission intensity gradually from 5 to 15 mol% and then decreases for 20 and 25 mol% of Eu^{3+} ions due to increase in the non-radiative pathways. This mechanism is known as concentration quenching, where the adequately low distance between the neighboring RE ions elevates non-radiative energy transfer pathways between the ions, which consequently give reduced emission. Inset of Figure 4C discloses the emission intensity variation of electric dipole transition with concentration of Eu^{3+} ions in $NaYF_4$ nanophosphor. Hence, it is righteous to conclude that 15 mol% of Eu^{3+} ions are the optimized amount in $NaYF_4$ host prepared by co-precipitation method. Furthermore, critical distance (R_c) can be determined to identify if exchange interaction between RE ions is responsible for non-radiative transition using the formula given by Blasse as shown below[29]:

$$R_c = 2\left(\frac{3V}{4\pi X_c N}\right)^{1/3} \quad (1)$$

where N denotes occupancy number, X denotes the amount of dopant for which the luminescence intensity is maximum, and V denotes the volume of unit cell. The value of R_c is computed, which is higher than 5 Å, and validates that exchange interaction does not account for non-radiative energy transfer among RE ions.

17447402, 0, Downloaded from https://ceramics.onlinelibrary.wiley.com/doi/10.1111/ijac.14805 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

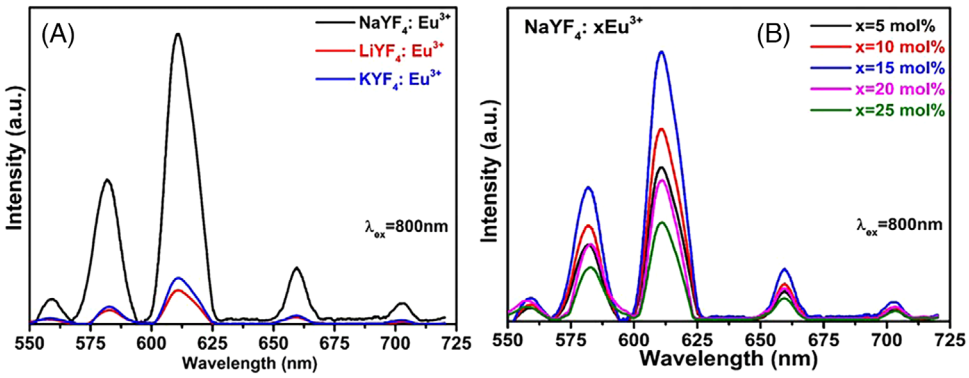


FIGURE 5 (A) Emission spectra of AYF_4:0.05Eu^{3+} (A: Li/Na/K) nanophosphor. (B) Emission spectra of $NaYF_4$:xEu^{3+} (x = 0.05, 0.10, 0.15, 0.20, and 0.25 mol) nanophosphor under 800 nm.

Consequently, the multipolar interaction between RE ions leads to non-radiative energy transfer.

In order to understand the bio-compatibility of the as-prepared nanophosphor, $NaYF_4$:Eu^{3+} was dissolved in DMSO and PVP solution (.1 g in 5 mL ethanol). It was observed that both the solutions were clear, suggesting complete solubility of the said nanophosphor. PL spectra of the two solutions were recorded and are shown in Figure S3. All the emission peaks are similar to those observed for the powder nanophosphor under 463 nm excitation. The result obtained signifies that the said nanophosphor has potential for biophotonic applications.[30,31]

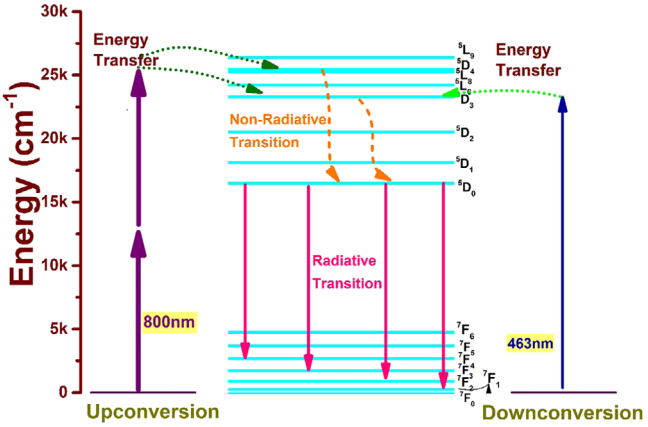


FIGURE 6 Partial energy level diagram of Eu^{3+}-doped nanophosphor representing upconversion and downconversion processes.

3.5 | UC PL studies

Figure 5A represents the UC PL spectra of AYF_4:Eu^{3+} (A: Li/Na/K) nanophosphors achieved by irradiating the samples at 800 nm. UC emission manifests similar emission peaks with some variation in the relative intensity of peaks. The distinguished emission peaks perceived at 559, 582, 611, 659, and 703 nm are arising from $^5D_0$ to $^7F_0$, $^7F_1$, $^7F_2$, $^7F_3$, and $^7F_4$ transitions, respectively.[32,33] All the emission peaks are similar with those observed in DC luminescence spectra recorded for the same samples. However, there is relative change in the intensity ratio for red (611 nm) to orange (582 nm) peak as the magnetic dipole (582 nm peak) intensity is relatively higher than the one observed in DC spectra, while the dominance of electric dipole (611 nm peak) transition is maintained in the UC spectra recorded under 800 nm. The most conceivable explanation for such a different luminescent behavior can be associated with the change in symmetries of sites around dopant ions.[34] It has been observed that the asymmetry ratio (1.839–2.019) for UC luminescence decreases when compared with asymmetry ratio (10.16–11.50) for DC luminescence.[35] Furthermore, the emission profiles of as-prepared three alkali metal-based yttrium halides are showing the similar trend of luminescence, where the lowest luminescence is observed from $LiYF_4$:Eu^{3+} and highest luminescence is observed from $NaYF_4$:Eu^{3+}, and luminescence of KYF_4:Eu^{3+} lies between $LiYF_4$:Eu^{3+} and $NaYF_4$:Eu^{3+}. While recording the emission for concentration variant of Eu^{3+} in $NaYF_4$ nanophosphor, as depicted in Figure 5B, the intensity variation follows the same trend as in DC observed above. The only difference seen in DC and UC processes is the FWHM value, which is 5.77 for the former emission and 11.25 for the latter one. The differences in FWHM between UC and DC spectra of the same sample can arise from various factors, including differences in photon energies, energy transfer mechanisms, and excited state dynamics. The concentration of Eu^{3+} ion for which the luminescence is maximum is quantified as 15 mol%, same as in case of DC luminescence of the said nanophosphor. The UC spectra observed suggested that the prepared nanophosphor can be used for biophotonic applications.

Figure 6 shows possible transitions in Eu^{3+}-doped nanophosphor for UC and DC processes. During UC

17447402, 0, Downloaded from https://ceramics.onlinelibrary.wiley.com/doi/10.1111/ijac.14805 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

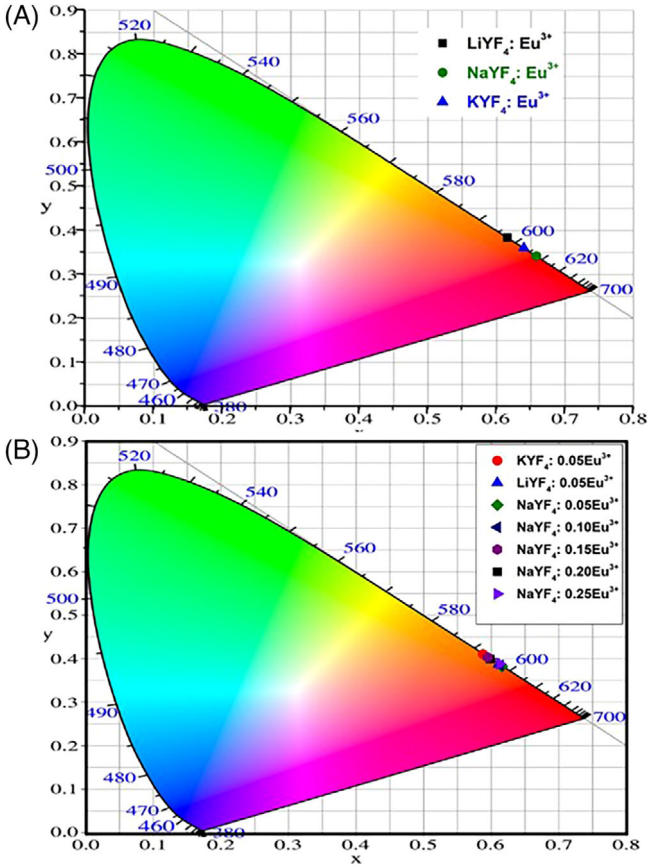


FIGURE 7 (A) CIE diagram of AYF_4:.05Eu^{3+} (A: Li/Na/K) nanophosphor under 463 nm excitation. (B) CIE diagram of $NaYF_4$:xEu^{3+} (x = .05, .10, .15, .20, and .25 mol) nanophosphor under 800 nm excitation.

process, there is no energy level corresponding to 800 nm excitation wavelength, which means that it might take up more than one photon to reach up to ^5D levels of the Eu^{3+} ions, from where after non-radiative transition, it reaches to $^5D_0$ lasing energy level.[36] It is observed from literature that $^5D_0$ level is the radiative transition level, which means that the transition from this level corresponds to radiative emission in the host. The same has been observed in the present studies. It is anticipated that during DC process, the Eu^{3+} ions got excited to $^5D_3$ level from ground state, which after non-radiative transition to $^5D_0$ level possess radiative emission lines, as discussed in the DC luminescent studies.

3.6 | Colorimetric analysis

In Figure 7A,B, the as computed CIE coordinates for the chromatic profiles for DC and UC luminescence are revealed. The CIE coordinates of AYF_4:Eu^{3+} (A: Li/Na/K) nanophosphors under blue excitation show transition in color from orange to red for Li to Na in AYF_4:Eu^{3+} (A: Li/Na/K) nanophosphors, as illustrated in Figure 7A. The CIE coordinates of $NaYF_4$:Eu^{3+} nanophosphors yield pure red color as the coordinates are close to the perimeter of the CIE diagram and for higher concentration, it is nearly the same in hue cycle. The CIE coordinates calculated for AYF_4:Eu^{3+} (A: Li/Na/K) nanophosphors and Eu^{3+} variation in $NaYF_4$ nanophosphors under 800 nm show that pure red emission is achievable for $NaYF_4$ host incorporated with Eu^{3+} ions, as exemplified in Figure 7B. Furthermore, the color purity of $NaYF_4$ host incorporated with optimized Eu^{3+} ions is evaluated using the following equation:

$$\text{Color purity} = \frac{\sqrt{(x - x_{wp})^2 + (y - y_{wp})^2}}{\sqrt{(x_d - x_{wp})^2 + (y_d - y_{wp})^2}} \times 100\% \quad (2)$$

where x and y indicate the CIE coordinates of the sample, x_{wp} and y_{wp} are the coordinates for white point, and x_d and y_d are the coordinates for dominant wavelength. The values of (x_d, y_d) for DC and UC are taken as (.661, .338) and (.618, .383), respectively, and the color purities under DC and UC are reckoned as 99.4% and 99.6%, respectively. These colorimetric analyses help to understand the ability of these nanophosphors as red emitter under DC as well as UC phenomenon, which is required for lighting and biophotonic applications, respectively.

3.7 | PL decay analysis

Figure 8A–E shows the decay curves of $NaYF_4$:xEu^{3+} (x = .05, .10, .15, .20, and .25 mol) nanophosphors under blue excitation for $^5D_0$ energy state. All the decay curves possess non-exponential behavior, which is fitted with bi-exponential equation, as mentioned elsewhere.[37] The fitted parameters were used to assess the decay lifetime values, which are estimated as 5.991, 3.008, 2.905, 2.789, and 2.597 ns for $NaYF_4$:xEu^{3+} (x = .05, .10, .15, .20, and .25 mol) nanophosphors, respectively. Figure 8F shows the lifetime values when the concentration of Eu^{3+} ions is altered. It can be noted that the lifetime values follow a declining trend with the upsurge of the amount of dopant ions. The transfer of energy that is being non-radiative in nature among the activators and concentration quenching accounts for the fall in lifetime values.[38] This result is in tune with the luminescence spectra as the luminescence decline due to energy transfer being non-radiative in nature occurs between Eu^{3+} ions. Moreover, the decrease in lifetime values suggests that there is more of non-radiative relaxation. To explore more information pertaining to declined lifetime values and to identify intrinsic lifetime of the excited energy state, Auzel's model

17447402, 0, Downloaded from https://ceramics.onlinelibrary.wiley.com/doi/10.1111/ijac.14805 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

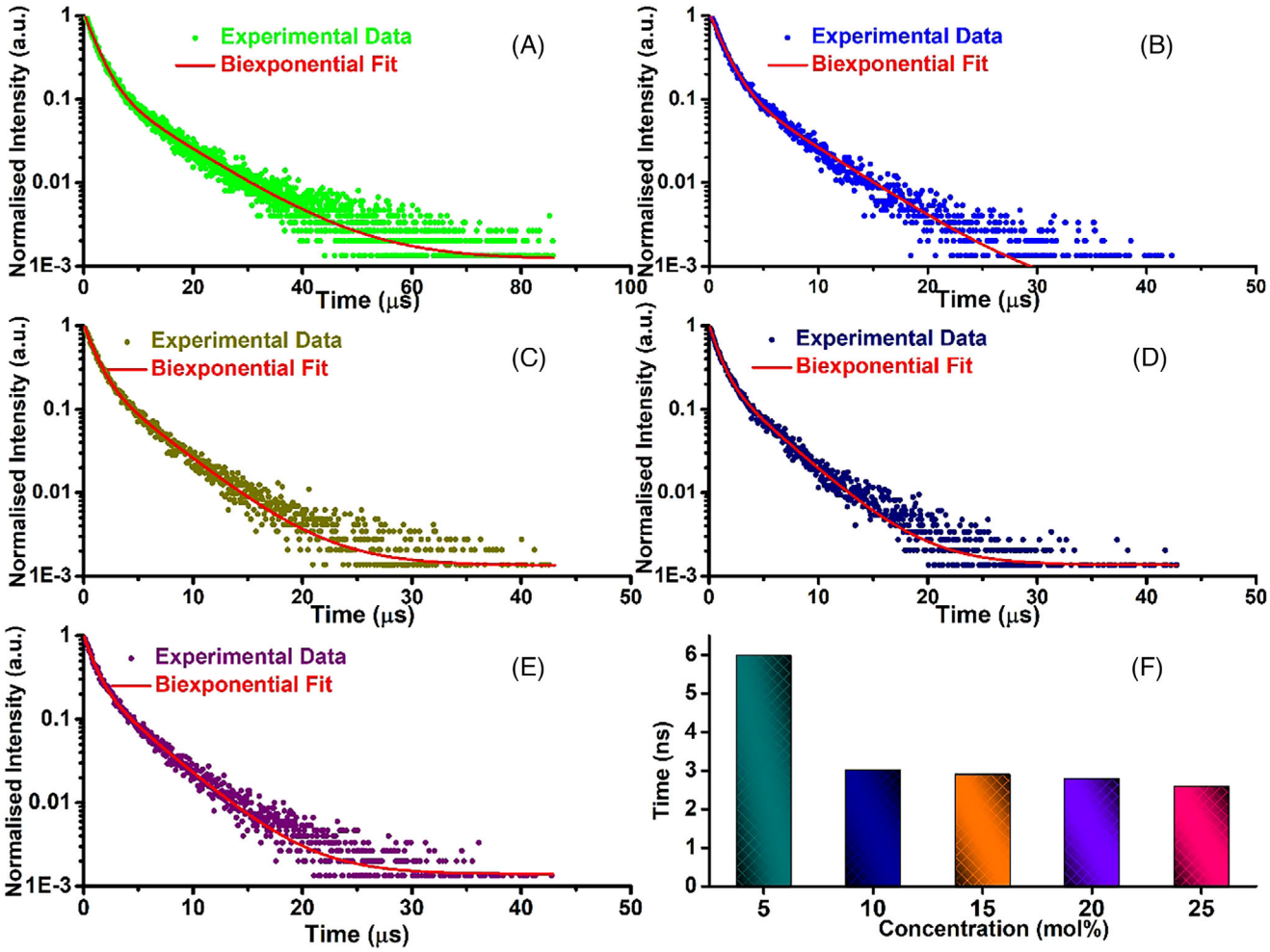


FIGURE 8 (A–E) Decay curves of $NaYF_4$:xEu^{3+} (x = .05, .10, .15, .20, and .25 mol) nanophosphor under 463 nm excitation and 613 nm emission wavelength. (F) Lifetime variation with concentration of Eu^{3+} ions.

is applied on the decay data by using the following equation[39]:

$$t_c = \frac{t_0}{1 + \frac{c}{c_0}\exp(-N/3)} \quad (3)$$

where t_c and t_0 are the lifetime values at concentration c and intrinsic value of the lifetime of the excited state, c_0 is a constant, and N denotes the phonon number. Therefore, a graph is plotted between lifetime t_c and concentration c of Eu^{3+} ions in $NaYF_4$ host nanophosphor. Figure 9 shows the data points on which Auzel's model is employed to conduct fitting. From the fitting, certain parameters were observed, including phonon number, $N = 6$, and intrinsic lifetime, $t_0 = 10.18$ ns for $^5D_0$-excited energy state. The phonon number obtained from the fitting parameters is consistent with the phonon energy of the host $NaYF_4$.[40] The intrinsic lifetime value along with experimentally observed lifetime values help to determine

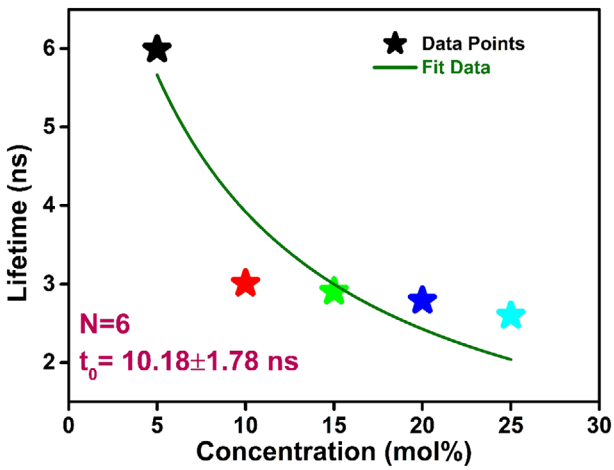


FIGURE 9 Lifetime versus concentration plot with Auzel's fit.

probability of non-radiative transitions rate (A_{NR}) and quantum efficiency of the investigated nanophosphor. The

17447402, 0, Downloaded from https://ceramics.onlinelibrary.wiley.com/doi/10.1111/ijac.14805 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

below mentioned equations are followed to compute A_{NR} and QE for $NaYF_4$ doped with Eu^{3+} ions:

$$\frac{1}{t_c} = \frac{1}{t_0} + A_{NR} \tag{4}$$

$$QE = \frac{t_c}{t_0} = \frac{A_R}{A_R + A_{NR}} \tag{5}$$

The computed values of non-radiative transition rates (A_{NR}) are .0446, .2100, .2218, .2362, and .2627, and quantum efficiency (QE) are 73.2, 36.7, 35.5, 34.0, and 31.7% for $NaYF_4$:xEu^{3+} (x = .05, .10, .15, .20, and .25 mol) nanophosphors, respectively. Basically, A_{NR} indicates depopulation of the excited state via non-radiative relaxation process and this value increases with increase in concentration of Eu^{3+} ions. And if A_{NR} increases with concentration, then the radiative emission transition rate gets suppressed by non-radiative transition rate, which leads to reduction in quantum efficiency. The low value of A_{NR} and good QE suggest that the prepared $NaYF_4$ host doped with Eu^{3+} ions can be an alternate for its utility in biophotonic applications.

3.8 | Judd–Ofelt parameters and quantum efficiency

The spectral transitions observed in Eu^{3+} ion-doped luminescent materials can be related with the Judd–Ofelt (J–O) intensity parameters (Ω_2, Ω_4, and Ω_6) calculated from the emission spectra. In RE ion-doped materials, certain transitions are hypersensitive whose peak position and intensity of emission strongly depend on the environment around the ion. Such hypersensitive transitions have relatively higher intensity when compared with transitions observed in emission spectrum. The J–O intensity parameters (Ω_2) dictates that the local surrounding, nature of bonding, rigidity, and viscosity are explained on the basis of Ω_4 and Ω_6. Using the emission spectra, the J–O intensity parameters are determined by employing the following equation[41]:

$$\Omega_J = \frac{S_{MD}\left(V_1^3\right)}{e^2\left(V_J^3\right)} \frac{9n^3}{n(n^2+2)^2} \frac{\int I_1(V_1)}{\langle J|U^J|J'\rangle^2 \int I_J(V_J)} \tag{6}$$

Here, S_{MD} (9.6×10^{-42} esu^2 cm^2) denotes the magnetic dipole line strength, and V_1 and V_J denote the transition frequency of $^7F_1$ and ^7F_J (J = 2, 4, and 6), Similarly, I_1 and I_J denote the corresponding intensities of the aforementioned transitions. "e" represents the elementary charge whose value is 4.803×10^{-10} esu, n denotes refractive index, which is estimated from band gap energy, as given in Supporting information, and $\langle J|U^J|J'\rangle^2$ represents the double-reduced matrix elements of the unit tensor operators, where U^J is .0032, .0023, and .0002 for J = 2, 4, and 6, respectively. The Ω_6 parameter is not evaluated as the transition $^5D_0$ to $^7F_6$ was not observed in the emission spectra. Furthermore, the spontaneous transition probability was calculated using the following equation[35,42]:

$$A_{J-J'} = \frac{64\pi^4\nu^3}{3h(2J+1)}\left(\chi_{ED}S_{ED} + \chi_{MD}S_{MD}\right) \tag{7}$$

where h is Planck's constant, ν is the transition wavenumber (cm^{-1}), S_{ED} similar to S_{MD} denotes the line strength corresponding to electric dipole transition, which can be calculated as:

$$S_{ED}(J, J') = e^2 \sum_{J = 2,4,6} \Omega_J \langle J|U^J|J'\rangle^2 \tag{8}$$

The χ_{ED} and χ_{MD} denote local field corrections for electric and magnetic dipole transitions, respectively, and can be determined using following equations:

$$\chi_{ED} = \frac{n\left(n^2+2\right)^2}{9} \tag{9}$$

$$\chi_{MD} = n^3 \tag{10}$$

The J–O intensity parameters (Ω_2 and Ω_4) and total transition probability (A) for different concentrations of Eu^{3+} ions in $NaYF_4$ host are shown in Table 1. The trend of intensity parameters ($\Omega_2 > \Omega_4$) shows an increase in covalency between Eu^{3+} ions and oxygen ligands and low site symmetry for different amounts of dopant ions as the higher value of Ω_2 suggest, whereas Ω_4 parameter indicates higher rigidity in the prepared nanophosphor.[43,44]

Also, the quantum efficiency has been calculated using the following equation[41]:

$$\eta = \frac{\tau}{\tau_r} = \frac{\frac{1}{A_r + A_{nr}}}{\frac{1}{A_r}} \tag{11}$$

where τ is the lifetime determined from decay curves and can be written as inverse of sum of radiative and non-radiative transition probabilities and τ_r is obtained from transition probability calculated above. The calculated quantum efficiency is reported in Table 1 and values are comparable with those evaluated using Auzel's model.

3.9 | TDPL investigations

It is well known that the emission color and output efficiency of the luminescent material get greatly influenced by its thermal stability. Thus, it is important to investigate the thermal stability of the nanophosphor for practical application. Also, the temperature aspects

17447402, 0, Downloaded from https://ceramics.onlinelibrary.wiley.com/doi/10.1111/ijac.14805 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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 Judd–Ofelt (J–O) intensity parameters (Ω_2 and Ω_4), total transition probability (A), radiative lifetime (τ_r), and quantum efficiency (η) as calculated from J–O theory.

$NaYF_4$ with x mol% Eu^{3+} ions	Ω_2 ($\times 10^{-20}$ cm^2)	Ω_4 ($\times 10^{-20}$ cm^2)	A ($\times 10^8$ s^{-1})	τ_r (ns)	Quantum efficiency, η (%)
5	4.422	.0468	1.309	7.637	78.4
10	4.836	.0512	1.363	7.336	41.0
15	4.855	.0514	1.375	7.271	39.9
20	4.862	.0515	1.382	7.231	38.5
25	4.874	.0516	1.402	7.129	36.4

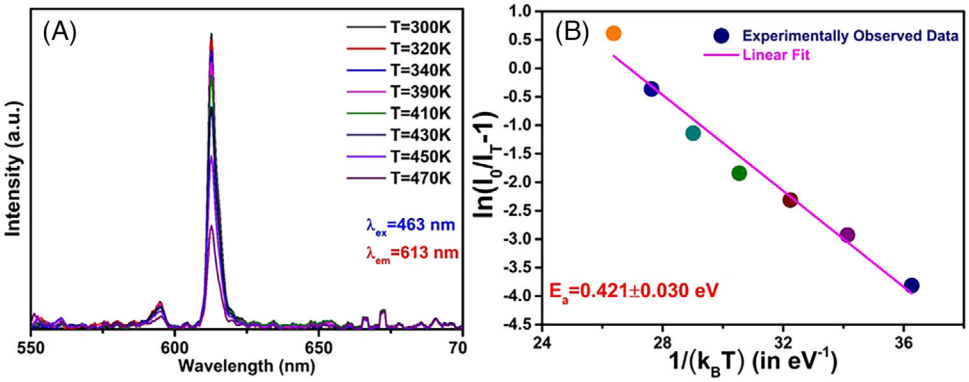


FIGURE 10 (A) Temperature-dependent photoluminescence of $NaYF_4$:.15Eu^{3+} nanophosphor monitored at 463 nm excitation. (B) Activation energy calculation for $NaYF_4$:.15Eu^{3+} nanophosphor.

of the nanophosphor for temperature sensing biophotonic applications can be understood by proclaiming temperature-dependent luminescence. The TDPL of the $NaYF_4$:.15Eu^{3+} nanophosphors were recorded by monitoring excitation wavelength at 463 nm in the temperature range of 300–470 K, as presented in Figure 10A. It is clear from Figure 10A that the emission intensity decreases to 90% at 390 K, which is the ambient operating temperature for practical applications. However, with increase in temperature, due to non-radiative phonon relaxation from upper energy levels via crossover process, the fall in emission is considerable, that is, 65% at 470 K due to thermal luminescence quenching effect. At higher temperature, there is more interaction between electron and phonon that tends to dissipate the excited state thermally and results in fall of emission at high temperature.[45] Furthermore, the well-established Arrhenius equation given below is applied on the temperature-dependent emission spectra of the said nanophosphor to investigate the thermal stability by evaluating activation energy[46,47]:

$$\ln\left(\frac{I_0}{I_T}-1\right)=\ln A-\frac{E_a}{k_BT} \tag{12}$$

where I_0 is the initial intensity at 300 K, I_T is the intensity at temperature $t = T$ K, and k_B is the Boltzmann constant. A graph is plotted between $\ln(I_0/I_T - 1)$ versus $1/k_BT$ to compute the activation energy, as illustrated in Figure 10B. The activation energy is also termed as trap depth, which is the energy required to release the captured electrons from the trap center.[48] The linear fitting of the plot gives slope as −.421 eV and negative of the slope gives activation energy as .421 eV. This value is higher than the earlier reported $Ca_{5.93-m}Sr_mBa(PO_4)_4O$:.07$Eu^{2+}$ (.276 and .292 eV for 0 and .1 value of m, respectively), $Ca_2GdZr_2Al_3O_{12}$:Mn^{4+} (.397 eV), $GdPO_4$:Eu^{3+} (.256 eV), $Ba_3Ce(PO_4)_3$:Eu^{3+} (.25 eV), $Zn_3V_2O_8$ (.14 eV), $BaSi_2O_5$:Eu^{2+} (.12 eV), and $Ca_9La_{.94}(PO_4)_5(SiO_4)F_2$:.05$Ce^{3+}$, .01$Tb^{3+}$ (.255 eV) hosts.[46,47,49–53] Hence, it can be inferred from the value of activation energy that the prepared nanophosphor exhibits good thermal stability and validate brilliance of Eu^{3+}-doped $NaYF_4$ nanophosphor for practical usage in biophotonic applications.

4 | CONCLUSIONS

Well-prepared AYF_4:.05Eu^{3+} (A: Li/Na/K) nanophosphor was compared for structural and luminescence behaviors. XPS study confirms the elemental composition of the prepared samples. It was found that $NaYF_4$:$x$$Eu^{3+}$ nanophosphor gives highest luminescence traits; therefore, $NaYF_4$:$x$$Eu^{3+}$ (x = .05, .10, .15, .20, and .25 mol) nanophosphors were prepared and detailed

17447402, 0, Downloaded from https://ceramics.onlinelibrary.wiley.com/doi/10.1111/ijac.14805 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

studies of luminescence properties were conducted on these nanophosphors. DC and UC studies show intense red emission under ultraviolet (n-UV) and 800 nm excitation. The PL analysis of phosphor solution in DMSO and PVP suggests the potential of the said nanophosphor for biophotonic applications. Furthermore, colorimetric studies proclaim pure red emission of the prepared nanophosphor under n-UV and infrared irradiations. Thermal analysis of $NaYF_4$:.15Eu^{3+} nanophosphor was conducted and the activation energy was observed as .421 eV, which suggests good thermal stability of the nanophosphor. The decay analysis shows bi-exponential fit for the observed data and fall in lifetime values. The fall in lifetime of the emitting state is related to the non-radiative transitions dominance over radiative transition for higher doping amount as validated by employing Auzel's model and have good quantum efficiency. The quantum efficiency was also calculated by using J–O theory and found in comparison with the values obtained from Auzel's model. Thus, Eu^{3+}-doped $NaYF_4$ nanophosphor are quite suitable for its usage in biophotonic devices.

ACKNOWLEDGMENTS

Dr. Sumandeep Kaur is grateful to Council of Scientific and Industrial Research, New Delhi, for the award of Research Associate fellowship (file no. 08/133(0053)/2020-EMR-I). The authors are thankful to the Delhi Technological University, Delhi, for providing the facilities to carry out the present work.

ORCID

A. Srinivasa Rao https://orcid.org/0000-0001-8199-6505

REFERENCES

1. Lee HL, Park JH, Choe HS, Lee MS, Park JM, Harada N, et al. Upconverting oil-laden hollow mesoporous silica microcapsules for anti-stokes-based biophotonic applications. ACS Appl Mater Interfaces. 2019;11(30):26571–80.
2. Sharma RK, Ghora M, Chouryal YN, Ganguly T, Acharjee D, Mondal DJ, et al. Multifunctional lanthanide-doped binary fluorides and graphene oxide nanocomposites via a task-specific ionic liquid. ACS Omega. 2022;7(20):16906–16.
3. Zheng B, Bai Y, Chen H, Pan H, Ji W, Gong X, et al. Near-infrared light-excited upconverting persistent nanophosphors in vivo for imaging-guided cell therapy. ACS Appl Mater Interfaces. 2018;10(23):19514–22.
4. Ilhan M, Ekmekci MK, Mergen A, Yaman C. Photoluminescence characterization and heat treatment effect on luminescence behavior of $BaTa_2O_6$:Dy^{3+} phosphor. Int J Appl Ceram Technol. 2017;14(6):1134–43.
5. Zhang X, Guan A, Zhou L, Gong M. Synthesis and luminescence study of $Zn_3V_2O_8$:Bi^{3+} yellow phosphor for solar spectral conversion. Int J Appl Ceram Technol. 2017;14(3):392–98.
6. Gan X, Tian B, Wang Z, Xu Y, Li L, Wang L, et al. Fabrication of Er^{3+}:$YAlO_3$ fibers by electrospinning method for upconversion luminescence applications. Int J Appl Ceram Technol. 2023;21(2):1199–207.
7. Chen G, Yang C, Prasad PN. Nanophotonics and nanochemistry: controlling the excitation dynamics for frequency up- and down-conversion in lanthanide-doped nanoparticles. Acc Chem Res. 2013;46(7):1474–86.
8. Ghosh P, Patra A. Influence of crystal phase and excitation wavelength on luminescence properties of Eu^{3+}-doped sodium yttrium fluoride nanocrystals. J Phys Chem C. 2008;112(49):19283–92.
9. Marin BC, Hsu SW, Chen L, Lo A, Zwissler DW, Liu Z, et al. Plasmon-enhanced two-photon absorption in photoluminescent semiconductor nanocrystals. ACS Photon. 2016;3(4):526–31.
10. Ren Y, Winter H, Rosch JG, Jung K, Duross AN, Landry MR, et al. PEGylated β-$NaGdF_4$/Tb@CaF_2 core/shell nanophosphors for enhanced radioluminescence and folate receptor targeting. ACS Appl Nano Mater. 2019;2(6):3718–27.
11. Wu X, Chen G, Shen J, Li Z, Zhang Y, Han G. Upconversion nanoparticles: a versatile solution to multiscale biological imaging. Bioconjug Chem. 2015;26(2):166–75.
12. Sharma RK, Nigam S, Chouryal YN, Nema S, Bera SP, Bhargava Y, et al. Eu-doped BaF_2 nanoparticles for bioimaging applications. ACS Appl Nano Mater. 2019;2(2):927–36.
13. Mclean A, Kanetidis M, Gogineni T, Ukani R, Mclean R, Cooke A, et al. Au nanobead chains with tunable plasmon resonance and intense optical scattering: scalable green synthesis, Monte Carlo assembly kinetics, discrete dipole approximation modeling, and nano-biophotonic application. Chem Mater. 2021;33(8):2913–28.
14. Kumar KN, Vijayalakshmi L, Choi J. Investigation of upconversion photoluminescence of Yb^{3+}/Er^{3+}:$NaLaMgWO_6$ noncytotoxic double-perovskite nanophosphors. Inorg Chem. 2019;58(3):2001–11.
15. Lingeshwar Reddy K, Srinivas V, Shankar KR, Kumar S, Sharma V, Kumar A, et al. Enhancement of luminescence intensity in red emitting $NaYF_4$:Yb/Ho/Mn upconversion nanophosphors by variation of reaction parameters. J Phys Chem C. 2017;121(21):11783–93.
16. Hong AR, Kim Y, Lee TS, Kim S, Lee K, Kim G, et al. Intense red-emitting upconversion nanophosphors (800 nm-driven) with a core/double-shell structure for dual-modal upconversion luminescence and magnetic resonance in vivo imaging applications. ACS Appl Mater Interfaces. 2018;10(15):12331–40.
17. Suyver JF, Grimm J, Van Veen MK, Biner D, Krämer KW, Güdel HU. Upconversion spectroscopy and properties of $NaYF_4$ doped with Er^{3+}, Tm^{3+} and/or Yb^{3+}. J Lumin. 2006;117(1):1–12.
18. Veronesi S, Parisi D, Faoro R, Tonelli M. Excitation and temperature dependence of the ^3P_J manifolds emission in KYF_4:Pr^{3+} single crystal. J Lumin. 2012;132(9):2307–11. https://doi.org/10.1016/j.jlumin.2012.03.059
19. Chauhan V, Deshmukh P, Satapathy S, Pandey PC. Greenish-yellow emission from rare-earth free Li^+ doped zinc vanadate phosphor. Results Phys. 2022;39:105689. https://doi.org/10.1016/j.rinp.2022.105689
20. Kumari C, Kumar A, Sharma S., Manam J. Sr_3LiSbO_6:Er^{3+} phosphors for green LEDs and solar cell applications. Vacuum. 2023;207:111599.

17447402, 0, Downloaded from https://ceramics.onlinelibrary.wiley.com/doi/10.1111/ijac.14805 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

21. Sharma VD, Khajuria P, Prakash R, Choudhary R. X-ray photoemission and optical investigation of novel Sm^{3+} activated $NaZr_2(PO_4)_3$ phosphor. Optik. 2022;267:169711.
22. Jha K, Jayasimhadri M. Effective sensitization of Eu^{3+} and energy transfer in Sm^{3+}/Eu^{3+} co-doped ZPBT glasses for CuPc based solar cell and w-LED applications. J Lumin. 2018;194:102–7. https://doi.org/10.1016/j.jlumin.2017.09.049
23. Rodriguez-Liviano S, Aparicio FJ, Rojas TC, Hungría AB, Chinchilla LE, Ocaña M. Microwave-assisted synthesis and luminescence of mesoporous RE-doped YPO_4 (RE = Eu, Ce, Tb, and Ce + Tb) nanophosphors with lenticular shape. Cryst Growth Des. 2012;12(2):635–45.
24. Fouad R, Saif M. Synthesis, spectroscopic and photoluminescence studies of novel Eu^{3+} nanophosphor complex as fluorescent sensor for highly sensitive detection of latent fingerprints and anti-counterfeiting. J Mol Struct. 2020;1217:128472. https://doi.org/10.1016/j.molstruc.2020.128472
25. Manasa P, Ran F, Basavapoornima C, Depuru SR, Jayasankar CK. Optical characteristics of (Eu^{3+},Nd^{3+}) co-doped lead fluorosilicate glasses for enhanced photonic device applications. J Lumin. 2020;223:117210. https://doi.org/10.1016/j.jlumin.2020.117210
26. Reisfeld R, Zigansky E, Gaft M. Europium probe for estimation of site symmetry in glass films, glasses and crystals. Mol Phys. 2004;102(11–12 special issue):1319–30.
27. Chang YS, Shi ZR, Tsai YY, Wu S, Chen HL. The effect of Eu^{3+}-activated $InVO_4$ phosphors prepared by sol–gel method. Opt Mater (Amst). 2011;33(3):375–80.
28. Dalal J, Dalal M, Devi S, Dhankhar P, Hooda A, Khatkar A, et al. Structural and Judd–Ofelt intensity parameters of a down-converting $Ba_2GdV_3O_{11}$:Eu^{3+} nanophosphors. Mater Chem Phys. 2020;243:122631. https://doi.org/10.1016/j.matchemphys.2020.122631
29. Blasse G. Energy transfer in oxidic phosphors. Phys Lett A. 1968;28(6):444–45.
30. Li Z, Zhang Y. Monodisperse silica-coated polyvinylpyrrolidone/$NaYF_4$ nanocrystals with multicolor upconversion fluorescence emission. Angew Chemie Int Ed. 2006;45(46):7732–35.
31. Wang M, Mi CC, Liu JL, Wu XL, Zhang YX, Hou W, et al. One-step synthesis and characterization of water-soluble $NaYF_4$:Yb,Er/polymer nanoparticles with efficient up-conversion fluorescence. J Alloys Compd. 2009;485(1–2):24–27.
32. Morkan AU, Gul E. Synthesis, structural characterization, and photoluminescence properties of Eu^{3+} -doped $Mg_{3-x}Eu_x(BO_3)_2$ phosphor. Int J Appl Ceram Technol. 2018;15AD:1–10.
33. Yao S, Li Y, Xue L, Yan Y. Synthesis and luminescent properties of a novel ultraviolet light-emitting diodes. Int J Appl Ceram Technol. 2011;8:701–8.
34. Patra A, Ghosh P, Chowdhury PS, Alencar MARC, Whualkuer LB, Rakov N, et al. Red to blue tunable upconversion in Tm^{3+}-doped ZrO_2 nanocrystals. J Phys Chem B. 2005;109(20):10142–46.
35. İlhan M, Ekmekçi MK, Keskin İÇ. Judd–Ofelt parameters and X-ray irradiation results of MNb_2O_6:Eu^{3+} (M = Sr, Cd, Ni) phosphors synthesizedviaa molten salt method. RSC Adv. 2021;11(18):10451–62.
36. Bajaj R, Rao AS, Prakash GV. Linear and nonlinear photoluminescence from thermally stable KYF_4:Eu^{3+} cubic nanocrystals. J Alloys Compd. 2021;885:160893. https://doi.org/10.1016/j.jallcom.2021.160893
37. Muh'd IB, Talib ZA, Baqiah H. Temperature-dependent photoluminescence of milled $Cd_{0.3}Zn_{0.7}Se$ nanocrystals. Optik (Stuttg). 2020;220:165117. https://doi.org/10.1016/j.ijleo.2020.165117
38. Wang L, Qiao J, Liu Y, Huang P, Shi Q, Tian Y, et al. Tunable luminescence of the full-color-emitting $LiGd_5P_2O_{13}$:Bi^{3+},Eu^{3+} phosphor based on energy transfers. Opt Mater (Amst). 2017;67:78–83. https://doi.org/10.1016/j.optmat.2017.03.048
39. Auzel F. A fundamental self-generated quenching center for lanthanide-doped high-purity solids. J Lumin. 2002;100(1–4):125–30.
40. Phogat P, Taxak VB, Malik RK. Er^{3+} doped $Ba_2GdV_3O_{11}$: synthesis and characterization for crystal and photoluminescent features of bright green emitting nanophosphor. Solid State Sci. 2022;133:107013. https://doi.org/10.1016/j.solidstatesciences.2022.107013
41. İlhan M, Güleryüz LF. Boron doping effect on the structural, spectral properties and charge transfer mechanism of orthorhombic tungsten bronze β-$SrTa_2O_6$:Eu^{3+} phosphor. RSC Adv. 2023;13(18):12375–85.
42. Suresh Kumar J, Pavani K, Sasikala T, Sreenivasa Rao A, Giri NK, Rai SB, et al. Photoluminescence and energy transfer properties of Sm^{3+} doped CFB glasses. Solid State Sci. 2011;13(8):1548–53. https://doi.org/10.1016/j.solidstatesciences.2011.05.020
43. Prakashbabu D, Ramalingam HB, Hari Krishna R, Nagabhushana BM, Chandramohan R, Shivakumara C, et al. Charge compensation assisted enhancement of photoluminescence in combustion derived Li+ co-doped cubic ZrO_2:Eu^{3+} nanophosphors. Phys Chem Chem Phys. 2016;18(42):29447–57. Available from: http://xlink.rsc.org/?DOI=C6CP04633A
44. Deopa N, Saini S, Kaur S, Prasad A, Rao AS. Spectroscopic investigations on Dy^{3+} ions doped zinc lead alumino borate glasses for photonic device applications. J Rare Earths. 2019;37(1):52–59. https://doi.org/10.1016/j.jre.2018.04.013
45. Liang H, Luo Z, Wu K, Tong J, Zhou Z, Lu A. Study of concentration and temperature-dependent photoluminescence properties in DyF_3-doped fluorophosphate glasses for thermal stable photonic device. Opt Mater (Amst). 2023;135:113256. https://doi.org/10.1016/j.optmat.2022.113256
46. Zhang Y, Chen J, Zhang J, An S. Intense emissions in Ce^{3+}/Eu^{2+} doped $Ca_{5.93-m}Sr_mBa(PO_4)_4O$ phosphors via host adjustment and energy transfer. Spectrochim Acta—Part A Mol Biomol Spectrosc. 2021;251:119431. https://doi.org/10.1016/j.saa.2021.119431
47. Wang L, Liu Y, Yu H, Mi R, Bu X, Mei L, et al. Far-red-emitting Mn^{4+}-doped $Ca_2GdZr_2Al_3O_{12}$ phosphors for growing indoor plants. J Lumin. 2022;251:119132. https://doi.org/10.1016/j.jlumin.2022.119132
48. Daniel DJ, Pandey IR, Karuppasamy P, Kim HJ, Lee MH. Optical and mechanical properties of $Li_2Mg_2(MoO_4)_3$ crystal grown by Czochralski method. Optik (Stuttg). 2020;207:164430. https://doi.org/10.1016/j.ijleo.2020.164430
49. Halappa P, Devakumar B, Shivakumara C. Effect of Ca^{2+} ion co-doping on radiative properties via tuning the local symmetry around the Eu^{3+} ions in orange red light emitting $GdPO_4$:Eu^{3+} phosphors. New J Chem. 2019;43(1):63–71.

17447402, 0, Downloaded from https://ceramics.onlinelibrary.wiley.com/doi/10.1111/ijac.14805 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

50. Li T, Li P, Wang Z, Xu S, Bai Q, Yang Z. Substituting different cations in tuning of the photoluminescence in $Ba_3Ce(PO_4)_3$. Inorg Chem. 2016;55(17):8758–69.
51. Qian T, Fan B, Wang H, Zhu S. Structure and luminescence properties of $Zn_3V_2O_8$ yellow phosphor for white light emitting diodes. Chem Phys Lett. 2019;715:34–39. https://doi.org/10.1016/j.cplett.2018.11.022
52. Zhang Q, Wang Q, Wang X, Ding X, Wang Y. Luminescence properties of Eu^{2+}-doped $BaSi_2O_5$ as an efficient green phosphor for light-emitting devices and wide color gamut field emission displays Qiang. New J Chem. 2016;40:8549–55.
53. Mei L, Xie J, Liu H, Liao L, Zhang Y, Li M. Luminescence properties and energy transfer of Ce^{3+}/Tb^{3+} co-doped $Ca_9La(PO_4)_5(SiO_4)F_2$ phosphor. Opt Commun. 2015;335:90–93.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kaur S, Arora S, Rao AS. Comparative study of luminescence in alkali-metal-based yttrium fluoride nanophosphor for biophotonic applications. Int J Appl Ceram Technol. 2024;1–13. https://doi.org/10.1111/ijac.14805

17447402, 0, Downloaded from https://ceramics.onlinelibrary.wiley.com/doi/10.1111/ijac.14805 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/376790144

Cosmic evolution in f (Q , T) gravity: Exploring a higher-order time-dependent function of deceleration parameter with observational constraints

Article *in* Physics of the Dark Universe · December 2023

DOI: 10.1016/j.dark.2023.101408

CITATIONS
5

READS
122

5 authors, including:



Madhur Khurana
Delhi Technological University
8 PUBLICATIONS **21** CITATIONS
SEE PROFILE



Himanshu Chaudhary
Delhi Technological University
46 PUBLICATIONS **163** CITATIONS
SEE PROFILE



Saadia Mumtaz
University of the Punjab
41 PUBLICATIONS **269** CITATIONS
SEE PROFILE



Shibesh Kumar Jas Pacif
Pacif Institute of Cosmology and Selfology (PICS)
75 PUBLICATIONS **859** CITATIONS
SEE PROFILE

All content following this page was uploaded by Himanshu Chaudhary on 31 December 2023.

The user has requested enhancement of the downloaded file.

Physics of the Dark Universe 43 (2024) 101408

Contents lists available at ScienceDirect

Physics of the Dark Universe

journal homepage: www.elsevier.com/locate/dark

Full length article

Cosmic evolution in $f(Q,T)$ gravity: Exploring a higher-order time-dependent function of deceleration parameter with observational constraints

Madhur Khurana [a], Himanshu Chaudhary [b,c,d], Saadia Mumtaz [e,1], S.K.J. Pacif [c], G. Mustafa [f,g,*]

[a] Department of Applied Physics, Delhi Technological University, Delhi 110042, India
[b] Department of Applied Mathematics, Delhi Technological University, Delhi 110042, India
[c] Pacif Institute of Cosmology and Selfology (PICS), Sagara, Sambalpur 768224, Odisha, India
[d] Department of Mathematics, Shyamlal College, University of Delhi, Delhi 110032, India
[e] Institute of Chemical Engineering and Technology, University of the Punjab, Quaid-e-Azam Campus, Lahore 54590, Pakistan
[f] Department of Physics, Zhejiang Normal University, Jinhua 321004, People's Republic of China
[g] New Uzbekistan University, Movarounnahr street 1, Tashkent 100000, Uzbekistan

ARTICLE INFO

Keywords:
Time-dependent deceleration parameter
f(Q,T) gravity
Hubble parameter

ABSTRACT

In this research paper, we explore a well-motivated parametrization of the time-dependent deceleration parameter, characterized by a cubic form in cosmic time t, in the context of late-time cosmic acceleration. The whole analysis of our work is within the framework of $f(Q,T)$ gravity theory, where we consider the background metric as the homogeneous and isotropic one represented by the Friedmann–Lemaître–Robertson–Walker (FLRW) metric. We solve the field equations with the considered functional form of $d(t)$. In order to find some best fit values of the model parameters and validate our obtained model, we constrain the model using some recent observational datasets, including cosmic chronometer (CC), Supernovae (SNIa), Baryon Acoustic Oscillation (BAO), Cosmic Microwave Background Radiation (CMB), Gamma-ray Burst (GRB), and Quasar (Q) datasets. The joint analysis of these datasets results in tighter constraints for the model parameters, enabling us to discuss both the physical and geometrical aspects of the model. Moreover, we determine the present values of the deceleration parameter (q_0), the Hubble parameter (H_0), and the transition redshift (z_t) from deceleration to acceleration ensuring consistency with some recent results of Planck 2018. Our statistical analysis yields highly improved results, surpassing those obtained in previous investigations. Overall, this study presents valuable insights into the higher order $q(t)$ model and its implications for late-time cosmic acceleration, shedding light on the nature of the late universe.

1. Introduction

In contemporary times, one of the most significant breakthroughs is the discovery of cosmic accelerated expansion, which has been confirmed through a range of observational techniques [1,2]. This expansion is accompanied by a mysterious energy component known as "dark energy" (DE), characterized by substantial negative pressure. Despite its enigmatic nature, DE constitutes about 70% of the cosmic content and plays a crucial role in maintaining the overall energy density of the universe consistent with predictions from inflationary theory. The investigation into the dominant constituents of the universe, namely dark energy and dark matter (DM), stands as a formidable challenge in modern physics, representing nearly 95% of the universe's imperceptible composition. Dark matter, an unseen type of matter exhibiting weak interactions with electromagnetic radiation, is detectable primarily through its gravitational effects on nearby ordinary matter. The existence of DM is supported by various observable phenomena, including rotation curves and mass discrepancies.

In the absence of any substantial evidence supporting the dark sources, alternative avenues have been explored extensively. The investigation of enigmatic approaches to these exotic terms has been delineated through two distinct methods: modifying matter sources or introducing additional degrees of freedom into the gravitational action. The initial approach involves the alteration of the matter sector within the Einstein–Hilbert Lagrangian density through the incorporation of various proposals. Numerous models have been proposed to explain DE, some of which correlate with observational evidence. Among these, the

* Corresponding author at: Department of Physics, Zhejiang Normal University, Jinhua 321004, People's Republic of China.
E-mail addresses: K.madhur2000@gmail.com (M. Khurana), himanshuch1729@gmail.com (H. Chaudhary), saadia.icet@pu.edu.pk (S. Mumtaz), shibesh.math@gmail.com (S.K.J. Pacif), gmustafa3828@gmail.com (G. Mustafa).
[1] Second corresponding author.

https://doi.org/10.1016/j.dark.2023.101408
Received 26 September 2023; Received in revised form 13 December 2023; Accepted 18 December 2023
Available online 23 December 2023
2212-6864/© 2023 Elsevier B.V. All rights reserved.

ΛCDM model is widely embraced due to its alignment with observations, although it also presents certain ambiguities such as fine-tuning and coincidence issues [3–7]. To address these limitations, alternative DE models like quintessence [8–10], phantom [11], k-essence [12–14], and Chaplygin gas [15] have been introduced, aiming to provide effective explanations for various cosmic inquiries and to explore diverse aspects of this exotic energy. Local (model-independent) H_0 values are biased to larger values than Planck ΛCDM. This disfavors Quintessence, K-essence, and any DE model in the Quintessence regime $w_{DE}(z) > -1$, while favoring phantom models, [16–20] H_0 tension is telling us new things about alternative DE models. Preceding the cosmic acceleration phase, the universe underwent a deceleration phase during its early epochs, where the impact of DE was relatively minor. It is believed that density perturbations during this period played a pivotal role in shaping cosmic structures. As a result, understanding the complete evolutionary timeline requires a cosmological model capable of describing both acceleration and deceleration phases.

The second approach involves an extension of the gravitational component in GR by introducing a DE source while keeping the matter sector unchanged. As the first category holds intriguing implications, it has not gained as much support due to certain ambiguities. In contrast, modified gravity frameworks have proven to be quite valuable due to their effective execution in the field of cosmology. Einstein formulated the concept of geometry-matter coupling whose adaptation has been integral to the development of General Relativity (GR). Several alternative theories of gravity, which incorporate additional curvature terms in the gravitational action, have been thoroughly examined in the scientific literature. Some well-known examples of these modified gravity theories encompass $f(R)$ gravity [21], Gauss–Bonnet gravity [22], $f(R,T)$ gravity [23], the scalar-tensor theory [24], $f(T)$ gravity [25], $f(T,T_G)$ gravity [26], $f(Q)$ gravity [27], etc. These alterations are rooted in the metric tensor g_{ij} being considered a dynamic variable. Interestingly, an alternative approach to GR has been gaining momentum in the scientific literature, known as $f(Q,T)$ gravity [28] appearing as an extension of symmetric teleparallel gravity, where Q and T correspond to the non-metricity and the trace of the energy–momentum tensor, respectively. This novel theory has sparked significant interest in investigating the late universe and has been the subject of different researchers in various contexts [29–32].

The Einstein field equations within the framework of General Relativity (GR) are notoriously intricate, comprising a set of nonlinear differential equations that pose significant challenges in terms of finding analytical solutions. To simplify these complexities, physicists often make certain physical assumptions, such as establishing relationships between the pressure and energy density of the universe's contents. However, when introducing components like dark energy, these equations become even more tangled. This complexity also extends to modified gravity (MG) theories, which incorporate higher-order derivative terms. To obtain tractable solutions for the field equations, whether within GR or MG theories, researchers employ a variety of techniques, including dynamical system analysis, autonomous systems, and notably, the model-independent approach. The model-independent approach, which typically involves the parametrization of any cosmological parameter, which is generally a functional form of the parameter. While the parametrization of cosmological parameters may initially appear adhoc, it proves to be a legitimate and powerful method when considering the broad spectrum of possible evolutions for the geometrical and physical parameters from a mathematical standpoint. This idea has garnered increasing interest in the realm of mathematical cosmology. This concept of cosmological parametrization has been extensively explored in the scientific literature, with comprehensive discussions by Pacif [33,34], providing valuable insights into its application and significance in cosmological studies. There are also various studies on the model-independent approach by Eric V. Linder within some cosmological contexts. Specifically, he delved into the intricate studies of cosmological parametrization, a crucial facet of his research, which aimed to comprehensively study and understand a wide range of dark energy models that play a pivotal role in understanding the universe's accelerating expansion [35–39]

Modern cosmology advocates the investigation of kinematic quantities, often referred to as "Cosmography" or "Cosmo-kinetics". This approach relies on observational data and sidesteps prior assumptions about gravity theory or specific cosmological models. Cosmography's adoption of symmetry principles offers a means to navigate debates surrounding DE, dark matter, and related topics without invoking Einstein field equations (Friedmann equations). While cosmography does not directly involve the scale factor, it allows for some inference into its evolutionary history. The foundation of modern cosmology traces back to the work by Sandage [40], who introduced fundamental cosmographic parameters. These parameters include the Hubble parameter (H_0), which dictates the rate of cosmic expansion, and the deceleration parameter (q_0), responsible for accounting for the effects of gravity that slow down the expansion rate. Despite the emergence of DE altering the cosmological landscape, the dynamics of cosmic evolution remain intimately connected to the deceleration parameter (q), defined as $q = -\frac{a\ddot{a}}{\dot{a}^2}$, where $a(t)$ represents the scale factor. A positive value of q ($\ddot{a} < 0$) indicates deceleration, while a negative value implies acceleration. The Hubble parameter characterizes the linear temporal evolution of $a(t)$, and a non-linear correction term (q_0) introduces local instabilities and chaotic behavior [41]. The deceleration parameter (q) also plays a vital role in examining the dynamics of observable galaxy numbers. Dark energy gains empirical support through various modifications to fit observational data, making its parameterization as a function of the scale factor ($a(t)$) or redshift (z) a valuable approach.

In the current paper, we have considered a parametrization of the deceleration parameter given in [42] within the classical gravity and extend the analysis in $f(Q,T)$ gravity, which is also a special case of the parametrization considered in [33]. A plethora of parametric forms for the deceleration parameter have been explored, each with its own set of limitations. While some diverge at large cosmic times, others remain well-behaved at low redshifts ($z \ll 1$) [43–81]. The adoption of parametric assumptions can occasionally lead to misinterpretations of the nature of DE, prompting an interest in non-parametric models that directly infer evolutionary mechanisms from observational data [82–86]. Nevertheless, the parameterization of q proves to be a more effective strategy for investigating the transition from cosmic deceleration to acceleration.

Highlighting the versatility and gravitational-theory independence of the parameterized q approach. In this study, we have made some analysis of the model produced by the cubic parametrization of the deceleration parameter and found tighter constraints to the model parameters involved. Additionally, we have done some cosmographic tests in order to differentiate a better-performing model when compared with cosmological data.

The paper is organized in the following manner. In Section 1, we provide an overview of the current state of modern cosmology, highlighting the various scenarios proposed to explain the late-time cosmic acceleration of the universe. We also outline the scope and objectives of the paper. In Section 2, we introduce the $f(Q,T)$ gravity framework and present the associated field equations. We derive the modified Friedmann equations within this context. In Section 3, we discuss a model-independent approach to parametrize the cosmological model. This section lays the foundation for our subsequent analyses. Section 4, focuses on characterizing the model through dynamical variables, providing insights into its behavior and evolution. In Section 5, we derive all cosmological and physical parameters in terms of redshift, enhancing our understanding of the model's evolution. Section 6 employs the Markov chain Monte Carlo (MCMC) method to constrain the model's parameters. We determine the best-fit values of model parameters by comparing them with the present Hubble rate function across different datasets. In Section 7, we validate the

model's predictions by comparing them with observational data, Section 8 delves into the kinematic cosmographic parameters, including the deceleration and jerk parameters, providing a detailed analysis of their implications within the model. Sections 9 and 10 are dedicated to the statefinder and *Om* diagnostic tests, offering further insights into the model's behavior and compatibility with observational data. In 11, we conduct a comprehensive statistical analysis to rigorously assess the model's performance and reliability. Sections 11 and 12 present the results of our study and draw conclusions based on our findings, summarizing the contributions and implications of this research.

2. Cosmological equations in $f(Q,T)$ gravity

From the modification in Einstein–Hilbert action of general relativity by introducing a function of two scalar invariants, Q and T, which are constructed from the non-metricity and the trace of the energy–momentum tensor. The action in $f(Q,T)$ is given by [87].

$$S = \int \left[\frac{1}{16\pi} f(Q,T) + \mathcal{L}_M\right] \sqrt{-g}\, d^4x \tag{1}$$

where, $g \equiv det(g_{\mu\nu})$, and $\mathcal{L}_M$ is Lagrangian density.

In Riemannian geometry, the metric tensor is always symmetric. However, for our research, we take the $f(Q,T)$ gravity, in which, we can take the non-symmetric part of the metric tensor called non-metricity. Which can be defined as

$$Q \equiv -g^{\mu\nu}(L^{\alpha}_{\beta\mu}L^{\beta}_{\nu\alpha} - L^{\alpha}_{\beta\alpha}L^{\beta}_{\mu\nu}) \tag{2}$$

where

$$L^{\alpha}_{\beta\gamma} \equiv -\frac{1}{2}g^{\alpha\lambda}(\nabla_{\gamma}g_{\beta\lambda} + \nabla_{\beta}g_{\lambda\gamma} - \nabla_{\lambda}g_{\beta\gamma}) \tag{3}$$

The trace of nonmetricity tensor is expressed as,

$$Q_{\alpha} \equiv Q_{\alpha}{}^{\mu}{}_{\mu},\ \tilde{Q}_{\alpha} \equiv Q^{\mu}_{\alpha\mu} \tag{4}$$

The trace of the energy–momentum tensor and modification in the metric tensor are respectively

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta(\sqrt{-g}\mathcal{L}_M)}{\delta g^{\mu\nu}} \tag{5}$$

$$\Theta_{\mu\nu} = g^{\alpha\beta}\frac{\delta T_{\alpha\beta}}{\delta g^{\mu\nu}} \tag{6}$$

Finding the variation of action of the field equation Eq. (1) with respect to metric tensors.

$$8\pi T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\nabla_{\alpha}(f_Q\sqrt{-g}P^{\alpha}_{\mu\nu} - \frac{1}{2}f g_{\mu\nu} + f_T(T_{\mu\nu} + \Theta_{\mu\nu}) - f_Q(P_{\mu\alpha\beta}Q^{\alpha\beta}_{\nu} - 2Q^{\alpha\beta}{}_{\mu}P_{\alpha\beta\nu})) \tag{7}$$

Where super-momentum

$$P^{\alpha}_{\mu\nu} \equiv \frac{1}{4}\left[-Q^{\alpha}_{\mu\nu} + 2Q^{\alpha}_{(\mu\ \nu)} + Q^{\alpha}g_{\mu\nu} - \tilde{Q}^{\alpha}g_{\mu\nu} - \delta^{\alpha}_{(\mu}Q_{\nu)}\right] = -\frac{1}{2}L^{\alpha}_{\mu\nu} + \frac{1}{4}\left(Q^{\alpha} - \tilde{Q}^{\alpha}\right)g_{\mu\nu} - \frac{1}{4}\delta^{\alpha}_{(\mu}Q_{\nu)} \tag{8}$$

Taking the FLRW metric as follows,

$$ds^2 = -N(t)^2dt^2 + a(t)^2(dx^2 + dy^2 + dz^2), \tag{9}$$

where N(t) is the lapse function and a(t) is the scale factor. Hence, $Q = 6H^2/N^2$. We assume the value of N(t) = 1, for a standard case. Hence, $Q = 6H^2$.

To find the generalized Friedmann equations, assuming the matter content as the perfect fluid with the energy–momentum tensor $T^{\mu}_{\nu} = diag(-\rho, p, p, p)$. The tensor Θ^{μ}_{ν} becomes,

$$\Theta^{\mu}_{\nu} = \delta^{\mu}_{\nu}p - 2T^{\mu}_{\nu} = diag(2\rho + p, -p, -p, -p) \tag{10}$$

For simplicity, taking $F \equiv f_Q = dF/dt$ and $8\pi\tilde{G} \equiv f_T = dF/dt$ the Friedmann equations we derived as follows,

$$8\pi\rho = \frac{f}{2} - 6FH^2 - \frac{2\tilde{G}}{1+\tilde{G}}(\dot{F}H + F\dot{H}) \tag{11}$$

$$8\pi p = -\frac{f}{2} + 6FH^2 + 2(\dot{F}H + F\dot{H}) \tag{12}$$

From Eqs. (11) and (12), modified Einstein's field equations are derived

$$3H^2 = 8\pi\rho_{eff} = \frac{f}{4F} - \frac{4\pi}{F}[(1+\bar{G})\rho + \bar{G}p] \tag{13}$$

$$2\dot{H} + 3H^2 = -8\pi p_{eff} = \frac{f}{4F} - \frac{2\dot{F}H}{F} + \frac{4\pi}{F}\left[(1+\bar{G})\rho + (2+\bar{G})p\right] \tag{14}$$

From Eqs. (13) and (14) and the derivative of Eq. (13) we derive the continuity equation

$$\dot{\rho}_{eff} + 3H\left(\rho_{eff} + p_{eff}\right) = 0 \tag{15}$$

Although there are several forms of the function $f(Q,T)$ is considered in the literature [88], we here only confined to the linear and additive form of $f(Q,T)$ function [87,89] in the form

$$f(Q,T) = \mu Q + \nu T \tag{16}$$

where, μ and ν are the non zero model constants. Hence the first derivatives $f_Q = \mu$ and $8\pi\tilde{G} = f_T = \nu$. Solving the modified Friedmann equations and applying the barotropic equation of states $p = \omega\rho$ we can find the equation of state parameter as follows

$$\omega = \frac{3H^2(8\pi+\nu) + \dot{H}(16\pi+3\nu)}{\nu\dot{H} - 3H^2(8\pi+\nu)} \tag{17}$$

Hence the energy density equation turns out to be

$$\rho = \frac{-3H^2\mu(8\pi+\nu) + \mu\nu\dot{H}}{2(4\pi+\nu)(8\pi+\nu)} \tag{18}$$

To find the value of $\dot{H}$ we use the relation $a_0/a = 1+z$ we can define a new relation between z and t.

$$\frac{d}{dt} = \frac{dz}{dt}\frac{d}{dz} = -(1+z)H(z)\frac{d}{dz} \tag{19}$$

normalizing the equation by taking the value of the scale factor as $a_0 = a(t=t_0) = 1$, t_0 refers the present time. Hence we can write the derivation Hubble parameter with respect to time in terms of red-shift as,

$$\dot{H} = -(1+z)H(z)\frac{dH}{dz} \tag{20}$$

3. The model

The modified field equations, denoted as Eqs. (11) and (12) are two independent equations describing the dynamics of the model. These fundamental equations intertwined with three unknown variables: a, ρ, and p. This implies that, in order to completely characterize and solve this system, we are in need of an additional equation. To address this, researchers in the field have adopt a widely accepted approach *the model independent way* study of model, often considers a scheme of parameterization of a cosmological parameter. In contemporary scientific literature, this approach is gaining prominence, as it is capable of solving the field equations in a model independent way that do not affect the background physics but provide a solution in simple way. It involves making an initial assumption regarding the scale factor, which can be done either directly or by expressing it in terms of cosmological parameters and their time derivatives. This strategic choice not only simplifies the equations but also allows for a more comprehensive exploration of the model's dynamics. To aid researchers in this endeavor, a wealth of parametrizations for various cosmological parameters has been meticulously compiled and organized in Refs. [33,90]. These compilations serve as invaluable resources, facilitating the selection of appropriate parameterization schemes based on the specific characteristics and goals of a given study. There are a few intriguing models of dark energy and modified gravity based on various parametrization

schemes of some geometrical parameters [91–95]. One noteworthy aspect of this concept is its ability to reconstruct the cosmic history and also the fate, while offering solutions to certain problems of standard cosmology.

In our own investigation, we opt to employ the parameterization of the deceleration parameter to effectively close the aforementioned system of equations. This choice aligns with established practices, streamlining our analytical framework while enabling us to delve deeply into the dynamics of our model. We employ here a scheme of parametrization of higher order time-dependent deceleration parameter ($q(t)$) in the form,

$$q(t) = -1 + \frac{m}{n} - \frac{4}{n}t^3, \tag{21}$$

where $m > 0$ and $n > 0$ are two arbitrary constants. Similar form is also considered in the Ref. [42] in the context of $f(R,T)$ gravity. By using the definition of the deceleration parameter in the form of Hubble parameter, $q = \frac{d}{dt}\left(\frac{1}{H}\right) - 1$, we may obtain an explicit form of Hubble function as

$$H(t) = \frac{n}{t\left(m - t^3\right)} \tag{22a}$$

Note that this explicit form of the Hubble parameter can also be interpreted as a particular form of the parametrization given in Ref. [33] as a special case. Using the definition of the Hubble parameter in terms of scale factor will give the explicit form of scale factor as,

$$a = \beta\left(\frac{t^3}{m - t^3}\right)^{\frac{n}{3m}}, \tag{23}$$

The deceleration parameter plays a crucial role in characterizing the dynamics of the universe's expansion. It provides insights into whether the expansion is slowing down or accelerating. Specifically, when the deceleration parameter is greater than zero ($q > 0$), it signifies a phase of decelerating expansion. Conversely, when the deceleration parameter is less than zero ($q < 0$), it indicates an accelerating expansion phase. A particularly interesting point occurs when the deceleration parameter equals zero ($q = 0$), marking a significant phase transition in the expansion of the universe. This phase transition at $q = 0$ occurs at a specific time, denoted as t_{tr}, which can be calculated in this considered case as, $t_{tr} = \sqrt[3]{\frac{m-n}{4}}$. This implies, m must be greater than n. This moment represents a critical juncture in the universe's evolution, where the expansion dynamics shift fundamentally. We can observe in Eqs. (23) and (22a), for $t = m^{1/3}$, $H \longrightarrow \infty$ and $a \longrightarrow \infty$, implying the existence of a big rip singularity in this model, which is anticipated to happen in the near future, precisely at $t = t_{end} = m^{1/3}$. The universe's expansion reaches a point of extreme instability and divergence, characterized by this singularity.

4. Characterizing the model through dynamical variables

This section is dedicated to providing a comprehensive analysis of the physical behavior and properties inherent in our model. Within this context, we aim to elucidate the model's physical dynamics by examining key parameters, including energy density (ρ), pressure (p), and the equation of state parameter (ω). These parameters are essential in understanding the behavior of our model and the physical processes it encapsulates. The temporal or redshift evolution of these parameters serves as a window into the dynamic nature of our model, shedding light on various aspects of the Universe's evolution, particularly during its late stages and with regard to stability considerations.

Our discussion in this section serves as the bedrock upon which a deeper comprehension of our model is built. It lays the essential groundwork for subsequent analyses and discussions. To initiate our exploration, we leverage the field equations, denoted as Eqs. (11) and (12). These equations provide us with a means to express the energy density (ρ) and pressure (p) in specific mathematical forms. This mathematical representation is pivotal in uncovering the intricate details of our model's physical behavior and will serve as the basis for our further investigations and interpretations. The explicit expressions for ρ and p using Eqs. (23) and (22a) are found in this case as;

$$\rho(t) = \frac{-n\mu[\nu(m - 4t^3) + 3n(8\pi + \nu)]}{2(-mt + t^4)^2(4\pi + \nu)(8\pi + \nu)} \tag{24}$$

and

$$p(t) = \frac{n\mu[8\pi(-2m + 3n + 8t^3) + 3\nu(-m + n + 4t^3)]}{2(-mt + t^4)^2(4\pi + \nu)(8\pi + \nu)} \tag{25}$$

The concept of the equation of state parameter denoted as ω, holds significant importance in the realm of cosmology. It serves as a fundamental tool for characterizing the relationship between pressure (p) and energy density (ρ) within a given system. This parameter is defined as the ratio of pressure to energy density, and it plays a crucial role in describing the thermodynamic behavior of fluids within the Universe. The value of ω offers critical insights into how a substance or component reacts to variations in volume or temperature. These insights, in turn, have profound implications for the overall dynamics and fate of the Universe. In this context, we have already derived expressions for both energy density and pressure specific to our model.

Now, by employing Eqs. (24) and (25), we can deduce the expression for the equation of state parameter ω, as follows:

$$\omega(t) = \frac{8\pi(2m - 3n - 8t^3) + 3\nu(m - n - 4t^3)}{\nu(m - 4t^3) + 3n(8\pi + \nu)} \tag{26}$$

5. Deriving cosmological parameters in terms of redshift

To effectively constrain our model parameters, we need to express all the obtained cosmological variables in terms of the redshift parameter, denoted as z. To achieve this, we utilize the relationship between the redshift (z) and the scale factor (a), given by

$$1 + z = \frac{a_0}{a}, \tag{27}$$

where a_0 is the present value of the scale factor and generally normalized to be $a_0 = 1$. The $t - z$ relationship in our case can be established as;

$$t = \sqrt[3]{k_1}\left\{1 + [\zeta(1 + z)]^3\frac{m}{n}\right\}^{-\frac{1}{3}} \tag{28}$$

Now, the Hubble parameter in terms of redshift z for the Model is,

$$H(z) = H_0\left(1 + \zeta^{3\eta}\right)^{-\frac{4}{3}}(1 + z)^{-3\eta}\left\{1 + [\zeta(1 + z)]^{3\eta}\right\}^{\frac{4}{3}}, \tag{29}$$

where $\frac{m}{n} = \eta$. Now, the deceleration parameter, energy density, pressure, and the equation of state parameter can be rewritten in terms of redshift z as follows (see Box I).

To comprehensively explore the evolution of both geometric and physical parameters, as well as to rigorously assess the validity of our derived model, it becomes imperative to acquire precise values for the model parameters that play a pivotal role in our theoretical framework. Thus, we perform the data analysis in the subsequent section to derive some precise values of the model parameter. Through this data analysis, we aim to gain a deeper understanding of the dynamics governing our cosmological model and evaluate its consistency with observation too.

6. Data analysis

In this section, we conduct an extensive comparison between our proposed cosmological model and a wide range of available cosmological data. Our aim is to understand the fundamental characteristics of the model through various datasets, including the Cosmic Chronometers (CC), type Ia supernovae (SNIa), Gamma Ray Bursts (GRBs), Quasars (Q), Baryon Acoustic Oscillation (BAO) and Cosmic Microwave Background (CMB) This rigorous investigation aims to determine the optimal values of key model parameters, such as η, ζ. These

$$q(z) = -1 + \eta - \frac{4\eta}{\left\{1 + [\zeta(1+z)]^{3\eta}\right\}}, \tag{30}$$

$$p(z) = -\frac{\mu H_0^2 \left(1 + (\zeta(1+z))^{3\eta}\right)^{5/3} (16\pi + 3\nu)\eta \left(-3 + (\zeta(1+z))^{3\eta}\right) - 3(8\pi + \nu)\left(1 + (\zeta(1+z))^{3\eta}\right)}{2(4\pi + \nu)(8\pi + \nu)(1 + \zeta^2(1+z)^{3\eta})^{8/3}} \tag{31}$$

$$\rho(z) = -\frac{\left((1+z)^{-6\eta}\mu H_0^2 \left(1 + ((1+z)\zeta)^{3\eta}\right)^{5/3} \left(3(8\pi + \nu - \nu\eta) + (3(8\pi + \nu) + \nu\eta)((1+z)\zeta)^{3\eta}\right)\right)}{2(4\pi + \nu)(8\pi + \nu)\left(1 + \zeta^3\eta^8\right)^{1/3}} \tag{32}$$

$$\omega(z) = \frac{(16\pi + 3\nu)\eta\left(-3 + ((1+z)\zeta)^{3\eta}\right) - 3(8\pi + \nu)\left(1 + ((1+z)\zeta)^{3\eta}\right)}{3(8\pi + \nu - \nu\eta) + 3(8\pi + \nu) + \nu\eta((1+z)\zeta)^{3\eta}} \tag{33}$$

Box I.

datasets together describe how our model behaves in the context of the $f(Q,T)$ gravity framework. Importantly, we also account for the present-day Hubble function, denoted as H_0, which plays a crucial role in shaping our results. To obtain best fit values our cosmological model we employ a robust Bayesian statistical methodology. This method relies on likelihood functions and a well-accepted technique called Markov Chain Monte Carlo (MCMC). Within this Bayesian framework, we construct a probabilistic assessment of how likely certain combinations of model parameters are based on actual observations. Our investigation reveals hidden aspects of our cosmological model, offering valuable insights into how it deeply connects with the observable universe.

6.1. Methodology

Constraining the Hubble function through observational data involves a process known as parameter estimation or model fitting. In our case, our goal is to determine the optimal values of key model parameters, using various datasets, including Cosmic Chronometers (CC), Type Ia Supernovae (SNIa), Gamma-Ray Bursts (GRBs), Quasars (Q), Baryon Acoustic Oscillations (BAO), and the Cosmic Microwave Background (CMB). The initial step is to establish a likelihood function that measures the agreement between our model predictions and the observed data.

$$\mathcal{L}(\theta) = \exp\left(-\frac{1}{2}\sum_{i=1}^{N}\frac{(O_i - M_i(\theta))^2}{\sigma_i^2}\right) \tag{34}$$

Where O_i is the observed data point for the ith data point, $M_i(\theta)$ is the model prediction for the ith data point based on the parameters θ, and σ_i represents the uncertainty associated with the observed data point. Next, we perform Bayesian parameter estimation [96]. This requires defining prior distributions for each parameters we want to constrain, which should encapsulate any existing knowledge or constraints. If strong prior information is lacking, relatively flat priors can be used. The posterior distribution, proportional to the likelihood function multiplied by the prior distribution, is then computed as:

$$P(\theta|D) \propto \mathcal{L}(\theta) \times \pi(\theta) \tag{35}$$

In the subsequent steps, Markov Chain Monte Carlo (MCMC) is employed to explore the posterior distribution and derive parameter constraints [97]. This widely employed sampling technique generates an extensive set of samples from the posterior distribution, allowing for the extraction of key statistical measures. These measures include mean, median, standard deviation, and credible intervals for each parameter, providing optimal parameter values and their associated uncertainties. Subsequently, we rigorously evaluate our model's performance against the Cosmic Chronometers (CC) and type Ia supernovae (SNIa) datasets. Visual comparisons and quantitative metrics, such as chi-squared values or the Akaike Information Criterion (AIC), are used to assess the model's fit. The ensuing discussions delve into the implications of the derived parameter constraints, emphasizing the model's alignment with observational data. We explore potential physical interpretations and ramifications, enriching our understanding of the underlying cosmological dynamics.

6.2. Data description

6.2.1. Cosmic chronometers

In our analysis, we utilize thirty-one data points acquired through the cosmic chronometers (CC) technique for the determination of the Hubble parameter. This approach allows us to directly extract information about the Hubble function at various redshifts, extending up to $z \lesssim 2$. The selection of CC data is motivated by its reliability, as it primarily involves measurements of the age difference between two passively evolving galaxies that originated at the same time but have a slight separation in redshift. This technique enables us to compute $\Delta z/\Delta t$, making CC data preferable to methods based on absolute age determinations for galaxies [98]. Our chosen CC data points were sourced from independent [99–105]. Importantly, these references are not influenced by the Cepheid distance scale or any specific cosmological model. Nevertheless, it is worth noting that they do depend on the modeling of stellar ages, which is established using robust stellar population synthesis techniques (for more details, see [101,103,106–109] for analyses related to CC systematics). We evaluate the goodness of fit using the χ^2_{CC} estimator, which is expressed as follows:

$$\chi^2_{CC}(\Theta) = \sum_{i=1}^{31}\frac{\left(H\left(z_i,\Theta\right) - H_{\rm obs}\left(z_i\right)\right)^2}{\sigma_H^2\left(z_i\right)}, \tag{36}$$

Here, $H\left(z_i,\Theta\right)$ represents the theoretical Hubble parameter values at redshift z_i with model parameters denoted as Θ. The observational data for the Hubble parameter at z_i is given by $H_{\rm obs}\left(z_i\right)$, with an associated observational error of $\sigma_H\left(z_i\right)$.

6.2.2. Type Ia supernovae (SNIa)

Over the years, a multitude of supernova datasets has been established, including references such as [110–114]. Recently, a refreshed version of the Pantheon dataset, referred to as Pantheon+, has been introduced [115]. This updated compilation comprises 1701 data points of type Ia supernovae (SNIa), spanning the redshift interval $0.001 < z < 2.3$. SNIa observations have played a pivotal role in unveiling the phenomenon of the universe's accelerating expansion. These observations serve as crucial tools for investigating the nature of the driving component behind this expansion, owing to SNIa's status as luminous

astrophysical objects. These objects, often treated as standard candles, enable the measurement of relative distances based on their intrinsic brightness. The Pantheon+ dataset stands as a valuable resource, offering insights into the accelerating universe's characteristics. The chi-square statistic serves as a fundamental tool for comparing theoretical models with observational data. In the context of the Pantheon+ dataset, chi-square values are computed using the subsequent equation

$$\chi^2_{\text{Pantheon+}} = \vec{D}^T \cdot \mathbf{C}^{-1}_{\text{Pantheon+}} \cdot \vec{D} \tag{37}$$

Here, $\vec{D}$ represents the difference between the observed apparent magnitudes m_{Bi} of SNIa and the expected magnitudes given by the cosmological model. M represents the absolute magnitude of SNIa, and μ_{model} is the corresponding distance modulus predicted by the assumed cosmological model. The term $\mathbf{C}_{\text{Pantheon+}}$ denotes the covariance matrix provided with the Pantheon+ data, which includes both statistical and systematic uncertainties. The distance modulus is a measure of the distance to an object, defined as:

$$\mu_{\text{model}}(z_i) = 5\log_{10}\left(\frac{D_L(z_i)}{(H_0/c)\ \text{Mpc}}\right) + 25 \tag{38}$$

Here, $D_L(z)$ represents the luminosity distance, which is calculated for a flat homogeneous and isotropic FLRW universe as:

$$D_L(z) = (1+z)H_0 \int_0^z \frac{dz'}{H(z')} \tag{39}$$

The Pantheon+ dataset differs from the previous Pantheon sample as it breaks the degeneracy between the absolute magnitude M and the Hubble constant H_0. This is achieved by rewriting the vector $\vec{D}$ in terms of the distance moduli of SNIa in the Cepheid hosts. The distance moduli in the Cepheid hosts, denoted as μ_i^{Ceph}, are measured independently using Cepheid calibrators. This allows for the independent constraint of the absolute magnitude M. The modified vector $\vec{D}'$ is defined as:

$$\vec{D'}_i = \begin{cases} m_{Bi} - M - \mu_i^{\text{Ceph}} & \text{if } i \text{ is in Cepheid hosts} \\ m_{Bi} - M - \mu_{\text{model}}(z_i) & \text{otherwise} \end{cases} \tag{40}$$

With this modification, the chi-square equation for the Pantheon+ dataset can be rewritten as:

$$\chi^2_{\text{SN}} = \vec{D'}^T \cdot \mathbf{C}^{-1}_{\text{Pantheon+}} \cdot \vec{D'} \tag{41}$$

This revised formulation allows for improved constraints on the absolute magnitude M and the cosmological parameters.

We have also expanded our investigation to include a subset of 162 Gamma Ray Bursts (GRBs) [116], spanning a redshift range of $1.44 < z < 8.1$. In this context, we define the χ^2 function as:

$$\chi^2_{\text{GRB}}(\phi^{\nu}_{\text{g}}) = \mu_{\text{g}} \mathbf{C}^{-1}_{\text{g,cov}} \mu^T_{\text{g}}, \tag{42}$$

Here, μ_{g} denotes the vector encapsulating the differences between the observed and theoretical distance moduli for each individual GRB. Similarly, for our examination of 24 compact radio quasar observations [117], spanning redshifts in the range of $0.46 \leq z \leq 2.76$, we establish the χ^2 function as:

$$\chi^2_{\text{Q}}(\phi^{\nu}_{\text{q}}) = \mu_{\text{q}} \mathbf{C}^{-1}_{\text{q,cov}} \mu^T_{\text{q}}, \tag{43}$$

In this context, μ_{q} represents the vector capturing the disparities between the observed and theoretical distance moduli for each quasar.

6.2.3. Baryon acoustic oscillations

To study Baryon Acoustic Oscillations (BAO), we utilize a dataset consisting of 333 measurements [118–129]. However, considering the potential error arising from data correlations, we select a smaller dataset of 17 BAO measurements for our analysis (please see table 1 of this work [130]). This selection helps to reduce errors and improve the accuracy of our results. One of the key measurements obtained from BAO studies in the transverse direction is the quantity $D_H(z)/r_d$, where $D_H(z)$ represents the comoving angular diameter distance. It is related to the following expression [131,132]:

$$D_M = \frac{c}{H_0} S_k\left(\int_0^z \frac{dz'}{E(z')}\right), \tag{44}$$

where $S_k(x)$ is defined as:

$$S_k(x) = \begin{cases} \frac{1}{\sqrt{\Omega_k}} \sinh\left(\sqrt{\Omega_k}x\right) & \text{if} \quad \Omega_k > 0 \\ x & \text{if} \quad \Omega_k = 0 \\ \frac{1}{\sqrt{-\Omega_k}} \sin\left(\sqrt{-\Omega_k}x\right) & \text{if} \quad \Omega_k < 0. \end{cases} \tag{45}$$

Additionally, we consider the angular diameter distance $D_A = D_M/(1+z)$ and the quantity $D_V(z)/r_d$. The latter is a combination of the coordinates of the BAO peak and r_d, representing the sound horizon at the drag epoch. Furthermore, we can directly obtain "line-of-sight" or "radial" observations from the Hubble parameter using the expression:

$$D_V(z) \equiv \left[z D_H(z) D_M^2(z)\right]^{1/3}. \tag{46}$$

By studying these BAO measurements, we gain insights into the cosmological properties and evolution of the universe, while minimizing potential errors and considering relevant distance measures and observational parameters.

6.2.4. Cosmic microwave background

The CMB distant prior measurements are taken [133]. The distance priors offer useful details about the CMB power spectrum in two ways: the acoustic scale l_A characterizes the CMB temperature power spectrum in the transverse direction, causing the peak spacing to vary, and the "shift parameter" R influences the CMB temperature spectrum along the line-of-sight direction, affecting the peak heights, which are defined as follows:

$$l_A = (1+z_d)\frac{\pi D_A(z)}{r_s}, \tag{47}$$

$$R(z) = \frac{\sqrt{\Omega_m} H_0}{c}(1+z_d) D_A(z) \tag{48}$$

The observables that [133] reports are: $R_z = 1.7502 \pm 0.0046$, $l_A = 301.471 \pm 0.09$, $n_s = 0.9649 \pm 0.0043$ and r_s is an independent parameter, with an associated covariance matrix. (see table I in [133]). The points represent the inflationary observables as well as the CMB epoch expansion rate. In addition to the CMB points, we also take into account other data from the late Universe. The result is a successful test of the model in relation to the data.

The contour plots derived from various datasets, including the CC dataset, BAO dataset, SNIa dataset, and the combined dataset (CC + SNIa + GRB + Q + BAO + CMB), are presented in Fig. 1, Fig. 2, Fig. 3, and Fig. 4, respectively. To provide a comprehensive overview, we have overlaid all four figures from these diverse datasets in Fig. 5. In Table 1, we have tabulated the best-fit values of the model parameters η and ζ, along with the present-day Hubble function H_0, complete with their corresponding error bars and $q(z=0)$, $j(z=0)$, and χ^2_{min} with respect to different datasets.

7. Observational and theoretical comparisons of the hubble function and distance modulus function

After obtaining the best-fit values for our cosmological model parameters, it is crucial to compare our model with the widely accepted ΛCDM model. The ΛCDM model has consistently aligned with various observational datasets and stands as a robust framework for understanding the Universe's evolution. This comparative analysis deepens our comprehension of differences between the two models, shedding light on the implications of these disparities in cosmology. By examining deviations between our model and the ΛCDM model, we pinpoint

Table 1
We present constraints on cosmological parameters for both the standard ΛCDM and proposed models at a 95% confidence level (CL). These constraints are derived from the CC dataset, Type Ia supernovae (SNIa), and Baryon Acoustic Oscillations (BAO) datasets. Additionally, the 'JOINT' dataset combines the 'CC' dataset with data from Type Ia supernovae, gamma-ray bursts, quasars, BAO, and CMB distant prior measurements. The table includes values for $q(z=0)$, $j(z=0)$, and $\chi^2_{\rm min}$ with respect to different datasets.

Model	Dataset	H_0	η	ζ	$q(z=0)$	$j(z=0)$
ΛCDM Model	CC	$68.080661^{+2.121480}_{-2.121480}$	–	–	$-0.52658^{+0.070940}_{-0.095155}$	1
	SNIa	$74.174328^{+1.267662}_{-1.267662}$	–	–	$-0.56965^{+0.019986}_{-0.040265}$	1
	BAO	$69.089209^{+4.396874}_{-4.396874}$	–	–	$-0.60418^{+0.014706}_{-0.027917}$	1
	Joint	$69.854848^{+1.259100}_{-1.259100}$	–	–	$-0.59092^{+0.018813}_{-0.035450}$	1
Model	CC	$71.819162^{+3.670608}_{-5.331030}$	$1.306004^{+0.092016}_{-0.171492}$	$1.341120^{+0.142478}_{-0.216258}$	$-1.30926^{+0.144385}_{-0.218206}$	$5.1455^{+0.096127}_{-0.178301}$
	SNIa	$74.945259^{+1.405723}_{-2.176923}$	$1.210586^{+0.052141}_{-0.099643}$	$1.451716^{+0.065137}_{-0.088348}$	$-1.17116^{+0.143365}_{-0.212348}$	$6.24952^{+0.092561}_{-0.176606}$
	BAO	$68.882349^{+4.806346}_{-7.794739}$	$1.241916^{+0.041110}_{-0.077956}$	$1.459481^{+0.116204}_{-0.202248}$	$-1.15407^{+0.006608}_{-0.009882}$	$6.70014^{+0.028481}_{-0.072066}$
	Joint	$70.090386^{+1.244084}_{-2.612879}$	$1.253896^{+0.017497}_{-0.037139}$	$1.401909^{+0.021304}_{-0.039846}$	$-1.02231^{+0.006802}_{-0.010784}$	$7.44366^{+0.029072}_{-0.076997}$

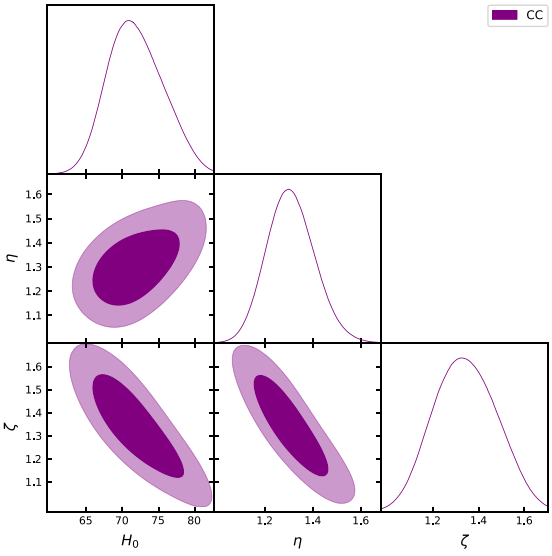


Fig. 1. The above figure shows the MCMC confidence contours at 1σ and 2σ obtained from the CC dataset.

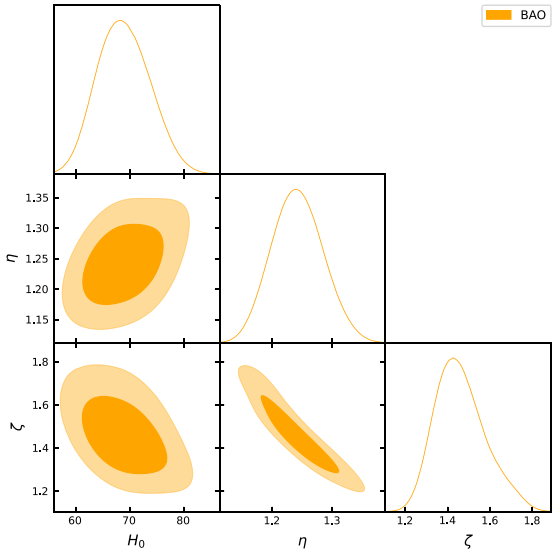


Fig. 3. The above figure shows the MCMC confidence contours at 1σ and 2σ obtained from BAO dataset.

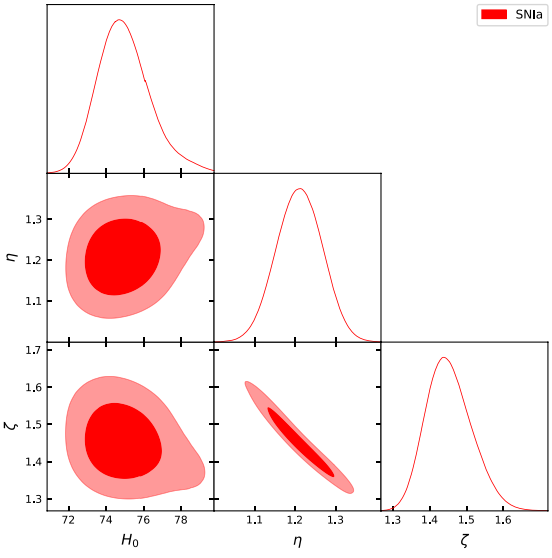


Fig. 2. The above figure shows the MCMC confidence contours at 1σ and 2σ obtained from the SNIa dataset.

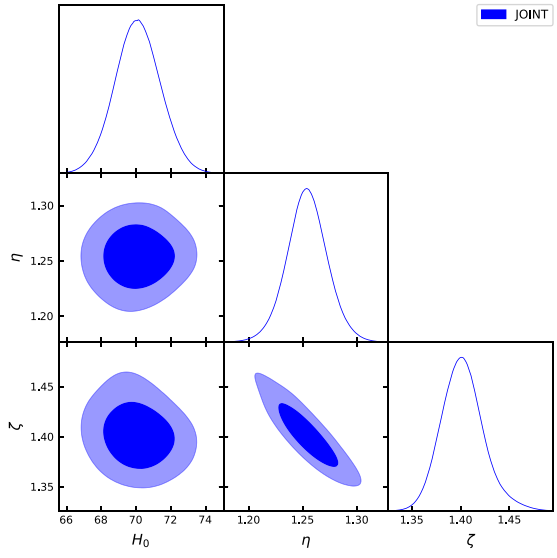


Fig. 4. The above figure shows the MCMC confidence contours at 1σ and 2σ obtained from the CC + SNIa + GRB + Q + BAO + CMB dataset.

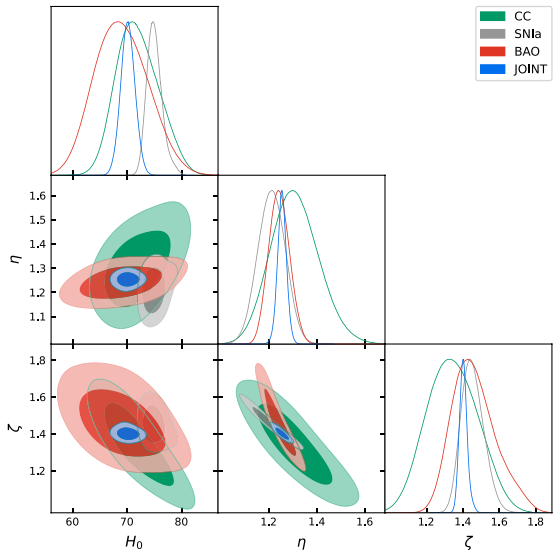


Fig. 5. The above figure shows the MCMC confidence contours at 1σ and 2σ with respect to All dataset.

the specific features that distinguish our parametrized model, such as the Universe's dynamics. This exploration provides valuable insights into our model's strengths and limitations, enriching our understanding of the cosmos. This comparison with the ΛCDM model acts as a benchmark, allowing us to assess the goodness-of-fit to observational data and gauge the alignment between our parametrized model and the established ΛCDM framework.

7.1. Comparison with the CC data points

We evaluate our model's compatibility with observational data by conducting a comparative analysis with the Cosmic Chronometers dataset, consisting of 31 data points represented by orange dots, each accompanied by error bars represented by purple dots. This comparison is illustrated in Fig. 6, where our model's predictions are depicted by the purple line. To provide a reference point, we also include the well-established ΛCDM model, represented by the black line, with cosmological parameters $\Omega_{\mathrm{m0}} = 0.3$ and $\Omega_{\Lambda} = 0.7$. The results of this analysis reveal a remarkable correspondence between our model's predictions and the observed data points, as indicated by their close alignment. This alignment underscores the ability of our model to effectively capture the inherent features and trends within the Cosmic Chronometers dataset. Consequently, our model demonstrates its capability to reproduce the expansion history of the universe, as inferred from the Hubble data.

7.2. Comparison with the type Ia supernova dataset

In this analysis, we carefully studied the $\mu(z)$ distance modulus function for our Model alongside the data from type Ia supernovae, which includes a significant 1701 data points. We also compared our Model with standard ΛCDM Model. The results of this comparison are represented in Fig. 7. The figures clearly illustrate that our model and the ΛCDM model exhibit a commendable level of agreement with the type Ia supernova dataset. This agreement indicates that these models effectively capture and replicate the observed distance measurements, signifying their consistency with empirical astronomical data.

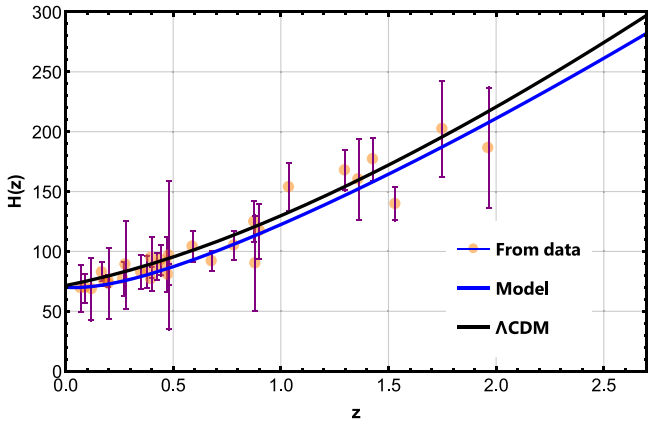


Fig. 6. Comparative analysis of our model (Blue Line) with 31 CC measurements (Orange dots) and ΛCDM model (black line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

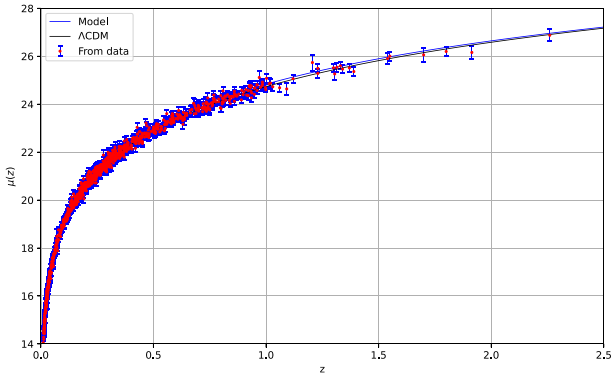


Fig. 7. Comparative analysis of our model (Blue Line) with 1701 type Ia supernova measurements (Orange dots) and ΛCDM model (black line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

7.3. Relative difference between model and ΛCDM

The Fig. 8 illustrates the relative difference between our Model and the standard ΛCDM model. It is evident from the Figure that our Model exhibits distinct behavior from the typical ΛCDM model, particularly noticeable start form low redshift values ($z > 0$). As the redshift increases, these disparities between our model and the ΛCDM Model become more pronounced.

8. Cosmography parameters

8.1. The deceleration parameter

The deceleration parameter [134], denoted as "q," is a fundamental cosmological parameter used in the study of the expansion dynamics of the universe. It plays a crucial role in understanding the past and future evolution of the cosmos. Introduced by Edwin Hubble in the early 20th century, this parameter provides insights into whether the expansion of the universe is slowing down or accelerating. The deceleration parameter is defined in terms of the second derivative of the scale factor of the universe, which describes how the universe's size changes with time. Mathematically, it can be expressed as:

$$q = -\frac{a\ddot{a}}{\dot{a}^2}, \tag{49}$$

where q is the deceleration parameter. $a(t)$ represents the scale factor of the universe as a function of time. $\dot{a}$ represents the first derivative of

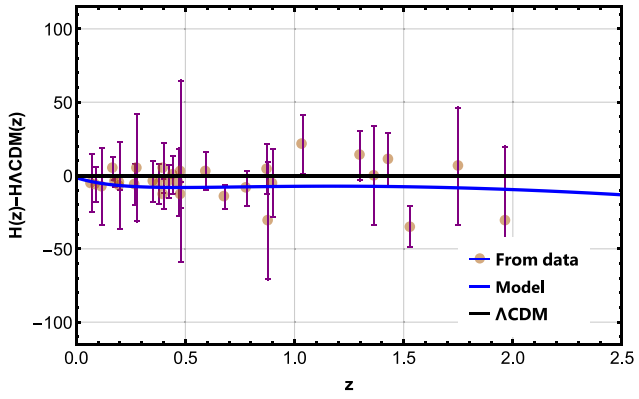


Fig. 8. Comparative analysis of our model (Blue Line) with 31 CC measurements (Orange dots) and ΛCDM model (black line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

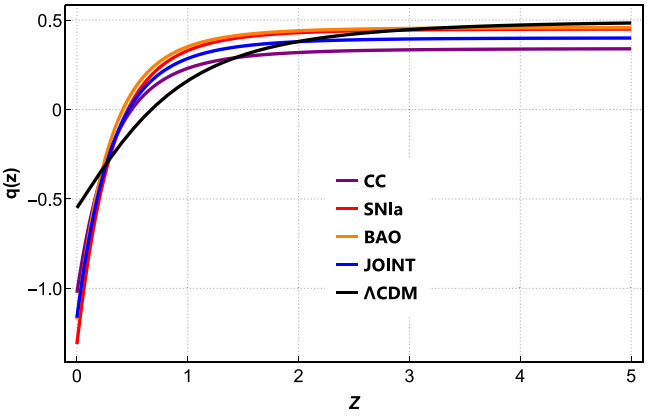


Fig. 9. Evolution of the deceleration parameter with respect to redshift across various datasets, depicted alongside the ΛCDM model represented by the black line, characterized by $\Omega_{m0} = 0.3$ and $\Omega_{\Lambda} = 0.7$.

the scale factor with respect to time. ä represents the second derivative of the scale factor with respect to time. The value of the deceleration parameter is indicative of the nature of the cosmic expansion:

1. $q > 0$ (Decelerating Universe): If the deceleration parameter is positive, it implies that the expansion of the universe is slowing down over time. In the past, this was the prevailing belief when the gravitational attraction of matter was thought to dominate the cosmic dynamics.
2. $q = 0$ (Critical Universe): A deceleration parameter of zero suggests that the expansion is proceeding at a constant rate. In this scenario, the universe's expansion is neither accelerating nor decelerating, often referred to as a "critical universe".
3. $q < 0$ (Accelerating Universe): A negative value for the deceleration parameter signifies that the universe's expansion is accelerating. This phenomenon gained significant attention in the late 20th century when the discovery of dark energy provided a compelling explanation for this observed acceleration.

In recent years, the study of the deceleration parameter has become increasingly important in cosmology, particularly in the context of understanding dark energy and the fate of the universe. It is a key parameter used in observational cosmology to probe the nature of the components of the universe, such as dark matter and dark energy, and to determine the overall geometry of the universe.

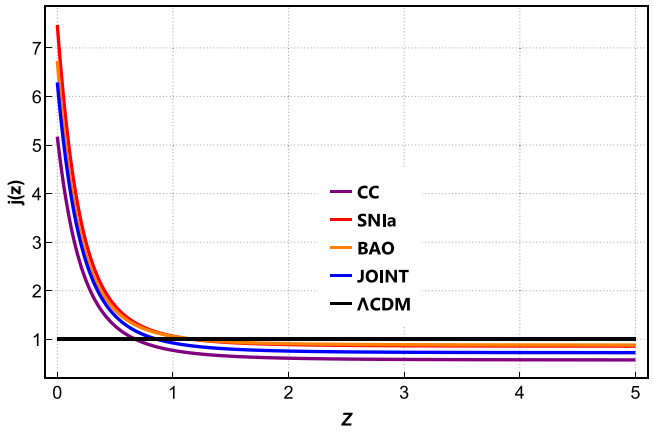


Fig. 10. Evolution of the jerk parameter with respect to redshift across various datasets, depicted alongside the ΛCDM model represented by the black line, characterized by $\Omega_{m0} = 0.3$ and $\Omega_{\Lambda} = 0.7$.

8.2. The jerk parameter

In the realm of cosmology, understanding the dynamics of the universe's expansion is of paramount importance. While the Hubble constant and the deceleration parameter have been essential tools in characterizing this expansion, a more nuanced parameter known as the "jerk parameter" has emerged as a valuable addition to our cosmological toolkit. The jerk parameter, denoted as "j," provides a deeper level of insight into the cosmic acceleration, complementing the information offered by the deceleration parameter [135]. The jerk parameter represents the third time derivative of the scale factor of the universe, building upon the concepts embodied by the Hubble parameter and the deceleration parameter. Mathematically, it can be expressed as:

$$j = \frac{1}{a}\frac{d^3a}{d\tau^3}\left[\frac{1}{a}\frac{da}{d\tau}\right]^{-3} = q(2q+1) + (1+z)\frac{dq}{dz}, \tag{50}$$

where: j is the jerk parameter, $a(t)$ represents the scale factor of the universe as a function of time, $\dot{a}$ represents the first derivative of the scale factor, ä represents the second derivative of the scale factor, $\dddot{a}$ represents the third derivative of the scale factor, z represents redshift. In the context of the Taylor expansion of the scale factor around a reference time t_0, the expansion takes the form:

$$\frac{a(t)}{a_0} = 1 + H_0(t-t_0) - \frac{1}{2}q_0H_0^2(t-t_0)^2 + \frac{1}{6}j_0H_0^3(t-t_0)^3 + O\left[(t-t_0)^4\right]. \tag{51}$$

where the subscript 0 denotes current values. Here, H_0 represents the Hubble constant at the reference time t_0, q_0 is the deceleration parameter, and j_0 is the jerk parameter at the same reference time. In contemporary cosmology, the jerk parameter has played a crucial role in refining our knowledge of the universe's evolution.

9. Statefinder diagnostic

In the field of cosmology, the study of the universe's evolution, demands a comprehensive understanding of dark energy (DE) and its influence on cosmic expansion. To analyze this cosmic dynamic without bias toward any specific DE model, cosmologists employ a valuable tool known as the statefinder diagnostic parameter. This mathematical tool, pioneered by researchers [136–139], employs higher derivatives of the cosmic scale factor to characterize the universe's expansion. Its primary purpose is to distinguish and compare different DE models effectively. What sets the statefinder diagnostic apart is its model-independent nature, enabling the exploration of various cosmological scenarios,

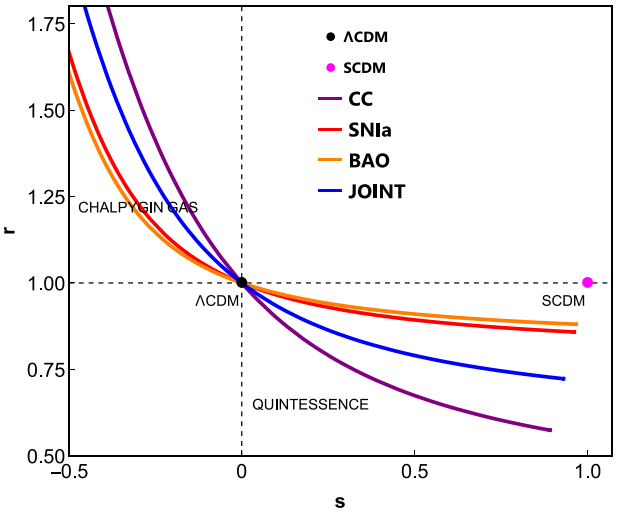


Fig. 11. Evolution of the $\{r, s\}$ profile with respect to redshift across various datasets.

including those with diverse forms of dark energy. The statefinder diagnostic is encapsulated in a parameter pair, denoted as $\{r, s\}$. These parameters are defined as follows:

$$r = \frac{\dddot{a}}{aH^3}, \quad s = \frac{r-1}{3\left(q-\frac{1}{2}\right)}. \tag{52}$$

These parameters leverage higher-order derivatives of the scale factor, the Hubble parameter H, and the deceleration parameter q to provide insights into cosmic expansion. Various possibilities in the $\{r, s\}$ and $\{q, r\}$ planes are exhibited to depict the temporal evolution of various DE models. With the assistance of the statefinder diagnostics pair. In these cases, some specific pairs typically correlate to classic DE models such as $\{r, s\} = \{1, 0\}$ represents ΛCDM model and $\{r, s\} = \{1, 1\}$ indicates standard cold dark matter Model (SCDM) in FLRW background. Also, $(-\infty, \infty)$ yields static Einstein Universe. In the $r-s$ plane, $s > 0$ and $s < 0$ define a quintessence-like model and phantom-like model of the DE, respectively. Moreover, the evolution from phantom to quintessence can be observed by deviation from $r, s = 1, 0$. On the other hand, $\{q, r\} = \{-1, 1\}$ corresponds to the ΛCDM model while $\{q, r\} = \{0.5, 1\}$ shows SCDM model. It is important to note that on a $r-s$ plane if the DE model's trajectories deviate from these standard values, the resulting model differs from the normal cosmic model.

10. Om diagnostic

In the realm of cosmology, the $Om(z)$ diagnostic [140–143] parameter plays a pivotal role in unraveling the cosmic mysteries. It is an essential tool for gauging the relative contribution of matter to the universe's total energy density, shedding light on the universe's overall geometry. ω_m, the heart of this diagnostic, tells us about the ratio of current matter density to the critical density required for a flat universe. If ω_m is less than 1, it hints at an open universe, while a value greater than 1 suggests a closed one. This parameter serves as a geometric yardstick and plays a crucial role in testing the ΛCDM model, which posits three primary components: dark matter, ordinary matter (baryonic), and enigmatic dark energy. One of the remarkable aspects of ω_m is its ability to discern different dark energy models from the standard ΛCDM model. By altering the slope of $Om(z)$, this diagnostic efficiently distinguishes between various models. A positive slope signifies a quintessence model, while a negative slope points to a phantom model. A constant slope, on the other hand, aligns with the cosmological constant. This feature allows us to explore the intricate balance among these cosmic components. Much like the statefinder

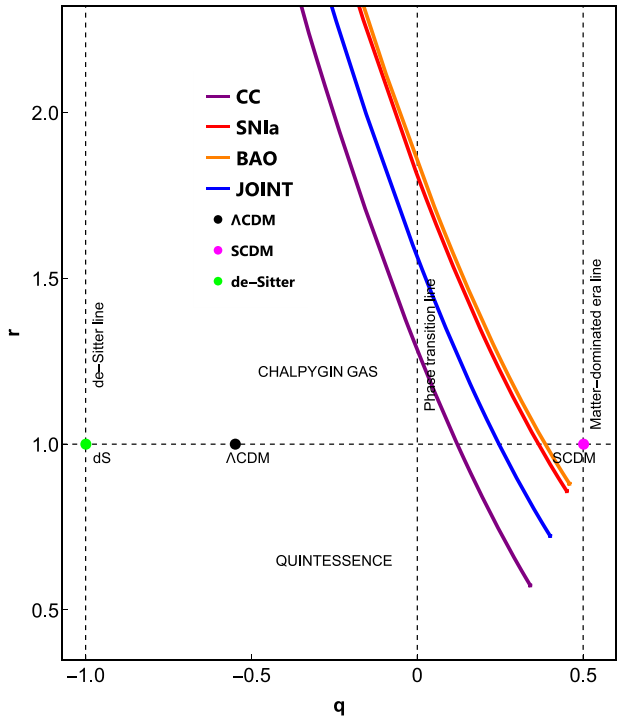


Fig. 12. Evolution of the $\{r, q\}$ profile with respect to redshift across various datasets.

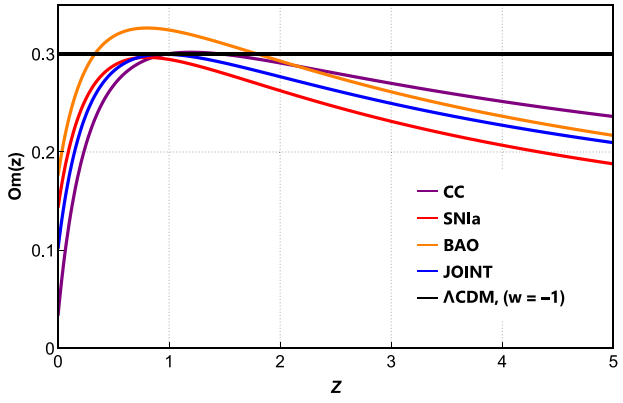


Fig. 13. Evolution of the $Om(z)$ profile with respect to redshift across various datasets.

diagnostic, $Om(z)$ serves as a powerful tool for understanding cosmic evolution, geometry, and the interplay of matter in shaping the universe's dynamics. In a flat universe, $Om(z)$ is defined as:

$$Om(z) = \frac{\left(\frac{H(z)}{H_0}\right)^2 - 1}{(1+z)^3 - 1}. \tag{53}$$

11. Physical parameters

In cosmology, physical parameters such as pressure, equations of state, and the density parameter (denoted as ρ) play crucial roles in describing the behavior and evolution of the universe. Here's a general overview of these concepts along with their associated formulas:

11.1. Pressure density p

Pressure density influences the overall behavior of the universe's expansion. The pressure associated with cosmic components, such as

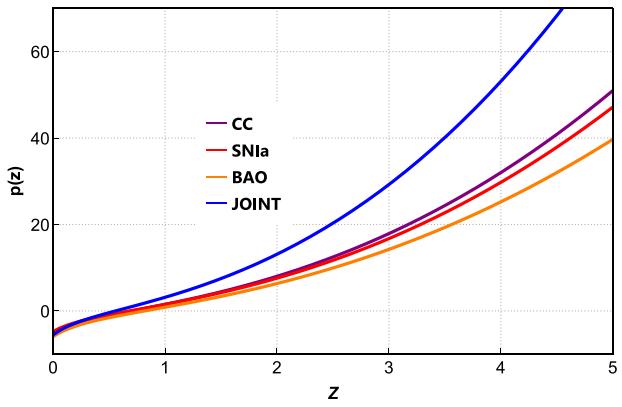


Fig. 14. Evolution of the Pressure density with respect to redshift across various datasets.

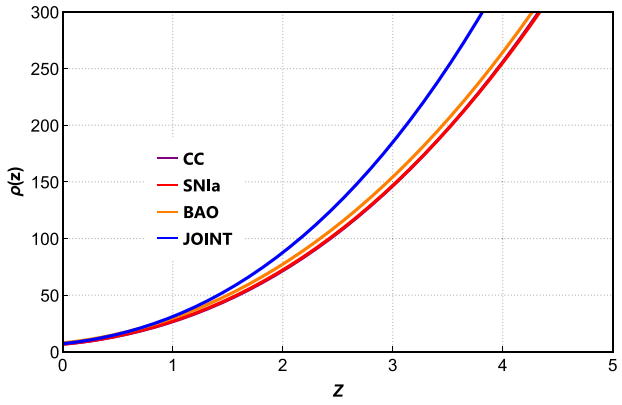


Fig. 15. Evolution of the Energy density with respect to redshift across various datasets.

matter, radiation, and dark energy, affects how they contribute to the expansion rate. Positive pressure (like that of matter) tends to slow down the expansion, while negative pressure (like that of dark energy) can accelerate it. The following Fig. 14 shows the transition from early positive values to the late negative values showing the accelerating expansion of the late universe. It is well described the early positive pressure contributes to the structure formation in the universe.

11.2. Energy density ρ

Energy density refers to the amount of energy contained within a given volume of space. In cosmology, it quantifies the total energy content of a specific cosmic component. Energy density is a crucial parameter as it determines the gravitational effects of that component on the overall cosmic dynamics. From the following Fig. 15, we can interpret the earlier high density universe to the present low values as expected.

11.3. Equation of state ω

The equation of state in cosmology establishes a relationship between the pressure (P) and energy density (ρ) of a cosmic component, effectively describing how pressure and density interrelate. In its general form, this equation is expressed as $P = \omega\rho$, where "(ω)" serves as the parameter characterizing the equation of state. Different values of "(ω)" correspond to distinct cosmic components. For instance, when $w = 0$, it corresponds to non-relativistic matter (e.g., cold dark matter). A value of $\omega = 1/3$ corresponds to radiation (e.g., photons), while $\omega \approx -1$ aligns with dark energy, exhibiting properties akin to a cosmological

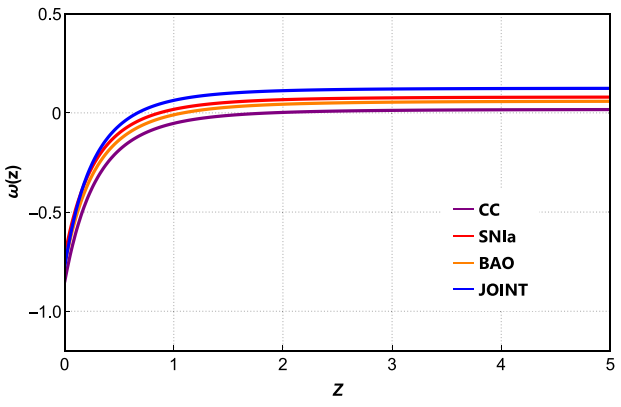


Fig. 16. Evolution of the EoS Parameter with respect to redshift across various datasets.

constant (Λ). These varying values of "(ω)" play a critical role in characterizing the behavior and properties of different cosmic components within the framework of cosmological models. The following Fig. 16 shows the evolution of the equation of state parameter in the due course of time. The universe enters into the quintessence phase in the recent past and remain in this phase in the near future too. The early positive values of ω shows the matter dominated era in the universe.

12. Statistical analysis

To determine the optimal model for our analysis, it is crucial to consider the number of free parameters associated with each model, in addition to the $\chi^2_{\rm min}$ value obtained. While choosing among the various information criteria available in the literature is not a straightforward task [144], we opt for the most commonly used ones. One of these is the Akaike Information Criterion (AIC) [144–146], defined as:

$$AIC \equiv -2\ln\mathcal{L}_{\rm max} + 2p_{\rm tot} = \chi^2_{\rm min} + 2p_{\rm tot} \tag{54}$$

Here, $p_{\rm tot}$ represents the total number of free parameters in the specific model, and $\mathcal{L}_{\rm max}$ signifies the maximum likelihood of the model being considered. Additionally, we incorporate the Bayesian Information Criterion (BIC), introduced by [144–146], which is defined as:

$$BIC \equiv -2\ln\mathcal{L}_{\rm max} + p_{\rm tot}\ln\left(N_{\rm tot}\right) \tag{55}$$

By utilizing the definitions (54) and (55), we compute the discrepancies $\triangle AIC$ and $\triangle BIC$ relative to the ΛCDM model in question. In accordance with [147], if $0 < |\triangle AIC| \leq 2$, it implies that the compared models may be viewed as compatible with each other. Conversely, if $|\triangle AIC| \geq 4$, it indicates that the model with the higher AIC value is not supported by the data. Similarly, for $0 < |\triangle BIC| \leq 2$, the model exhibiting the higher BIC value is marginally less favored by the data. In cases where $2 < |\Delta BIC| \leq 6$ ($|\triangle BIC| > 6$), the model with the higher BIC values is significantly (highly) less favored. The specific distinctions among the investigated cosmological models are detailed in 2.

In cosmology, the terms "P-value" and "L-statistic" are commonly used in statistical analyses to assess the significance of observations and test hypotheses.

- P-value (Probability Value) The P-value is a statistical measure that quantifies the evidence against a null hypothesis. It tells you the probability of observing data as extreme or more extreme than what you have, assuming that the null hypothesis is true. In cosmology, P-values are often used in the context of hypothesis testing [148–150]. For example, when analyzing cosmic microwave background (CMB) data, cosmologists may calculate P-values to assess whether the observed CMB power spectrum matches the predictions of a particular cosmological model. A low P-value suggests that the data is inconsistent with the model,

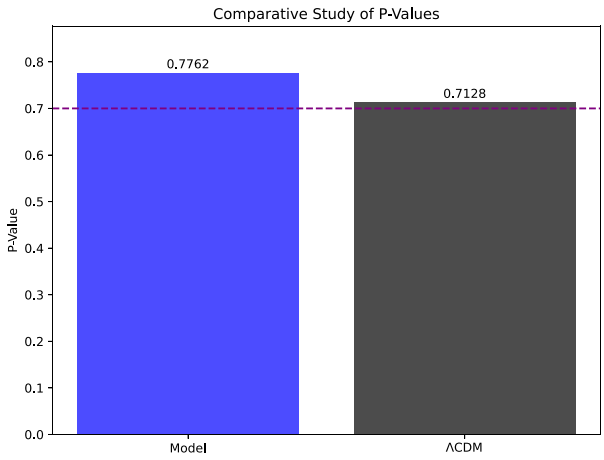


Fig. 17. The Comparative Study of P-Values of joint dataset.

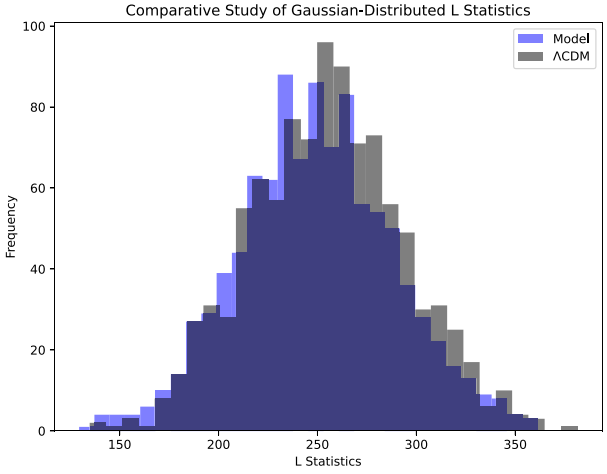


Fig. 18. The comparative histogram plot of the distribution of L statistics for both models of joint dataset.

while a high P-value suggests that the data is consistent with the model.

- L-statistic (Likelihood Statistic) The L-statistic, often referred to as the likelihood ratio [151–153], is a statistical measure used to compare the likelihood of observing data under two different hypotheses. In cosmology, the L-statistic is frequently used when comparing different cosmological models or parameter values. For instance, in Bayesian cosmological analyses, the likelihood function quantifies how well a given model explains the observed data. The L-statistic is used to compare the likelihoods of different models, and models with higher likelihoods are considered more compatible with the data.

In practical terms, when cosmologists perform statistical analyses in cosmology, they often calculate likelihoods and P-values to evaluate the goodness of fit between observational data and theoretical models. These statistical tools help cosmologists make inferences about the parameters of the universe, such as the density of dark matter, dark energy properties, and the geometry of the universe (see Fig. 17).

13. Results

13.1. Deceleration parameter

Fig. 9 represents the redshift evolution of the deceleration parameter (q) across various cosmological epochs, denoted by the redshift (z),

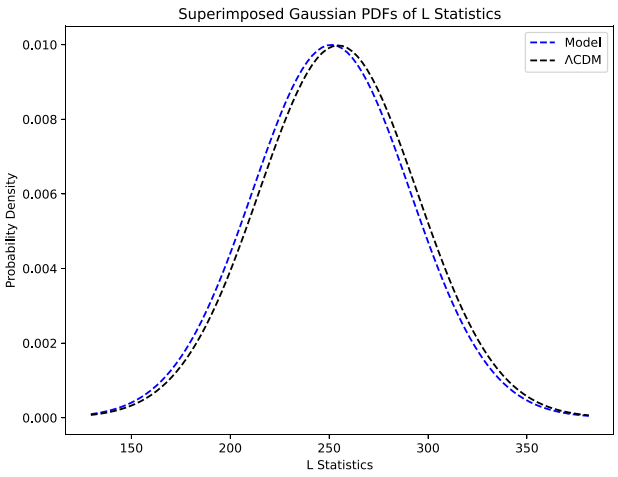


Fig. 19. The superimposed overlays of Gaussian Probability Density Functions (PDFs) on the L statistics for both models of joint dataset.

as constrained by different datasets, including Cosmic Chronometers (CC), Type Ia Supernovae (SNIa), Baryon Acoustic Oscillations (BAO), and a Joint Dataset encompassing CC + SNIa + Gamma-Ray Bursts (GRB) + Quasars (Q) + BAO and ΛCD paradigm. At high redshifts, corresponding to earlier epochs in the universe's history, the model exhibits a noteworthy convergence toward a consistent value of q across all analyzed datasets, specifically around 0.2763. As the universe evolves and redshift decreases, it undergoes a phase transition from a decelerating phase to an accelerating one. Nonetheless, the long-term behavior of the two models starkly diverges. An intriguing and noteworthy observation is that the model predicts an acceleration at half-time higher than the ΛCDM model with an approximate value of $q(0) \approx -1$, whereas the ΛCDM model predicts a value of approximately $q(0) \approx -0.5$. This deceleration parameter maintains consistency across all the datasets employed in our comprehensive analysis.

13.2. Jerk parameter

Fig. 10 illustrates the redshift-dependent behavior of the jerk parameter across the various datasets. At high redshifts, the value of j in the proposed model is across all datasets and the ΛCDM model consistently converges to approximately 1. However, important differences do exist at lower redshifts, with a very significant difference appearing at $z = 0$, where the proposed model predicts a value higher than the ΛCDM value of $j = 1$ across all datasets. Determining the current value of j through observations can be a crucial test for the proposed model cosmological model.

13.3. $\{r, s\}$ profile

The $\{r, s\}$ profile of our obtained cosmological model, as depicted in Fig. 11, reveals a fascinating cosmological evolution. At the initial stages, the model exhibits values in the range $r < 1$ and $s > 0$, which correspond to characteristics akin to the quintessence region In cosmology, quintessence is a dynamic form of dark energy with a varying energy density and negative pressure, potentially explaining the universe's accelerated expansion. Further by crossing intermediate ΛCDM fixed point $\{0, 1\}$, it transitions into a different region where $r > 1$ and $s < 0$, represent Chaplygin gas. Chaplygin gas is a cosmological model featuring exotic matter with unique properties, used to explain cosmic expansion. Remarkably, this $\{r, s\}$ profile consistently exhibits the same behavior with respect to all the datasets used in our analysis.

Table 2
Summary of $\chi^{2\ min}_{tot}$, χ^2_{red}, AIC_c, ΔAIC_c, **BIC**, Δ**BIC**, P-value and L-statistic for ΛCDM and $f(Q,T)$ model across all dataset.

Model	Dataset	$\chi^{2\ min}_{tot}$	χ^2_{red}	$\mathcal{K}_f$	AIC_c	ΔAIC_c	BIC	ΔBIC	P-value	L-statistic
ΛCDM Model	CC	15.5126	0.9053	3	21.5126	0	25.8146	0	0.560256	26.140535
	SNIa	1036.50	0.9356	3	1042.51	0	1057.43	0	0.989391	203.623840
	BAO	26.4431	0.9182	3	32.4431	0	34.9427	0	0.550075	10.811795
	Joint	1107.35	0.9645	3	1113.35	0	1128.83	0	0.712755	254.538726
Model	CC	15.6871	0.8953	3	21.6871	0.17	25.9891	0.17	0.554023	26.140535
	SNIa	1033.91	0.9298	3	1039.91	−2.59	1054.84	−2.29	0.991874	203.635102
	BAO	27.3445	0.9177	3	33.3445	0.90	35.8441	0.90	0.664331	9.446289
	Joint	1108.32	0.9634	3	1114.32	0.97	1129.80	0.97	0.776201	251.087532

13.4. $\{r, q\}$ profile

The $\{r, q\}$ profile of our derived cosmological model, as shown in Fig. 12, reveals a fascinating evolution. Initially, the model exhibits values within the region $q > 0$ and $r < 1$, which resembles characteristics commonly linked with the quintessence. As the Universe evolves, a distinctive transition occurs, leading the model into a different region characterized by $q < 0$ and $r > 1$, represented by the Chaplygin gas scenario. Notably, this $\{r, q\}$ profile consistently demonstrates the same behavior across all the datasets integrated into our analysis.

13.5. Om diagnostic

Fig. 13 illustrates the evolution of $Om(z)$ at various redshifts (z) for our model across different datasets. Notably, for the CC, SNIa, and Joint datasets, $Om(z)$ consistently remains below the current matter density parameter Ω_{m0}. This observation suggests that the model consistently resides in the phantom region throughout the entire evolution of the Universe at all redshifts. On the other hand, for the BAO dataset, $Om(z)$ consistently remains below the current matter density parameter for redshifts $z > 1.8$. This suggests that the model falls within the phantom domain. As the Universe evolves, $Om(z)$ transitions to being higher than the current matter density parameter in the redshift range of $0.34 > z > 1.8$, indicating that the model enters the quintessence domain. Finally, $Om(z)$ decreases below the current matter density parameter for redshifts $z < 0.34$, signifying that the model reverts to the phantom domain.

13.6. Physical parameters

Figs. 14, 15, and 16 provide the necessary illustration into the dynamic evolution of physical parameters, pressure (p), energy density (ρ), and the equation of state parameter (ω) at different redshifts. These figures meticulously portray how these parameters change as a function of redshift, with each curve representing a distinct set of model parameters derived from diverse datasets. The beauty of these plots lies in their ability to unveil the intricate phases of evolution exhibited by these physical parameters at different points in cosmic history. In particular, Fig. 14 unveils the transition in the pressure parameter: it shifts from being positive in the early universe to becoming increasingly negative as time progresses, hinting at the onset of late-time cosmic acceleration. However, what truly sets this observation apart from conventional cosmological models like the ΛCDM model is the intriguing revelation that the pressure does not just remain negative. Meanwhile, Fig. 15 charts the path of energy density (ρ), illustrating how it initially diminishes from its higher values in the past to its present, smaller magnitude. Fig. 16 provides a glimpse into the cosmic temperament of the Universe through the equation of state parameter. For the region $z > 1$, the model's ω falls within $0 < w < \frac{1}{3}$, aligning with the behavior expected in the Matter-dominated era, where pressure is negligible compared to energy density. As the Universe evolves and enters the epoch $z < 1$, the ω value shifts into the range $-1 < w < -\frac{1}{3}$. This shift signifies a transition to a quintessence field domain, suggesting the presence of a dynamic form of dark energy with negative pressure, contributing to the observed cosmic acceleration. We can witness the matter dominated era in the early universe to the late dark energy-dominated era in the quintessence realm.

13.7. Statistical analysis

Based on the Table 2 furnishes a detailed comparison between the $f(Q,T)$ model and the standard ΛCDM model based on various statistical indicators. Starting with $\chi^{2\ min}_{tot}$, which represents the minimum total chi-square, the $f(Q,T)$ model yields slightly higher values across all datasets compared to the ΛCDM model. This suggests that, when attempting to minimize the discrepancies between the model and observational data, ΛCDM achieves a marginally better fit. The reduced chi-square (χ^2_{red}) accounts for the number of degrees of freedom, providing a more normalized measure. Here, too, the $f(Q,T)$ model exhibits slightly higher values, indicating that ΛCDM offers a slightly improved fit when considering the complexity of the models. Moving on to information criteria, ΔAIC_c and ΔBIC are crucial for model comparison. In both metrics, positive values for the $f(Q,T)$ model signify a weaker degree of support compared to ΛCDM. Typically, ΔAIC values in the range of $(0, 2)$ suggest strong support, while values in $(4, 7)$ are less supported. Values beyond 10 indicate models without any meaningful support. Therefore, the positive ΔAIC_c values for the $f(Q,T)$ model suggest that, in terms of information criteria, ΛCDM is more strongly supported. Considering P-values and L-statistic, which provide insights into the overall goodness of fit, the $f(Q,T)$ model generally yields higher P-values and lower L-statistic values compared to ΛCDM. This implies that the $f(Q,T)$ model has a comparatively poorer fit to the observational data. Fig. 18 provides the comparative histogram distribution of L statistics for ΛCDM model and our Model. The x-axis represents the L statistics values, while the y-axis represents the frequency or count of data points within each bin. You can observe how the L statistics are distributed for both models. Look for differences in the central tendencies (peaks) and spreads (widths) of the distributions. Since our model's histogram is similar to the ΛCDM histogram, it suggests that our model is a reasonable fit for the data, Fig. 19 provides the superimposed plot overlays Gaussian Probability Density Functions (PDFs) on the L statistics for both models. The x-axis represents the L statistics values, and the y-axis represents the probability density. Again the Gaussian PDF of our model is closely aligned with the histogram of L statistics, it suggests a better match between the ΛCDM model and the data.

14. Conclusion

In this study, we conducted a comprehensive and rigorous examination of the $f(Q,T)$ cosmological model. We extend the analysis conducted in classical gravity by incorporating a parametrization of the deceleration parameter, as outlined in [42]. Our investigation encompasses the realm of $f(Q,T)$ gravity, allowing us to explore its implications and effects on the parametrization. This model was compared against a range of cosmological observations, encompassing 31 Cosmic Chronometers, 1071 type Ia supernovae measurements, 162 Gamma Ray Bursts (GRBs), 24 measurements from compact radio quasars, 17 Baryon Acoustic Oscillation (BAO) measurements, and CMB distant prior. To determine the best-fit value for the model parameters, we employed the Markov Chain Monte Carlo (MCMC) methodology, allowing us to derive the optimal fit for these parameters. Subsequently, utilizing these best-fit values, the data-fitting process yielded

exceptionally accurate outcomes for both the CC and SNIa datasets. We observed the redshift evolution of the deceleration parameter (q) across different cosmological epochs using various datasets, the model exhibits a consistent convergence toward a value of approximately 0.2763 at high redshifts, indicative of the early universe's behavior. As the universe evolves and redshift decreases, a transition from a decelerating to an accelerating phase occurs. Notably, the proposed model predicts an acceleration at a rate higher than the ΛCDM model, with an approximate value of $q(0) \approx -1$, while the ΛCDM model predicts a value of approximately $q(0) \approx -0.5$. Recent studies have provided varied estimates for the present-day deceleration parameter ($q(0)$), deviating from the standard ΛCDM model. In the cited research [154], the authors found that the value of $q(0)$ is approximately 0.05641. Similarly, in the study [77], a different estimate of $q(0) \approx 0.84331$ was derived. In the referenced research [155], $q(0) \approx -0.73$. In the study [156], the authors obtained $q(0) \approx -0.34$. In the investigation by Xu et al. [157], three models were introduced, each suggesting distinct values. In Model 1, $q(0) \approx -0.657$, in Model 2, $q(0) \approx -0.982$, and in Model 3, $q(0) \approx -0.410$. These non-standard values indicate a departure from the expected ΛCDM prediction and emphasize the growing recognition of variations in the cosmological dynamics, prompting further investigation into alternative models and theoretical frameworks. This divergence is maintained consistently across all datasets utilized in our comprehensive analysis. Moving to the jerk parameter, at high redshifts, both the proposed model and the ΛCDM model converge to a value of approximately 1 across all datasets. However, significant differences emerge at lower redshifts, particularly at $z = 0$, where the proposed model consistently predicts a higher value for j compared to the ΛCDM value of $j = 1$ across all datasets. The determination of the current value of j through observations emerges as a crucial test for validating the proposed cosmological model. This presents a crucial test for the proposed model. The analysis of the $\{r, s\}$ profile revealed a fascinating cosmological evolution. Our model exhibited characteristics resembling quintessence and Chaplygin gas phases, maintaining consistency across all datasets. The $\{r, q\}$ profile depicted a dynamic evolution, transitioning from quintessence-like behavior to a Chaplygin gas scenario. This behavior remained consistent across all datasets. The $Om(z)$ analysis indicated that our model consistently positions $Om(z)$ below Ω_{m0} for CC, SNIa, and Joint datasets, indicating a sustained presence in the phantom region. The observed transition in the BAO dataset from the phantom to quintessence domains introduces complexity to the model's dynamic evolution. The analysis of pressure (p), energy density (ρ), and the equation of state parameter (ω) across varying redshifts reveals deviations from standard cosmological models like the ΛCDM model. The EoS reveals a nuanced cosmic temperament, shifting from matter domination to quintessence, offering a unique perspective distinct from conventional models. The comparative analysis between our proposed model and the ΛCDM Model across various statistical metrics, including chi-square values, information criteria, and goodness-of-fit indicators, the results consistently favor the ΛCDM model over the $f(Q,T)$ model. The statistical analyses collectively suggest that ΛCDM provides a slightly better balance between goodness of fit and model. In conclusion, while our proposed model provides intriguing insights and predicts unique cosmic behaviors, such as super acceleration, the standard ΛCDM Model remains favored by statistical criteria. The differences observed in various cosmological parameters and profiles underscore the complexity of cosmological models and the need for ongoing research to refine our understanding of the universe's evolution. Further observations and refined datasets may shed more light on these intriguing cosmological phenomena.

CRediT authorship contribution statement

Madhur Khurana: Conceptualization, Data curation, Formal analysis, Investigation, Project administration, Resources. **Himanshu Chaudhary:** Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Writing – original draft, Writing – review & editing. **Saadia Mumtaz:** Data curation, Formal analysis, Funding acquisition, Validation, Visualization, Writing – original draft, Writing – review & editing. **S.K.J. Pacif:** Conceptualization, Data curation, Funding acquisition, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. **G. Mustafa:** Conceptualization, Data curation, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgment

Author SKJP thanks IUCAA, Pune for hospitality and other facilities under its IUCAA associateship program, where a large part of work has been done. G. Mustafa is very thankful to Prof. Gao Xianlong from the Department of Physics, Zhejiang Normal University, for his kind support and help during this research. Further, G. Mustafa acknowledges Grant No. ZC304022919 to support his Postdoctoral Fellowship at Zhejiang Normal University, China.

References

[1] Saul Perlmutter, Goldhaber Aldering, Gerson Goldhaber, R.A. Knop, Peter Nugent, Patricia G. Castro, Susana Deustua, Sebastien Fabbro, Ariel Goobar, Donald E. Groom, et al., Measurements of Ω and Λ from 42 high-redshift supernovae, Astrophys. J. 517 (2) (1999) 565.

[2] Adam G. Riess, Alexei V. Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M. Garnavich, Ron L. Gilliland, Craig J. Hogan, Saurabh Jha, Robert P Kirshner, et al., Observational evidence from supernovae for an accelerating universe and a cosmological constant, Astron. J. 116 (3) (1998) 1009.

[3] Miguel Quartin, Mauricio O. Calvao, Sergio E. Joras, Ribamar RR Reis, Ioav Waga, Dark interactions and cosmological fine-tuning, J. Cosmol. Astropart. Phys. 2008 (05) (2008) 007.

[4] Sean M. Carroll, Heywood Tam, Unitary evolution and cosmological fine-tuning, 2010, arXiv preprint arXiv:1007.1417.

[5] Hermano E.S. Velten, R.F. Vom Marttens, Winifried Zimdahl, Aspects of the cosmological "coincidence problem", Eur. Phys. J. C 74 (2014) 1–8.

[6] Navin Sivanandam, Is the cosmological coincidence a problem? Phys. Rev. D 87 (8) (2013) 083514.

[7] Varun Sahni, The cosmological constant problem and quintessence, Classical Quantum Gravity 19 (13) (2002) 3435.

[8] Ivaylo Zlatev, Limin Wang, Paul J. Steinhardt, Quintessence, cosmic coincidence, and the cosmological constant, Phys. Rev. Lett. 82 (5) (1999) 896.

[9] Philippe Brax, Jerome Martin, Robustness of quintessence, Phys. Rev. D 61 (10) (2000) 103502.

[10] T. Barreiro, Edmund J. Copeland, N.J. a Nunes, Quintessence arising from exponential potentials, Phys. Rev. D 61 (12) (2000) 127301.

[11] Robert R. Caldwell, Marc Kamionkowski, Nevin N. Weinberg, Phantom energy: dark energy with w<- 1 causes a cosmic doomsday, Phys. Rev. Lett. 91 (7) (2003) 071301.

[12] Jaume Garriga, Viatcheslav F. Mukhanov, Perturbations in k-inflation, Phys. Lett. B 458 (2–3) (1999) 219–225.

[13] Takeshi Chiba, Takahiro Okabe, Masahide Yamaguchi, Kinetically driven quintessence, Phys. Rev. D 62 (2) (2000) 023511.

[14] Christian Armendariz-Picon, V. Mukhanov, Paul J. Steinhardt, Dynamical solution to the problem of a small cosmological constant and late-time cosmic acceleration, Phys. Rev. Lett. 85 (21) (2000) 4438.

[15] Vittorio Gorini, Alexander Kamenshchik, Ugo Moschella, Can the Chaplygin gas be a plausible model for dark energy? Phys. Rev. D 67 (6) (2003) 063509.

[16] Aritra Banerjee, Haiying Cai, Lavinia Heisenberg, E Ó Colgáin, Mohammad Mahdi Sheikh-Jabbari, Tao Yang, Hubble sinks in the low-redshift swampland, Phys. Rev. D 103 (8) (2021) L081305.

[17] Bum-Hoon Lee, Wonwoo Lee, Eoin Ó Colgáin, M.M. Sheikh-Jabbari, Somyadip Thakur, Is local H 0 at odds with dark energy EFT? J. Cosmol. Astropart. Phys. 2022 (04) (2022) 004.

[18] Rocco D'Agostino, Rafael C. Nunes, Cosmographic view on the h 0 and σ 8 tensions, Phys. Rev. D 108 (2) (2023) 023523.

[19] Sunny Vagnozzi, Seven hints that early-time new physics alone is not sufficient to solve the Hubble tension, Universe 9 (9) (2023) 393.

[20] Sunny Vagnozzi, New physics in light of the H 0 tension: an alternative view, Phys. Rev. D 102 (2) (2020) 023518.

[21] Thomas P. Sotiriou, Valerio Faraoni, f (R) theories of gravity, Rev. Modern Phys. 82 (1) (2010) 451.

[22] Guido Cognola, Emilio Elizalde, Shin'ichi Nojiri, Sergei D. Odintsov, Sergio Zerbini, Dark energy in modified Gauss-Bonnet gravity: Late-time acceleration and the hierarchy problem, Phys. Rev. D 73 (2006) 084007.

[23] Tiberiu Harko, Francisco SN Lobo, Shin'ichi Nojiri, Sergei D Odintsov, f (R, T) gravity, Phys. Rev. D 84 (2) (2011) 024020.

[24] Yasunori Fujii, Kei-ichi Maeda, The Scalar-Tensor Theory of Gravitation, Cambridge University Press, 2003.

[25] Asifa Ashraf, G. Mustafa, Mushtaq Ahmad, Ibrar Hussain, Lorentz distributed wormhole solutions in f (T) gravity with off-diagonal tetrad under conformal motions, Modern Phys. Lett. A 35 (29) (2020) 2050240.

[26] G. Mustafa, G. Abbas, T. Xia, Wormhole solutions in F (T, TG) gravity under Gaussian and Lorentzian non-commutative distributions with conformal motions, Chinese J. Phys. 60 (2019) 362–378.

[27] Zinnat Hassan, Ghulam Mustafa, Pradyumn Kumar Sahoo, Wormhole solutions in symmetric teleparallel gravity with noncommutative geometry, Symmetry 13 (7) (2021) 1260.

[28] Yixin Xu, Guangjie Li, Tiberiu Harko, Shi-Dong Liang, Regular article-theoretical physics, Eur. Phys. J. C 79 (2019) 708.

[29] Simran Arora, J.R.L. Santos, P.K. Sahoo, Constraining f (Q, T) gravity from energy conditions, Phys. Dark Univ. 31 (2021) 100790.

[30] Salim H. Shekh, Aylin Caliskan, Dr G. Mustafa, Ertan Gudekli, Anirudh Pradhan, Sunil Kumar Maurya, Observational Constraints on Parameterized Deceleration Parameter with F (Q, T) Gravity. Available at SSRN 4384140.

[31] Antonio Nájera, Amanda Fajardo, Cosmological perturbation theory in f (Q, T) gravity, J. Cosmol. Astropart. Phys. 2022 (03) (2022) 020.

[32] S.A. Narawade, M. Koussour, B. Mishra, Constrained f (q, t) gravity accelerating cosmological model and its dynamical system analysis, Nuclear Phys. B 992 (2023) 116233.

[33] SKJ Pacif, Dark energy models from a parametrization of H: a comprehensive analysis and observational constraints, Eur. Phys. J. Plus 135 (10) (2020) 1–34.

[34] Shibesh Kumar Jas Pacif, Ratbay Myrzakulov, Shynaray Myrzakul, Reconstruction of cosmic history from a simple parametrization of H, Int. J. Geom. Methods Mod. Phys. 14 (07) (2017) 1750111.

[35] Eric V. Linder, Exploring the expansion history of the universe, Phys. Rev. Lett. 90 (9) (2003) 091301.

[36] Eric V. Linder, Cosmological parametrization of gamma ray burst intensity distribution, Astron. Astrophys. 326 (1997) 29–33.

[37] Eric V. Linder, Dragan Huterer, How many dark energy parameters? Phys. Rev. D 72 (4) (2005) 043509.

[38] Eric V. Linder, The dynamics of quintessence, the quintessence of dynamics, Gen. Relativity Gravitation 40 (2008) 329–356.

[39] Eric V. Linder, Mapping the cosmological expansion, Rep. Progr. Phys. 71 (5) (2008) 056901.

[40] Allan Sandage, Beginnings of observational cosmology in Hubble's time: historical overview, Hubble Deep Field (1998) 1–26.

[41] Yu L. Bolotin, V.A. Cherkaskiy, O.A. Lemets, D.A. Yerokhin, LG Zazunov, Cosmology in terms of the deceleration parameter. Part I, 2015, arXiv preprint arXiv:1502.00811.

[42] D. Sofuoğlu, H. Baysal, R.K. Tiwari, Observational constraints on the cubic parametrization of the deceleration parameter in f (R, T) gravity, Eur. Phys. J. Plus 138 (6) (2023) 1–14.

[43] Salvatore Capozziello, Rocco D'Agostino, Orlando Luongo, Model-independent reconstruction of f (T) teleparallel cosmology, Gen. Relativity Gravitation 49 (2017) 1–21.

[44] Salvatore Capozziello, Rocco D'Agostino, Orlando Luongo, Kinematic model-independent reconstruction of Palatini f (R) cosmology, Gen. Relativity Gravitation 51 (2019) 1–19.

[45] Salvatore Capozziello, Rocco D'Agostino, Orlando Luongo, Thermodynamic parametrization of dark energy, Phys. Dark Univ. 36 (2022) 101045.

[46] Salvatore Capozziello, Rocco D'Agostino, Orlando Luongo, Cosmographic analysis with Chebyshev polynomials, Mon. Not. R. Astron. Soc. 476 (3) (2018) 3924–3938.

[47] S. Capozziello, R. D'Agostino, O. Luongo, High-redshift cosmography: auxiliary variables versus Padé polynomials, Mon. Not. R. Astron. Soc. 494 (2) (2020) 2576–2590.

[48] Orlando Luongo, Cosmography with the Hubble parameter, Modern Phys. Lett. A 26 (2011) 1459–1466.

[49] Orlando Luongo, Dark energy from a positive jerk parameter, Modern Phys. Lett. A 28 (2013) 1350080.

[50] Orlando Luongo, Hernando Quevedo, A unified dark energy model from a vanishing speed of sound with emergent cosmological constant, Internat. J. Modern Phys. D 23 (2014) 1450012.

[51] Orlando Luongo, Giovanni Battista Pisani, Antonio Troisi, Cosmological degeneracy versus cosmography: a cosmographic dark energy model, Internat. J. Modern Phys. D 26 (03) (2016) 1750015.

[52] Orlando Luongo, Marco Muccino, Speeding up the universe using dust with pressure, Phys. Rev. D 98 (10) (2018) 103520.

[53] Alejandro Aviles, Christine Gruber, Orlando Luongo, Hernando Quevedo, Cosmography and constraints on the equation of state of the Universe in various parametrizations, Phys. Rev. D 86 (2012) 123516.

[54] Alejandro Aviles, Alessandro Bravetti, Salvatore Capozziello, Orlando Luongo, Precision cosmology with Padé rational approximations: Theoretical predictions versus observational limits, Phys. Rev. D 90 (4) (2014) 043531.

[55] Christine Gruber, Orlando Luongo, Cosmographic analysis of the equation of state of the universe through Padé approximations, Phys. Rev. D 89 (10) (2014) 103506.

[56] Alejandro Aviles, Jaime Klapp, Orlando Luongo, Toward unbiased estimations of the statefinder parameters, Phys. Dark Univ. 17 (2017) 25–37.

[57] Alejandro Aviles, Jaime Klapp, Orlando Luongo, Toward unbiased estimations of the statefinder parameters, Phys. Dark Univ. 17 (2017) 25–37.

[58] Sergio del Campo, Ivan Duran, Ramon Herrera, Diego Pavon, Three thermodynamically-based parameterizations of the deceleration parameter, Phys. Rev. D 86 (2012) 083509.

[59] J.V. Cunha, Jose Ademir Sales de Lima, Transition redshift: new kinematic constraints from supernovae, Mon. Not. R. Astron. Soc. 390 (1) (2008) 210–217.

[60] J.V. Cunha, Kinematic constraints to the transition redshift from SNe Ia union data, Phys. Rev. D 79 (2009) 047301.

[61] Adam G. Riess, et al., Supernova Search Team Collaboration, Type Ia supernova discoveries at z > 1 from the Hubble Space Telescope: Evidence for past deceleration and constraints on dark energy evolution, Astrophys. J. 607 (2004) 665–687.

[62] Lix-In Xu, Cheng-Wu Zhang, Bao-Rong Chang, Hong-Ya Liu, Constraints to deceleration parameters by recent cosmic observations, Modern Phys. Lett. A 23 (2008) 1939–1948.

[63] Lixin Xu, Jianbo Lu, Cosmic constraints on deceleration parameter with Sne Ia and CMB, Modern Phys. Lett. A 24 (2009) 369–376.

[64] Remya Nair, Sanjay Jhingan, Deepak Jain, Cosmokinetics: a joint analysis of standard candles, rulers and cosmic clocks, J. Cosmol. Astropart. Phys. 2012 (01) (2012) 018.

[65] Ozgur Akarsu, Tekin Dereli, Suresh Kumar, Lixin Xu, Probing kinematics and fate of the Universe with linearly time-varying deceleration parameter, Eur. Phys. J. Plus 129 (2014) 22.

[66] Beethoven Santos, Joel C. Carvalho, Jailson S. Alcaniz, Current constraints on the epoch of cosmic acceleration, Astropart. Phys. 35 (1) (2011) 17–20.

[67] Yun-Gui Gong, Anzhong Wang, Reconstruction of the deceleration parameter and the equation of state of dark energy, Phys. Rev. D 75 (2007) 043520.

[68] Michael S. Turner, Adam G. Riess, Do SNe Ia provide direct evidence for past deceleration of the universe? Astrophys. J. 569 (2002) 18.

[69] Abdulla Al Mamon, Sudipta Das, A divergence free parametrization of deceleration parameter for scalar field dark energy, Internat. J. Modern Phys. D 25 (03) (2016) 1650032.

[70] Amine Bouali, Himanshu Chaudhary, Ujjal Debnath, Tanusree Roy, G Mustafa, Constraints on the parameterized deceleration parameter in FRW universe, 2023, arXiv preprint arXiv:2301.12107.

[71] Himanshu Chaudhary, Amine Bouali, Ujjal Debnath, Tanusree Roy, Ghulam Mustafa, Constraints on the parameterized deceleration parameter in FRW universe, Phys. Scr. (2023).

[72] Amine Bouali, Himanshu Chaudhary, Ujjal Debnath, Alok Sardar, G Mustafa, Data Analysis of three parameter models of deceleration parameter in FRW Universe, 2023, arXiv preprint arXiv:2304.13137.

[73] Amine Bouali, B.K. Shukla, Himanshu Chaudhary, Rishi Kumar Tiwari, Mahvish Samar, G Mustafa, Cosmological tests of parametrization q=α- β H in f (Q) FLRW cosmology, Int. J. Geom. Methods Mod. Phys. (2023) 2350152.

[74] Amine Bouali, Himanshu Chaudhary, Saadia Mumtaz, G. Mustafa, SK Maurya, Observational constraining study of new deceleration parameters in frw universe, Fortschr. Phys. (2023) 2300033.

[75] Amine Bouali, B.K. Shukla, Himanshu Chaudhary, Rishi Kumar Tiwari, Marco San Martin, Cosmographic studies of q=α- β H parametrization in f (T) framework, Int. J. Geom. Methods Mod. Phys. (2023).

[76] Himanshu Chaudhary, Aditya Kaushik, Ankita Kohli, Cosmological test of $\sigma\theta$ as function of scale factor in f (R, T) framework, New Astron. 103 (2023) 102044.

[77] Amine Bouali, Himanshu Chaudhary, Amritansh Mehrotra, SKJ Pacif, Model-independent study for a quintessence model of dark energy: Analysis and Observational constraints, 2023, arXiv preprint arXiv:2304.02652.

[78] Abdulla Al Mamon, Sudipta Das, A parametric reconstruction of the deceleration parameter, Eur. Phys. J. C 77 (7) (2017) 495.

[79] Abdulla Al Mamon, Constraints on a generalized deceleration parameter from cosmic chronometers, Modern Phys. Lett. A 33 (10n11) (2018) 1850056.

[80] Madhur Khurana, Himanshu Chaudhary, Ujjal Debnath, Niyaz Uddin Molla, G Mustafa, Cosmological test of dark energy parametrizations in Horava-Lifshitz gravity, 2023, arXiv preprint arXiv:2310.07410.

[81] Madhur Khurana, Himanshu Chaudhary, Ujjal Debnath, Alok Sardar, G Mustafa, Exploring late-time cosmic acceleration with EoS parameterizations in Horava-Lifshitz gravity via baryon acoustic oscillations, 2023, arXiv preprint arXiv:2310.11413.

[82] Bijan Berenji, Jennifer Gaskins, Manuel Meyer, Constraints on axions and axionlike particles from Fermi Large Area Telescope observations of neutron stars, Phys. Rev. D 93 (4) (2016) 045019.

[83] Tonatiuh Matos, L. Arturo Urena-Lopez, Further analysis of a cosmological model with quintessence and scalar dark matter, Phys. Rev. D 63 (6) (2001) 063506.

[84] Tonatiuh Matos, L. Arturo Urena-Lopez, Quintessence and scalar dark matter in the universe, Classical Quantum Gravity 17 (13) (2000) L75.

[85] Tonatiuh Matos, F. Siddhartha Guzmán, Dario Nunez, Spherical scalar field halo in galaxies, Phys. Rev. D 62 (6) (2000) 061301.

[86] L. Arturo Urena-Lopez, Tonatiuh Matos, New cosmological tracker solution for quintessence, Phys. Rev. D 62 (8) (2000) 081302.

[87] Yixin Xu, Guangjie Li, Tiberiu Harko, Shi-Dong Liang, f (Q, T) gravity, Eur. Phys. J. C 79 (2019) 1–19.

[88] Yixin Xu, Tiberiu Harko, Shahab Shahidi, Shi-Dong Liang, Weyl type f (Q, T) gravity, and its cosmological implications, Eur. Phys. J. C 80 (5) (2020) 1–22.

[89] Simran Arora, P.K. Sahoo, Energy conditions in f (Q, T) gravity, Phys. Scr. 95 (9) (2020) 095003.

[90] Matt Visser, General relativistic energy conditions: The Hubble expansion in the epoch of galaxy formation, Phys. Rev. D 56 (12) (1997) 7578.

[91] M. Koussour, S.K.J. Pacif, M. Bennai, P.K. Sahoo, Dynamical dark energy models from a new Hubble parameter in $f(Q)$ gravity, 2022, arXiv preprint arXiv:2208.04723.

[92] S.K.J. Pacif, Md Salahuddin Khan, L.K. Paikroy, Shalini Singh, An accelerating cosmological model from a parametrization of Hubble parameter, Modern Phys. Lett. A 35 (05) (2020) 2050011.

[93] S.K.J. Pacif, Simran Arora, P.K. Sahoo, Late-time acceleration with a scalar field source: Observational constraints and statefinder diagnostics, Phys. Dark Univ. 32 (2021) 100804.

[94] Ritika Nagpal, Shibesh Kumar Jas Pacif, Cosmological aspects of f (R, T) gravity in a simple model with a parametrization of q, Eur. Phys. J. Plus 136 (8) (2021) 875.

[95] Ritika Nagpal, S.K.J. Pacif, J.K. Singh, Kazuharu Bamba, A Beesham, Analysis with observational constraints in Λ Λ-cosmology in f (R, T) gravity, Eur. Phys. J. C 78 (2018) 1–17.

[96] Roberto Trotta, Bayesian methods in cosmology, 2017, arXiv preprint arXiv:1701.01467.

[97] Joël Akeret, Sebastian Seehars, Adam Amara, Alexandre Refregier, André Csillaghy, CosmoHammer: Cosmological parameter estimation with the MCMC Hammer, Astron. Comput. 2 (2013) 27–39.

[98] Raul Jimenez, Abraham Loeb, Constraining cosmological parameters based on relative galaxy ages, Astrophys. J. 573 (1) (2002) 37.

[99] Cong Zhang, Han Zhang, Shuo Yuan, Siqi Liu, Tong-Jie Zhang, Yan-Chun Sun, Four new observational H (z) data from luminous red galaxies in the Sloan Digital Sky Survey data release seven, Res. Astron. Astrophys. 14 (10) (2014) 1221.

[100] Raul Jimenez, Licia Verde, Tommaso Treu, Daniel Stern, Constraints on the equation of state of dark energy and the hubble constant from stellar ages and the cosmic microwave background, Astrophys. J. 593 (2) (2003) 622.

[101] Michele Moresco, Lucia Pozzetti, Andrea Cimatti, Raul Jimenez, Claudia Maraston, Licia Verde, Daniel Thomas, Annalisa Citro, Rita Tojeiro, David Wilkinson, A 6% measurement of the Hubble parameter at z 0.45: direct evidence of the epoch of cosmic re-acceleration, J. Cosmol. Astropart. Phys. 2016 (05) (2016) 014.

[102] Joan Simon, Licia Verde, Raul Jimenez, Constraints on the redshift dependence of the dark energy potential, Phys. Rev. D 71 (12) (2005) 123001.

[103] Michele Moresco, Andrea Cimatti, Raul Jimenez, Lucia Pozzetti, Gianni Zamorani, Micoll Bolzonella, James Dunlop, Fabrice Lamareille, Marco Mignoli, Henry Pearce, et al., Improved constraints on the expansion rate of the universe up to z 1.1 from the spectroscopic evolution of cosmic chronometers, J. Cosmol. Astropart. Phys. 2012 (08) (2012) 006.

[104] Daniel Stern, Raul Jimenez, Licia Verde, Marc Kamionkowski, S Adam Stanford, Cosmic chronometers: constraining the equation of state of dark energy. I: H (z) measurements, J. Cosmol. Astropart. Phys. 2010 (02) (2010) 008.

[105] Michele Moresco, Raising the bar: new constraints on the Hubble parameter with cosmic chronometers at z 2, Month. Not. R. Astron. Soc. Lett. 450 (1) (2015) L16–L20.

[106] Adrià Gómez-Valent, Luca Amendola, H0 from cosmic chronometers and Type Ia supernovae, with Gaussian Processes and the novel Weighted Polynomial regression method, J. Cosmol. Astropart. Phys. 2018 (04) (2018) 051.

[107] M. López-Corredoira, A. Vazdekis, Impact of young stellar components on quiescent galaxies: deconstructing cosmic chronometers, Astron. Astrophys. 614 (2018) A127.

[108] M. López-Corredoira, A. Vazdekis, C.M. Gutiérrez, N Castro-Rodríguez, Stellar content of extremely red quiescent galaxies at z> 2, Astron. Astrophys. 600 (2017) A91.

[109] Licia Verde, Pavlos Protopapas, Raul Jimenez, The expansion rate of the intermediate universe in light of Planck, Phys. Dark Univ. 5 (2014) 307–314.

[110] Marek Kowalski, David Rubin, Greg Aldering, RJ Agostinho, A. Amadon, R. Amanullah, C. Balland, K. Barbary, G. Blanc, P.J. Challis, et al., Improved cosmological constraints from new, old, and combined supernova data sets, Astrophys. J. 686 (2) (2008) 749.

[111] Rahman Amanullah, Chris Lidman, D. Rubin, G. Aldering, P. Astier, K. Barbary, M.S. Burns, A. Conley, K.S. Dawson, S.E. Deustua, et al., Spectra and Hubble Space Telescope light curves of six type Ia supernovae at 0.511< z< 1.12 and the Union2 compilation, Astrophys. J. 716 (1) (2010) 712.

[112] N. Suzuki, D. Rubin, C. Lidman, G. Aldering, R Amanullah, K. Barbary, L.F. Barrientos, J. Botyanszki, M. Brodwin, N. Connolly, et al., The Hubble Space Telescope cluster supernova survey. V. Improving the dark-energy constraints above z> 1 and building an early-type-hosted supernova sample, Astrophys. J. 746 (1) (2012) 85.

[113] M.E.A. Betoule, R. Kessler, J. Guy, J. Mosher, D. Hardin, R. Biswas, P. Astier, P. El-Hage, M. Konig, S. Kuhlmann, et al., Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples, Astron. Astrophys. 568 (2014) A22.

[114] Daniel Moshe Scolnic, D.O. Jones, A. Rest, Y.C. Pan, R. Chornock, R.J. Foley, M.E. Huber, R. Kessler, Gautham Narayan, A.G. Riess, et al., The complete light-curve sample of spectroscopically confirmed SNe Ia from Pan-STARRS1 and cosmological constraints from the combined pantheon sample, Astrophys. J. 859 (2) (2018) 101.

[115] Dan Scolnic, Dillon Brout, Anthony Carr, Adam G. Riess, Tamara M. Davis, Arianna Dwomoh, David O. Jones, Noor Ali, Pranav Charvu, Rebecca Chen, et al., The Pantheon+ Type Ia Supernova sample: the full dataset and light-curve release, 2021, arXiv preprint arXiv:2112.03863.

[116] Marek Demianski, Ester Piedipalumbo, Disha Sawant, Lorenzo Amati, Cosmology with gamma-ray bursts-I. The Hubble diagram through the calibrated Ep, i–Eiso correlation, Astron. Astrophys. 598 (2017) A112.

[117] Carl Roberts, Keith Horne, Alistair O. Hodson, Alasdair Dorkenoo Leggat, Tests of \ Λ CDM and Conformal Gravity using GRB and Quasars as Standard Candles out to z \ sim 8, 2017, arXiv preprint arXiv:1711.10369.

[118] Will J. Percival, Beth A. Reid, Daniel J. Eisenstein, Neta A. Bahcall, Tamas Budavari, Joshua A. Frieman, Masataka Fukugita, James E. Gunn, Zeljko Ivezi'c, Gillian R. Knapp, et al., Baryon acoustic oscillations in the Sloan Digital Sky Survey data release 7 galaxy sample, Mon. Not. R. Astron. Soc. 401 (4) (2010) 2148–2168.

[119] Florian Beutler, Chris Blake, Matthew Colless, D Heath Jones, Lister Staveley-Smith, Lachlan Campbell, Quentin Parker, Will Saunders, Fred Watson, The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant, Mon. Not. R. Astron. Soc. 416 (4) (2011) 3017–3032.

[120] Timothee Delubac, James Rich, Stephen Bailey, Andreu Font-Ribera, David Kirkby, J-M Le Goff, Matthew M. Pieri, Anze Slosar, Éric Aubourg, Julian E. Bautista, et al., Baryon acoustic oscillations in the Lyα forest of BOSS quasars, Astron. Astrophys. 552 (2013) A96.

[121] Lauren Anderson, Eric Aubourg, Stephen Bailey, Dmitry Bizyaev, Michael Blanton, Adam S. Bolton, Jon Brinkmann, Joel R. Brownstein, Angela Burden, Antonio J. Cuesta, et al., The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample, Mon. Not. R. Astron. Soc. 427 (4) (2012) 3435–3467.

[122] Hee-Jong Seo, Shirley Ho, Martin White, Antonio J Cuesta, Ashley J. Ross, Shun Saito, Beth Reid, Nikhil Padmanabhan, Will J. Percival, Roland De Putter, et al., Acoustic scale from the angular power spectra of SDSS-III DR8 photometric luminous galaxies, Astrophys. J. 761 (1) (2012) 13.

[123] Ashley J. Ross, Lado Samushia, Cullan Howlett, Will J. Percival, Angela Burden, Marc Manera, The clustering of the SDSS DR7 main Galaxy sample–I. A 4 per cent distance measure at z=0.15, Mon. Not. R. Astron. Soc. 449 (1) (2015) 835–847.

[124] Rita Tojeiro, Ashley J. Ross, Angela Burden, Lado Samushia, Marc Manera, Will J. Percival, Florian Beutler, J Brinkmann, Joel R. Brownstein, Antonio J. Cuesta, et al., The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: galaxy clustering measurements in the low-redshift sample of Data Release 11, Mon. Not. R. Astron. Soc. 440 (3) (2014) 2222–2237.

[125] Julian E. Bautista, Mariana Vargas-Magaña, Kyle S Dawson, Will J. Percival, Jonathan Brinkmann, Joel Brownstein, Benjamin Camacho, Johan Comparat, Hector Gil-Marín, Eva-Maria Mueller, et al., The SDSS-IV extended baryon oscillation spectroscopic survey: baryon acoustic oscillations at redshift of 0.72 with the DR14 luminous red galaxy sample, Astrophys. J. 863 (1) (2018) 110.

[126] E. De Carvalho, A. Bernui, G.C. Carvalho, C.P. Novaes, HS Xavier, Angular Baryon Acoustic Oscillation measure at z=2.225 from the SDSS quasar survey, J. Cosmol. Astropart. Phys. 2018 (04) (2018) 064.

[127] Metin Ata, Falk Baumgarten, Julian Bautista, Florian Beutler, Dmitry Bizyaev, Michael R. Blanton, Jonathan A. Blazek, Adam S. Bolton, Jonathan Brinkmann, Joel R. Brownstein, et al., The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2, Mon. Not. R. Astron. Soc. 473 (4) (2018) 4773–4794.

[128] T.M.C. Abbott, F.B. Abdalla, A. Alarcon, S. Allam, F. Andrade-Oliveira, J. Annis, S. Avila, Mandakranta Banerji, N. Banik, K. Bechtol, et al., Dark Energy Survey Year 1 Results: Measurement of the Baryon Acoustic Oscillation scale in the distribution of galaxies to redshift 1, Mon. Not. R. Astron. Soc. 483 (4) (2019) 4866–4883.

[129] Z. Molavi, A. Khodam-Mohammadi, Observational tests of Gauss-Bonnet like dark energy model, Eur. Phys. J. Plus 134 (6) (2019) 254.

[130] David Benisty, Denitsa Staicova, Testing late-time cosmic acceleration with uncorrelated baryon acoustic oscillation dataset, Astron. Astrophys. 647 (2021) A38.

[131] Natalie B. Hogg, Matteo Martinelli, Savvas Nesseris, Constraints on the distance duality relation with standard sirens, J. Cosmol. Astropart. Phys. 2020 (12) (2020) 019.

[132] Matteo Martinelli, Carlos Jose Amaro Parente Martins, S. Nesseris, Domenico Sapone, I. Tutusaus, Anastasios Avgoustidis, Stefano Camera, Carmelita Carbone, S. Casas, Stéphane Ilić, et al., Euclid: Forecast constraints on the cosmic distance duality relation with complementary external probes, Astron. Astrophys. 644 (2020) A80.

[133] Lu Chen, Qing-Guo Huang, Ke Wang, Distance priors from Planck final release, J. Cosmol. Astropart. Phys. 2019 (02) (2019) 028.

[134] Matt Visser, Cosmography: Cosmology without the Einstein equations, Gen. Relativity Gravitation 37 (2005) 1541–1548.

[135] Matt Visser, Jerk, snap and the cosmological equation of state, Classical Quantum Gravity 21 (11) (2004) 2603.

[136] Ujjaini Alam, Varun Sahni, Tarun Deep Saini, AA Starobinsky, Exploring the expanding universe and dark energy using the Statefinder diagnostic, Mon. Not. R. Astron. Soc. 344 (4) (2003) 1057–1074.

[137] Kazuharu Bamba, Salvatore Capozziello, Shin'ichi Nojiri, Sergei D Odintsov, Dark energy cosmology: the equivalent description via different theoretical models and cosmography tests, Astrophys. Space Sci. 342 (2012) 155–228.

[138] Varun Sahni, Tarun Deep Saini, Alexei A. Starobinsky, Ujjaini Alam, Statefinder—a new geometrical diagnostic of dark energy, J. Exp. Theor. Phys. Lett. 77 (2003) 201–206.

[139] Edmund J. Copeland, Mohammad Sami, Shinji Tsujikawa, Dynamics of dark energy, Internat. J. Modern Phys. D 15 (11) (2006) 1753–1935.

[140] Matteo Martinelli, Carlos Jose Amaro Parente Martins, S. Nesseris, Domenico Sapone, I. Tutusaus, Anastasios Avgoustidis, Stefano Camera, Carmelita Carbone, S. Casas, Stéphane Ilić, et al., Euclid: Forecast constraints on the cosmic distance duality relation with complementary external probes, Astron. Astrophys. 644 (2020) A80.

[141] Ujjaini Alam, Varun Sahni, Tarun Deep Saini, Alexei A Starobinsky, Is there supernova evidence for dark energy metamorphosis? Mon. Not. R. Astron. Soc. 354 (1) (2004) 275–291.

[142] Varun Sahni, Tarun Deep Saini, Alexei A. Starobinsky, Ujjaini Alam, Statefinder—a new geometrical diagnostic of dark energy, J. Exp. Theor. Phys. Lett. 77 (2003) 201–206.

[143] Ujjaini Alam, Varun Sahni, Tarun Deep Saini, AA Starobinsky, Exploring the expanding universe and dark energy using the Statefinder diagnostic, Mon. Not. R. Astron. Soc. 344 (4) (2003) 1057–1074.

[144] Andrew R. Liddle, How many cosmological parameters, Mon. Not. R. Astron. Soc. 351 (3) (2004) L49–L53.

[145] Gideon Schwarz, Estimating the dimension of a model, Ann. Stat. (1978) 461–464.

[146] Savvas Nesseris, Juan Garcia-Bellido, Is the Jeffreys' scale a reliable tool for Bayesian model comparison in cosmology? J. Cosmol. Astropart. Phys. 2013 (08) (2013) 036.

[147] Harold Jeffreys, The Theory of Probability, OuP Oxford, 1998.

[148] Martin Kerscher, Jochen Weller, On model selection in cosmology, in: SciPost Physics Lecture Notes, 2019, p. 009.

[149] Nathan J. Secrest, Sebastian von Hausegger, Mohamed Rameez, Roya Mohayaee, Subir Sarkar, A challenge to the standard cosmological model, Astrophys. J. Lett. 937 (2) (2022) L31.

[150] Yashar Akrami, Tomi S. Koivisto, Marit Sandstad, Cosmological constraints on ghost-free bigravity: background dynamics and late-time acceleration, in: The Thirteenth Marcel Grossmann Meeting: On Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories, World Scientific, 2015, pp. 1252–1254.

[151] Christopher M. Hirata, Uroš Seljak, Analyzing weak lensing of the cosmic microwave background using the likelihood function, Phys. Rev. D 67 (4) (2003) 043001.

[152] E. O'Mongain, Application of statistics to results in gamma ray astronomy, Nature 241 (5389) (1973) 376–379.

[153] Alma X. Gonzalez-Morales, Robert Poltis, Blake D. Sherwin, Licia Verde, Are priors responsible for cosmology favoring additional neutrino species? 2011, arXiv preprint arXiv:1106.5052.

[154] Joseph Ryan, Constraints on power law cosmology from cosmic chronometer, standard ruler, and standard candle data, J. Cosmol. Astropart. Phys. 2021 (08) (2021) 051.

[155] Özgür Akarsu, Tekin Dereli, Cosmological models with linearly varying deceleration parameter, Internat. J. Theoret. Phys. 51 (2012) 612–621.

[156] Suresh Kumar, Observational constraints on Hubble constant and deceleration parameter in power-law cosmology, Mon. Not. R. Astron. Soc. 422 (3) (2012) 2532–2538.

[157] Lixin Xu, Hongya Liu, Constraints to deceleration parameters by recent cosmic observations, Modern Phys. Lett. A 23 (23) (2008) 1939–1948.

View publication stats

Information and Inference: A Journal of the IMA (2021) **00**, 1–33
https://doi.org/dpae012

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

Cross-efficiency aggregation by ordered visibility graph averaging: method, and application in portfolio selection

Reenu Kumari*
Department of Applied Sciences, Maharaja Surajmal Institute of technology, Janakpuri, 110058, Delhi, India

Department of Applied Mathematics, Delhi Technological University, Rohini, 110042, Delhi, India
Abha Aggarwal
University School of Basic and Applied Sciences, Guru Gobind Singh Indraprastha University, Dwarka, 110078, Delhi, India

and

Anjana Gupta
Department of Applied Mathematics, Delhi Technological University, Rohini, 110042, Delhi, India
*Corresponding author: reenu_kumari@msit.in

[Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year]

In research and practice of Data Envelopment Analysis (DEA), the arithmetic average is commonly used to aggregate cross-efficiency scores. For this, each decision-making unit (DMU) contributes an equal weight, and many essential decision-making details are lost in the final aggregated cross-efficiency. We propose a novel application of the ordered visibility graph averaging (OVGA) operator for DEA cross-efficiency aggregation and apply the proposed method to study the portfolio selection problem. When solving this problem, several practical concerns, such as a budget, cardinality, buy-in requirements, and restrictions against short selling are also considered. The proposed OVGA aggregated cross-efficiency approach is explained through a numerical example, followed by the formulation of optimal portfolios based on these cross-efficiencies. The suggested method is also tested using empirical data from the Indian banking industry. The results of this study can be used to create the most acceptable portfolio in stock companies, financial institutions, and businesses in the public and private sectors.

Keywords: Data Envelopment Analysis; Decision Making Units; Aggregation method; Portfolio Optimization

1. Introduction

Portfolio selection refers to an investorâĂŹs decision to diversify their financial assets to reduce investment risk while maintaining a specific level of projected return. Portfolio selection offers guidance to an investor for allocating resources among various investment opportunities. The major breakthrough in

© The Author(s) 2024. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

UNCORRECTED MANUSCRIPT

the field came in 1952 when Harry Markowitz published his modern portfolio theory commonly known as mean-variance portfolio theory. Markowitz (1952) mean-variance portfolio selection model is based on the trade-off between risk and return under various economic conditions. To alleviate the limitations of the mean-variance framework, various alternative risk measures have been proposed in the literature. Although variance is widely accepted as a risk measure, it has limitations. Semi-variance is another risk measure, originally introduced by Markowitz (1959), and used in the mean-semi-variance portfolio selection model in Guo and Ching (2021).

Data envelopment analysis (DEA) is a non-parametric empirical approach for evaluating the relative efficiency of decision-making units (DMUs) in the presence of multiple inputs and multiple outputs. DEA model was first introduced by Farrell (1957), and then further developed by Charnes et al. (1978) (called CCR model) and Banker et al. (1984) (called BCC model). Due to the lack of translation invariance, the CCR and BCC models cannot handle the negative data values in inputs and outputs unless the data gets transformed into positive values. The directional distance function (DDF) based DEA models have been described in the literature to handle the non-positive data values without converting them into positive ones.

One of the applications of DEA in finance is portfolio selection. Few articles in the literature have studied the problem of portfolio selection using DEA. Chen (2008) presented an application of DEA in portfolio selection. Dia (2009) introduced a four-step methodology based on DEA for portfolio selection. Chen et al. (2018) constituted fuzzy portfolio assessment models based on DEA using various risk indicators to address the uncertainty involved in the financial market. Zhou et al. (2018) offered a DEA frontier improvement technique as a rebalancing strategy for investors. Gupta et al. (2020) introduced credibilitic fuzzy portfolio selection models based on DEA. Zhu and Zhu (2023) pioneered the concept of eco-efficiency of industrial investment (EEII) in terms of energy consumption, economic benefits, and environmental impact by using a slack-based DEA model. Omid et al. (2023) introduced a two-stage multiobjective optimization problem incorporating network data envelopment analysis. Soltanifar et al. (2023) developed DEA-R models for merger analysis that can deal with negative data and apply it to the Iranian banking sector. Ghiyasi and Zhu (2020) proposed an inverse semi-oriented DEA model to assess the efficiency of Chinese commercial banks following the global financial crisis, when negative outputs were present. Amirteimoori et al. (2022) developed the chance-constrained multistage DEA model for the relative performances of supply chains. Emrouznejad et al. (2023) recently presented a review of the eco-efficiency evaluation using DEA. Recently, Lim et al. (2014) introduced a novel way of portfolio selection using the DEA cross-efficiency. They consider Markowitz's mean-variance optimization model, which identifies a portfolio consisting of DMUs with minimum variation in their cross-efficiency scores and having a desirable expected mean efficiency.

Conventional DEA approaches permit each DMU to assess its efficiency using the most favorable weights, which might lead to exaggerated weight schemes. One of the most prominent ways of dealing with limited discriminative power is the cross-efficiency assessment, which presented citesexton1986data as an expansion of DEA and further improved by Doyle and Green (1994). These authors initiated aggressive and benevolent strategies as secondary objectives to resolve the problem of nonunique optimum weights in DEA. Davtalab-Olyaie and Asgharian (2021) discussed the problem of a non-Pareto optimum solution in CE evaluation. They proposed a multiobjective program to obtain a set of Pareto-optimal CE scores. Chen et al. (2020a) highlighted the problem of non-uniqueness in CE assessments and developed a meta-frontier analysis framework to develop a novel CE assessment method. The approach presented in the paper Oukil and Amin (2015) derives the importance of each DMU from a common consensus among all DMUs, which is then reflected via a preference voting matrix. Because self and peer evaluation scores represent DMU's mutual appreciation for one another,

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

the CE matrix is viewed as an election framework in which each DMU is both a candidate and a voter. Oukil (2020) introduced a new method based on collective weight profile (CWP) for each DMU by utilizing the preference voting system embedded within the weight matrix, which views the assessing DMUs as voters and the input/output factors as candidates. For a more detailed study of cross-efficiency innovations, limitations, and extensions, see Wu et al. (2021).

DEA cross-efficiency is helpful for portfolio selection problems since it removes the disadvantage of shifting weights in traditional DEA models. To achieve the maximum efficiency value, the DMU can give some inputs and outputs exceptionally high weights. As a result, portfolio selection that uses the traditional DEA technique is not recommended. To resolve this issue, Lim et al. (2014) proposed an MV-DEA portfolio selection method that selects a portfolio of DMUs with the least volatility in their CE scores that meets a specified anticipated return. Mashayekhi and Omrani (2016) applied the suggested MV-DEA cross-efficiency portfolio selection by assuming fuzzy returns. Recently, Amin and Hajjami (2021) examined the effect of the alternative optimal solution in portfolio selection and indicated that significant improvement in portfolio selection is obtained when alternate optimum solutions are used to obtain a cross-efficiency matrix (CEM). Chen et al. (2021) describes a portfolio selection method using fuzzy DEA cross-efficiency assessment, which considers undesirable ambiguous inputs and outputs. Deng and Fang (2019) developed a novel mean-variance-maverick DEA cross-efficiency approach based on the prospect of selecting a fuzzy portfolio. Gong et al. (2021) presented a method based on regret theory, in which they applied a fuzzy multiobjective portfolio optimization model with four objectives, namely, mean, variance, skewness, and efficiency, bounded by several realistic constraints. Essid et al. (2018) offered an innovative framework for portfolio optimization that combines the Maverick index with a DEA game cross-efficiency model.

Mehlawat et al. (2018) looked at the portfolio selection issue by adding additional facts about the non-normality of asset returns using skewness and kurtosis. The efficiency evaluation is also a critical factor to consider in the portfolio selection problem.

One of the drawbacks of the DEA cross-efficiency approach in portfolio selection is that all DMUs/stocks in the optimal portfolio are given identical weights. Amin and Hajjami (2021) suggested the development of a novel DEA cross-efficiency portfolio selection model to obtain more reasonable weights for DMUs/stocks. Although many scholars have realized the importance of aggregation methods for cross-efficiency assessment, there is still room for improvement in the existing aggregation methods and their applications in portfolio optimization problems. This paper uses the ordered visibility graph averaging (OVGA) operator to develop a cross-efficiency aggregation technique to analyze the units more thoroughly. The weights in the OVGA operator are determined by the relevance of the data in the obtained visibility graph Jiang et al. (2016). The OVGA operator is vital in aggregating CE scores since it allows values to be given varied weights depending on the importance of CE scores of different DMUs.

1.1. *Literature review*

The cross-efficiency technique, an extended form of DEA, considers both self and peer appraisal, resulting in a unique ranking of DMUs and removing the problem of artificial weight selections. Most cross-efficiency studies emphasize the cross-evaluation technique, and the arithmetic average is frequently used as a CE aggregation method in the existing research. The arithmetic average technique has certain drawbacks. First, self-evaluated efficiency does not play a sufficient role in the efficiency

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

aggregation procedure, as each DMU has only one self-evaluated efficiency value with multiple peer-evaluated efficiency values. Second, there is no other methodology to consider the decision-maker's subjective preferences, as the arithmetic average approach provides equal weight to each efficiency score (Oukil (2019), Wang and Chin (2011)). To remedy these issues, Song et al. (2017), and Li et al. (2018) devised aggregation methods that are motivated by the Shannon entropy weight and balanced adjustment weights, respectively, which provide both a solution for the inadequacies of the average arithmetic technique and an acceptable theoretical justification for CE aggregation. Wang and Chin (2011) introduced the OWA (ordered weighted averaging) operator weight to cross-efficiency aggregation. Using weights based on the OWA operator for CE aggregation permits the decision-maker's optimism level toward the best relative efficiencies. Liu and Chen (2022) pointed out that different efficiency scores affect integration outcomes differently. The relative importance of each efficiency value is not considered in an average aggregation of cross-efficiency. Liu and Chen (2022) first applied regret theory to the CE aggregation process to understand the regret aversion behavior of DMs, which plays a crucial role in the decision-making process. To determine the regret-based aggressive and benign cross-efficiency, weight matrices are built into their study. Wu et al. (2012) suggested a common weight technique to aggregate the cross-efficiency by considering DMUs as players in a cooperative game sense and calculating their Shapley values to address these limitations. Chen et al. (2020b) developed prospect theory to represent DMs' subjective preferences in their innovative technique to aggregate CE scores based on the prospect-consensus procedure. Li et al. (2022) outlined a novel strategy for generating aggregation weights from subjective and objective viewpoints. To reflect DM preferences from a subjective standpoint, the authors established prospect theory. Their method offers reference points as intervals that can be chosen based on the DMs' decision aims and preferences. Chen et al. (2020b) introduced a novel method based on prospect consensus (APC) theory for CE aggregation. The APC method keeps as much of the cross-efficiency matrix's decision-making information as feasible and then aggregates the cross-efficiency that suits decision-making demands. Combining cumulative prospect theory, Fang and Yang (2019) proposed an ordering method for interval CE aggregation.

All the studies mentioned above have different methodologies, indicators, and characteristics, requiring deep knowledge to understand the various procedures. Additionally, the above-discussed methods require many parameters in the calculation process that can be very tedious in handling large-scale data.

1.2. *Research motivation and novelty of the proposed approach*

Although several research studies have addressed DEA cross-efficiency aggregation under various frameworks, to the best of our knowledge, there is still some room for a new aggregation method that requires only the data of the cross-efficiency matrix, we will still be able to evaluate the aggregated CE scores efficiently. Also, we find fewer studies in the literature where DEA cross-efficiency has been applied to portfolio optimization problems. The proposed study makes the following contributions:

1. Compared to the aggregated method existing in the literature (León et al. (2014), Oukil (2019), Wang and Chin (2011)), our proposed OVGA method aggregates the cross-efficiencies with unique weights and provides weightage according to the importance of the data value, while the weights depend upon different characteristics such as DMs' optimism level and orness degree value, which is uncertain; furthermore, different orness degrees provide different sets of weights, as seen in Wang and Chin (2011) and León et al. (2014).

2. Compared with different aggregating operators, the OVGA operator, fully considers the information of orders and the correlation between the values. The main advantage of OVGA is that it is

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

completely data-driven, and the aggregated results can be produced from a series of original values.
3. We develop a new application of OVGA compared to that seen in Jiang et al. (2016, 2017) and Wang et al. (2022) in regard to DEA cross-efficiency aggregation.
4. Furthermore, compared to Deng and Fang (2019), Lim et al. (2014) and Essid et al. (2018), we integrate the MV-DEA cross-efficiency model with OVGA cross-efficiency scores and deal with the existence of an alternate optimal solution in CE for portfolio optimization problems.

1.3. *Organization of the paper*

The remainder of this paper is structured as follows. Section 2 is separated into two subsections, its subsequent subsections review the fundamental concepts of the DEA models and OVGA operator. Section 3 first presents the terminology of the OVGA operator for cross-efficiency aggregation, and then introduces the portfolio performance evaluation model. In Section 4, a numerical data set from the literature is used to demonstrate the applicability of the proposed model. Further, a case study of the Indian banking sector is taken care-off and the Managerial implications given the proposed model are also discussed here. Some concluding remarks are furnished in Section 5.

2. Preliminaries

This section presents the DEA models existing in the literature, used to investigate the efficiency and cross-efficiency of a DMU. Further, the ordered visibility graph weighted averaging (OVGWA) operator, developed by Wang et al. (2015), is also presented here. For basic definitions and notations regarding DEA models and OVGA operator, we may refer to Charnes et al. (1978) and Wang et al. (2015) respectively.

2.1. *DEA models for efficiency evaluation*

Suppose we have n *DMUs* to be evaluated with m inputs and s outputs. Let j^{th} DMU, DMU_j, $j = 1,2,\ldots,n$ uses x_{ij}, $i = 1,2,\ldots,m$ inputs and produce y_{rj}, $r = 1,2,\ldots,s$ outputs. Let DMU_k, $k = 1,2,\ldots,n$ be the DMU that is under evaluation. Let ω_{ik} $i = 1,2,\ldots,m$ and μ_{rk} $r = 1,2,\ldots,s$ be the weight given by the DMU_k, $k = 1,2,\ldots,n$ to the ith input and r^{th} outputs, respectively. The best relative self-efficiency $e^{\star}_{kk}$, under the assumption of constant return to scale, can be determined by using the following input-oriented CCR model (Charnes et al. (1978)).

Model I

$$\max e_{kk} = \sum_{r=1}^{s} \mu_{rk} y_{rk}$$

subject to,

$$\sum_{i=1}^{m} x_{ik}\omega_{ik} = 1,$$

$$\sum_{r=1}^{s} \mu_{rk} y_{rj} - \sum_{i=1}^{m} \omega_{ik} x_{ij} \leq 0, \qquad j = 1,2...n,$$

$$\omega_{ik} \geq 0,\ \mu_{rk} \geq 0, \qquad i = 1,2,\ldots,m,\ r = 1,2,\ldots,s.$$

Let $\mu^{\star}_{rk}$ $r = 1,2,\ldots,s$ and $\omega^{\star}_{ik}$ $i = 1,2,\ldots,m$ be the optimal solution of Model I, then $e^{\star}_{kk} = \sum_{r=1}^{s} \mu^{\star}_{rk} y_{rk}$ is the optimal efficiency that DMU_k can attain, which is also referred to the CCR-efficiency of DMU_k under self-evaluation. If $e^{\star}_{kk} = 1$ and all the optimal weights $\mu^{\star}_{rk}$ $r = 1,2,\ldots,s$

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

and $\omega_{ik}^{\star}\ i=1,2,\ldots,m$ are positive, then DMU_k is CCR-efficient, otherwise, it is CCR-inefficient.

The DEA cross-efficiency assessment approach uses peer review in place of sole self-evaluation to assess efficiency. Thus, the peer evaluation of $DMU_j,\ j=1,2,\ldots,n$ by the DMU_k is defined as follows:

$$e_{kj}=\frac{\sum_{r=1}^{s}\mu_{rk}^{\star}y_{rj}}{\sum_{i=1}^{m}\omega_{ik}^{\star}x_{ij}},\quad (k=1,2,\ldots,n). \tag{2.1}$$

It should be noted that the optimal weights derived from model I are rarely unique. As a result, the cross-efficiency score defined in (2.1) will not produce a unique set of cross-efficiency values. To deal with this issue Oral et al. (2015) introduced the maximum resonated appreciative model, which was further discussed in Amin and Hajjami (2021) and Amin and Oukil (2019). In Oral et al. (2015), authors have shown that the cross-efficiency score of any DMU_j with respect to any $DMU_k,\ k=1,2,\ldots,n$ can take any value between the interval. So the minimum and the maximum cross-efficiency score of DMU_j, evaluated by $DMU_k,\ k=1,2,\ldots,n$ can be computed by those optimal weights for which the self-efficiency score value will remain unchanged. They presented the following two formulations Models II and Model III, to determine the cross-efficiency interval of DMU_j evaluated by $DMU_k,\ k=1,2,\ldots,n$.

Model II

$$\begin{aligned}
\min\quad & \hat{e}_{kj}^{L}=\sum_{r=1}^{s}\mu_{rk}y_{rj}\\
\text{subject to,}\quad & \\
& \sum_{i=1}^{m}\omega_{ik}x_{ij}=1\\
& \sum_{r=1}^{s}\mu_{rk}y_{rj}-\sum_{i=1}^{m}\omega_{ik}x_{ij}\le 0,\quad j=1,2...n,\ j\neq k\\
& \sum_{r=1}^{s}\mu_{rk}y_{rk}-e_{kk}^{\star}\sum_{i=1}^{m}\omega_{ik}x_{ik}=0,\\
& \mu_{rk},\omega_{ik}\ge 0,\qquad r=1,2,\ldots,s;\ i=1,2,\ldots,m,
\end{aligned}$$

Let $\mu_{rk}^{\star}\ r=1,2,\ldots,s$ and $\omega_{ik}^{\star}\ i=1,2,\ldots,m$ be the optimal solutions of Model II, then $\hat{e}_{kj}^{L}=\sum_{r=1}^{s}\mu_{rk}^{\star}y_{rj}$ is the optimal efficiency that DMU_j can attain when evaluated by DMU_k. We can find out the Min cross-efficiency matrix for n $DMUs$ corresponding to Model II. The n efficiency values constitute a cross-efficiency matrix, as shown in Table 1.

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

Model III

$$\max \quad \hat{e}^U_{kj} = \sum_{r=1}^{s} \mu_{rk} y_{rj}$$

subject to,

$$\sum_{i=1}^{m} \omega_{ik} x_{ij} = 1$$

$$\sum_{r=1}^{s} \mu_{rk} y_{rj} - \sum_{i=1}^{m} \omega_{ik} x_{ij} \leq 0, \quad j = 1,2...n,\ j \neq k$$

$$\sum_{r=1}^{s} \mu_{rk} y_{rk} - e^{\star}_{kk} \sum_{i=1}^{m} \omega_{ik} x_{ik} = 0,$$

$$\mu_{rk}, \omega_{ik} \geq 0, \qquad r = 1,2,\ldots,s;\ i = 1,2,\ldots,m,$$

Let $\mu^{\star}_{rk}$, $r = 1,2,\ldots,s$ and $\omega^{\star}_{ik}$, $i = 1,2,\ldots,m$ be the optimal solutions of Model III, then $\hat{e}^U_{kj} = \sum_{r=1}^{s} \mu^{\star}_{rk} y_{rj}$ is the optimal efficiency that DMU_j can attain when evaluated by DMU_k.

TABLE 1 *Cross efficiency matrix for n DMUs corresponding to Model II*

Target DMU_K	Cross Efficiency of DMU_j					
	1	2	…	l	…	n
1	$\hat{e}^L_{11}$	$\hat{e}^L_{12}$	…	$\hat{e}^L_{1l}$	…	$\hat{e}^L_{1n}$
2	$\hat{e}^L_{21}$	$\hat{e}^L_{22}$	…	$\hat{e}^L_{2l}$	…	$\hat{e}^L_{2n}$
3	$\hat{e}^L_{31}$	$\hat{e}^L_{32}$	…	$\hat{e}^L_{3l}$	…	$\hat{e}^L_{3n}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	$\hat{e}^L_{n1}$	$\hat{e}^L_{n2}$	…	$\hat{e}^L_{nl}$	…	$\hat{e}^L_{nn}$
Average cross efficiency	$\frac{\sum_{k=1}^{n} \hat{e}^L_{k1}}{n}$	$\frac{\sum_{k=1}^{n} \hat{e}^L_{k2}}{n}$	…	$\frac{\sum_{k=1}^{n} \hat{e}^L_{kl}}{n}$	…	$\frac{\sum_{k=1}^{n} \hat{e}^L_{kn}}{n}$

Across all possible optimal solutions of CCR, Model II aims to determine the worst cross-efficiency score for DMU_k, and Model III aims to determine the best cross-efficiency score for DMU_k.

Remark 1 *The alternative optimal solutions of CCR model I corresponding to DMU_k are utilized to solve models II and III. It is obvious that $\hat{e}^L_{kj} < \hat{e}^U_{kj}$ only holds if the corresponding self-evaluation DEA model has different optimal solutions. In Oral et al. (2015) have stated that model III takes into consideration the issue of multiple optimal solutions from the CCR modeling the evaluation of cross-efficiency. The results can be greatly impacted by selecting the values from these intervals. One can select extreme*

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

cross-efficiency (Min cross-efficiency or Max cross-efficiency) for pessimistic and optimistic decisions. We investigate the results of the OVGA aggregated min cross-efficiency score in this work.

Similarly, we can find out the Max cross-efficiency matrix for n *DMUs* corresponding to Model III. The most popular method is averaging each row or column's efficiency in the cross-efficiency matrix with equal weights to determine a DMU's overall performance measurement. Assigning identical weights for cross-efficiency aggregation is not the best option because the self-evaluated efficiency score contributes less than peer-evaluated efficiencies do. In the next section, to follow we have the notion of the Ordered Visibility Graph Averaging operator for aggregation of efficiency of $DMU's$.

2.2. *Ordered Visibility Graph Weighted Averaging operator*

A Visibility Graph (VG) is an undirected graph of n nodes. The visibility graph method converts a time series into a graph. The associated graph derived from a time series has the following properties:

1. Connected: a node can see its nearest neighbors.
2. Undirected: the associated graph extracted from a time series
3. Invariant under affine transformations of the series data: Although rescale both horizontal and vertical axes, the visibility criterion is invariant.

Further, Wang et al. (2015) introduced an algorithm for ordered visible graphs (OVG). The VG method is to convert a time series into a graph, while the OVG maps a set of ordered data into a network.

Definition 1 (Wang et al. (2015)) *Let $O = \{o_1, o_2, \ldots, o_n\}$ represents a set of n ordered data, where, o_j is the jth largest element of the set. An ordered value o_j and its order j make up the coordinate (j, o_j) and represent a node j in the graph VG.*

Definition 2 (Wang et al. (2015)) *Two data values $(i, o_i), (j, o_j)$ have visibility, if any other value (k, o_k) is placed between them fulfills*

$$o_k < o_j + (o_i - o_j)\frac{j-k}{j-i}.$$

These argument values are connected in a network of connections by the visibility criteria. These connections are made depending on both the argument values themselves and the order in which they are presented.

The visibility criteria of OVG are similar to VG and are defined as follows:

Definition 3 (Wang et al. (2015)) *Two ordered data values $(i, o_i), (j, o_j)$ have visibility, if any other value (k, o_k) is placed between them fulfills*

$$o_k < o_j + (o_i - o_j)\frac{j-k}{j-i}.$$

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

Definition 4 *Let $\{(1,o_1),(2,o_2),\ldots,(n,o_n)\}$ be the set of n ordered data values. The adjacency matrix $A=[a_{ij}]_{n\times n}$ on the basis of visibility between the node i and node j is defined as*

$$a_{ij}=\begin{cases}1 & \text{if there is visibility from node } i \text{ to node } j\ , i\neq j\\ 0 & \text{otherwise.}\end{cases}$$

Let d_i be the degree of node i, and is given by $d_i=\sum_j a_{ij}$

Definition 5 (Ordered visibility graph weighted aggregation (OVGWA) operator) *Let $\{a_1,a_2,\ldots,a_n\}$ be the set of n ordered values, where $a_i\in R$. OVGWA operator is a mapping $F: I^n\longrightarrow I$, $I\in R$, such that $F(a_1,a_2,\ldots,a_n)=w_1a_1+w_2a_2+....+w_na_n$, where a_i is the ith value and w_i is the corresponding weight of the value a_i, satisfying the conditions $w_i\in[0,1]$, $\sum_{i=1}^n w_i=1$, and is defined as follows:*

$$w_i=\frac{d_i}{\sum_{i=1}^{n} d_i},\quad i=1,2,\ldots,n,$$

where d_i is the degree of the ith node. Further the aggregated value of the ordered set $\{a_1,a_2,\ldots,a_n\}$ is $F(a_1,a_2,\ldots,a_n)=w_1a_1+w_2a_2+....+w_na_n$.

The Python program that calculates the adjacency matrix, degree, and weight of an ordered set, on the lines of Definitions 3, 4, and 5 is described as Algorithm 1 in "Appendix". To give more clarity, we consider an example considered in Wang et al. (2015).

Example 1 (Wang et al. (2015)) *Assume there are 8 ordered values, $O=\{85,75,70,70,55,50,45,40\}$. In the OVG, every node represents the corresponding element in the ordered set, in the same order.*

From Algorithm 1, the corresponding adjacency matrix A, degree, and weight of an ordered set O (depicted in Table 2) are as follows:

$$A=\begin{bmatrix}0&1&1&1&0&0&0&0\\1&0&1&1&0&0&0&0\\1&1&0&1&0&0&0&0\\1&1&1&0&1&1&1&1\\0&0&0&1&0&1&0&0\\0&0&0&1&1&0&1&0\\0&0&0&1&0&1&0&1\\0&0&0&1&0&0&1&0\end{bmatrix}$$

According to the visibility criteria and on the lines of the adjacency matrix, Fig. 1 depicts the OVG so obtained.

Further, the aggregated value of the ordered data is

$$\begin{aligned}F(85,75,70,70,55,50,45,40) &= 85\times 3/26+75\times 3/26+70\times 3/26+70\times 6/26\\ &\quad +55\times 2/26+50\times 3/26+45\times 3/26+40\times 2/26\\ &= 60.96\end{aligned}$$

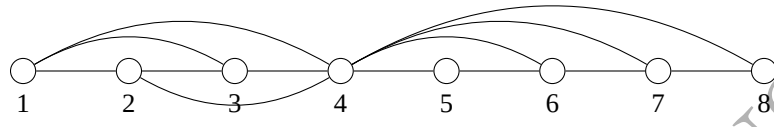
UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

TABLE 2 *Degree and weights of ordered set O*

Nodes (i)	1	2	3	4	5	6	7	8
degree (d_i)	3	3	3	7	2	3	3	2
weight (w_i)	$\frac{3}{26}$	$\frac{3}{26}$	$\frac{3}{26}$	$\frac{6}{26}$	$\frac{2}{26}$	$\frac{3}{26}$	$\frac{3}{26}$	$\frac{2}{26}$

FIG. 1. OVG of ordered set *O*



It should be noted that if a vertex is connected to another vertex by an edge, it is visible from that vertex since the two vertices have a direct line of sight. On the other hand, if a vertex is not connected to another vertex by an edge, it means that there is an obstruction in the way and the vertex cannot be seen from that specific vertex. In a visible network, the vertex with the highest degree is the one with the greatest number of connections to other vertices. It resembles a graph's center that is visible from several other spots. This vertex is crucial to the network as a whole because it is connected to a larger number of other vertices.

3. Portfolio optimization model based on the cross-efficiency matrix: Proposed model

Lim et al. (2014) studied Markowitz (1952) portfolio selection model based on the efficiencies of their assets. Lim et al. (2014) considered return and risk aspects of Markowitz's model in terms of mean efficiencies and the co-variance between the cross-efficiency scores of all *DMU′s* respectively. Before introducing the suggested model, it is necessary to comprehend the terminology from the previous part in terms of DMU efficiency.

Definition 6 (Adjacency matrix and degree of DMU_p with respect to DMU_j corresponding to min-cross-s-efficiency) *Let $\hat{E}_j^L = \{\hat{e}_{1j}^L, \hat{e}_{2j}^L, \ldots, \hat{e}_{pj}^L, \ldots, \hat{e}_{nj}^L\}$ be the set of n ordered cross-efficiencies of a DMU_j, where $\hat{e}_{ij}^L$ is the ith largest element of the set $\hat{E}_j^L$. Let $\{(1,\hat{e}_{1j}^L),(2,\hat{e}_{2j}^L),\ldots,(p,\hat{e}_{pj}^L),\ldots,(n,\hat{e}_{nj}^L)$ be the corresponding set of n ordered data nodes. The adjacency matrix for DMU_j, denoted by $A_j^L = [a_{pq}^j]_{n\times n}$ on the basis of visibility between DMU_p, $p = 1,2,\ldots,n$ and DMU_q, $q = 1,2,\ldots,n$ becomes as follows*

$$a_{pq}^j = \begin{cases} 1, & \text{if there is visibility between the data values } (p,\hat{e}_{pj}^L) \text{ and } (q,\hat{e}_{qj}^L),\ p \neq q \\ 0, & \text{otherwise.} \end{cases}$$

Let d_p^j be the degree of DMU_p under the consideration of DMU_j, then $d_p^j = \sum_q a_{pq}^j$

Definition 7 (OVGA min-efficiency of DMU_j) *Let $\hat{E}_j^L = \{\hat{e}_{1j}^L, \hat{e}_{2j}^L, \ldots, \hat{e}_{kj}^L, \ldots, \hat{e}_{nj}^L\}$ be the set of n ordered min-cross-efficiencies of a DMU_j, where $\hat{e}_{ij}^L$ is the ith largest element of the set $\hat{E}_j^L$. Then OVGA*

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

min-efficiency of DMU_j, denoted by $\tilde{e}_j^L$ is defined as

$$\tilde{e}_j^L = w_{1j}\hat{e}_{1j}^L + w_{2j}\hat{e}_{2j}^L + \ldots w_{pj}\hat{e}_{pj}^L + \ldots w_{nj}\hat{e}_{nj}^L,$$

where,

$$w_{pj} = \frac{d_p^j}{\sum_{p=1}^{n} d_p^j},\ p = 1,2,\ldots,n.$$

Definition 8 (Average of min-cross efficiency) *Let $\hat{E}_j^L = \{\hat{e}_{1j}^L, \hat{e}_{2j}^L, \ldots, \hat{e}_{kj}^L, \ldots, \hat{e}_{nj}^L\}$ be the set of n cross-efficiencies of a DMU_j, where $\hat{e}_{kj}^L$ is the min-cross-efficiency of DMU_j evaluated with respect to DMU_k. Then the average of min-cross-efficiencies of DMU_j, denoted by $\bar{e}_j$ is defined as*

$$\bar{e}_j^L = \frac{\hat{e}_{1j}^L + \hat{e}_{2j}^L + \ldots, \hat{e}_{kj}^L + \ldots \hat{e}_{nj}^L}{n}$$

Definition 9 (Adjacency matrix and degree of DMU_p with respect to DMU_j corresponding to max-cross-efficiency) *Let $\hat{E}_j^U = \{\hat{e}_{1j}^U, \hat{e}_{2j}^U, \ldots, \hat{e}_{pj}^U, \ldots, \hat{e}_{nj}^U\}$ be the set of n ordered max-cross-efficiencies of a DMU_j, where $\hat{e}_{ij}^U$ is the ith largest element of the set $\hat{E}_j^U$. Let $\{(1,\hat{e}_{1j}^U), (2,\hat{e}_{2j}^U), \ldots, (p,\hat{e}_{pj}^U), \ldots, (n,\hat{e}_{nj}^U)$ be the corresponding set of n ordered data nodes. The adjacency matrix for DMU_j, denoted by $B_j^U = [b_{pq}^j]_{n\times n}$ on the basis of visibility between DMU_p, $p = 1,2,\ldots,n$ and DMU_q, $q = 1,2,\ldots,n$ becomes as follows*

$$b_{pq}^j = \begin{cases} 1, & \text{if there is visibility between the data values} (p,\hat{e}_{pj}^U) \text{and} (q,\hat{e}_{qj}^U),\ p \neq q \\ 0, & \text{otherwise.} \end{cases}$$

Let c_p^j be the degree of DMU_p under the consideration of DMU_j, then $c_p^j = \sum_q b_{pq}^j$

Definition 10 (OVGA max-efficiency of DMU_j) *Let $\hat{E}_j^U = \{\hat{e}_{1j}^U, \hat{e}_{2j}^U, \ldots, \hat{e}_{kj}^U, \ldots, \hat{e}_{nj}^U\}$ be the set of n ordered max-cross-efficiencies of a DMU_j, where $\hat{e}_{ij}^U$ is the ith largest element of the set $\hat{E}_j^U$. Then OVGA max-efficiency of DMU_j, denoted by $\tilde{e}_j^U$ and is defined as*

$$\tilde{e}_j^U = w_{1j}\hat{e}_{1j}^U + w_{2j}\hat{e}_{2j}^U + \ldots w_{pj}\hat{e}_{pj}^U + \ldots w_{nj}\hat{e}_{nj}^U,$$

where,

$$w_{pj} = \frac{c_p^j}{\sum_{p=1}^{n} c_p^j},\ p = 1,2,\ldots,n.$$

Definition 11 (Average of max-cross efficiency) *Let $\hat{E}_j^U = \{\hat{e}_{1j}^U, \hat{e}_{2j}^U, \ldots, \hat{e}_{kj}^U, \ldots, \hat{e}_{nj}^U\}$ be the set of n cross-efficiencies of a DMU_j, where $\hat{e}_{kj}^U$ is the max-cross-efficiency of DMU_j evaluated with respect to DMU_k. Then the average of max-cross-efficiency of DMU_j, denoted by $\bar{e}_j^U$ is defined as*

$$\bar{e}_j^U = \frac{\hat{e}_{1j}^U + \hat{e}_{2j}^U + \ldots, \hat{e}_{kj}^U + \ldots \hat{e}_{nj}^U}{n}$$

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

Remark 2 *In the visibility graph, the weight of each node is determined by the degree distribution. Thus if a node links with more nodes than others, its degree will be greater and it will be more significant. The following are the advantages of the OVGA operators over other methods for cross-efficiency aggregation:-*

1. *In the OWA method, we obtain different aggregated results for different optimism levels and it is very difficult for decision-makers to select which value to make final decisions.*
2. *Compared with the classical OWA operator, the proposed OVGA operator not only takes the information of orders but also the argument values themselves into consideration to determine the weights. The OVGA operator is inspired by the complex network theory and Newton's law of universal gravitation.*
3. *Other cross-efficiency aggregation methods, such as prospect and regret theory, require many calculation parameters that will be very difficult to handle and will be very calculative for the large-scale data set. Prospect theory cross-efficiency aggregation requires at least four parameters, including risk attitude coefficients, the concave degree of the value function within the gain/loss region, and a psychological preference value. Similarly, the regret theory cross-efficiency aggregation method utilizes the regret-rejoice function based on the regret aversion coefficient.*
4. *The primary benefit of the OVGA operator is that it is based on the idea of human intuition. In simpler terms, a vertex with the highest degree can be considered a popular person who knows many people and has a lot of influence. Their connections and interactions provide valuable insights into the relationships and dynamics of the time series data.*

Portfolio selection model corresponding to min-cross-efficiency

Let us consider a portfolio Ω constituted with n *DMUs* say DMU_i, $i = 1,2,\ldots,n$. On the same lines Lim et al. (2014), in this paper, the return and risk characteristics of Markowitz's model (Markowitz (1952)), mean and variance of the model are described by mean cross-efficiency scores and variance of its cross-efficiencies, respectively.

Let $[\hat{e}^L_{ij}]_{n\times n}$ be the min-cross-efficiency matrix on n DMUs taken under consideration for portfolio selection. Let x_i be the proportion of the total efficiency invested on DMU_i. The portfolio Ω of n *DMUs* by the OVGA operator is obtained by solving the following cross-efficiency portfolio optimization model:

Model IV

$$\min \quad V_\Omega = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \tilde{\sigma}^L_{ij}$$

subject to,

$$E_\Omega = \sum_{i=1}^{n} x_i \tilde{e}^L_i \geq (1-\gamma)\tilde{E}^b_\Omega$$

$$\sum_{i=1}^{n} x_i = 1,$$

$$\textstyle\sum_{i=1}^{n} z_i = d, \quad i = 1,2,\ldots,n,$$

$$\varepsilon_i z_i \leq x_i \leq \delta_i z_i, \quad i = 1,2,\ldots,n,$$

$$x_i \in \{0,1\}, \quad i = 1,2,\ldots,n,$$

$$x_i \geq 0, \quad i = 1,2,\ldots,n,$$

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

where, $\tilde{e}_i^L$ is the mean cross- efficiency score of DMU_i by OVGA operator, defined in Definition 7 and $\tilde{\sigma}_{ij}^L = \frac{1}{n}\sum_{k=1}^{n}(\hat{e}_{ki}^L - \tilde{e}_i^L)(\hat{e}_{kj}^L - \tilde{e}_j^L)$ is the covariance between the cross-efficiency scores of DMU_i and DMU_j. Here, γ is the trade-off parameter between mean efficiency and co-variance of their efficiency. Further, for a better selection of a DMU, the budget constraint $\sum_{i=1}^{n} x_i = 1$, cardinality constraint $\sum_{i=1}^{n} z_i = d, z_i \in \{0,1\}, \quad i = 1,2,\ldots,n,$, buy-in threshold $\varepsilon_i z_i \leq x_i \leq \delta_i z_i, \quad i = 1,2,\ldots,n,$, and no short-selling criteria $x_i \geq 0,\ i = 1,2,\ldots,n$, are also considered here. Here, E_Ω^b is the maximum OVGA efficiency achievable by the portfolio under the budget constraint, cardinality constraint, buy-in threshold, and no short-selling criteria.

Let $\tilde{E}_{\Omega^\star}$ and $\tilde{V}_{\Omega^\star}$ be the optimal mean efficiency and optimal risk corresponding to the optimal portfolio $\Omega^\star$ so obtained.

Similarly, the mean-variance cross-efficiency model for simple average efficiency as defined in Definition 8 is as follows:

Model V

$$\begin{aligned}
\min \quad & V_\Omega = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \bar{\sigma}_{ij}^L \\
\text{subject to,} \quad & \\
& \sum_{i=1}^{n} x_i \bar{e}_i^L \geq (1-\gamma)\bar{E}_\Omega^b \\
& \sum_{i=1}^{n} x_i = 1 \\
& \textstyle\sum_{i=1}^{n} z_i = d, \quad i = 1,2,\ldots,n, \\
& \varepsilon_i z_i \leq x_i \leq \delta_i z_i, \quad i = 1,2,\ldots,n, \\
& z_i \in \{0,1\}, \quad i = 1,2,\ldots,n, \\
& x_i \geq 0, \quad i = 1,2,\ldots,n,
\end{aligned}$$

where, $\bar{e}_i^L$ is the mean cross efficiency score of DMU_i defined in Definition 8 and $\bar{\sigma}_{ij}^L = \frac{1}{n}\sum_{k=1}^{n}(e_{ki}^L - \bar{e}_i^L)(e_{kj}^L - \bar{e}_j^L)$ is the covariance between the cross efficiency scores of DMU_i and DMU_j. E_Ω^b is the maximum mean efficiency, achievable by the portfolio E_Ω. Further, γ is the trade-off parameter between mean efficiency and co-variance of their efficiency. Further, other constraints have the same interpretation as explained in Model IV. Let $\tilde{E}_{\Omega^\star}$ and $\tilde{V}_{\Omega^\star}$ be the optimal mean efficiency and optimal risk corresponding to the optimal portfolio $\Omega^\star$ so obtained.

Remark 3 *Now on the lines of Definitions 9, 10, and 11, we can extend Model IV and Model V to study the portfolio selection for max-cross-efficiency of DMUs.*

Remark 4 *It is imperative to bear in mind that the selection of the Minimum or Maximum cross-efficiencies, or the extraction of efficiencies from the interval between the min and max cross-efficiencies, will have a significant influence on the portfolio selection outcomes. In this work, we solely address the ideal portfolio derived from model IV. We can explore the OVGA max cross-efficiency and its impact on portfolio outcomes more thoroughly in future studies.*

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

4. Numerical Illustration

In this section, we apply our proposed method to a numerical example consisting of 22 DMUs to demonstrate the empirical effectiveness of the method. The numerical dataset is taken from Lim et al. (2014) which was further analyzed by Amin and Hajjami (2021). The input and output data for the DMUs are presented in Table 3.

TABLE 3 *Input and output data values for DMUs*

DMUs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Input 1	5	6	4	8	5	8	4.4	2.6	3.4	3.6	2	3	3	2.6	4	5	6	4	7	6	8	9
Input 2	4	5	5	5	6	3	4.4	8	8	4.4	7	7	5.6	5	4	3.2	4	3.5	3	2.5	2	2
Output 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

On solving the Model I, for DMU_1, we get $\mu_{11}^{\star} = 0.85$, $\omega_{11}^{\star} = 0.0769$, $\omega_{21}^{\star} = 0.1538$, with the optimal value of the objective function as $e_{11}^{\star} = 0.85$. Similarly on solving Model I, for other *DMUs*, the self-efficiency of respective *DMUs* can be acquired as follows:

TABLE 4 *Self-efficiency score $e_{kk}^{\star}$ of DMU_k*

k	1	2	3	4	5	6	7	8	9	10	11
$e_{kk}^{\star}$	0.85	0.69	0.84	0.61	0.69	0.80	0.85	0.82	0.71	0.94	1
k	12	13	14	15	16	17	18	19	20	21	22
$e_{kk}^{\star}$	0.80	0.88	1	0.94	0.96	0.79	1	0.85	1	1	1

For cross-efficiency interval evaluation of DMU_j, we have to solve Model II and Model III respectively for different values of k for a fixed value of j. For example if we take $k = 2$ and $j = 1$, we can find $\tilde{e}_{21}^{L}$ and $\bar{e}_{21}^{U}$. As in the cross-efficiency matrix, DMU_j is evaluated by DMU_k, so to determine the weights given by DMU_j to DMU_k to find out its aggregated efficiency by the OVGA operator, we need re-ordered minimum cross-efficiency matrix.

To get more clarity, let us calculate the weights given by DMU_1 to DMU_k, $k = 1, 2, \ldots, n$ for the OVGWA operator to get its aggregated efficiency. In a similar fashion, we can find out the weights given by other *DMUs* to DMU_k, $k = 1, 2, \ldots, n$. The set of ordered cross-efficiencies of DMU_1, $\hat{e}_{k1}^{L}$, $k =$

UNCORRECTED MANUSCRIPT

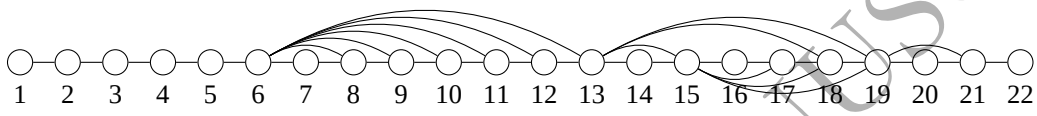
Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

$1,2,\ldots,22$ (as per Table 7) for DMU_1 is

$$\hat{E}_1^L = \{0.85,\ 0.85,\ 0.85,\ 0.85,\ 0.85,\ 0.85,\ 0.83,\ 0.83,\ 0.83,\ 0.83,\ 0.83,\ 0.83,\ 0.83,\ 0.76,\ 0.76,\ 0.66,\ 0.66,\ 0.66,\ 0.66,\ 0.50,\ 0.50,\ 0.40.\}\ .$$

To calculate the weights given by DMU_1 to DMU_k, $k = 1,2,\ldots,22$, for evaluating its OVGWA efficiency, the below-given Fig.1 depicts the ordered visibility graph of all $DMUs$ with respect to ordered cross efficiency set $\hat{E}_1^L$.

FIG. 2. Ordered visibility graph of ordered efficiency set $\hat{E}_1^L$



Now on lines of Definition 6, from Algorithm 1, the corresponding adjacency matrix is

$$A_1^L = \begin{bmatrix}
0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&0&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&0&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&0&0&1&0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&0&0&0&1&0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&0&0&0&0&1&0&1&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&0&0&0&0&0&1&0&1&1&0&0&1&1&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&1&1&1&1&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&1&0&1&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&1&0&1&1&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0
\end{bmatrix}$$

So the aggregated cross-efficiency by OVGWA operator and by simple average is $\tilde{e}_1^L = \sum_{p=1}^{22} w_p^1 \hat{e}_{p1}^L = 0.76$ and $\bar{e}_1^L = \sum_{p=1}^{22} (1/n)\hat{e}_{p1}^L = 0.75$, respectively. These values are the same as given in Table 9, and

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

TABLE 5 *degree and weights of DMU_p w.r.t. DMU_1*

p	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
d_p^1	1	2	2	2	2	8	2	3	3	3	3	3	6	2	6	2	3	4	5	2	3	1
w_p^1	$\frac{1}{68}$	$\frac{2}{68}$	$\frac{2}{68}$	$\frac{2}{68}$	$\frac{2}{68}$	$\frac{8}{68}$	$\frac{2}{68}$	$\frac{3}{68}$	$\frac{3}{68}$	$\frac{3}{68}$	$\frac{3}{68}$	$\frac{3}{68}$	$\frac{6}{68}$	$\frac{2}{68}$	$\frac{6}{68}$	$\frac{2}{68}$	$\frac{3}{68}$	$\frac{4}{68}$	$\frac{5}{68}$	$\frac{2}{68}$	$\frac{3}{68}$	$\frac{1}{68}$

in a similar manner, we can calculate the aggregated cross-efficiency scores of other DMUs. For this, we first need to solve Model II, for $k, j = 1, 2, \ldots, n$ the minimum cross-efficiency matrix is depicted in Table 9. Further, Table 7 presents the re-ordered cross-efficiency matrix. As per the requirement of the calculation Table 8 illustrates the weights given by DMU_j to DMU_p for $j, p = 1, 2, \ldots, 22$.

Portfolio selection

Here, optimal portfolios are selected in two ways:

Case I: the portfolio selection is performed with equal weights

The decision maker wishes to have an equal proportion of investment on the DMUs for the optimal portfolios. To present the advantages of the proposed portfolio selection model, we first obtain portfolios using the proposed model for different cardinality constraints i.e., d=5, d=10, and d=7. The obtained portfolio strategies are given in Table 10. To compare the portfolios obtained from the traditional method, i.e., after solving Model V, the portfolio strategies under the different cardinality constraints i.e., d=5, d=10, and d=15 are presented in Table 11. Here it is to be noted that, the portfolios obtained through the proposed model have a considerable reduction in variance (risk) and a slight reduction in mean. For example, the portfolio obtained through the proposed method, presented in Table 10 corresponding to d=5, and $\gamma = 0.1$ shows the 1.73% decrease in mean in comparison to the 33.4% decrease in variance.

Case II: Optimal portfolio selection with unequal weights

Here it is to be noted that different portfolios are generated by varying values of $d, \gamma, \varepsilon_i, \delta_i$. The optimal portfolio strategies corresponding to the proposed model i.e., after solving Model IV, are presented in Table 12, and with respect to the traditional method, i.e., after solving Model V, the optimal strategies are presented in Table 13. As discussed above, the portfolios obtained through the proposed model have a considerable reduction in variance (risk) against the slight reduction in mean. For example, the portfolio obtained through a proposed method, presented in Table 12 corresponding to $d = 5, \gamma = 0.1, \varepsilon_i = 0.03$ and $\delta = 0.4$ shows the 1.1% decrease in mean against the considerable decrease 27.5% in variance.

It is to be noted that, a bigger value of γ can also result in a more significant decrease in risk for a given expected return. The effectiveness of the proposed approach for portfolio selection is empirically supported by the acquired portfolio selection strategies. The suggested model can generate portfolios based on the investor's choices corresponding to γ.

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May

TABLE 6 *The cross-efficiency matrix* $\tilde{e}^L_{kj}$

k \ j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	0.85	0.69	0.79	0.61	0.65	0.79	0.83	0.59	0.57	0.89	0.69	0.65	0.77	0.87	0.92	0.96	0.79	1.00	0.85	1.00	0.92	0.85
2	0.85	0.69	0.79	0.61	0.65	0.79	0.83	0.59	0.57	0.89	0.69	0.65	0.77	0.87	0.92	0.96	0.79	1.00	0.85	1.00	0.92	0.85
3	0.83	0.68	0.84	0.57	0.69	0.67	0.85	0.72	0.67	0.94	0.85	0.76	0.88	1.00	0.94	0.91	0.75	1.00	0.74	0.87	0.74	0.67
4	0.85	0.69	0.79	0.61	0.65	0.79	0.83	0.59	0.57	0.89	0.69	0.65	0.77	0.87	0.92	0.96	0.79	1.00	0.85	1.00	0.92	0.85
5	0.83	0.68	0.84	0.57	0.69	0.67	0.85	0.72	0.67	0.94	0.85	0.76	0.88	1.00	0.94	0.91	0.75	1.00	0.74	0.87	0.74	0.67
6	0.76	0.62	0.67	0.57	0.55	0.80	0.73	0.46	0.45	0.75	0.53	0.52	0.63	0.71	0.80	0.90	0.73	0.89	0.84	1.00	1.00	0.94
7	0.83	0.68	0.84	0.57	0.69	0.67	0.85	0.72	0.67	0.94	0.85	0.76	0.88	1.00	0.94	0.91	0.75	1.00	0.74	0.87	0.74	0.67
8	0.66	0.55	0.75	0.43	0.60	0.46	0.72	0.82	0.71	0.83	1.00	0.80	0.88	1.00	0.79	0.69	0.57	0.81	0.52	0.61	0.48	0.43
9	0.66	0.55	0.75	0.43	0.60	0.46	0.72	0.82	0.71	0.83	1.00	0.80	0.88	1.00	0.79	0.69	0.57	0.81	0.52	0.61	0.48	0.43
10	0.83	0.68	0.84	0.57	0.69	0.67	0.85	0.72	0.67	0.94	0.85	0.76	0.88	1.00	0.94	0.91	0.75	1.00	0.74	0.87	0.74	0.67
11	0.40	0.33	0.50	0.28	0.40	0.28	0.45	0.77	0.59	0.56	1.00	0.67	0.67	0.77	0.50	0.40	0.33	0.50	0.29	0.33	0.25	0.22
12	0.66	0.55	0.75	0.43	0.60	0.46	0.72	0.82	0.71	0.83	1.00	0.80	0.88	1.00	0.79	0.69	0.57	0.81	0.52	0.61	0.48	0.43
13	0.83	0.68	0.84	0.57	0.69	0.67	0.85	0.72	0.67	0.94	0.85	0.76	0.88	1.00	0.94	0.91	0.75	1.00	0.74	0.87	0.74	0.67
14	0.66	0.55	0.75	0.43	0.60	0.46	0.72	0.72	0.67	0.83	0.85	0.76	0.88	1.00	0.79	0.69	0.57	0.81	0.52	0.61	0.48	0.43
15	0.83	0.68	0.84	0.57	0.69	0.67	0.85	0.72	0.67	0.94	0.85	0.76	0.88	1.00	0.94	0.91	0.75	1.00	0.74	0.87	0.74	0.67
16	0.85	0.69	0.79	0.61	0.65	0.79	0.83	0.59	0.57	0.89	0.69	0.65	0.77	0.87	0.92	0.96	0.79	1.00	0.85	1.00	0.92	0.85
17	0.85	0.69	0.79	0.61	0.65	0.79	0.83	0.59	0.57	0.89	0.69	0.65	0.77	0.87	0.92	0.96	0.79	1.00	0.85	1.00	0.92	0.85
18	0.83	0.68	0.79	0.57	0.65	0.67	0.83	0.59	0.57	0.89	0.69	0.65	0.77	0.87	0.92	0.91	0.75	1.00	0.74	0.87	0.74	0.67
19	0.85	0.69	0.79	0.61	0.65	0.79	0.83	0.59	0.57	0.89	0.69	0.65	0.77	0.87	0.92	0.96	0.79	1.00	0.85	1.00	0.92	0.85
20	0.76	0.62	0.67	0.57	0.55	0.79	0.73	0.46	0.45	0.75	0.53	0.52	0.63	0.71	0.80	0.90	0.73	0.89	0.84	1.00	0.92	0.85
21	0.50	0.40	0.40	0.40	0.33	0.67	0.45	0.25	0.25	0.45	0.29	0.29	0.36	0.40	0.50	0.63	0.50	0.67	0.67	0.80	1.00	0.94
22	0.50	0.40	0.40	0.40	0.33	0.67	0.45	0.25	0.25	0.45	0.29	0.29	0.36	0.40	0.50	0.63	0.50	0.57	0.67	0.80	1.00	1.00

UNCORRECTED MANUSCRIPT

TABLE 7 *Re-ordered cross-efficiency matrix* $\tilde{e}^{L}_{kj}$

k \ j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	0.85	0.69	0.84	0.61	0.69	0.80	0.85	0.82	0.71	0.94	1	0.80	0.88	1	0.94	0.96	0.79	1	0.85	1	1	1
2	0.85	0.69	0.84	0.61	0.69	0.79	0.85	0.82	0.71	0.94	1	0.80	0.88	1	0.94	0.96	0.79	1	0.85	1	1	0.94
3	0.85	0.69	0.84	0.61	0.69	0.79	0.85	0.82	0.71	0.94	1	0.8	0.88	1	0.94	0.96	0.79	1	0.85	1	1	0.94
4	0.85	0.69	0.84	0.61	0.69	0.79	0.85	0.77	0.67	0.94	1	0.76	0.88	1	0.94	0.96	0.79	1	0.85	1	0.92	0.85
5	0.85	0.69	0.84	0.61	0.69	0.79	0.85	0.72	0.67	0.94	0.85	0.76	0.88	1	0.94	0.96	0.79	1	0.85	1	0.92	0.85
6	0.85	0.69	0.84	0.61	0.69	0.79	0.85	0.72	0.67	0.94	0.85	0.76	0.88	1	0.94	0.96	0.79	1	0.85	1	0.92	0.85
7	0.83	0.68	0.79	0.57	0.65	0.79	0.83	0.72	0.67	0.89	0.85	0.76	0.88	1	0.92	0.91	0.75	1	0.84	1	0.92	0.85
8	0.83	0.68	0.79	0.57	0.65	0.79	0.83	0.72	0.67	0.89	0.85	0.76	0.88	1	0.92	0.91	0.75	1	0.84	1	0.92	0.85
9	0.83	0.68	0.79	0.57	0.65	0.67	0.83	0.72	0.67	0.89	0.85	0.76	0.88	1	0.92	0.91	0.75	1	0.74	0.87	0.92	0.85
10	0.83	0.68	0.79	0.57	0.65	0.67	0.83	0.72	0.67	0.89	0.85	0.76	0.88	1	0.92	0.91	0.75	1	0.74	0.87	0.92	0.85
11	0.83	0.68	0.79	0.57	0.65	0.67	0.83	0.72	0.59	0.89	0.85	0.67	0.77	0.87	0.92	0.91	0.75	1	0.74	0.87	0.74	0.67
12	0.83	0.68	0.79	0.57	0.65	0.67	0.83	0.59	0.57	0.89	0.69	0.65	0.77	0.87	0.92	0.91	0.75	1	0.74	0.87	0.74	0.67
13	0.83	0.68	0.79	0.57	0.65	0.67	0.83	0.59	0.57	0.89	0.69	0.65	0.77	0.87	0.92	0.91	0.75	1	0.74	0.87	0.74	0.67
14	0.76	0.62	0.75	0.57	0.6	0.67	0.73	0.59	0.57	0.83	0.69	0.65	0.77	0.87	0.8	0.9	0.73	0.89	0.74	0.87	0.74	0.67
15	0.76	0.62	0.75	0.57	0.6	0.67	0.73	0.59	0.57	0.83	0.69	0.65	0.77	0.87	0.8	0.9	0.73	0.89	0.747	0.87	0.74	0.67
16	0.66	0.55	0.75	0.43	0.6	0.67	0.72	0.59	0.57	0.83	0.69	0.65	0.77	0.87	0.79	0.69	0.57	0.81	0.67	0.8	0.74	0.67
17	0.66	0.55	0.75	0.43	0.6	0.67	0.72	0.59	0.57	0.83	0.69	0.65	0.77	0.87	0.79	0.69	0.57	0.81	0.67	0.8	0.74	0.67
18	0.66	0.55	0.67	0.43	0.55	0.46	0.72	0.59	0.57	0.75	0.69	0.65	0.67	0.77	0.79	0.69	0.57	0.81	0.52	0.61	0.48	0.43
19	0.66	0.55	0.67	0.43	0.55	0.46	0.72	0.46	0.45	0.75	0.53	0.52	0.63	0.71	0.79	0.69	0.57	0.81	0.52	0.61	0.48	0.43
20	0.5	0.4	0.5	0.4	0.4	0.46	0.45	0.46	0.45	0.56	0.53	0.52	0.63	0.71	0.5	0.63	0.5	0.57	0.52	0.61	0.48	0.43
21	0.5	0.4	0.4	0.4	0.33	0.46	0.45	0.25	0.25	0.45	0.29	0.29	0.36	0.4	0.5	0.63	0.5	0.57	0.52	0.61	0.48	0.43
22	0.4	0.33	0.4	0.25	0.33	0.25	0.45	0.25	0.25	0.45	0.29	0.29	0.36	0.4	0.5	0.4	0.33	0.5	0.29	0.33	0.25	0.22
$\bar{e}_j$	0.75	0.61	0.73	0.53	0.60	0.66	0.76	0.63	0.58	0.83	0.75	0.66	0.77	0.87	0.83	0.83	0.68	0.89	0.71	0.84	0.76	0.70

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/maman/advance-article/doi/10.1093/maman/dpae012/7671391 by Delhi Technological University user on 22 May

TABLE 8 *The weight w_p^j given by DMU_j to DMU_p on the basis of re-ordered cross-efficiency matrix $\hat{e}_{pj}^L$*

p	w_p^1	w_p^2	w_p^3	w_p^4	w_p^5	w_p^6	w_p^7	w_p^8	w_p^9	w_p^{10}	w_p^{11}	w_p^{12}	w_p^{13}	w_p^{14}	w_p^{15}	w_p^{16}	w_p^{17}	w_p^{18}	w_p^{19}	w_p^{20}	w_p^{21}	w_p^{22}
1	0.015	0.014	0.015	0.014	0.015	0.092	0.014	0.012	0.012	0.015	0.014	0.012	0.016	0.016	0.014	0.015	0.015	0.017	0.016	0.016	0.014	0.073
2	0.029	0.029	0.029	0.029	0.029	0.026	0.027	0.024	0.024	0.029	0.029	0.024	0.032	0.032	0.027	0.029	0.030	0.033	0.031	0.032	0.028	0.024
3	0.029	0.029	0.029	0.029	0.029	0.039	0.027	0.098	0.095	0.029	0.029	0.095	0.032	0.032	0.027	0.029	0.030	0.033	0.031	0.032	0.111	0.110
4	0.029	0.029	0.029	0.029	0.029	0.039	0.027	0.098	0.024	0.029	0.114	0.024	0.032	0.032	0.027	0.029	0.030	0.033	0.031	0.032	0.028	0.024
5	0.029	0.029	0.029	0.029	0.029	0.039	0.027	0.024	0.036	0.029	0.029	0.036	0.032	0.032	0.027	0.029	0.030	0.033	0.031	0.032	0.042	0.037
6	0.118	0.114	0.118	0.143	0.118	0.039	0.108	0.049	0.036	0.118	0.043	0.036	0.032	0.032	0.108	0.132	0.121	0.033	0.047	0.032	0.042	0.037
7	0.029	0.029	0.029	0.029	0.029	0.039	0.027	0.049	0.036	0.029	0.043	0.036	0.032	0.032	0.027	0.029	0.030	0.033	0.031	0.032	0.042	0.049
8	0.044	0.043	0.044	0.043	0.044	0.145	0.041	0.049	0.036	0.044	0.043	0.036	0.032	0.032	0.041	0.044	0.045	0.033	0.141	0.129	0.042	0.049
9	0.044	0.043	0.044	0.043	0.044	0.026	0.041	0.049	0.036	0.044	0.043	0.036	0.032	0.032	0.041	0.044	0.045	0.033	0.031	0.032	0.042	0.049
10	0.044	0.043	0.044	0.043	0.044	0.039	0.041	0.049	0.119	0.044	0.043	0.119	0.129	0.129	0.041	0.044	0.045	0.033	0.047	0.048	0.125	0.122
11	0.044	0.043	0.044	0.043	0.044	0.039	0.041	0.122	0.095	0.044	0.129	0.095	0.032	0.032	0.041	0.044	0.045	0.033	0.047	0.048	0.028	0.024
12	0.044	0.043	0.044	0.043	0.044	0.039	0.041	0.024	0.036	0.044	0.029	0.036	0.048	0.048	0.041	0.044	0.045	0.033	0.047	0.048	0.042	0.037
13	0.088	0.086	0.088	0.043	0.088	0.039	0.108	0.037	0.048	0.088	0.043	0.048	0.048	0.048	0.108	0.059	0.061	0.100	0.047	0.048	0.042	0.037
14	0.029	0.029	0.029	0.043	0.029	0.039	0.027	0.037	0.048	0.029	0.043	0.048	0.048	0.048	0.027	0.029	0.030	0.033	0.047	0.048	0.042	0.037
15	0.088	0.086	0.044	0.114	0.044	0.039	0.081	0.037	0.048	0.044	0.043	0.048	0.048	0.048	0.081	0.118	0.106	0.100	0.063	0.065	0.042	0.037
16	0.029	0.029	0.044	0.029	0.044	0.039	0.041	0.037	0.048	0.044	0.043	0.048	0.048	0.048	0.041	0.029	0.030	0.033	0.031	0.032	0.042	0.037
17	0.044	0.043	0.059	0.043	0.059	0.079	0.054	0.037	0.048	0.059	0.043	0.048	0.081	0.081	0.054	0.044	0.045	0.067	0.094	0.097	0.083	0.073
18	0.059	0.057	0.029	0.043	0.029	0.026	0.054	0.049	0.060	0.029	0.057	0.060	0.048	0.048	0.054	0.044	0.045	0.067	0.031	0.032	0.028	0.024
19	0.074	0.086	0.074	0.057	0.074	0.039	0.081	0.024	0.024	0.074	0.029	0.024	0.048	0.048	0.081	0.059	0.061	0.100	0.047	0.048	0.042	0.037
20	0.029	0.029	0.044	0.043	0.044	0.039	0.027	0.049	0.048	0.044	0.057	0.048	0.081	0.081	0.027	0.029	0.030	0.033	0.047	0.048	0.042	0.037
21	0.044	0.043	0.044	0.057	0.044	0.039	0.041	0.024	0.024	0.044	0.029	0.024	0.032	0.032	0.041	0.059	0.061	0.050	0.047	0.048	0.042	0.037
22	0.015	0.029	0.044	0.014	0.044	0.013	0.027	0.024	0.024	0.044	0.029	0.024	0.032	0.032	0.027	0.015	0.015	0.033	0.016	0.016	0.014	0.012

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/maman/advance-article/doi/10.1093/maman/dpae012/7671391 by Delhi Technological University user on 22 May

TABLE 9 *The aggregated minimum cross-efficiency score by simple average operator $\bar{e}_j^L$ and by OVGA operator $\tilde{e}_j^L$ of target DMU_j*

j	1	2	3	4	5	6	7	8	9	10	11
$\bar{e}_j^L$	0.75	0.61	0.73	0.53	0.6	0.66	0.76	0.63	0.58	0.83	0.75
$\tilde{e}_j^L$	0.76	0.62	0.74	0.54	0.60	0.69	0.76	0.66	0.60	0.82	0.77
j	12	13	14	15	16	17	18	19	20	21	22
$\bar{e}_j^L$	0.66	0.77	0.87	0.83	0.83	0.68	0.89	0.71	0.84	0.76	0.7
$\tilde{e}_j^L$	0.68	0.77	0.87	0.84	0.85	0.70	0.88	0.72	0.85	0.80	0.76

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

TABLE 10 *The optimal portfolio strategies corresponding to Model IV for equal weights*

	$\gamma = 0.1$			$\gamma = 0.2$		
DMU_k	$\varepsilon = \delta = 0.2$ d=5	$\varepsilon = \delta = 0.1$ d=10	$\varepsilon = \delta =$ 0.1429 d=7	$\varepsilon = \delta = 0.2$ d=5	$\varepsilon = \delta = 0.1$ d=10	$\varepsilon = \delta =$ 0.1429 d=7
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0.10	0
5	0	0	0	0	0	0
6	0	0	0	0	0.10	0.1429
7	0	0	0	0	0	0
8	0	0.10	0.1429	0.20	0.10	0.1429
9	0	0	0	0.20	0.10	0.1429
10	0	0	0	0	0	0
11	0.20	0.10	0.1429	0.20	0.10	0.1429
12	0	0	0	0	0.10	0.1429
13	0.20	0.10	0	0	0	0
14	0.20	0.10	0.1429	0	0.10	0
15	0	0.10	0.1429	0	0	0
16	0	0	0	0	0	0
17	0	0	0	0	0	0
18	0	0.10	0	0	0	0
19	0	0.10	0	0	0.10	0
20	0	0.10	0.1429	0	0	0
21	0.20	0.10	0.1429	0.20	0.10	0.1429
22	0.20	0.10	0.1429	0.20	0.10	0.1429
$\overline{E}_\Omega$	0.794	0.792	0.793	0.718	0.709	0.709
$\overline{V}_\Omega$	0.0064	0.0094	0.0078	0.0046	0.0063	0.0052

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May

TABLE 11 *The optimal portfolio strategies corresponding to Model V for equal weights*

	$\gamma = 0.1$			$\gamma = 0.2$		
DMU_k	$\varepsilon = \delta = 0.2, d = 5$	$\varepsilon = \delta = 0.1, d = 10$	$\varepsilon = \delta = 0.0667, d = 15$	$\varepsilon = \delta = 0.2, d = 5$	$\varepsilon = \delta = 0.1, d = 10$	$\varepsilon = \delta = 0.0667, d = 15$
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0.10	0.1429
7	0	0	0	0	0	0
8	0	0	0	0.20	0.10	0.1429
9	0	0	0	0	0.10	0
10	0.20	0.10	0.1429	0	0	0
11	0.20	0.10	0.1429	0.20	0.10	0.1429
12	0	0	0	0	0.10	0.1429
13	0	0.10	0	0.20	0.20	0
14	0.20	0.10	0.1429	0	0.10	0.1429
15	0	0.10	0	0	0	0
16	0	0.10	0	0	0	0
17	0	0	0	0	0	0
18	0	0.10	0.1429	0	0	0
19	0	0	0	0	0.10	0
20	0	0.10	0.1429	0	0	0
21	0.20	0.10	0.1429	0.20	0.10	0.1429
22	0.20	0.10	0.1429	0.20	0.10	0.1429
$\overline{E}_\Omega$	0.808	0.807	0.805	0.721	0.709	0.718
$\overline{V}_\Omega$	0.0096	0.0116	0.0102	0.0043	0.0063	0.0052

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May

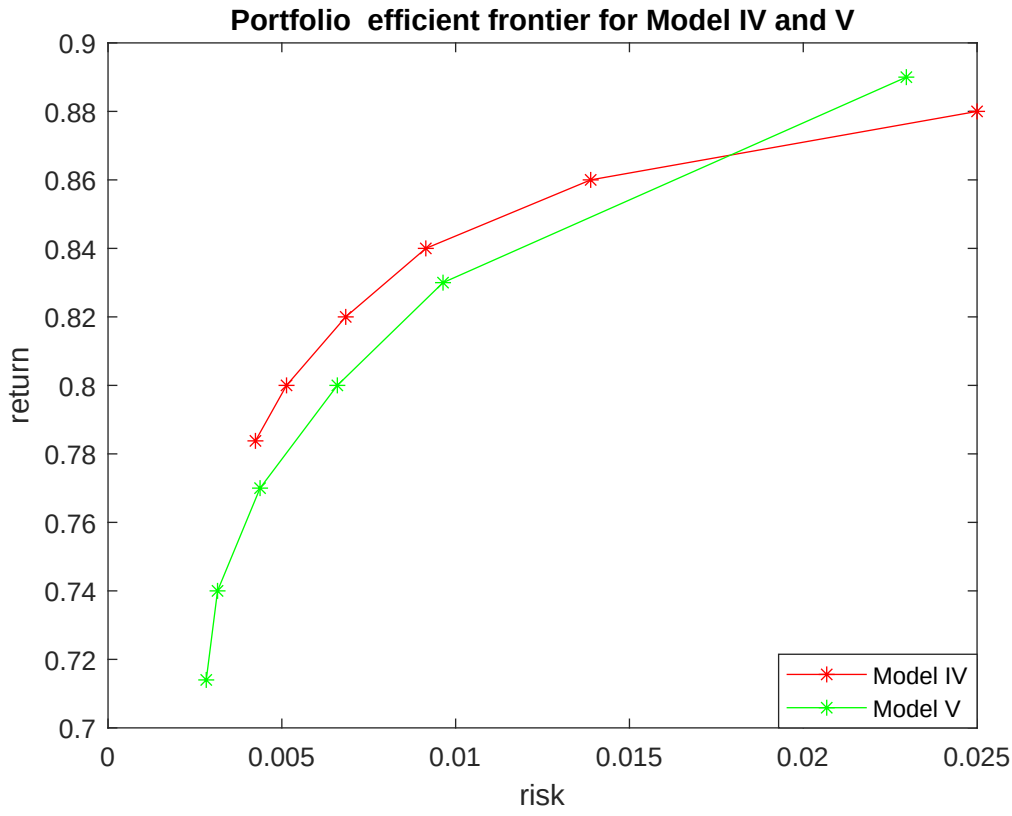


FIG. 3. Efficient frontier of Model IV and Model V

Remark 5 *For risk-averse investors, Fig.3 depicts that the return produced from the suggested model is higher than the standard model already present in the literature.*

UNCORRECTED PROOF

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

TABLE 12 *The optimal portfolio strategies corresponding to Model IV for unequal weights*

	$\gamma = 0.1,\ \varepsilon = 0.03,\ \delta = 0.4$			$\gamma = 0.2,\ \varepsilon = 0.02,\ \delta = 0.3$		
DMU_k	d=5	d=10	d=7	$\varepsilon = \delta = 0.2$ d=5	$\varepsilon = \delta = 0.1$ d=10	$\varepsilon = \delta =$ 0.1429 d=7
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0.0529	0.02
4	0	0	0	0	0	0
5	0	0	0	0	0.037	0
6	0.03	0.03	0	0	0	0.02
7	0	0	0	0	0.02	0
8	0	0	0.03	0.2101	0.0316	0.1343
9	0	0	0	0.0282	0.09	0.1149
10	0	0.03	0	0	0	0
11	0.40	0.2886	0.40	0.2981	0.2819	0.30
12	0	0.03	0	0	0.021	0
13	0	0.03	0	0	0.047	0
14	0.1282	0.1423	0.0685	0	0	0
15	0	0	0	0	0	0
16	0	0.03	0.03	0	0	0
17	0	0	0	0	0	0
18	0	0	0.0397	0	0	0
19	0	0.03	0	0	0	0
20	0	0	0.	0	0	0
21	0.400	0.3591	0.400	0.1635	0.1093	0.1108
22	0.0418	0.03	0.318	0.30	0.2996	0.30
$\overline{E}_\Omega$	0.792	0.792	0.792	0.744	0.739	0.734
$\overline{V}_\Omega$	0.0050	0.0061	0.0052	.0045	0.0050	0.0046

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May

TABLE 13 *The optimal portfolios strategies corresponding to Model V for unequal weights*

	$\gamma = 0.1,\ \varepsilon = 0.03,\ \delta = 0.4$			$\gamma = 0.2,\ \varepsilon = 0.02,\ \delta = 0.3$		
DMU_k	d=5	d=10	d=7	d=5	d=10	d=7
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0.0204	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0.1924	0.0248	0.1809
7	0	0	0	0	0	0
8	0	0.03	0	0.1220	0.1479	0.1177
9	0	0	0	0	0.02	0
10	0	0.03	0.03	0	0.036	0.02
11	0.1569	0.1077	0.2052	0.30	0.2966	0.2977
12	0	0	0	0	0.0248	0
13	0	0.03	0	0	0	0.028
14	0.3583	0.3565	0.3253	0.0857	0	0
15	0	0.03	0	0	0	0
16	0	0.03	0	0	0	0
17	0	0	0	0	0	0
18	0	0	0.03	0	0	0
19	0	0	0	0	0.02	0
20	0.03	0.03	0.0509	0	0	0
21	0.3959	0.3250	0.3285	0	0.1093	
22	0.0319	0.03	0.03	0.30	0.30	0.30
$\overline{E}_\Omega$	0.801	0.801	0.801	0.712	0.712	0.712
$\overline{V}_\Omega$	0.0069	0.0080	0.0073	0.0042	0.0037	0.0043

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May

In the subsequent section to follows, we demonstrate the benefit of the proposed method by taking a real-world application problem.

4.1. *A Case Study of the Indian Banking Sector*

We present a case study with 20 banking companies listed on the National Stock Exchange as a practical example of the suggested methodology. The data for the financial year 2021 is used here for the case study. This is just for illustrative purposes; the analysis may easily be continued for several years.

In our analysis, we have randomly selected the following twenty Banks listed in trade brains portal (https://portal.tradebrains.in).

1. SBIN (B_1)
2. AXIS (B_2)
3. PNB (B_3)
4. HDFC (B_4)
5. ICICI (B_5)
6. KOTAK (B_6)
7. FEDERAL (B_7)
8. CANARA (B_8)
9. UNION (B_9)
10. UNCO (B_{10})
11. BARODA (B_{11})
12. INDIAN (B_{12})
13. BANDHAN (B_{13})
14. MAHA (B_{14})
15. IDBI (B_{15})
16. INDIAN (B_{16})
17. CUB (B_{17})
18. DCB (B_{18})
19. DHAN (B_{19})
20. SOUTH (B_{20})

We denote the Banks by $B_k,\ k = 1,2,\ldots,20$, in order of the above listing. The description of input and output variables considered here are as follows:

Inputs:

1. Input 1: Ratio of Market price to the earning per share (I_1)
2. Input 2: Ratio of the Market price to the book value per share (I_2)

Outputs

1. Output 1: Ratio of net income to stockholders equity (O_1)
2. Output 2: percentage of a firmâĂŹs ability to generate a return from its assets (O_2)

Table 14 presents the input and output values of the 20 banking stocks and their self-evaluated DEA efficiency. Further, from Definition 7 and Definition 8, the minimum OVGWA cross-efficiency $\tilde{e}_j^L$ and the minimum average cross-efficiency $\bar{e}_j^L$ for Bank $B_j,\ j = 1,2,\ldots,20$ are also presented there. Similarly, from Definitions 9, 10 and 11, the maximum OVGWA cross-efficiency and maximum mean cross-efficiency $\tilde{e}_j^U$ and $\bar{e}_j^U$ of respective Bank's are also provided therein. We set $\varepsilon_i = 0.02$, $\delta_i = 0.6$, to

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

create portfolios with unequal weights. When we solve the proposed models IV and V for a range of d and gamma values, we found that a small increase in portfolio return (from 0.72 to 0.75) would result in a sizable increase in risk (from 0.0128 to 0.0145). The results are shown in Table15

4.2. *Managerial implications:*

The proposed two-stage methodology has significant managerial implications in the field of DEA cross-efficiency aggregation and portfolio selection in terms of their efficiency. First, the developed approach can help managers to gain a better understanding of portfolio selection. Second, the approach's results are simple to grasp for managers. It can help investors and portfolio managers to build an efficient portfolio by incorporating only the best-performing stocks. Based on the implications discussed above, the proposed approach can be used in a financial market but also in other industries to examine the impact of operational and long-term decisions on business performance. Furthermore, the problem addressed in this study is in its early stages of investigation, and further research can be conducted based on the findings of this paper. Some options are as follows:

1. A similar study can be conducted to deal with stochastic data.
2. In this study, the proposed model was applied to a problem in the banking industry. However, the same model could be used for a variety of other selection-based problems like R& D and project selection, and the stock market. The methodology applies to other asset allocation problems, mutual fund portfolio selection, multi-period portfolio problems, and other case studies.
3. Furthermore, comparative studies in different industries or countries using the model and methods from this study would be necessary to validate its claims.
4. The OVGA aggregated efficiencies can be used in the different ranking problems and the obtained results can be further compared with other aggregation methods in terms of complexity.

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

TABLE 14 *Detailed description of the Banks*

B_k	(I_1)	(I_2)	(O_1)	(O_2)	$e^{\star}_{kk}$	$\bar{e}^{L}_{k}$	$\bar{e}^{U}_{k}$	$\tilde{e}^{L}_{k}$	$\tilde{e}^{U}_{k}$
B_1	14.1	1.29	10.13	0.54	0.3092	0.2361	0.2561	0.2489	0.2616
B_2	29.7	2.06	7.64	0.75	0.1877	0.1442	0.1568	0.1473	0.1610
B_3	14.99	0.45	2.98	0.2	0.2292	0.1569	0.1734	0.1533	0.1684
B_4	25.86	3.92	16.5	1.88	0.4737	0.2660	0.3112	0.2626	0.2909
B_5	21.87	2.6	14.85	1.38	0.4111	0.2732	0.3081	0.2799	0.2985
B_6	34.94	4.12	13.12	2.15	0.4009	0.2305	0.2704	0.2198	0.2457
B_7	9.09	0.92	10.52	0.85	0.6093	0.4619	0.5069	0.4849	0.5038
B_8	8.67	0.46	6.07	0.28	0.4494	0.3032	0.3361	0.3079	0.3238
B_9	7.62	0.37	6.25	0.35	0.5753	0.4185	0.4551	0.4008	0.4888
B_{10}	65.26	0.63	0.98	0.07	0.0573	0.0310	0.0367	0.0288	0.0374
B_{11}	24.78	0.5	1.98	0.12	0.1349	0.0822	0.0925	0.0702	0.0816
B_{12}	10.68	0.55	5.54	0.32	0.343	0.2585	0.2795	0.2536	0.2658
B_{13}	24.74	3.13	13.53	2.13	0.561	0.3157	0.3713	0.3059	0.3543
B_{14}	23.88	1.23	5.5	0.31	0.1523	0.1133	0.1228	0.1110	0.1165
B_{15}	27.38	1.32	4.85	0.48	0.1875	0.1251	0.1372	0.1320	0.1391
B_{16}	4.17	0.33	9.69	0.64	1	0.9511	1.0000	0.9592	1.0000
B_{17}	19.43	1.97	10.64	1.15	0.3856	0.2651	0.2980	0.2505	0.3163
B_{18}	9.5	0.9	10.01	0.86	0.5898	0.4537	0.4972	0.4572	0.4967
B_{19}	10.09	0.52	5.26	0.29	0.3445	0.2531	0.2751	0.2482	0.2618
B_{20}	28.02	0.32	1.16	0.06	0.1235	0.0642	0.0758	0.0520	0.0639

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May

TABLE 15 *The optimal portfolio strategies in the case study of Banks Corresponding to Model IV and Model V*

	$\gamma = 0.25,\ \varepsilon = 0.02,\ \delta = 0.6,\ d = 5$		$\gamma = 0.2,\ \varepsilon = 0.02,\ \delta = 0.75, d = 7$		$\gamma = 0.3, \varepsilon = 0.01\ \delta = 0.5, d = 10$	
B_k	Model IV	Model V	Model IV	Model V	Model IV	Model V
B_1	0	0	0	0	0.01	0
B_2	0	0	0.02	0.02	0.01	0
B_3	0	0	0	0	0.01	0
B_4	0	0	0	0	0	0
B_5	0	0	0	0	0.01	0.01
B_6	0	0	0	0	0	0.
B_7	0.1335	0.313	0	0	0.0192	0.3068
B_8	0	0.02	0	0	0	0.01
B_9	0.0712	0.0469	0.02	0.1180	0.1572	0.1232
B_{10}	0	0	0	0	0	0
B_{11}	0	0	0	0.	0	0
B_{12}	0.1752	0	0.957	0.0296	0.0914	0.01
B_{13}	0	0	0	0	0	0.01
B_{14}	0	0	0.02	0	0	0
B_{15}	0	0	0.0802	0.02	0.01	0
B_{16}	0.60	0.60	0.7441	0.75	0.5	00.5
B_{17}	0	0	0	0	0	0.01
B_{18}	0	0.02	0	0.0424	0	0.01
B_{19}	0.02	0	0.02	0.02	0.01	0.01
B_{20}	0	0	0	0	0	0
$\overline{E}_\Omega$	0.72	0.75	0.764	0.80	0.669	0.699
$\overline{V}_\Omega$	0.0128	0.0141	0.0145	0.0160	0.0112	0.0127

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May

Algorithm 1 Algorithm 1

```
import pandas as pd
from fractions import Fraction
function FUNC(i, j, ls)
    temp ← 1
    for k in range(i + 1, j) do
        ok ← ls[k]
        rhs ← ls[j] + (ls[i] − ls[j]) · ((j−k)/(j−i))
        if ok ≥ rhs then
            return 0
        end if
    end for
    return 1
end function
adj ← [[−1 for _ in range(n)] for _ in range(n)]
for i in range(n) do
    for j in range(i, n) do
        if adj[i][j] ≠ −1 then
            print("asjndk", str(i), " : ", str(j))
            continue
        end if
        if i == j then
            adj[i][j] ← 0
            continue
        end if
        if i − j == 1 or j − i == 1 then
            adj[i][j] ← 1
            continue
        end if
        adj[i][j] ← func(i, j, ls)
    end for
end for
for j in range(n) do
    for i in range(j, n) do
        adj[i][j] ← adj[j][i]
    end for
end for
for i in range(n) do
    print(adj[i])
end for
```

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

5. Conclusion and Future Directions

In this study, we introduce a novel portfolio selection model that can assist investors in identifying well-diversified equities based on numerous evaluation criteria. The unique primary contributions of this paper can be seen in two ways. First, the use of OVGA in cross-efficiency aggregation may be viewed as an extension and refinement of the DEA cross-efficiency aggregation approaches. Meanwhile, this is the first time, that the OVGA aggregated cross-efficiencies have been successfully applied to portfolio management. It appears to be a promising tool for assessing financial assets. It is demonstrated herein that introducing the OVGA aggregation operator into the cross-efficiency portfolio selection yields superior portfolios with reduced variance and more significant or similar projected returns; at least for this particular study. To demonstrate the value of the suggested approach, a real-time example from the Indian banking industry is used. Further, the proposed strategy can also be completed by stock selection in larger stock markets.

It is worth noting that if asset return distributions are not symmetric, using variance as a risk measure is not recommended because doing so leads to portfolio predictions that are far from the actual target. Many researchers have utilized semi-variance and semi-absolute deviation as alternative risk metrics to qualify risk in such scenarios. Additionally, the OVGA cross-efficiency aggregation model may be extended to study portfolio selection problems in the fuzzy environment. Further, this study can be extended to study the efficiency of the decision-making units having undesirable inputs and outputs.

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical statement

This article does not contain any studies with human participants or animals performed by the author.

Funding

The second author was supported by Guru Gobind Singh Indraprastha University, New Delhi, India vide letter no. GGSIPU/RDC/FRGS/2023/1448/16.

Data availability statement

The data used for the numerical illustration in this study are included in this published article https://doi.org/10.1016/j.ejor.2013.12.002 The data from the Indian banking sector used for the case study are available on the Trade Brains portal (https://portal.tradebrains.in).

Acknowledgements

Authors would like to thank the Editor, Associative editor, and the anonymous reviewers for their valuable comments and detailed suggestions that have improved the presentation of this paper.

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

Appendix

The Python program that calculates the adjacency matrix, degree, and weight of an ordered set is described as Algorithm 1.

References

G. R. Amin and M. Hajjami. Improving DEA cross-efficiency optimization in portfolio selection. *Expert Systems with Applications*, 168:114280, 2021.

G. R. Amin and A. Oukil. Gangless cross-evaluation in dea: An application to stock selection. *RAIRO-Operations Research*, 53(2):645–655, 2019.

A. Amirteimoori, L. Khoshandam, S. Kordrostami, M. Jahani Sayyad Noveiri, and R. K. Matin. Performance analysis in a stochastic supply chain with reverse flows: a DEA-based approach. *IMA Journal of Management Mathematics*, 33(3):433–456, 2022.

R. D. Banker, A. Charnes, and W. W. Cooper. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, 30(9):1078–1092, 1984.

A. Charnes, W. W. Cooper, and E. Rhodes. Measuring the efficiency of decision making units. *European journal of operational research*, 2(6): 429–444, 1978.

H.-H. Chen. Stock selection using data envelopment analysis. *Industrial Management & Data Systems*, 108(9):1255–1268, 2008.

L. Chen, Y. Huang, M.-J. Li, and Y.-M. Wang. Meta-frontier analysis using cross-efficiency method for performance evaluation. *European Journal of Operational Research*, 280(1):219–229, 2020a.

L. Chen, Y.-M. Wang, and Y. Huang. Cross-efficiency aggregation method based on prospect consensus process. *Annals of Operations Research*, 288(1):115–135, 2020b.

W. Chen, Y. Gai, and P. Gupta. Efficiency evaluation of fuzzy portfolio in different risk measures via DEA. *Annals of Operations Research*, 269(1): 103–127, 2018.

W. Chen, S.-S. Li, M. K. Mehlawat, L. Jia, and A. Kumar. Portfolio selection using data envelopment analysis cross-efficiency evaluation with undesirable fuzzy inputs and outputs. *International Journal of Fuzzy Systems*, 23:1478–1509, 2021.

M. Davtalab-Olyaie and M. Asgharian. On pareto-optimality in the cross-efficiency evaluation. *European Journal of Operational Research*, 288(1): 247–257, 2021.

X. Deng and W. Fang. A novel mean-variance-maverick DEA prospect cross-efficiency approach for fuzzy portfolio selection. *Journal of Intelligent & Fuzzy Systems*, 37(6):8113–8130, 2019.

M. Dia. A portfolio selection methodology based on data envelopment analysis. *INFOR: Information Systems and Operational Research*, 47(1): 71–79, 2009.

J. Doyle and R. Green. Efficiency and cross-efficiency in DEA: Derivations, meanings and uses. *Journal of the operational research society*, 45(5): 567–578, 1994.

A. Emrouznejad, M. Marra, G.-l. Yang, and M. Michali. Eco-efficiency considering netzero and data envelopment analysis: a critical literature review. *IMA journal of management mathematics*, page dpad002, 2023.

H. Essid, J. Ganouati, and S. Vigeant. A mean-maverick game cross-efficiency approach to portfolio selection: An application to paris stock exchange. *Expert Systems with Applications*, 113:161–185, 2018.

L. Fang and J. Yang. An integrated ranking approach using cross-efficiency intervals and the cumulative prospect theory. *Computers & Industrial Engineering*, 136:556–574, 2019.

M. J. Farrell. The measurement of productive efficiency. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 120(3):253–281, 1957.

M. Ghiyasi and N. Zhu. An inverse semi-oriented radial data envelopment analysis measure for dealing with negative data. *IMA Journal of Management Mathematics*, 31(4):505–516, 2020.

X. Gong, C. Yu, L. Min, and Z. Ge. Regret theory-based fuzzy multi-objective portfolio selection model involving DEA cross-efficiency and higher moments. *Applied Soft Computing*, 100:106958, 2021.

S. Guo and W.-K. Ching. High-order markov-switching portfolio selection with capital gain tax. *Expert Systems with Applications*, 165:113915, 2021.

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

P. Gupta, M. K. Mehlawat, A. Kumar, S. Yadav, and A. Aggarwal. A credibilistic fuzzy DEA approach for portfolio efficiency evaluation and rebalancing toward benchmark portfolios using positive and negative returns. *International Journal of Fuzzy Systems*, 22(3):824–843, 2020.

W. Jiang, B. Wei, J. Zhan, C. Xie, and D. Zhou. A visibility graph power averaging aggregation operator: A methodology based on network analysis. *Computers & Industrial Engineering*, 101:260–268, 2016.

W. Jiang, B. Wei, Y. Tang, and D. Zhou. Ordered visibility graph average aggregation operator: An application in produced water management. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(2):023117, 2017.

T. León, N. Ramón, J. L. Ruiz, and I. Sirvent. Using induced ordered weighted averaging (OWA) operators for aggregation in cross-efficiency evaluations. *International Journal of Intelligent Systems*, 29(12):1100–1116, 2014.

F. Li, Q. Zhu, Z. Chen, and H. Xue. A balanced data envelopment analysis cross-efficiency evaluation approach. *Expert Systems with Applications*, 106:154–168, 2018.

M.-J. Li, J.-C. Lu, and L. Chen. A method to determine the integrated weights of cross-efficiency aggregation. *Soft Computing*, pages 1–13, 2022.

S. Lim, K. W. Oh, and J. Zhu. Use of DEA cross-efficiency evaluation in portfolio selection: An application to korean stock market. *European Journal of Operational Research*, 236(1):361–368, 2014.

D. Liu and Q. Chen. A regret cross-efficiency ranking method considering consensus consistency. *Expert Systems with Applications*, 208:118192, 2022.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

H. Markowitz. Portfolio selection: Efficient diversification of investments, wiley, New York. 1959.

Z. Mashayekhi and H. Omrani. An integrated multi-objective markowitz-DEA cross-efficiency model with fuzzy returns for portfolio selection problem. *Applied Soft Computing*, 38:1–9, 2016.

M. K. Mehlawat, A. Kumar, S. Yadav, and W. Chen. Data envelopment analysis based fuzzy multi-objective portfolio selection model involving higher moments. *Information sciences*, 460:128–150, 2018.

A. Omid, A. Azar, and M. Taleb. Assessing environmental and operational efficiencies: a multi-objective optimisation problem in a two-stage network data envelopment analysis. *IMA Journal of Management Mathematics*, page dpad025, 2023.

M. Oral, G. R. Amin, and A. Oukil. Cross-efficiency in dea: A maximum resonated appreciative model. *Measurement*, 63:159–167, 2015.

A. Oukil. Embedding OWA under preference ranking for DEA cross-efficiency aggregation: Issues and procedures. *International Journal of Intelligent Systems*, 34(5):947–965, 2019.

A. Oukil. Exploiting value system multiplicity and preference voting for robust ranking. *Omega*, 94:102048, 2020.

A. Oukil and G. R. Amin. Maximum appreciative cross-efficiency in dea: A new ranking method. *Computers & Industrial Engineering*, 81:14–21, 2015.

M. Soltanifar, M. Ghiyasi, and H. Sharafi. Inverse dea-r models for merger analysis with negative data. *IMA Journal of Management Mathematics*, 34(3):491–510, 2023.

M. Song, Q. Zhu, J. Peng, and E. D. S. Gonzalez. Improving the evaluation of cross efficiencies: A method based on shannon entropy weight. *Computers & Industrial Engineering*, 112:99–106, 2017.

D. Wang, F. Tian, and D. Wei. An improved ordered visibility graph aggregation operator for madm. *IEEE Access*, 10:93464–93474, 2022.

H. Wang, H. Mo, R. Sadiq, Y. Hu, and Y. Deng. Ordered visibility graph weighted averaging aggregation operator: A methodology based on network analysis. *Computers & Industrial Engineering*, 88:181–190, 2015.

Y.-M. Wang and K.-S. Chin. The use of OWA operator weights for cross-efficiency aggregation. *Omega*, 39(5):493–503, 2011.

J. Wu, J. Sun, and L. Liang. DEA cross-efficiency aggregation method based upon shannon entropy. *International Journal of Production Research*, 50(23):6726–6736, 2012.

J. Wu, J. Sun, and L. Liang. Methods and applications of DEA cross-efficiency: Review and future perspectives. *Frontiers of Engineering Management*, 8(2):199–211, 2021.

Z. Zhou, H. Xiao, Q. Jin, and W. Liu. DEA frontier improvement and portfolio rebalancing: An application of china mutual funds on considering sustainability information disclosure. *European Journal of Operational Research*, 269(1):111–131, 2018.

C. Zhu and N. Zhu. Assessing the eco-efficiency of industrial investment in china: a dea approach. *IMA Journal of Management Mathematics*, 34 (1):143–163, 2023.

UNCORRECTED MANUSCRIPT

Downloaded from https://academic.oup.com/imaman/advance-article/doi/10.1093/imaman/dpae012/7671391 by Delhi Technological University user on 22 May 2024

Cyclic Heap Permutation Based Reconfiguration of Solar PV Array for Optimal Shade Dispersion

Viraj Jain
Department of Electrical Engineering
Delhi Technological University
Delhi, India
viraj.jain@hotmail.com

Vikas Kumar
Department of Electrical Engineering
Delhi Technological University
Delhi, India
vikasrajput1306@gmail.com

Shlok Garg
Department of Electrical Engineering
Delhi Technological University
Delhi, India
2003skgarg@gmail.com

Vinod Kumar Yadav
Department of Electrical Engineering
Delhi Technological University
Delhi, India
vinodkumar@dtu.ac.in

2024 11th International Conference on Signal Processing and Integrated Networks (SPIN) | 979-8-3503-0843-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/SPIN60856.2024.10511429

***Abstract*—Partial shading poses a critical challenge to optimizing the performance of photovoltaic (PV) systems, affecting the attainment of maximum power output and overall system efficiency. Inadequate management of partial shading conditions (PSCs) can threaten the stability and effectiveness of PV systems. This study presents an innovative reconfiguration approach grounded in a cyclic heap permutation strategy, to alleviate the detrimental effects of Partial Shading Conditions (PSCs). Through the strategic redistribution of shading effects, the proposed reconfiguration method improves the efficiency of PV systems across a spectrum of shading scenarios. Rigorous performance evaluations, considering various shading cases and configurations, reveal the efficacy of the proposed cyclic heap permutation-based reconfiguration technique. We also highlight its versatility and adaptability across different PV array sizes and shading conditions. This study adds to continuous endeavors aimed at fortifying the resilience and improving the efficiency of solar PV systems confronted with partial shading challenges.**

***Index Terms*—solar, partial shading, photovoltaic, reconfiguration, shade dispersion**

I. Introduction

In response to the detrimental environmental impacts linked with fossil fuels, there is a growing global demand for energy sources that are both environmentally friendly and sustainable [1]. This demand is further heightened by the expansion of the world's population and the depletion of fossil fuel reserves, creating an urgent requirement for continuous and sustainable energy production. Solar power, among diverse renewable energy sources, has emerged as a viable solution to meet the rising demand for electricity.

While solar energy is acknowledged for its sustainability and cost-effectiveness, a significant challenge lies in the decrease in output resulting from variations in irradiation. These variations lead to non-uniformities and multiple peaks in the power-voltage (P-V) and current-voltage characteristics (I-V) of photovoltaic systems.

Partial shading (PS) manifests when distinct sections of a solar PV array experience disparate solar irradiance levels, often attributed to factors like cloud cover, the movement of birds, and the accumulation of dirt resulting from inadequate maintenance, among other causes [2]. The ramifications and challenges associated with partial shading conditions (PSCs) encompass several critical issues [3] [4]. Firstly, hotspot formation on solar PV modules can lead to thermal losses, potentially disrupting the overall functionality of the PV system [5]–[7]. Moreover, PSCs contribute to the ineffectiveness of conventional maximum power tracking methods, hindering the identification of the global maximum and leading to multiple peaks in the power-voltage (P-V) plot [8] [9]. This, in turn, hinders MPPT algorithms from tracking the global maxima, resulting in suboptimal power output and diminished system performance [10]. Considering the imperative to capture the maximum power output under different Photovoltaic System Conditions (PSCs), PV array reconfiguration emerges as a compelling solution. This process can be categorized into static reconfiguration, where the electrical interconnect remains unchanged, and dynamic reconfiguration, which involves alterations in the electrical interconnect [11] [12]. While the dynamic approach exhibits strong optimization capabilities, it comes with a significant cost and necessitates complex algorithms to operate at the optimum point. In contrast, the static approach necessitates a single configuration to achieve efficient shading dispersion throughout the PV array.

In tackling the difficulties presented by partial shading in photovoltaic (PV) arrays, prior research has proposed different configurations, including Series Parallel (SP) configuration, Total Cross Tied (TCT) configuration, Honey Comb (HC) configuration, and Bridge Link (BL) configuration [13] [14]. Moreover, various methods for maximizing Power Point Tracking (MPPT) have been investigated to enhance the output of solar arrays in the context of partial shading conditions [12].

This research introduces an innovative array reconfiguration technique named Cyclic Heap Permutation (CHP) as a novel solution to alleviate shading impacts on PV arrays. The performance of various PV array configurations, such as SP, BL, TCT, and HC, is assessed using a 5x5 PV array. The research

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:57:23 UTC from IEEE Xplore. Restrictions apply.

meticulously examines the outcomes across diverse partial shading patterns through Simulink (MATLAB) simulations. The noteworthy advancements of this work are outlined as follows:

1) This work addresses the optimization challenge under partial shading conditions (PSCs). Introducing a novel cyclic heap permutation-based reconfiguration technique, the study endeavors to enhance the effectiveness of PV systems in the presence of diverse shading scenarios.
2) The outcomes of this research validate the efficacy of the proposed cyclic heap permutation-based reconfiguration technique and underscore its versatility and adaptability across various PV array sizes and shading conditions. This highlights the applicability of the proposed method in various real-world scenarios.
3) The research adds to continuous endeavors aimed at fortifying the resilience and efficiency of solar PV systems confronted with partial shading challenges. The cyclic heap permutation strategy proves to be a compelling solution for mitigating shading impacts on PV arrays, thereby improving overall system performance.

II. Modeling of PV Array

Understanding the operational characteristics of a photovoltaic cell is pivotal, and its accurate modeling is the cornerstone for achieving this comprehension. The typical design of the photovoltaic cell is illustrated in the equivalent circuit shown in Fig. 1, and this representation extends to a photovoltaic (PV) module.

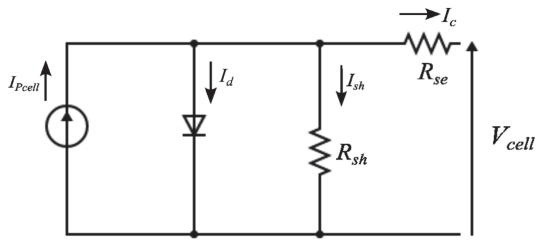


Fig. 1: Equivalent model of a Photovoltaic cell

$$I_c = I_{\text{cell}} - I_d \left[\exp\left(\frac{e(V_{\text{cell}} + R_{\text{se}} I_c)}{KBT} \right) - 1 \right] - \left[\frac{V_{\text{cell}} + R_{\text{se}} I_c}{R_{\text{sh}}} \right] \quad (1)$$

The functional characteristics of the photovoltaic (PV) cell are expressed through the mathematical formulation in Equation (1), where I_c represents the output current. I_{cell} denotes the photoelectric current generated by incoming photons, V_{cell} is the output voltage, R_{se} and R_{sh} are the series and parallel resistances across the diode, I_d is the saturation current, K is Boltzmann's constant, B is the ideality factor, and T is the module temperature in kelvin. The term e represents the fundamental charge of an electron. This equation encapsulates the intricate relationship among these parameters, defining the current-voltage characteristics of the PV cell.

III. Proposed Reconfiguration Algorithm: Cyclic Heap Permutation (CHP)

In this section, we introduce a novel reconfiguration algorithm, termed Cyclic Heap Permutation (CHP), designed for efficient and intricate matrix shuffling. The CHP algorithm draws inspiration from cyclic permutations, matrix transposition, Rubik's Cube moves, and Heap's Algorithm-based row permutations as shown in the flowchart in Fig. 3. The description of each operation performed is given below:

1) Cyclic Permutation
The foundation of the CHP algorithm lies in the cyclic permutation of elements within each row of the matrix. This step ensures the dispersion of elements originally situated in the same row across various rows in the shuffled matrix. The cyclic permutation introduces an initial level of complexity and serves as the cornerstone for subsequent transformations.
For a given matrix M of size $n \times n$, the cyclic permutation P_c is defined by:

$$P_c(p, q) = M[(p + q) \mod n][q] \quad (2)$$

where:
- $P_c(p, q)$ is the element at position (p, q) in the matrix after cyclic permutation.
- M is the original matrix.

2) Matrix Transpose
Following the cyclic permutation, the matrix undergoes a transpose operation. The transpose swaps rows and columns, further diversifying the arrangement of elements. This operation contributes to the overall intricacy of the shuffling process, promoting a more varied and layered transformation.
For a given matrix M of size $m \times n$, the transpose M^T is obtained by swapping its rows and columns:

$$M^T_{ij} = M_{ji} \quad (3)$$

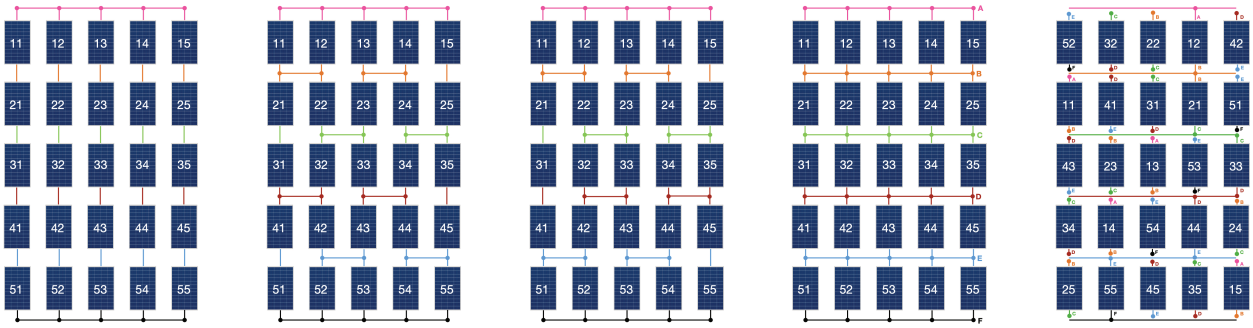
3) Rubik's Cube Move
Inspired by the movements of a Rubik's Cube, the algorithm incorporates a Rubik's Cube move to swap specific elements within each row. This step introduces a puzzle-solving aspect to the reconfiguration, creating a distinctive and dynamic arrangement of matrix elements. The Rubik's Cube move $RCM(p, q, r)$ is defined as swapping the elements at positions (p, q) and (p, r) within each row:

$$RCM(p, q, r) : M_{pq} \leftrightarrow M_{pr} \quad (4)$$

where, M_{pq} is the element at position (p, q) in the matrix, and M_{pr} is the element at position (p, r)in the matrix.

4) Heap's Algorithm for Row Permutation
To enhance the complexity of the shuffled matrix, the CHP algorithm leverages Heap's Algorithm to generate permutations of row indices. A selected permutation

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:57:23 UTC from IEEE Xplore. Restrictions apply.



(a) SP (b) BL (c) HC (d) TCT (e) CHP

Fig. 2: PV array configurations

is applied to reorder the rows, providing additional variability to the final configuration.

Heap's Algorithm generates permutations recursively:

$$\begin{aligned} &\pi(i) \leftrightarrow \pi(n) \\ \text{if } i < n: \quad &\text{swap}(\pi(i), \pi(n)); \\ &\text{generate}(i+1); \\ &\text{swap}(\pi(i), \pi(n)) \end{aligned} \tag{5}$$

The permutation π is then applied to reorder the rows of the matrix.

The Cyclic Heap Permutation algorithm represents a holistic approach to matrix reconfiguration, incorporating diverse mathematical concepts and puzzle-solving strategies. The synergy of cyclic permutation, matrix transpose, Rubik's Cube moves, and Heap's Algorithm-based row permutations results in a versatile and efficient method for achieving complex matrix shuffling.

IV. Result and Discussion

To evaluate and compare the effectiveness of Cyclic Heap Permutation (CHP) in conjunction with other photovoltaic (PV) configurations, we conducted a comprehensive analysis using a 250W PV module from the SunPower Solar brand, characterized by an open circuit voltage V_{oc} of 50.93 V and short circuit current I_{sc} of 6.2 A, under diverse Partial Shading Conditions (PSCs). In photovoltaic (PV) arrays, various configurations are employed to optimize energy production and system reliability.

The series-parallel configuration entails linking solar panels in both series and parallel to optimize power output by balancing voltage and current. The bridge-linked configuration utilizes bridges to interconnect panels, enhancing overall system robustness and minimizing energy losses. The honeycomb configuration arranges solar modules in a hexagonal pattern, promoting the uniform distribution of sunlight and improving overall efficiency. The total cross-tied configuration connects every solar panel in the array to multiple inverters, providing redundancy and reducing the impact of shading or individual panel failures on the entire system. The configurations are shown in Fig. 2 alongside the CHP-reconfigured array.

The assessment covered four shading scenarios, each characterized by unique irradiance levels ranging from 250 W/m² to 1000 W/m². For a thorough assessment, the analysis incorporated four key metrics: quantifying Power Loss (PL), assessing Mismatch Loss (ML), identifying the Global Maximum Power Point (GMPP), and evaluating the Fill Factor (FF).

The simulations were meticulously executed using the MATLAB/Simulink framework, incorporating diverse insolation limitations while upholding a consistent temperature of 25°C. This rigorous evaluation offers valuable insights into the relative performance of CHP compared to other PV configurations under realistic shading scenarios.

A. Analysis of Shade Case 1

In the initial shading scenario, referred to as Shade Case 1 (SC1) or the short and narrow shading pattern, the 5×5 solar photovoltaic array experiences varying levels of solar irradiation. This variation is visually depicted in Fig. 4, showcasing the reconfigured PV array using the proposed Cyclic Heap Permutation (CHP) and illustrating the resultant dispersion of shading. The simulation results are presented in the P-V curve in Fig. 5, where CHP attains a peak power of 5512 watts in this shading scenario, surpassing the performance of all other configurations discussed. This exceeds the average peak power of the other four configurations by 10.3%. The PV array reconfigured using CHP demonstrates superior performance compared to alternative reconfiguration techniques in effectively mitigating the adverse effects of shading.

B. Analysis of Shade Case 2

In the second shading scenario (Shade Case 2), characterized by the short and wide shading pattern, the 5×5 solar PV array is exposed to varying solar irradiation levels. Fig. 6 illustrates the short wide shading pattern, the CHP-reconfigured PV array, and provides a visual representation of the dispersion of shading. In the P-V curve shown in Fig. 7, CHP outperforms other configurations, achieving a peak power of 4809 watts. On average, CHP achieved 30% higher

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:57:23 UTC from IEEE Xplore. Restrictions apply.

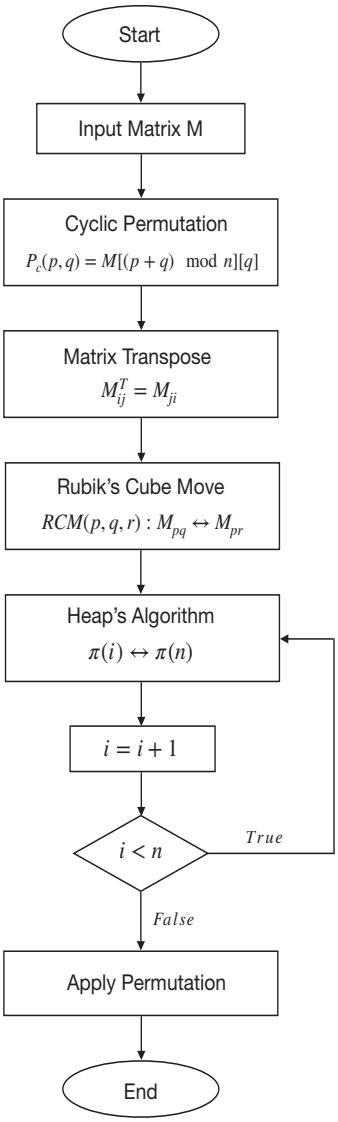


Fig. 3: Flowchart of CHP Algorithm

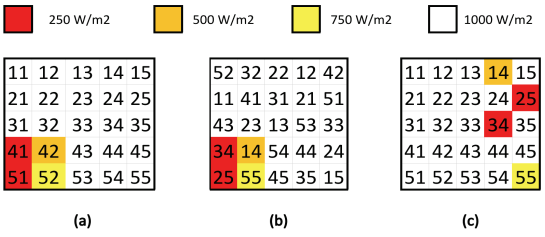


Fig. 4: Representation of Shade Case 1: (a) Shading pattern (b) Cyclic Heap Permutation (CHP) (c) Shade dispersion by CHP

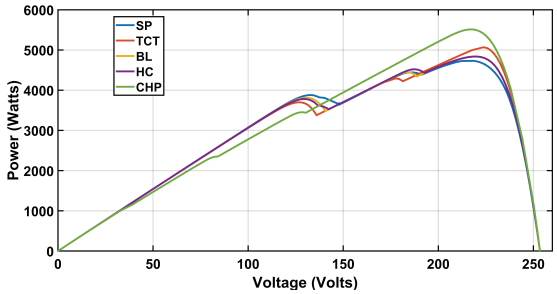


Fig. 5: Photovoltaic Performance Characteristics in the Presence of Shade Case 1

peak power compared to the other four configurations. This outcome emphasizes the effectiveness of CHP in addressing shading challenges associated with the Short Wide pattern.

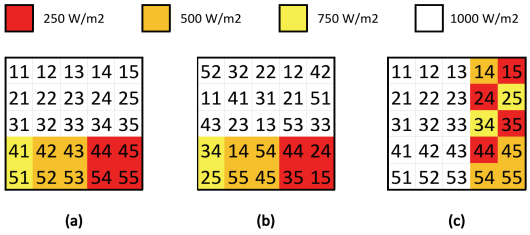


Fig. 6: Representation of Shade Case 2: (a) Shading pattern (b) Cyclic Heap Permutation (CHP) (c) Shade dispersion by CHP

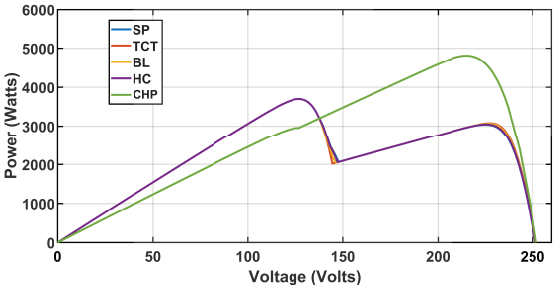


Fig. 7: Photovoltaic Performance Characteristics in the Presence of Shade Case 2

C. Analysis of Shade Case 3

In the third shading scenario (Shade Case 3), characterized as the long and wide shading pattern, the 5×5 solar PV array undergoes varying solar irradiation levels. Fig. 8 shows the long wide shading pattern along with the PV array reconfigured by CHP and the achieved effective shade dispersion. Upon examining the simulation results through the P-V curve in Fig. 9, CHP distinguishes itself by achieving a peak power of 3979 watts, surpassing the performance of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:57:23 UTC from IEEE Xplore. Restrictions apply.

alternative configurations. Notably, CHP's peak power exceeds the average peak power of the other four configurations by 12.11%, highlighting its adaptability in effectively addressing shading challenges associated with the Long Wide pattern.

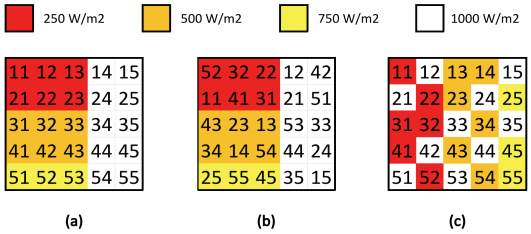


Fig. 8: Representation of Shade Case 3: (a) Shading pattern (b) Cyclic Heap Permutation (CHP) (c) Shade dispersion by CHP

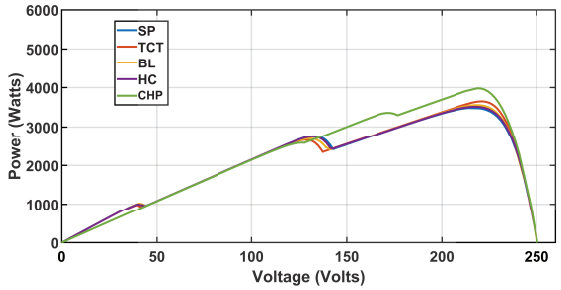


Fig. 9: Photovoltaic Performance Characteristics in the Presence of Shade Case 3

D. Analysis of Shade Case 4

In the Random Partial Shading Case (SC4), the 5×5 solar PV array encounters varied solar irradiation levels. Fig. 10 represents the shading pattern, the PV array reconfigured by CHP, and offers insights into the effective dispersion of shading. Analyzing simulation results through the P-V curve in Fig. 11, CHP once again exhibits superior performance with a peak power of 4552 watts, and an average peak power that surpasses the average peak power of the other four configurations by 32.21%, showcasing its versatility in mitigating the effects of randomly distributed shading.

E. GMPP and Power Loss

In contrast to alternative configurations, the proposed Cyclic Heap Permutation (CHP) consistently demonstrates the highest Global Maximum Power Point (GMPP) across all shading scenarios, as depicted in Fig. 12(a).

The Power Loss (PL) of the PV array is quantified by the equation:

$$PL(\%) = \frac{GMPP_{STC} - GMPP_{PSC}}{GMPP_{STC}} \times 100 \qquad (6)$$

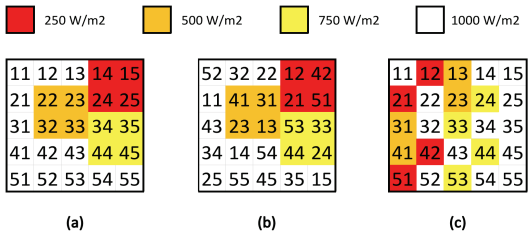


Fig. 10: Representation of Shade Case 4: (a) Shading pattern (b) Cyclic Heap Permutation (CHP) (c) Shade dispersion by CHP

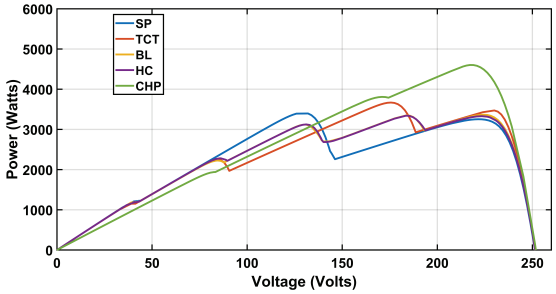


Fig. 11: Photovoltaic Performance Characteristics in the Presence of Shade Case 4

After computing and averaging the Power Loss across all shading scenarios, Table I shows that the Cyclic Heap Permutation technique proposed attains a Power Loss of 24.52%. In contrast, traditional configurations, including Bridge Link (BL) configuration, Honey Comb (HC) configuration, Series Parallel (SP) configuration, and Total-Cross-Tied (TCT) configuration, exhibit higher Power Loss values of 38.1%, 38.39%, 38.71%, and 35.6%, respectively. Fig. 12(b) depicts the comparison

Patterns	SC1	SC2	SC3	SC4
BL	22.48	40.76	43.06	46.09
HC	22.55	40.76	43.75	46.49
SP	24.24	40.76	44.25	45.61
TCT	18.85	40.76	41.57	41.25
CHP	**11.72**	**22.98**	**36.27**	**27.10**

TABLE I: Power Loss %

F. Fill Factor and Mismatch Loss

The efficiency and performance of PV arrays are significantly impacted by critical parameters such as Fill Factor (FF) and Mismatch Loss (ML). These metrics are quantified as percentages and defined by the following equations, where $P_{max,k}$ is the maximum power that can be produced by individual modules in shading condition:

$$ML(\%) = \frac{\sum_{k=1}^{m\times n} P_{max,k} - GMPP_{PSC}}{\sum_{k=1}^{m\times n} P_{max,k}} \times 100 \qquad (7)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:57:23 UTC from IEEE Xplore. Restrictions apply.

Patterns	SC1	SC2	SC3	SC4
BL	14.87	24.1	15.07	29.12
HC	14.95	24.1	16.1	29.64
SP	16.81	24.1	16.84	28.49
TCT	10.88	24.1	12.85	22.76
CHP	**3.06**	**1.33**	**4.94**	**4.14**

TABLE II: Mismatch Loss %

Patterns	SC1	SC2	SC3	SC4
BL	61.31	46.86	45.03	42.64
HC	61.26	46.86	44.49	42.32
SP	59.92	46.86	44.09	43.02
TCT	64.18	46.86	46.21	46.46
CHP	**69.82**	**60.92**	**50.4**	**57.66**

TABLE III: Fill Factor %

$$FF(\%) = \frac{P_{\text{max, PSC}}}{V_{\text{oc}} \times I_{\text{sc}}} \times 100 \tag{8}$$

Table II provides a comprehensive overview of the Mismatch Loss for all configurations under various shading scenarios. Notably, the proposed Cyclic Heap Permutation (CHP) technique demonstrates a substantial reduction in Mismatch Loss, averaging 3.36% whereas traditional configurations including Bridge Link (BL) configuration, Honey Comb (HC configuration), Series Parallel (SP) configuration, and Total-Cross-Tied (TCT) configuration exhibit higher mismatch losses of 20.79%, 21.19%, 21.56% and 17.64% respectively.

The outcomes, outlined in Table III, highlight the superior performance of the suggested CHP reconfiguration technique. On average, CHP exhibits an increment of 17.22%, 22.5%, 21.94%, and 23.16% in Fill Factor compared to TCT, HC, BL, and SP, respectively. This emphasizes the effectiveness of CHP in optimizing the array's power conversion capabilities. Fig. 12(c, d) shows the comparison of Fill Factor and Mismatch Loss in the PV configurations.

V. Conclusion

The superior performance of CHP is evident in both increased peak power output and a substantial decrease in power loss and mismatch losses, emphasizing its effectiveness in optimizing overall system performance. Additionally, the highlighted scalability of CHP enhances its practical applicability to PV systems of varying sizes, further solidifying its status as a promising and practical solution for addressing shading challenges in solar PV systems.

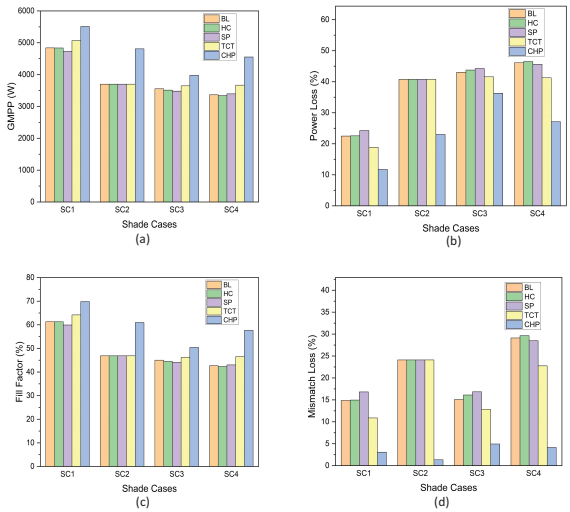


Fig. 12: Comparitive analysis of (a) Global Maximum Power Point (GMPP) in watts, (b) Power Loss in percentage, (c) Fill Factor in percentage, and (d) Mismatch Loss in percentage

References

[1] M. S. El-Dein, M. Kazerani, and M. Salama, "Optimal photovoltaic array reconfiguration to reduce partial shading losses," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 1, pp. 145–153, 2012.

[2] S. Anjum, V. Mukherjee, and G. Mehta, "Hyper sudoku-based solar photovoltaic array reconfiguration for maximum power enhancement under partial shading conditions," *Journal of Energy Resources Technology*, vol. 144, no. 3, p. 031302, 2022.

[3] T. S. Babu, J. P. Ram, T. Dragičević, M. Miyatake, F. Blaabjerg, and N. Rajasekar, "Particle swarm optimization based solar pv array reconfiguration of the maximum power extraction under partial shading conditions," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 1, pp. 74–85, 2017.

[4] R. Singh, V. K. Yadav, and M. Singh, "Optimal shade dispersion strategy for enhanced pv power extraction under pscs," *IEEE Transactions on Power Electronics*, 2023.

[5] S. Silvestre, A. Boronat, and A. Chouder, "Study of bypass diodes configuration on pv modules," *applied energy*, vol. 86, no. 9, pp. 1632–1640, 2009.

[6] R. Singh, V. K. Yadav, and M. Singh, "An improved hot spot mitigation approach for photovoltaic modules under mismatch conditions," *IEEE Transactions on Industrial Electronics*, 2023.

[7] H. S. Sahu and S. K. Nayak, "Extraction of maximum power from a pv array under nonuniform irradiation conditions," *IEEE Transactions on electron devices*, vol. 63, no. 12, pp. 4825–4831, 2016.

[8] A. Mäki, S. Valkealahti, and J. Leppäaho, "Operation of series-connected silicon-based photovoltaic modules under partial shading conditions," *Progress in photovoltaics: research and applications*, vol. 20, no. 3, pp. 298–309, 2012.

[9] H. Patel and V. Agarwal, "Matlab-based modeling to study the effects of partial shading on pv array characteristics," *IEEE transactions on energy conversion*, vol. 23, no. 1, pp. 302–310, 2008.

[10] B. Yang, T. Zhu, J. Wang, H. Shu, T. Yu, X. Zhang, W. Yao, and L. Sun, "Comprehensive overview of maximum power point tracking algorithms of pv systems under partial shading condition," *Journal of Cleaner Production*, vol. 268, p. 121983, 2020.

[11] B. Dhanalakshmi and N. Rajasekar, "A novel competence square based pv array reconfiguration technique for solar pv maximum power extraction," *Energy conversion and management*, vol. 174, pp. 897–912, 2018.

[12] B. Yang, H. Ye, J. Wang, J. Li, S. Wu, Y. Li, H. Shu, Y. Ren, and H. Ye, "Pv arrays reconfiguration for partial shading mitigation: Recent advances, challenges and perspectives," *Energy Conversion and Management*, vol. 247, p. 114738, 2021.

[13] O. Bingöl and B. Özkaya, "Analysis and comparison of different pv array configurations under partial shading conditions," *Solar Energy*, vol. 160, pp. 336–343, 2018.

[14] M. S. El-Dein, M. Kazerani, and M. Salama, "An optimal total cross tied interconnection for reducing mismatch losses in photovoltaic arrays," *IEEE transactions on sustainable energy*, vol. 4, no. 1, pp. 99–107, 2012.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:57:23 UTC from IEEE Xplore. Restrictions apply.

Check for updates

Received: 1 September 2023 | Revised: 12 March 2024 | Accepted: 16 April 2024

DOI: 10.1002/mop.34173

RESEARCH ARTICLE

WILEY

Design and simulation of solid-core octagonal photonic crystal fiber for terahertz wave propagation

Gurmeet Singh | **Shubham Sharma** | **Ajeet Kumar**

Department of Applied Physics, Advanced Photonics Simulation Research Laboratory, Delhi Technological University, New Delhi, Delhi, India

Correspondence
Ajeet Kumar, Advanced Photonics Simulation Research Laboratory, Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi, Delhi 110042, India.
Email: ajeetdph@dtu.ac.in

Abstract

A solid-core (SC) octagonal photonic crystal fiber (O-PCF) designed for operation in the terahertz regime aims to enhance its performance. The effective material loss, confinement loss, effective mode area, and nonlinear coefficient have been investigated and compared for two different media: silica glass and Teflon. The terahertz frequency range from 0.7 to 3.0 THz has been utilized to analyze all the results using a full-vectorial finite element method, with COMSOL Multiphysics software employed for the simulations. With a suitable selection of parameters, the proposed SC O-PCF demonstrates the potential to minimize losses and increase the effective mode area of the fundamental mode.

KEYWORDS

confinement loss, effective material loss, effective mode area, solid-core photonic crystal fiber, terahertz

1 | INTRODUCTION

Photonic crystal fibers (PCFs), also known as holey fibers, are made of a single material with numerous air holes in the cladding that remain periodically intact.[1] The low-index cladding around the core helps tune the dispersion slope and control confinement losses (CLs).[2] Regarding the guiding mechanism, PCFs are characterized in two ways: one has a solid core (SC) and a low index in the cladding region, which makes PCFs function like conventional fibers by guiding light through total internal reflection.[3,4] The other type is photonic band gap guidance PCF, equipped to control light guidance for any frequency band. PCFs find applications in various fields, such as PCF-based sensors, mid- and far-infrared guidance, terahertz guidance, and more.[5] The communication sector has recently evolved and undergone tremendous changes.[6] This recent transformation has brought about various applications in the terahertz (THz) region. On the electromagnetic spectrum, terahertz has a frequency range from 100 GHz to 10 THz, within which phonons, excitons, and cooper pairs are important material parameters that render the terahertz frequency a vital region.[7]

Unlike other terahertz PCFs, the benefit of the SC octagonal photonic crystal fiber (O-PCF) is that it allows operation in the fundamental mode, resulting in lower CL, a high effective mode area, and features of nonlinearity.[6] The design of O-PCF can be obtained by arranging air holes based on an isosceles triangle with a $\pi/4$ vertices angle. The arrangement of five rings can be formed, and by rotating the triangle at periods of $\pi/4$, $\pi/2$, $3\pi/4$, π, $5\pi/4$, $3\pi/2$, and $7\pi/4$, the octagonal design is achieved. If the number of rings in the cladding is decreased, radiation leakage from the core to the cladding will increase. Therefore, the rings are arranged in such a way that they can trap most of the radiation in the core and prevent any leakage into the cladding.[1] In this design, the finite element method (FEM) has been applied because, using this method, we can approach continuous functions as discrete models.[8]

In the existing PCF design, Dash et al.[9] reported on hexagonal PCF, in which at 1 THz, the CL, effective

 © 2024 Wiley Periodicals LLC.

mode area, and nonlinearity coefficient are 10^{-9}dB/cm, 5.62×10^{-6}m^2, and 3×10^{-8}rad/Wm, respectively. Kumar Paul et al.[10] reported effective material loss (EML), CL, and effective mode area of 6.6×10^{-2}/cm, 5.42×10^{-13}dB/cm, and 1.1×10^{-7}m^2, respectively. Moreover, in Islam et al.,[11] oligoporous-core (PC) PCF at 1 THz, CL, EML, and effective area are 7.24×10^{-7}/cm, 5.0×10^{-2}/cm, and 9.5×10^{-8}m^2, respectively. In 2019, Zhang et al.[12] proposed the porous core PCF. They evaluated EML, CL, and effective mode area of 0.08/cm, 3.2×10^{-13}/cm, and 1.20×10^{-7}m^2, respectively. Abdullah-Al-Shafi and Sen[13] suggested octagonal core PCF and calculated CL and effective mode area of 2.72×10^{-2}dB/m and 1.49×10^{-7}m^2, respectively. Sultana et al.[14] proposed the Zeonex-based PCF in the terahertz regime and observed EML, CL, and effective mode area as 0.05/cm, 7.79×10^{-12}/cm, and 2.0×10^{-7}m^2, respectively. Moreover, in 2020, Hossain and Sen[15] proposed optical chemical sensor-based PCF and culminated CL and effective mode area 2.70×10^{-6}dB/m and 1.25×10^{-7}m^2, respectively. The variation in the CL and EML, which is discussed above, is found to be several orders of magnitude. Such variation is expected due to novel designs presented causing tight confinement of light and less EML of background material.

In this paper, we have investigated the required parameters in the THz regime from 0.7 to 3 THz. We examined minimum losses and the maximum effective area for fused silica glass and Teflon materials. This paper is organized so that in Section 2, we introduce the geometrical structure of SC O-PCF and related parameters. Then in Section 3, required formulations are presented, and in Sections 4 and 5, numerical results and conclusions are reported.

2 | PROPOSED SC PCF DESIGN

Figure 1 illustrates the transverse cross-sectional view of the proposed SC O-PCF with the fundamental mode at 1 THz for *x*-polarization. The octagonal five-layer array of air holes is constructed in another material along the length of the fiber. In the cladding region, the air holes in the first, second, third, fourth, and fifth layers are 8, 16, 24, 32, and 40, respectively.

For comparison, two different media, namely, silica glass and Teflon, are utilized. In all five rings, the radius r remains constant. The distance between two air holes is the pitch Λ, and air filling fraction (AFF) is fixed at 0.5 which is calculated as

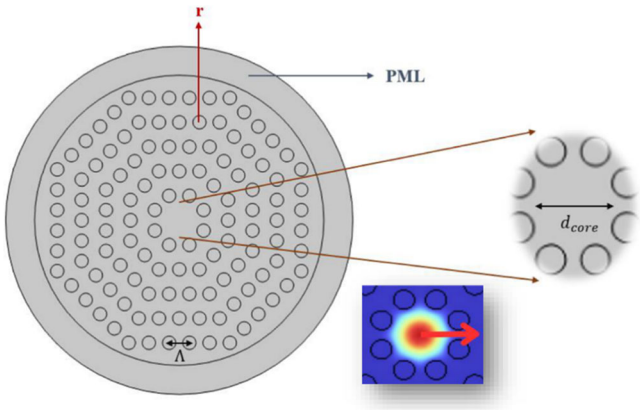


FIGURE 1 Proposed SC O-PCF with the fundamental mode at 1 THz for *x*-polarization. O-PCF, octagonal photonic crystal fiber; PML, perfectly matched layer; SC, solid core.

$$AFF = \frac{Diameter}{Pitch}. \tag{1}$$

The diameter of air holes and pitch remain constant throughout the simulation study. The perfectly matched layer (PML) is also employed to calculate the CL.

3 | FORMULATIONS

The COMSOL Multiphysics software, based on a full-vectorial FEM, evaluates the effective indices, CL, and nonlinear coefficient. Absorption loss or EML (α_{eff}) depicts the loss of the PCFs while operating at THz frequencies, and it can be calculated as[16]

$$\alpha_{eff} = \sqrt{\frac{\epsilon_0}{u_0}} \frac{\int_{mat} n_{mat}\alpha_{mat}|E|^2\,dA}{2\left|\int_{all} s_z\,dA\right|} \quad (\text{dB/cm}), \tag{2}$$

where α_{mat} is the material absorption loss, $\vec{s}_z = \frac{1}{2}(\vec{E} \times \vec{H})\hat{z}$ is the pointing vector of the z-component, here, E and H are electric and magnetic fields, n_{mat} is the effective refractive index of core material, and ϵ_o and μ_o are electric permittivity and magnetic permeability, respectively.

CL which is essential to find in terahertz applications of PCF, can be evaluated from the imaginary value of n_{eff} as

$$\begin{aligned} L\ (\text{dB/cm}) &= \frac{40\pi}{\ln(10)\lambda} Im(n_{eff}) \times 10^{-2} \\ &= 8.686 k_0 Im(n_{eff}) \times 10^{-2}, \end{aligned} \tag{3}$$

10982760, 2024, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mop.34173 by Delhi Technological University, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where $Im(n_{eff})$ is the imaginary part of the effective mode index of that mode. And $k_0 = \frac{2\pi}{\lambda}; \lambda$ is the operating wavelength of light in meters.[17]

In high-power transmission, PCFs with a large effective mode area are used. This characteristic also suppresses the unwanted nonlinear effects. A greater effective mode area will give lesser nonlinear effects and vice versa. It is calculated using[18]

$$A_{eff} = \frac{\left[\int I(r)\,dr\right]^2}{\left[\int I^2(r)\,dr\right]}, \tag{4}$$

where $I(r) = [E(t)]^2$ is the amplitude of the transverse electric field propagating through the fiber.

The nonlinear coefficient defines the interaction of the nonlinearity of light with an optical material. It can be found by

$$\gamma = \left(\frac{2\pi}{\lambda}\right)\left(\frac{n_2}{A_{eff}}\right), \tag{5}$$

where n_2 is the nonlinear refractive index of the material.[19]

4 | NUMERICAL RESULTS

We have investigated the EML, CL, effective mode area, and nonlinear coefficient using a full-vectorial FEM in the given structure. To correlate, fused silica and Teflon of refractive indices 1.45 and 1.38, respectively, have been used.

Initially, we have chosen the fixed structural parameters, $r = 0.95\,\text{mm}$, $\Lambda = 1.9\,\text{mm}$, or $AFF = 0.5$, $cladding = 10.4\,\text{mm}$, and outer layer 20% of cladding as PML. The fundamental modes at 1 THz frequency are observed for fused silica and Teflon, which are 1.44 and 1.37, respectively. Figure 1 (inset) shows the contour plot of the field intensity of one of the fundamental modes at 1 THz.

Figure 2 shows the variation in the EML for increasing frequency when the core-cladding media is set to fused silica. The integration of the numerator of Equation (2) is performed only over the solid material region, which indents EML as a function of frequency. The loss is reported as low as 2.07×10^{-2}/cm at 0.7 THz of frequency and reached as high as 2.08×10^{-2}/cm at 3 THz. The loss in material is reaching the value of bulk material absorption loss (α_{eff}) of silica glass.

In Figure 3, the comparison of CL between silica glass and Teflon is depicted. CL is directly dependent on the frequency, and it is clarified that at a lower frequency of

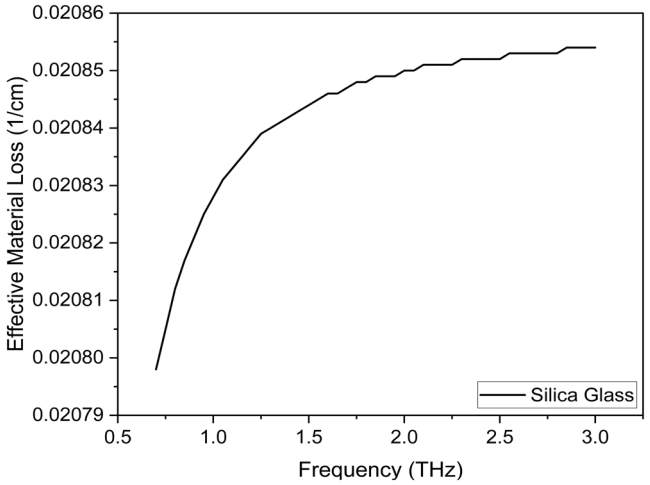


FIGURE 2 Variation of effective material loss with frequency.

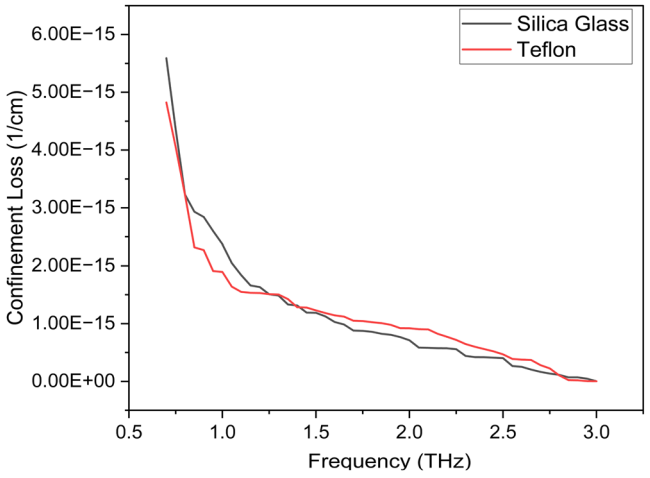


FIGURE 3 Variation of confinement loss with frequency (silica glass and Teflon).

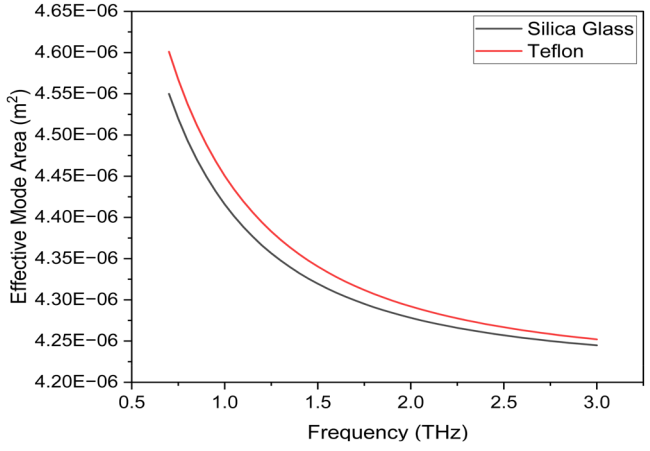


FIGURE 4 Variation of effective mode area with frequency (silica glass and Teflon).

10982760, 2024, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mop.34173 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

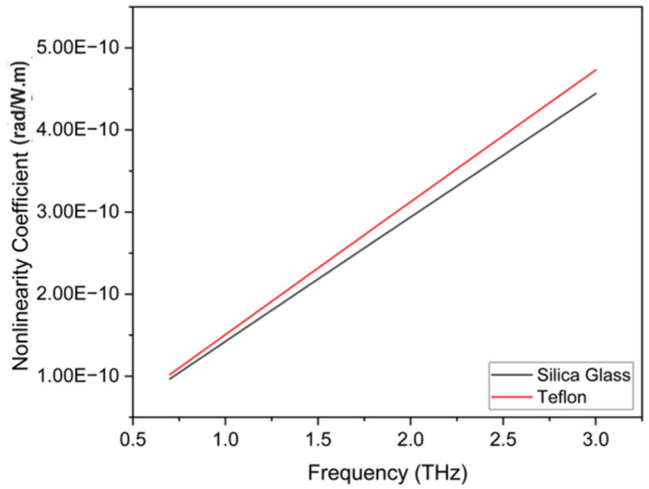


FIGURE 5 Variation of γ with frequency (silica glass and Teflon).

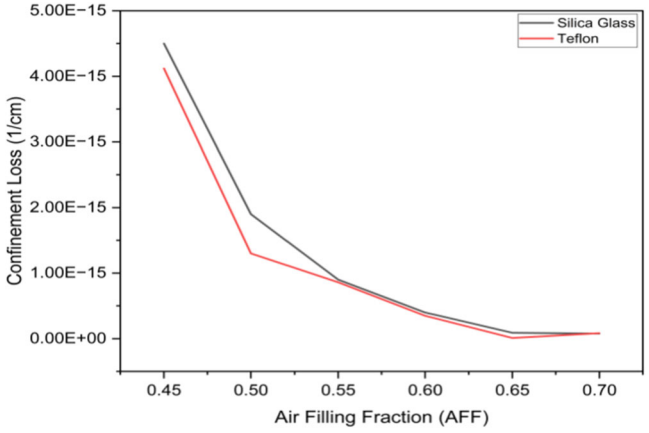


FIGURE 6 Variation of confinement loss with air filling fraction (silica glass and Teflon).

0.7 THz, it approximates 10^{-15}/cm. Subsequently, it decreases to 10^{-17}/cm for both materials in the fundamental mode. However, Teflon exhibits slightly better reliability in containing the light within the core at 1 THz frequency. This change in behavior can be attributed to a tightly packed core that prevents radiation from leaking.[20]

We have observed a significant order change between EML and CL. This behavior is primarily attributed to the core material's properties. In the case of EML, low values signify less bulk material absorption (α_{eff}) of radiation. This differs from CL, which is mainly govern by the confinement and leakage of radiation within the designed fiber.[21]

In Figure 4, large effective mode areas are obtained for lower frequencies in fundamental modes; however, it gradually decreases with an increase in frequency. Teflon shows a large effective mode area within the range [0.7, 3.0] THz among both materials. The reason for these variations is radiation confinement in the core at higher frequencies.

Theoretically, a low nonlinear coefficient will be obtained with a larger effective mode area. Figure 5 proves the theory; at lower frequencies, nonlinear coefficients are as low as 10^{-10} rad/Wm for both media, and it keeps increasing with an increase in frequency. Yet Teflon is the most effective compared to silica glass because of its high value of the nonlinear refractive index, which is 3.2×10^{-6} rad/Wm. In Figure 6, we have examined the variation in CL of silica glass and Teflon material with AFF. These findings indicate that increasing the AFF value results in lower CL within the proposed PCF. A higher AFF value brings the number of rings and air holes closer to each other, causing most of the radiation to trap within the core of the fiber and

TABLE 1 Feature of proposed SC O-PCF with the existing designs.

Ref.	EML (1/cm)	L	A_{eff} (m^2)	γ (rad/Wm)
[9]	–	10^{-9} dB/cm	5.62×10^{-6}	3.0×10^{-8}
[10]	6.60×10^{-2}	5.42×10^{-13} dB/cm	1.10×10^{-7}	–
[11]	5.0×10^{-2}	7.24×10^{-7}/cm	9.50×10^{-8}	–
[12]	8.0×10^{-2}	3.2×10^{-13}/cm	1.2×10^{-7}	–
[13]	–	2.27×10^{-2} dB/m	1.49×10^{-7}	–
[14]	5.0×10^{-2}	7.79×10^{-12}/cm	2.0×10^{-7}	–
[15]	–	2.70×10^{-6} dB/m	1.25×10^{-7}	–
Proposed (silica glass)	2.08×10^{-2}	2.0×10^{-15}/cm	4.42×10^{-6}	1.42×10^{-10}
Proposed (Teflon)	2.99×10^{-2}	2.63×10^{-15}/cm	4.45×10^{-6}	1.51×10^{-10}

Abbreviations: EML, effective material loss; O-PCF, octagonal photonic crystal fiber; SC, solid core.

10982760, 2024, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mop.34173 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

consequently reducing the CL. Specifically, it shows a decrease in CL from 4.5×10^{-15} to 7.8×10^{-17}/cm in silica glass and 4.12×10^{-15} to 8.3×10^{-17}/cm in Teflon as the AFF value increased from 0.45 to 0.7.

Table 1 compares previous results[9–15] and determined results of SC O-PCF in this paper. The proposed design shows better results in comparison to other existing models.

5 | CONCLUSION

In the proposed O-PCF, we have improved the required factors by changing the AFF in the cladding region. Making all five rings uniform plays a vital role and using materials like silica glass and Teflon reported low EML and CL due to their low material absorption loss. This paper obtains the fundamental mode of fused silica and Teflon at 1.44 and 1.37.

Both materials can reduce EML up to 2.08×10^{-2}/cm for fused silica 2.99×10^{-2}/cm for Teflon. Moreover, CLs are reduced from up to ~10^{-17}/cm for both the materials at larger frequency with the proposed design. This design will be helpful in several applications, such as telecommunication, spectroscopy, sensing, and so forth.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no data sets were generated or analyzed during the current study.

ORCID

Ajeet Kumar http://orcid.org/0000-0003-4632-2841

REFERENCES

1. Chiang JS, Wu TL. Analysis of propagation characteristics for an octagonal photonic crystal fiber (O-PCF). *Opt Commun*. 2006;258(2):170-176.
2. Habib MS, Habib MS, Razzak SA, Hossain MA. Proposal for highly birefringent broadband dispersion compensating octagonal photonic crystal fiber. *Opt Fiber Technol*. 2013;19(5):461-467.
3. Ademgil H. Highly sensitive octagonal photonic crystal fiber based sensor. *Optik*. 2014;125(20):6274-6278.
4. Ahmed K, Morshed M. Design and numerical analysis of microstructured-core octagonal photonic crystal fiber for sensing applications. *Sens Bio-Sens Res*. 2016;7:1-6.
5. Cerqueira SA. Recent progress and novel applications of photonic crystal fibers. *Rep Prog Phys*. 2010;73(2):024401.
6. Hossain MS, Sen S, Hossain MM. Performance analysis of octagonal photonic crystal fiber (O-PCF) for various communication applications. *Phys Scr*. 2021;96(5):055506.
7. Goto M, Quema A, Takahashi H, Ono S, Sarukura N. Teflon photonic crystal fiber as terahertz waveguide. *Jpn J Appl Phys*. 2004;43(2B):L317.
8. Tekkaya AE, Soyarslan C. Finite element method. In: Laperrière L, Reinhart G, eds. *The International Academy for Production Engineering*; 2018.
9. Dash S, Mathur V, Pandey N, Sinha RK. Terahertz wave propagation characteristics in graded Teflon based solid-core photonic crystal fibre. *J Phys: Conf Ser*. 2023;2426(1):012021.
10. Kumar Paul B, Haque MA, Ahmed K, Sen S. A novel hexahedron photonic crystal fiber in terahertz propagation: design and analysis. *Photonics*. 2019;6(1):32.
11. Islam MS, Sultana J, Dinovitser A, Ng BWH, Abbott D. A novel Zeonex based oligoporous-core photonic crystal fiber for polarization preserving terahertz applications. *Opt Commun*. 2018;413:242-248.
12. Zhang Y, Xue L, Qiao D, Guang Z. Porous photonic-crystal fiber with near-zero ultra-flattened dispersion and high birefringence for polarization-maintaining terahertz transmission. *Optik*. 2020;207:163817.
13. Abdullah-Al-Shafi M, Sen S. Design and analysis of a chemical sensing octagonal photonic crystal fiber (O-PCF) based optical sensor with high relative sensitivity for terahertz (THz) regime. *Sens Bio-Sens Res*. 2020;29:100372.
14. Sultana J, Islam MS, Ahmed K, Dinovitser A, Ng BWH, Abbott D. Terahertz detection of alcohol using a photonic crystal fiber sensor. *Appl Opt*. 2018;57(10):2426-2433.
15. Hossain MS, Sen S. Design and performance improvement of optical chemical sensor based photonic crystal fiber (PCF) in the terahertz (THz) wave propagation. *Silicon*. 2021;13(11):3879-3887.
16. Rana S, Ali S, Ahmed N, Islam R, Aljunid SA. Ultra-high birefringent and dispersion-flattened low loss single-mode terahertz wave guiding. *IET Commun*. 2016;10(13):1579-1583.
17. Saitoh K, Koshiba M, Hasegawa T, Sasaoka E. Chromatic dispersion control in photonic crystal fibers: application to ultra-flattened dispersion. *Opt Express*. 2003;11(8):843-852.
18. Hassani A, Dupuis A, Skorobogatiy M. Low loss porous terahertz fibers containing multiple subwavelength holes. *Appl Phys Lett*. 2008;92(7):071101.
19. Eid MMA, Habib MA, Anower MS, Rashed ANZ. Highly sensitive nonlinear photonic crystal fiber based sensor for chemical sensing applications. *Microsyst Technol*. 2021;27:1007-1014.
20. Singh G, Sharma S, Kumar A. Design and analysis of spider-web photonic crystal fiber in the terahertz regime. *Indian J Phys*. 2023. doi:10.1007/s12648-023-03038-7
21. Harun SW, Arof H, eds. *Current Developments in Optical Fiber Technology*. BoD–Books on Demand; 2013.

How to cite this article: Singh G, Sharma S, Kumar A. Design and simulation of solid-core octagonal photonic crystal fiber for terahertz wave propagation. *Microw Opt Technol Lett*. 2024;66:e34173. doi:10.1002/mop.34173

10982760, 2024, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mop.34173 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

International Journal of Advanced Research in Engineering and Technology (IJARET)
Volume 14, Issue 03, March-April 2023, pp. 1-13, Article ID: IJARET_14_03_001
Available online at https://iaeme.com/Home/issue/IJARET?Volume=14&Issue=3
ISSN Print: 0976-6480 and ISSN Online: 0976-6499
DOI: https://doi.org/10.17605/OSF.IO/HC46T

© **IAEME** Publication

OPEN ACCESS

DESIGNING OF A TWO STAGE TO ORBIT LAUNCH VEHICLE

Priyanshu Nailwal, Prakash N. Budhaulia, Rahul Kumar, Raj K. Singh
Department of Mechanical Engineering,
Delhi Technological University, Delhi, India

ABSTRACT

There is an increase in demand for cost effective and reliable launch vehicles. The commonly used and simplified vehicles are Two-Stage to Orbit (TSTO) launch vehicles which are opted for their relatively simpler design with respect to higher stage vehicles. The aim of this paper is to understand the significance and conceptual designing of a TSTO launch vehicle by studying various design parameters relating to gravity turn trajectory using a MATLAB model capable of creating a 2D simplified trajectory of the vehicle using the constraints of design and different design inputs. The designed TSTO of our consideration uses a liquid propellant engine for both the stages which has an advantage of reusability and easy manufacturing.

Keywords: Complexity, Design, Launch Vehicle, MATLAB, Physics, Reusability, Two Stage to Orbit (TSTO)

Cite this Article: Priyanshu Nailwal, Prakash N. Budhaulia, Rahul Kumar and Raj K. Singh, Designing of A Two Stage to Orbit Launch Vehicle, International Journal of Advanced Research in Engineering and Technology (IJARET), 14(3), 2023, pp. 1-13. https://iaeme.com/Home/issue/IJARET?Volume=14&Issue=3

1. INTRODUCTION

A TSTO launch vehicle is a type of rocket that uses two different stages in its trajectory to inject the payload into the desired orbit. In the first stage, the rocket is lifted to the atmospheric phase and the second stage, in which the uppermost stage of the vehicle is powered up with the help of its vacuum optimized engine, to inject the payload into orbit without carrying the dry mass of the first stage to space [1]. The advantage of having a two-stage launch vehicle is having lesser complexity compared to other multi-stage rockets as the minimum number of separation events minimizes the complexity, cost and risk of failure. Single stage to orbit vehicles could've been great for reusability, but they haven't been flown to orbit, so the next best option is having a two-stage vehicle.

A Reusable Launch Vehicle (RLV) is one which can have its parts recovered after flying on a space mission and reflown for the next mission. Rocket stages, boosters, or the spacecraft on top of the vehicle are the ones that are being reused, which makes it easier for the company to not constantly manufacture the same component for each launch. This reduces the cost of manufacturing, however, there is still the cost of recovery and refurbishments which is less compared to manufacturing.[2] Most common concept of RLV utilizes liquid propellant rocket engine, for its throttling capabilities.

There has been a significant rise in the number of rockets launches which led to a highly competitive market. Various space agencies have made their rockets and eventually, it has been realized that the best way to dominate this market is by having a low cost per kg of payload for orbital launch vehicles. This can be successfully achieved not just by having high performance but also the rocket can be manufactured with ease and they can be made partially reusable.

There has been a significant number of launch vehicles that employ two stages only, however, only a few are even partially reusable. Single stage to Orbit can be used in case of full reusability but as they consume more fuel they are not used. Dividing the rocket into multiple stages, each of which can be optimized to enhance performance leads to overall improved performance and a reduction in losses. Different engines can be used depending on external conditions, improving efficiency. When building multistage rockets, optimization of elements like mass distribution among stages, appropriate engine size, and propellant quantity are crucial. In this paper, we will study the Two Stage To Orbit Launch Vehicle physics and their basic conceptual designing.

2. LITERATURE REVIEW

M. A. Sharaf and L.A. Alaqal [3], and David W. Callaway [4] worked on basic gravity turn maneuver equations for the vehicle trajectory and created various results to understand the relations between various parameters. Their work has been used to understand the ascent trajectory based on gravity turn.

David Yaylalia developed simplified model to study the gravity turn maneuver in both 2D and 3D plane assuming Earth to be spherical and non-rotating.[5]

Dr Gerald M Gregorek calculated the drag force acting on the vehicle. Every component produces different values of drag for different types of drag. Due to complexity of certain type of drag, the value of drag coefficient was assumed constant for the conceptual designing phase.[6]

L. Brevault , M. Balesdent et.al: have developed dual-orbital two-stage rockets using the SABER air-breathing rocket engine and the LOx/CH4 rocket engine (for the second stage) to inject a 15-tonne payload into the Low Earth Orbit (LEO). They ran various simulations of the rocket's trajectory compiling it with various systems of the vehicle. In this work, the main focus was on the rocket trajectory.[7]

Ankit Kumar Mishra, Kshtij Gandhi et.al did an overview of a two-stage sounding rocket in their work. They emphasize the importance of multi staging and various important design parameters. The model of vehicle was developed in Open Rocket and engine was designed using Rocket Propulsion Analysis (RPA) software, which were latter analyzed in ANSYS Fluent using appropriate model of physics.[1]

Christie Alisa Maddock , Lorenzo Ricciardi et.al: have described the conceptual design and performance analysis of a partially reusable rocket for small payloads. The system was designed for a nominal mission of placing a 500 kg payload in a 600 km circular orbit with a polar orbit of 88.2 degrees. The goal of the system design was to develop an economically viable launch system with a short period of operation.

Therefore, the emphasis is on the efficient use of advanced Technology Readiness Level (TRL) technology and details on various parameters to consider when designing a spacecraft was also discussed.[9]

Byeong-Un Jo, Jaemyung Ahn have used a new method to calculate and minimize the total cost of reusable vehicles over their entire lifecycle using integrated optimization modules and cost model. The validity and effectiveness of the proposed algorithm are demonstrated through a case study on an optimal staging environment using a real reusable vehicle (Falcon 9). Case study results demonstrate that the minimum lift-off configuration achievable by conventional stage optimization does not necessarily result in the lowest life-cycle cost, justifying the usefulness of the proposed approach.[10]

3. DESIGN AND METHODOLOGIES USED

The methodology used to bring out the results was using software MATLAB for generating the results of the input equations and variable parameters and generating graphs for various dependable variables. MATLAB code was constructed with the equations and variables that need to be used to design a launch vehicle trajectory and various parameters that have a key role were used to study the physics and conceptual design of a launch vehicle. Certain key points which were needed to design the launch vehicle were Weight Distribution, Thrust parameter, Gravity Turn, and Vehicle design.

In weight distribution, we have to study the weight of the launch vehicle and optimize it in such a manner that maximum payload capacity can be achieved. Weight distribution was divided into three stages. Stage one where the structure mass and propellant mass were taken into account before the separation stage. Stage two where the structure mass and propellant mass were taken after the separation stage and the third stage was the payload meant to be injected into the orbit.

3.1. Phase of Flight

Initially, the vehicle is launched vertically from the launchpad. However, it is essential to have the vehicle inclined parallel to the local horizon. The vehicle must remain completely vertical for the initial seconds while it is still over the launchpad [11]. Once the vehicle has cleared off from the launch pad, the vehicle can initiate the gravity turn maneuver by providing a pitch kick at a specified altitude which makes the initially vertical vehicle, now inclined with some angle from vertical reference. The altitude at which the pitch kick is given and the angle at which it is given can be considered as design parameters.

This maneuver does not use the thrust of the vehicle but gravity to steer it onto its desired trajectory. Thrust is entirely used to accelerate the vehicle.

Primarily the thrust vector always points towards the direction of the velocity vector so that maximum acceleration is achieved by the rocket. Therefore, the vehicle has a very small or negligible angle of attack (hence this maneuver is also known as ‘zero lift trajectory’) that minimizes transverse aerodynamic stresses acting on the vehicle [3]. Thus, we can have lesser structure weight to improve performance.

Now due to gravity the rocket slightly rotates its orientation from vertical to horizontal. So along with the initial rest conditions of the vehicle at $t = 0$, in a gravity turn, we also have to decide two initial parameters mainly, the altitude at which gravity turns start and the angle given with respect to vertical at the start of maneuver. [5]

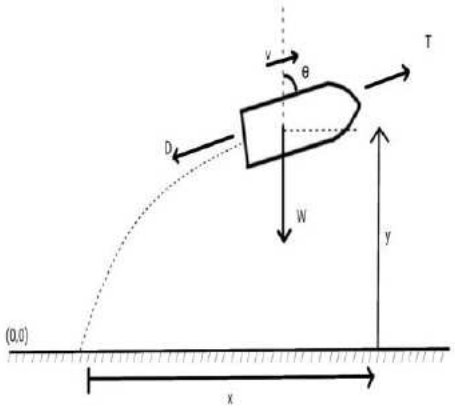


Fig: 1. Representation of parameters during gravity turn trajectory

$$\frac{dv}{dt} = \left(T - \frac{1}{2}\,\rho V^2 S C C_D\right)\frac{1}{m} - g\cos\theta \quad (1)$$

Equation (2) depicts the turn rate of the vehicle as it turns from vertical to horizontal inclination.

$$\frac{d\theta}{dt} = \frac{g\sin\theta}{V} \quad (2)$$

Equations (3) and (4) tell the velocity along horizontal and vertical direction which can provide the altitude and downrange of the vehicle at every instant.

$$\frac{dh}{dt} = v\cos\theta \quad (3)$$

$$\frac{dx}{dt} = v\sin\theta \quad (4)$$

Reduction of mass of vehicle due to propellant flow rate can be depicted in equation (5).

$$\frac{dm}{dt} = -\dot{m} * n_{engine} \quad (5)$$

h = Altitude; x = Downrange; v = velocity; m = Mass of vehicle; θ = Angle w.r.t vertical;

T = Thrust; D = Drag; g = Acceleration due to gravity; (All the parameters are at that instantaneous altitude)

ṁ = Mass flow rate from single engine;

n_{engine} = Number of engines used in the stage

After reaching a certain point during the gravity turn, the propellant in the first stage exhausts. After this, stage separation happens which removes the first stage from the vehicle and the second stage engine fires up which further pushes the payload. The altitude for stage separation should be optimally decided such that vacuum engines take the charge of operation before the sea level engine gets any more inefficient. The second stage engine which is optimized for vacuum-level performance starts operating and it will align the vehicle parallel to the local horizon. The vehicle after reaching the orbit might not necessarily have the required orbital speed to remain in the circular orbit. Thus, the vehicle is further accelerated at the fixed altitude to reach the required orbital speed.

3.2. Vehicle Dimension

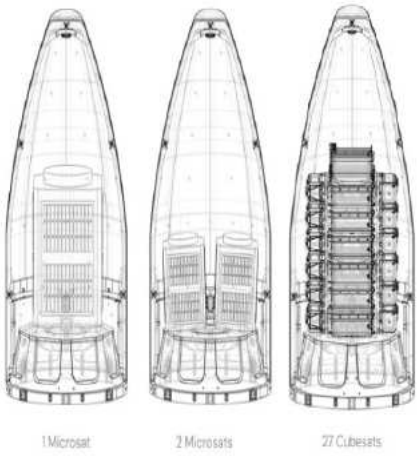


Fig: 2. Electron Rocket, Payload Configurations; Capability to put 300 kg to LEO [13]

In vehicle design, the radius of the vehicle is taken into consideration as it is important for drag calculation and the frontal area. While deciding the radius of the rockets we need to specify the payload that the vehicle has to carry with proper mounting options. The size of the payload and its weight will put constraints on the radius and payload capacity of the vehicle. The majority portion of length of vehicle can be simply obtained by converting the mass of propellants into volume

3.5. Modeling atmospheric and Gravitational effects

Various equations and assumptions we have taken to understand the conceptual design of the launch vehicle where drag is constant as 0.75. For modeling the effect of gravity, we consider Earth, spherically symmetrical and use the inverse square gravity method. [11,14]

$$\overline{Fg} = -\mu * \frac{\bar{R}}{||R^3||} \tag{6}$$

where $\mu = 3.986005 * 10^4$

For modeling the atmosphere, we used the International Standard Atmosphere (ISA). For the isothermal layer of the atmosphere, we use equations (7) & (8): [15,16]

$$\frac{P_1}{P_2} = e^{-(\frac{g_\circ}{RT})(h_2 - h_1)} \tag{7}$$

$$\frac{\rho_1}{\rho_2} = e^{-(\frac{g_\circ}{RT})(h_2 - h_1)} \tag{8}$$

For the gradient layer of the atmosphere, we use the following formulas, equations (9) & (10):

$$\frac{P}{P_1} = \left(\frac{T}{T_1}\right)^{-\left(\frac{g_\circ}{LR}\right)} \tag{9}$$

$$\frac{\rho}{\rho_1} = \left(\frac{T}{T_1}\right)^{-\left(\frac{g_\circ}{LR}+1\right)} \tag{10}$$

Were,

L = Lapse rate for the corresponding layer

g_0 = Acceleration due to gravity at ground level

R = Gas Constant

P_1, T_1, ρ_1= Pressure, Temperature, Density at the initial point of any gradient layer

3.4. Engine Parameters

For modeling the value of thrust generated by an engine, we use the basic thrust equation (11),

$$F = mv_e + (p_e - p_a)A_e \quad (11)$$

were,

P_e = Pressure at the exit of the nozzle

P_a = Pressure of the atmosphere immediately outside the nozzle

A_e = Nozzle exit area

This can be converted into another useful quantity called specific impulse. The specific impulse (I_{sp}) is the ratio of the thrust produced to the weight flow of the propellants which is used to measure the fuel efficiency of a rocket engine. [15,17]

$$I_{sp} = \frac{F}{\dot{m}g_\circ} \quad (12)$$

A rocket nozzle has optimum performance for the altitude where the exit pressure of the nozzle is equal to ambient pressure. However, since the vehicle has to ascend upwards, ambient pressure will keep reducing and thus the performance will vary with altitude. The nozzle will be designed to have optimum performance at a specific altitude and for the remaining altitude, we'll have to deal with the under-expansion of flow, which happens when ambient pressure is lower than exit pressure, and overexpansion of flow, where the ambient pressure is higher than exit pressure. The inefficiency of the nozzle is directly related to the difference between ambient and exit pressure. To have optimum performance for very low ambient pressure, we require a relatively large nozzle.

Since we have a two-stage launch vehicle, the lower-stage engines, which are referred to as sea-level engines, will have nozzles optimized for larger ambient pressure in comparison to the ones placed on the upper stage, referred to as vacuum-level engines since they operate predominantly in vacuum environments. In order to save research costs and reduce manufacturing processes, both sea level and vacuum engines are of the same design, except for the diverging section of the nozzle.

To have the engine operate optimally at a different ambient pressure, only the diverging section of the nozzle has to be changed as the dimensions of the combustion chamber and converging section are entirely similar. Sea-level engines have a small exit area, so a cluster of them can be placed but the vacuum engines are optimized for such small pressure, we cannot place the same number of engines in the upper stage. Having a multi-engine configuration and similar configuration of sea level and vacuum engine, except for a few changes due to the diverging portion, we can employ mass production of components. [18] In earlier times, larger and lesser engines were manufactured as even a single engine failure will lead to the failure of the entire mission. However, with the use of highly sophisticated computers, we can have multiple smaller engines on vehicles. Even if an engine fails, it'll be deactivated in time to safeguard the mission. Having multiple smaller engines is highly beneficial for low manufacturing costs. The specifications used in SpaceX's Falcon 9 vehicle which has 9 Merlin 1D sea level engines in octa-web arrangement and 1 Merlin 1D vacuum adapted engine. Single vacuum engine has a size comparable to that of 9 sea level engines combined.[19]

3.5. Physical Classification

To understand the effect of parameters, mission requirement is considered to deliver CubeSats of a total weight of 300 kg to 250 km. Based on the mission requirement, the diameter of the vehicle can be estimated to be 1.2 m by appropriately understanding the placement of CubeSats.

The vehicle has 2 stages and the following are the parameters set for both stages in table 1:

Table 1: Values of parameters at first stage and second stage

Parameter	First Stage	Second Stage
Structural Mass [ton]	0.9	0.3
Propellant Mass [ton]	10	2
Number of Engine	9(All sea-level engine)	1(Vacuum engine)

We consider the altitude for which pitch over maneuver initiates as 2 km and the angle provided by pitch kick as 10.2°. Till the altitude of 2 km, the vehicle accelerates upward vertically. At 2 km, a pitch kick is given to the vehicle which makes the vehicle's angle with the vertical reference line as 10.2°. This initiates the gravity turn maneuver.

The first stage has 9 sea-level engines which we can arrange in octagon arrangement. These engines are optimized for ambient pressure of 0.4 atm based on the Summerfield criterion. The vacuum engine cannot be entirely optimized for vacuum-level operation as the exit area would be significantly high [1]. Thus, we constrain the exit diameter of the vacuum level engine lower than the entire vehicle diameter and let there be slight losses due to under-expansion of flow. [20]

Using RPA software, the following parameters of engines were finalized in table 2. Since the combustion chamber assembly for both Sea-level and vacuum-optimized engines is the same, few parameters are the same. We consider the chamber pressure to be 7 MPa, the mass flow rate as 8 kg/s, and the propellant to be RP-1 + LOX. Due to the different diverging portions of both types of engines, the following parameters are separate for sea-level and vacuum-optimized engines.

Table 2: Values of different parameters obtained for sea level engine and vacuum engine

Engine Parameter	Sea Level Engine	Vacuum Engine
Thrust [kN]	24.23173	28.40706
Specific Impulse [s]	308.86886	362.08926
Exit Pressure [atm]	0.4	0.015
Exit Diameter [cm]	23.449	90.087
Exit Area [m^2]	0.0432	0.6374

4. RESULTS AND DISCUSSION

We consider the payload mass to be a part of the propellant weight in the code. Here the payload mass is targeted to be as 300kg and the propellant weight in the second stage is 2000kg, the total propellant weight as 2300 kg which the vehicle can consume till the second stage engine cut-off.

Table 3: Values of various parameters during various phases of flight

Important Phases of Flight	Time[s]	Altitude [km]	Ground Range [km]	Velocity [m/s]	Mass [kg]	Angle [degrees]
Lift-Off	0	0	0	0	13500	0
Pitch Kick	27.38	2	0	156.57	11528.64	10.2
First Stage Cut-Off	138.89	74.59	94.07	2975.91	2600	62.09
Stage Separation	138.89	74.59	94.07	2975.91	2600	62.09
Second Stage Ignites	138.89	74.59	94.07	2975.91	2600	62.09
Orbit Reached	368.81	254.51	1096.42	7041.5	760.64	90
Second Stage Cut-Off	385.54	254.51	1219.98	7752.11	626.8	90

As per the table 3, stage separation occurs at the altitude of 74.6 km, after which the vehicle attains an orbit at 254.5 km by getting parallel to the local horizon. The velocity at that point is lower than the orbital speed for that altitude by 710.6 m/s. Thus, it uses 133.84 kg of propellant further in order to attain the orbital velocity which will allow the vehicle to maintain that altitude. Figure 3 shows the entire trajectory of the vehicle and also highlights various important phases of flight.

In figure 4(right), showing variation of entire mass of vehicle at every instant, it is shown that after 385.5 seconds, the second stage vacuum engine shuts down, and the net weight of the vehicle is 626.8 kg and by removing the structural weight, the remaining weight becomes 326.8 kg which can be considered as the payload capacity to this altitude. In Figure 4 (left) shows the variation of all the 3 forces acting on the vehicle. It can be seen that the effect of gravity is always more significant than drag force. The drag force which initially increases due to increase of velocity, later starts declining due to lowering density at higher altitudes. Gravitational force is primarily reducing due to reducing weight. Thrust force sees a significant drop after stage separation due to the decrease of number of engines.

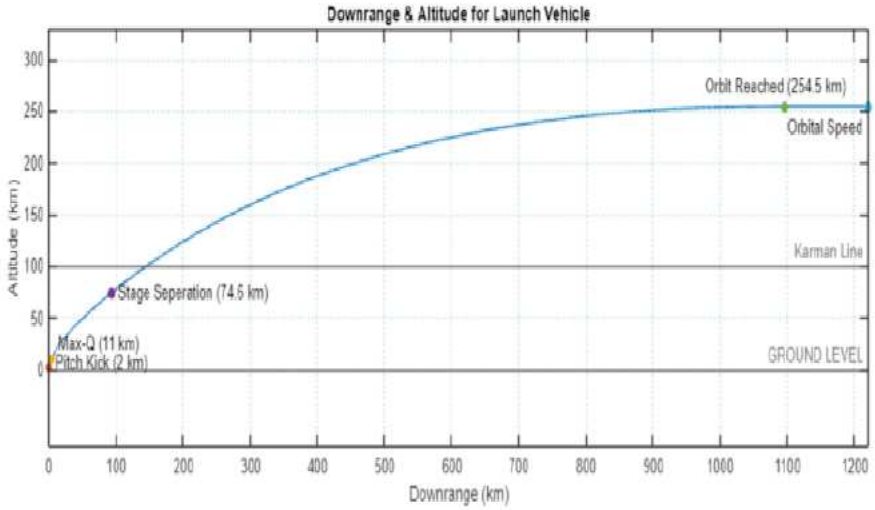


Fig: 3. Altitude vs Downrange (km)

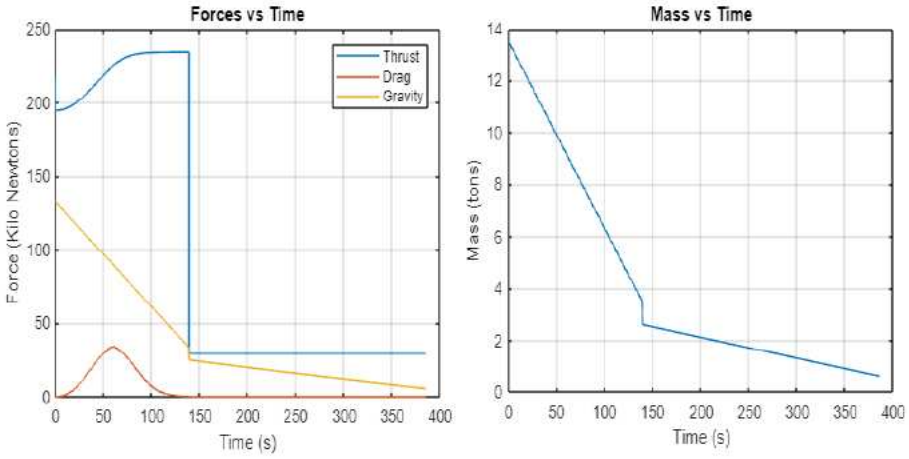


Fig: 4. Variation of forces acting on vehicle wrt time; Mass of vehicle wrt time

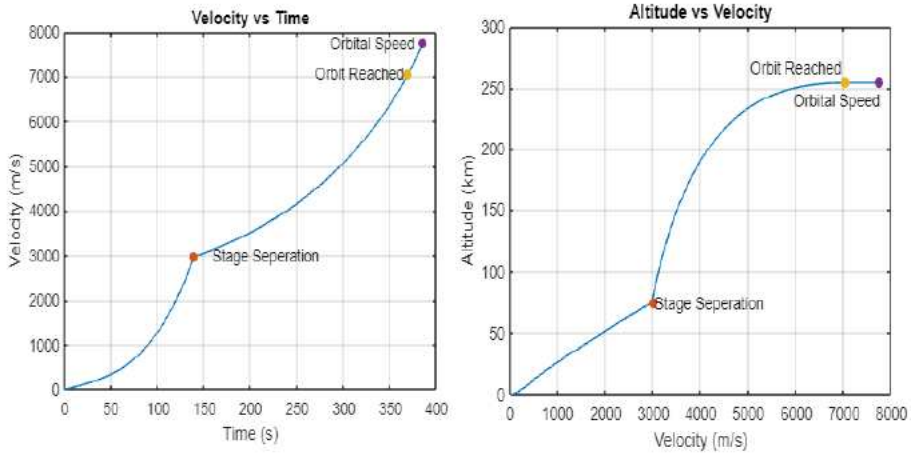


Fig: 5. Variation of velocity wrt time; Altitude achieved wrt velocity

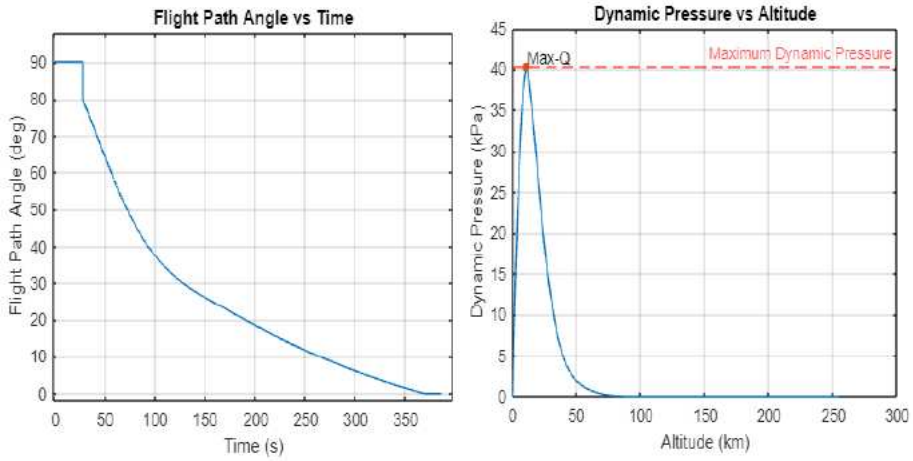


Fig: 6. Flight Path Angle vs Time; Dynamic Pressure vs Altitude

The altitude at which gravity turn initiates and the angle which is given to the vehicle with respect to vertical influences the end state of the vehicle. If the gravity turn is initiated at a higher altitude, more angle should be given to get a similar result. For example, in the table below, 2 combinations of altitude and angle were selected with comparable results.

In figure 5, variation of velocity of the vehicle has been shown with respect to the time (left) and altitude (right). In figure 6 (left), the flight path angle is 90° during the vertical lift-off phase but when pitch kick is given, it sees an instantaneous drop by a specified value. After this drop, the vehicle starts its gravity turn and reaches the value of 0° indicating that the vehicle is parallel to the local horizon.

The Max-Q point, in figure 6 (right) is attained at the altitude of 11 km which represents the phase where the vehicle experiences maximum dynamic pressure. The value of dynamic pressure for the current configuration is 40.373 kN which is very important in the design phase for structural strength and reliability of payload fairing during the first part of the trajectory.

The graph (figure 7) tells that if the angle given at pitch kick is less, while the altitude is constant, the turning of the vehicle be slower which leads to stage separation happening at a relatively higher altitude and allowing the vehicle to reach a higher orbit, though it uses relatively more fuel to reach an orbit and maintain an orbital speed. If we wish to start the pitch kick maneuver at higher altitude, the kick angle would have to be larger in value. A simple example has been highlighted in table 4 which shows that if the maneuver is initiated at 1.5 km, kick angle needed is 9.8^o to match the results of case where altitude to initiate is 1 km and angle taken as 7°.

The following plots have been created using the code which gives a deep insight into the variation of parameters during the entire phase of launch.

From these results, we can safely conclude that even if the altitude or time of pitch kick is decided arbitrarily, we can estimate the value of pitch kick accordingly depending on the final requirements of the mission.

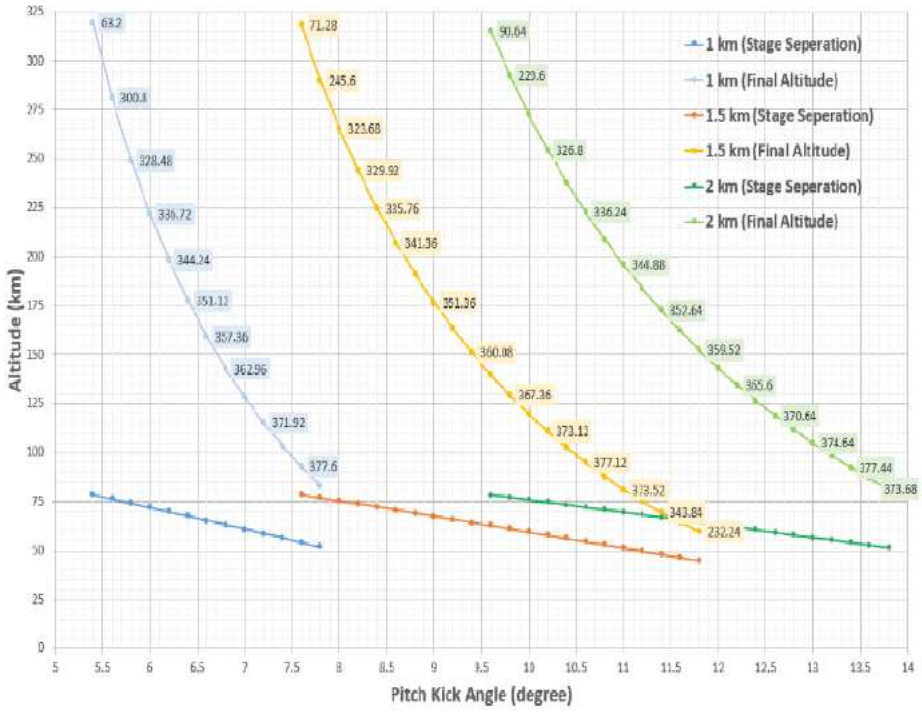


Fig: 7. Study of gravity turn parameters on the altitude of stage separation and final altitude. (Data labels correspond to the final payload mass in kg)

Table 4: Values for 2 combinations of kick angle & pitch kick altitude

Altitude for Pitch Kick [km]	Pitch Kick Angle [degrees]	Stage Separation Altitude [km]	Final Orbit Altitude [km]	Final Payload Mass [kg]
1	7	60.979	128.19	367.84
1.5	9.8	60.973	129.34	367.36

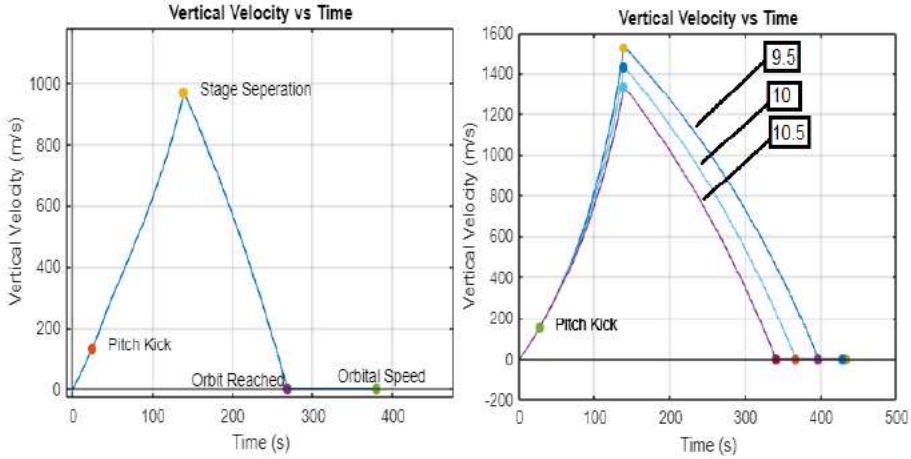


Fig: 8. Vertical Velocity vs time; Effect on change of pitch kick angle

Another parameter that we can see is vertical velocity variation with time (figure 8) which is sensitive to the kick angle. The vertical velocity increases till the sea level engine operates, and after reaching the peak, which is determined by kick angle, the value starts descending. If we create a plot for 3 different pitch kick angles, the plots are coincident till the pitch kick is reached (as for this study, the altitude for the pitch kick is the same). If the pitch angle is higher, the peak of the plot gets lower. [21]

5. CONCLUSION

In the above study we have designed the TSTO launch vehicle operating with a liquid propellant rocket engine, and the key equations which need to be kept in check while designing a vehicle. Relation between equations and various parameters were studied which plays a crucial role in deciding the design and weight of a rocket to inject a payload to desired orbit.

The paper considers various parameters and showcases their relation under a gravity turn maneuver. Important design aspects related to sea level and vacuum engines were also seen during the study and the values obtained using RPA software, were used in the MATLAB code to obtain a trajectory for the vehicle. The effect of gravity turns on the rocket with respect to the pitch kick angle and the altitude at which it is initiated was studied in the paper.

FUTURE SCOPE

The current model uses gravity turn trajectory throughout the flight for a stationary atmosphere and no effect of Earth's rotation. After testing gravity turn trajectory, we will see possibilities of using other trajectories, especially for the free-flight phase and also draw their comparison with the existing model. After understanding the basics of the parameters, the model will be further expanded to incorporate the effects of Earth's rotation, wind model, inefficiencies, optimum staging, stage separation. Another important aspect of designing any launch vehicle is to have a detailed understanding of engine design, specifically what parameters are varied while shifting from sea level engine to vacuum level engines.

REFERENCES

[1] Ankit Kumar Mishra, Kshtij Gandhi, Kuldeep Sharma, Neerumalla Sumanth, Y. Krishna Teja, Conceptual design and Analysis of Two Stage Sounding Rocket, International Journal of Science and Engineering.

[2] Dietrich E. Koelle. Economics of Small Fully Reusable Launch Systems, Acta Astronautica.

[3] M. A. Sharaf and L.A. Alaqal, Computational Algorithm for Gravity Turn Maneuver, Global Journal of Science Frontier Research Mathematics and Decision Sciences.

[4] David W. Callaway, Coplanar Air Launch with Gravity-Turn Launch Turn Launch Trajectories, Air Force Institute of Technology.

[5] David Yaylalia, Launch Dynamics and Gravity Turn Maneuvers, Air Force Institute of Technology.

[6] Dr Gerald M Gregorek, Aerodynamic Drag of Model Rockets, National Association of Rocketry.

[7] L.Brevault, M. Balesdent, R. Wuilbercq, N. Subra, S. Oriol, C. Bonnal, M. Dodds, N. Taylor and R. Varvill, Conceptual design of a Two-Stage-To-Orbit vehicle using SABRE engines, 8th European Conference for Aeronautics and Aerospace Sciences (EUCASS).

[8] Christie Alisa Maddock, Lorenzo Ricciardi, Michael West, Joanne West, Konstantinos Kontis, Sriram Rengarajan, David Evans, Andy Milne, Stuart McIntyre, Conceptual design analysis for a two-stage-to-orbit semi-reusable launch system for small satellites, Acta Astronautica

[9] Byeong-Un Jo, Jaemyung Ahn, Optimal staging of reusable launch vehicles for minimum life cycle cost, Aerospace Science and Technology

[10] Fernando Alonso Zotes and Matilde Santos Penas, Multi-criteria genetic optimization of the maneuvers of a two-stage launcher

[11] Willam E. Wiesel. Spaceflight Dynamics. Aphelion, third edition edition, 2010

[12] Electron Payload User's Guide, 2022

[13] Diogo Marques Gaspar, A Tool for Preliminary Design of Rockets, 2014

[14] Ashish Tewari. Atmospheric and Space Flight Dynamics. Birkhauser, 2006.

[15] US Standard Atmosphere 1976, US Government Printing Office, Washington, D.C., 1976.

[16] K Ramamurthi, Rocket Propulsion

[17] D. Huzel and D. Huang, Modern engineering for design of liquid-propellant rocket engines

[18] SpaceX Falcon User's Guide, 2021

[19] George P. Sutton and Oscar Biblarz, Rocket Propulsion Elements

[20] Cassandre Victoria Marie Pradon, Estimating launch vehicle trajectories and atmospheric emissions, 2021

[21] J. Huh and S. Kwon, A practical design approach for a single-stage sounding rocket to reach a target altitude

[22] Pedro Orgeira-Crespo 1, Guillermo Rey, Carlos Ulloa, Uxia Garcia-Luis, Pablo Rouco and Fernando Aguado-Agelet, Optimization of the Conceptual Design of a Multistage Rocket Launcher.

Citation: Priyanshu Nailwal, Prakash N. Budhaulia, Rahul Kumar and Raj K. Singh, Designing of A Two Stage to Orbit Launch Vehicle, International Journal of Advanced Research in Engineering and Technology (IJARET), 14(3), 2023, pp. 1-13

https://doi.org/10.17605/OSF.IO/HC46T

Article Link:
https://iaeme.com/MasterAdmin/Journal_uploads/IJARET/VOLUME_14_ISSUE_3/IJARET_14_03_001.pdf

Copyright: © 2023 Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Creative Commons license: Creative Commons license: CC BY 4.0

editor@iaeme.com

2024 International Conference on Global Aeronautical Engineering and Satellite Technology (GAST) | 979-8-3503-7159-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/GAST60528.2024.10520762

Detection and Classification of Pulsed Doppler Radar Signals Using Deep Learning

Suhani Grover
Department of Applied Physics
Delhi Technological University
New Delhi, India
suhani1077@gmail.com

Sujal Gupta
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
guptasujal2003@gmail.com

Anup Mandpura
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
amandpura@gmail.com

***Abstract*—Pulsed Doppler radars are commonly used in military applications, especially in aircraft and missile systems, as they are able to detect and track fast-moving targets, such as enemy aircraft, while filtering out ground clutter and slower-moving objects. Classification of modulation types in airborne radar signals is essential for ensuring efficient and reliable communication, target identification, and interference rejection. This research paper presents a comprehensive study on modulation classification techniques for radar signals on an airborne module. The transmitter identity message signal is coded with pseudo-random noise codes and modulated using five modulations, namely, BASK, BFSK, BPSK, QAM and QPSK. Additionally, it is also subjected to Additive White Gaussian Noise(AWGN) and Rayleigh distribution for multipath fading in wireless transmission environments. The paper proposes a modulation recognition scheme based on multi-class classification, hierarchical feature extraction using a convolutional neural network (CNN) and residual neural network (ResNet50) architecture. The simulated result indicates that the proposed method demonstrates adaptability to modulational recognition even in low Signal-to-Noise Ratio (SNR) environments.**

***Index Terms*—Signal Detection, Modulation Classification, Continuous Wavelet Transform (CWT), Convolutional Neural Networks (CNN), Residual Neural Network, Deep Learning**

I. Introduction

Pulse Doppler radars are essential for accurately monitoring the movement and position of objects in both civilian and military domains. Radar waveform recognition is vital function required in electronic warfare systems, such as electronic intelligence (ES/ELINT), electronic protection (EP), and electronic attack (EA) systems. Pulse Doppler radar systems are designed to minimise the risk of detection by employing techniques such as low-power transmissions and frequency spread. This is particularly crucial for surveillance and reconnaissance missions where maintaining stealth is of utmost importance.

In this paper, we conduct an extensive review of the latest advancements in deep learning techniques applied to modulation recognition of pulsed Doppler radars. We begin by examining conventional methods employed in this area and subsequently delve into diverse deep-learning approaches proposed for this task. In [1], the authors used a multi-resolution convolutional neural network (CNN), feature fusion with 128x128 inputs, obtaining a recognition rate of 95.2% at SNR of -8 dB. BPSK, LFM, Costas, Frank, P1, P2, P3, P4, T1, T2, T3, and T4 are considered with varying SNR values between -10 dB and 10 dB. The radar signals are converted into time-frequency images using Fourier-based Synchrosqueezing Transform (FSST), which shows good performance even at low SNRs. Whereas in [2], CWD time–frequency transform and image binarization is used to obtain a 2-dimensional time-frequency image of the signal. The preprocessed time–frequency image is then trained using CNN and sent to the tree structure-based machine learning process optimization classifier [25]. The experimental results show that the system achieves an overall recognition rate of 94.42% when the SNR is -4 dB. In [3], LPI radar signals (LFM, BPSK, Costas, Frank, and T1–T4) are pre-processed using the Choi-Williams distribution time-frequency transform to obtain a 2-D time-frequency image. The pre-processed image is sent to the pre-training model, i.e. Inception-v3 and ResNet-152 for feature extraction with Support Vector Machine (SVM) classifiers. The overall recognition rate of the ResNet-152-SVM system reaches 97.8%, and the Inception-v3-SVM system reaches 96.2% with an SNR of -2 dB. A CNN-based architecture is proposed in [4], where all the waveforms are collected only for a single period, and the input to the CNN is a grayscale time-frequency image. To process a large number of signal samples with reduced computational cost-effectively, a sample averaging technique (SAT) is deployed. The trained CNN is employed for classifying received radar signals, achieving a commendable recognition rate of 95.8% at an SNR of -6 dB. A novel method is proposed in [5] for monostatic, bi-static, and multistatic radar systems for the detection and tracking of targets. The transmitter identity code is embedded with the LFM signal and a Pseudo Noise (PN) sequence is used to code transmitter identity, modulated using BASK. Echo signals received are identified through spectral analysis and target parameters like delay, bistatic angle, and range pertaining to a single target are estimated geometrically with reference to the transmitter separately and consecutive ranges are fused using the maximum likelihood estimation technique. In this paper, we comprehensively explore and evaluate these advanced methodologies, shedding light on the efficacy of deep learning techniques in addressing the intricate challenges of LPI radar waveform recognition.

The paper is organised as follows: Section II provides an

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:13:10 UTC from IEEE Xplore. Restrictions apply.

elaborate review of the system model. In Section III, the signal decomposition and preprocessing is discussed. The deep learning model feature extraction classification are presented in Section IV. Section V discusses the simulation result and discussion namely sample creation, deep learning model training and hyperparameter tuning. Finally, the research is concluded, and future works are stated in Section VI of the research paper.

II. System Model

Bit wise modulated message signals are transmitted through a free space channel, taking into consideration the radar position, velocity and aerial target position. A Doppler shifted frequency and time delay is induced in the signal to account for Rayleigh fading due to interferences in wireless transmission. The signal attenuation is taken into consideration in the form of transmission loss, Rayleigh fading , free channel losses, and AWGN. In a multi-static scenario, we propose a method to introduce the transmitter identity code in the transmitted signal. The transmitter identity code is modulated with an LFM signal using binary amplitude shift keying (BASK) technique. The message signal consists of a bitstream of binary data grouped into four bits.

To encrypt the pulse Doppler radar signal with an identity before transmission, we convert it into a twelve and sixteen-bit binary code using repetition of bits [5]. For example, consider an original four-bit message signal 1011. On applying repetition thrice, we convert it into twelve bits, i.e., 111000111111. While applying repetition four times, we get a sixteen-bit code i.e. 1111000011111111. A pseudorandom noise code is then used to encrypt this message signal.

Consider the pseudorandom noise codes 101101000101 and 1011010001010110. The pseudorandom noise code is XOR-ed with the 12 and 16-bit codes giving 101001000000 and 0011001000100001. The transmitted identity code is embedded with a sinusoidal carrier wave.

Resultant bits are modulated according to Binary Amplitude Shift Keying(BASK), Binary Frequency Shift Keying(BFSK), Binary Phase Shift Keying (BPSK), Quadrature Amplitude Modulation(QAM), and Quadrature Phase Shift Keying(QPSK).

The input signal is a pulsed-doppler waveform with additive white Gaussian noise (AWGN) and Rayleigh multipath fading. The signal can be expressed mathematically as:

$$s(t) = x(t) \cdot h(t) + n(t), \quad 0 \leq t \leq T \tag{1}$$

where $j = \sqrt{-1}$ is the imaginary unit, $x(t)$ is the modulated signal with the phase law $\phi(t)$, $h(t)$ denotes the Rayleigh distribution, and $n(t)$ is the additive white Gaussian noise. .

If the receiver and transmitter are in relative motion, the characteristics of the transmitted signal are negatively influenced by the Doppler effect, depending on its angle of arrival. The simulation takes into account the radar target orientation, position, and velocity as well as the electrodynamics of backscattering for every target.

In any terrestrial environment, due to multipath propagation, the received pulse doppler signal is perceived as an infinite sum of attenuated, delayed and phase-shifted versions of the transmitted signal that may interfere constructively or destructively. The signals are vulnerable to time dispersive effects of the channel and consequently, the frequency response characteristic of the received signal gets distorted by the channel, depending on the amplitude relations and delay differences. This is called delay dispersion or delay spreading.

This signal has line-of-sight (LOS) as well as non-line-of-sight (NLOS) components. When the non-line-of-sight components are large, the faded signal is approximated as a sum of complex Gaussian processes, whose probability density function follows Rayleigh distribution, given by

$$f_{\text{signal}}(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right), \quad r \geq 0 \quad \textbf{[13]} \tag{2}$$

Where σ^2 is the variance of the in-phase and quadrature components, and r is the random variable corresponding to the signal amplitude.

It is assumed that all signals suffer nearly the same attenuation but arrive with different phases. The phase of the complex envelope of the received signal is normally assumed to be uniformly distributed in $[0, 2\pi]$.

III. Signal Decomposition and Preprocessing

In this section, the radar waveform, depicted as a time-frequency image, undergoes binarization and transformation to serve as suitable input for CNNs. The images are transformed into time series graphs, scalograms, and spectrograms to analyse the correlation between temporal and spectral characteristics.

a) Time Series Graph: Time series graphs provide a raw representation of signal amplitude variations over time, making it easy to identify patterns and anomalies.

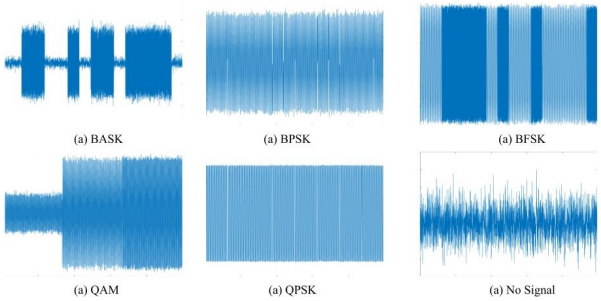


Fig. 1. Time Series Graph with different modulations

b) Scalogram: The scalogram is obtained through continuous wavelet transform (CWT) analysis. CWT is a generalised Fourier transform, designed to provide comparatively good time-frequency resolution. Wavelet functions used in the wavelet transform are chosen as

$$\psi_{k,q}(t) = 2^{k/2}\psi[2^k t - q] \tag{3}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:13:10 UTC from IEEE Xplore. Restrictions apply.

where $k = 0, 1, \ldots, K-1$ and $q = 0, 1, \ldots, 2K-1-1$.

Here, $\psi(t)$ is the mother wavelet. The parameter k introduces time scale compression of the mother wavelet by $2^0, 2^1, \ldots, 2^{K-1}$ times, and q accounts for time shifts of the compressed wavelets. The multiplier $2^{k/2}$ normalizes the square of the wavelet $\psi_{k,q}(t)$ to unity in the interval, considering that the square of the mother wavelet $\psi(t)$ is already normalized.

This enhances the effectiveness of the deep learning algorithms by extracting discriminative features from the radar data. The scalogram offers the advantage of adaptively capturing localised time-frequency features, allowing deep-learning models to learn and exploit the inherent patterns within the radar signal.

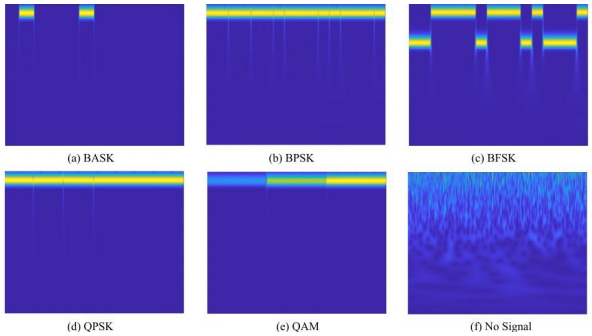


Fig. 2. Scalogram with different modulations

c) Spectrogram: The spectrogram employs a fast Fourier transform to depict the time-frequency representation of the signal. Spectrograms are effective for tasks such as target detection, classification, and anomaly detection. They provide a more intuitive representation of frequency changes over time and are particularly beneficial for capturing short-term variations in signal characteristics. This can enhance the effectiveness of deep learning models in classification, where the temporal evolution of the signal is essential for accurate training and performance.

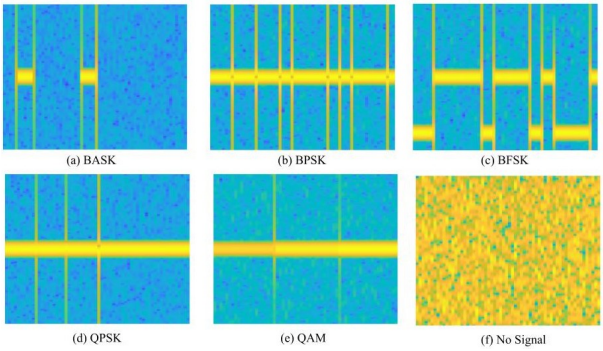


Fig. 3. Spectrogram with different modulations

IV. Deep Learning Model

In this section, we introduce a sophisticated method for improving pre-trained CNN models to enhance image analysis. Our methodology comprises two distinct sub-sections. The first section investigates image fusion to enrich input data and combines a pre-trained hierarchical CNN encoder and ResNet50 model to extract deep features at several levels of abstraction. The second subsection introduces a multi-layer dense classifier that refines decision boundaries and captures finer inter-class variations, culminating in improved classification accuracy.

A. Feature Extraction

The preprocessed images are concatenated into a single image to improve the performance of the CNN model by providing complementary information, enhancing feature learning, and reducing noise. Concatenation handles variability and generalisation, which enables multimodal learning and reduces overfitting. The papers by Zhang et al. (2017) [12], Wang et al. (2020) [14] and Wang et al. (2021) [14] provide a comprehensive overview of the use of image concatenation for deep learning and discuss the different ways in which it can be implemented.

Convolutional neural networks (CNNs) require large datasets to train, but collecting and labelling radar waveform data is time-consuming and expensive. Transfer learning with ResNet50, a pre-trained CNN model, can be used to mitigate this problem. ResNet50 has been shown to be effective for radar waveform identification in studies by [3], [14] and [24].

Extending a pre-trained model like ResNet50 with additional layers in a CNN architecture allows for a more task-specific adaptation, enabling the model to capture intricate patterns and features relevant to the targeted classification problem. Hence, three convolution layers are added after the ResNet50 model, to extract higher-level features, such as shape and texture. This makes the model more discriminative and improves the classification accuracy, thus fine-tuning our feature, for requirement-specific classification.

B. Classification

The features extracted from CNN layers are flattened and passed through dense layers. The flattening layer is used to convert the output of the CNN, which is a multidimensional tensor, into a one-dimensional vector. This vector is then passed through the dense layers, which learn to associate the features extracted by the CNN with the different categories of radar signals.

For classification, a categorical cross-entropy loss function, specifically designed for classification problems is employed. The loss function quantifies dissimilarity between the predicted probability distribution and the true probability distribution of the classes. The objective of the model is to minimise this loss function, achieved by adjusting the weights of the model through the training process.

The formula for the categorical cross-entropy loss function is as follows:

$$H(y, \hat{y}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{4}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:13:10 UTC from IEEE Xplore. Restrictions apply.

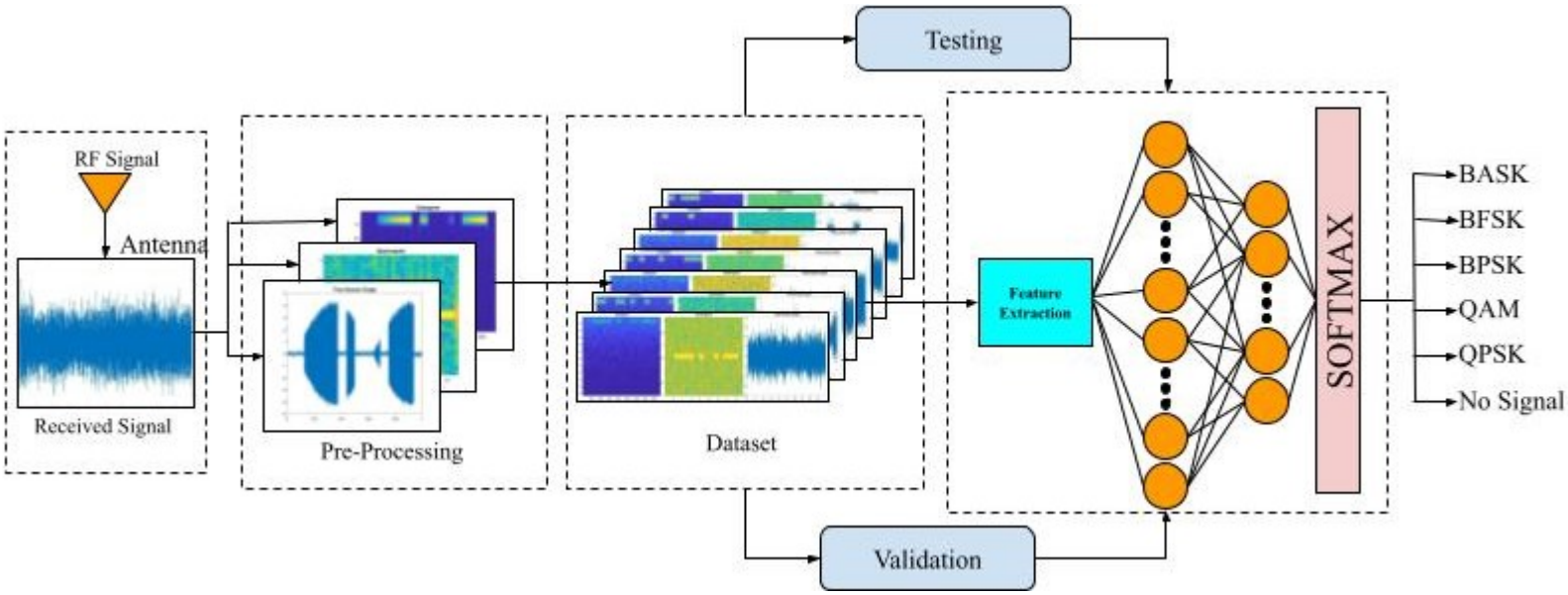


Fig. 4. Deep Learning Model

where y is the true probability distribution of the classes, $\hat{y}$ is the predicted probability distribution of the classes, C is the number of classes, and i is the index of the class.

The model attained high accuracy and demonstrated effective generalization to unseen data. In the following section, we will present the results of the simulation study conducted to evaluate the performance of the model.

V. Experimentation and Results

This section initiates an in-depth discussion of the sample creation and specific parameters selected for this paper. Subsequently, the outcomes of transfer learning on the concatenated images are explored, and hyperparameters employed for training purposes are discussed.

A. Dataset Generation

Creating training data is essential for conducting feature extraction [3] and training the classifier before the system executes classification and recognition tasks. The sample data generated is simulated on MATLAB 2022b. Within this simulation, each parameter of the signal is randomised within the designated range. Stochastic Gradient Descent (SGD) is used for optimisation of the loss calculated by the model. The specific parameters of the radar waveform are outlined in Table I.

In this paper, we simulate five modulation techniques for radar signals, namely BASK, BFSK, BPSK, QAM, and QPSK. The training and test sets are generated based on the parameter configurations of these five signal types. The data sets are created at 5 dB intervals from -20 dB to 20 dB, with every signal type producing 200 samples at each SNR. The Doppler velocity is varied from 0 to 300m/s while the range is kept between 0 to 10km. These samples are then divided into a training and test set in a ratio of 3:1 .The training set comprises

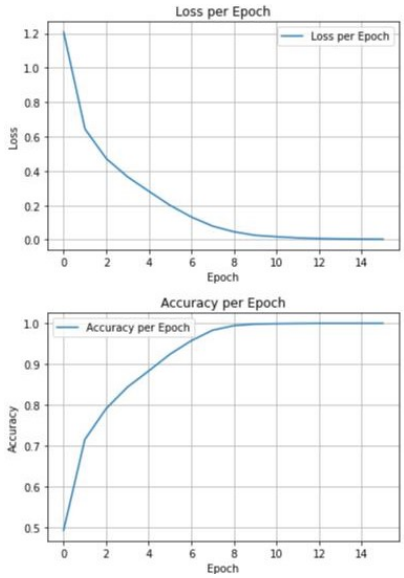


Fig. 5. Loss and Accuracy per Epoch

10,800 samples, and the test set comprises 3,600 samples. The model is constructed using Python's TensorFlow framework, and the GPU used is Nvidia P5000.

B. Deep learning model training and hyperparameter tuning

The proposed deep learning model was trained using the Adam optimizer with a learning rate of 0.0001. The model was trained for 100 epochs and stopped if the validation loss did not improve for 10 consecutive epochs. The following hyperparameters of the model were tuned using a grid search:

- Number of layers in model
- Number of neurons in each layer

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:13:10 UTC from IEEE Xplore. Restrictions apply.

TABLE I
PARAMETERS FOR SAMPLE GENERATION

Radar Signal	Parameter	Uniformly sampled from
BASK	*fc* *fs* *snr*	{*100Hz-5KHz*} [4**fc*] {*20dB, 10dB, -10dB, -20dB*}
BFSK	*fc* *fs* *snr* Δf	{*100Hz-5KHz*} [4**fc*] {*20dB, 10dB, -10dB, -20dB*} *0.75*fc*
BPSK	*fc* *fs* *snr*	{*100Hz-5KHz*} [4**fc*] {*20dB, 10dB, -10dB, -20dB*}
QPSK	*fc* *fs* *snr*	{*100Hz-5KHz*} [4**fc*] {*20dB, 10dB, -10dB, -20dB*}
QAM	*fc* *fs* *snr*	{*100Hz-5KHz*} [4**fc*] {*20dB, 10dB, -10dB, -20dB*}
No Signal	*fc* *fs* *snr*	{*100Hz-5KHz*} [4**fc*] {*20dB, 10dB, -10dB, -20dB*}

- Activation function for each layer
- Dropout rate

The most effective model comprised 5 layers, with 128 neurons in each layer. The activation function for each layer was ReLU, and the dropout rate was 0.5. The model was trained on a dataset of 10,800 samples, partitioned into a training set comprising 8,100 samples and a test set comprising 2,700 samples. The model demonstrated an accuracy of 98.2% on the test set.

TABLE II
PARAMETERS FOR TRAINING CNN

Name	Parameter	Value
SGD	Learning rate Decay Momentum	0.001 10^{-6} 0.9
Early Stop	Monitoring Min. change Patience	Validation loss 0.001 15 epochs
Other	Batch Size Epochs Validation data size Test data size	16 100 15% shuffled data 15% shuffled data

VI. RESULTS

The proposed deep learning model was evaluated on a test dataset of 3600 samples. The model achieved an accuracy of 98.2%, which is significantly higher than the accuracy of the previous model (95%). The proposed model was also able to generalise well to unseen data, as evidenced by its high accuracy on the test set. The model is also relatively simple to implement, which makes it a good choice for practical applications.

TABLE III
NUMBER OF MISCLASSIFIED IMAGES

Architecture	Confusion between modulations	Other confusions	Total
VGG16	*21*	*5*	*26*
ResNet50	*18*	*4*	*22*
Inception-Net	*24*	*7*	*31*
DenseNet	*7*	*4*	*11*
Proposed Model	19	2	21

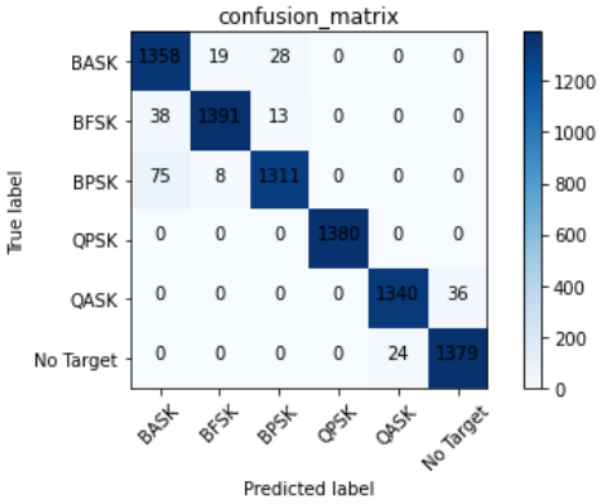


Fig. 6. Confusion Matrix of the results

A. Accuracy Comparison

The accuracy for the proposed model is calculated by the following expression:

$$Accuracy = \frac{True_p + True_n}{True_p + True_n + False_p + False_n} \quad (5)$$

Where p and n represent positive and negative respectively. Table IV displays the test dataset accuracy for both the proposed and previous models. The proposed model exhibits significantly improved test accuracy compared to existing models, indicating its enhanced capability in accurately classifying radar signals when compared to the previous model. The loss and time per epoch are also discussed in the table IV.

TABLE IV
RESULTS FOR TRAINING WITH TRANSFER LEARNING

Architecture	Test Accuracy(%)	Test Loss	Time Per Step(ms)
VGG16	90.30	0.229	194
ResNet50	97.36	0.084	118
Inception-Net	96.61	0.134	114
DenseNet	93.43	0.243	112
Proposed Model	98.83	0.036	121

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:13:10 UTC from IEEE Xplore. Restrictions apply.

B. Factors Contributing to Improved Performance

The improved performance of the proposed model can be attributed to a more sophisticated CNN architecture, specifically ResNet50. This facilitates the model in capturing intricate relationships within the features extracted from radar signals. Additionally, the incorporation of a training dataset comprising 10,800 samples, surpasses the size of the dataset used in training the previous model. The extended dataset provides a more comprehensive learning opportunity, contributing to its improved performance. The model's effectiveness is further substantiated through a comparative analysis with other deep learning models designed for radar signal classification. In this evaluation, it exhibited superior accuracy, outperforming both the VGG16 and AlexNet models.

VII. Conclusion

The results of this study suggest that the proposed deep learning model is a promising approach to radar signal classification. The model is able to achieve high accuracy on a variety of radar signals, and is able to generalise well to unseen data. The model is also relatively simple to implement, making it well-suited for real-world scenarios. Importantly, the model's potential for scalability suggests its adaptability to varying complexities within radar signal datasets.

In our research, a novel method for detecting and classifying pseudo-random noise encrypted and modulated pulsed Doppler radar signals is presented. A comparison to the accuracy results in Table IV is presented in order to demonstrate the performance enhancements of the altered method. This method is applied to detect the signal from an SNR of -20 to +20 dB, with a carrier frequency of 100 Hz to 5 kHz, and classify it based on five modulation schemes.

A. Limitations and Future Work

The proposed model's training relied on signals from a controlled environment, potentially impacting its effectiveness in real-world scenarios. This is addressed by training the model on a larger dataset of radar signals generated in a real-world environment. Additionally the model will be deployed on an FPGA-based architecture for expedited signal detection and modulation recognition.

References

[1] X. Ni, H. Wang, F. Meng, J. Hu and C. Tong, "LPI Radar Waveform Recognition Based on Multi-Resolution Deep Feature Fusion," in IEEE Access, vol. 9, pp. 26138-26146, 2021, doi: 10.1109/ACCESS.2021.3058305.

[2] Wan J, Yu X, Guo Q. LPI Radar Waveform Recognition Based on CNN and TPOT. Symmetry. 2019; 11(5):725. https://doi.org/10.3390/sym11050725

[3] Guo Q, Yu X, Ruan G. LPI Radar Waveform Recognition Based on Deep Convolutional Neural Network Transfer Learning. Symmetry. 2019; 11(4):540. https://doi.org/10.3390/sym11040540

[4] S. -H. Kong, M. Kim, L. M. Hoang and E. Kim, "Automatic LPI Radar Waveform Recognition Using CNN," in IEEE Access, vol. 6, pp. 4207-4219, 2018, doi: 10.1109/ACCESS.2017.2788942.

[5] J. Qumar, S. Christopher and R. Bhattacharjee, "Novel techniques for target detection with waveform identity in multi-static radar using bi-static configuration," 2016 International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, 2016, pp. 1-5, doi: 10.1109/SPCOM.2016.7746703.

[6] Zha X, Peng H, Qin X, Li G, Yang S. A Deep Learning Framework for Signal Detection and Modulation Classification. Sensors. 2019; 19(18):4042. https://doi.org/10.3390/s19184042

[7] C. Wang, J. Tian, J. Cao and X. Wang, "Deep Learning-Based UAV Detection in Pulse-Doppler Radar," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-12, 2022, Art no. 5105612, doi: 10.1109/TGRS.2021.3104907.

[8] F. Yavuz, "Radar Target Detection with CNN," 2021 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 2021, pp. 1581-1585, doi: 10.23919/EUSIPCO54536.2021.9616316.

[9] T. Huynh-The, C. -H. Hua, V. -S. Doan, Q. -V. Pham, T. -V. Nguyen and D. -S. Kim, "Deep Learning for Constellation-based Modulation Classification under Multipath Fading Channels," 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2020, pp. 300-304, doi: 10.1109/ICTC49870.2020.9289413.

[10] M. Sreekantamurthy, D. C. Popescu and R. P. Joshi, "Classification of digital modulation schemes in multipath environments using higher order statistics," SoutheastCon 2016, Norfolk, VA, USA, 2016, pp. 1-6, doi: 10.1109/SECON.2016.7506754.

[11] Zhang X, Li T, Gong P, Liu R, Zha X. Modulation Recognition of Communication Signals Based on Multimodal Feature Fusion. Sensors. 2022; 22(17):6539. https://doi.org/10.3390/s22176539

[12] Zha, Y., et al.: Intelligent identification technology for high-order digital modulation signals under low signal-to-noise ratio conditions. IET Signal Process. e12189 (2023). https://doi.org/10.1049/sil2.12189

[13] Awon, N. T., Islam, M. A., Rahman, M. M., & Islam, A. Z. M. (2012). Effect of AWGN & Fading (Raleigh & Rician) channels on BER performance of a WiMAX communication System. arXiv preprint arXiv:1211.4294.

[14] C. Wang, J. Wang, and X. Zhang, "Automatic radar waveform recognition based on time-frequency analysis and convolutional neural network," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Mar. 2017, pp. 2437–2441.

[15] X. Ni, H. Wang, F. Meng, J. Hu, and C. Tong, "LPI radar waveform recognition based on multi-resolution deep feature fusion," IEEE Access, vol. 9, pp. 26138–26146, 2021.

[16] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[17] Mahafza, B.R. (2022). Radar Systems Analysis and Design Using MATLAB (4th ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003051282

[18] Skolnik, R. (2012). Global Health 101 (2nd ed.). Burlington, MA: Jones & Bartlett Learning.

[19] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and¡ 0.5 MB model size. arXiv preprint arXiv:1602.07360.

[20] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016).Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[21] J. Qumar, S. Christopher and R. Bhattacharjee, "Design of Ambiguity Function for Multi-Dynamic Radar extension of Multi-Static Radar," 2017 14th IEEE India Council International Conference (INDICON), Roorkee, India, 2017, pp. 1-6, doi: 10.1109/INDICON.2017.8487717.

[22] Z. Geng, H. Yan, J. Zhang and D. Zhu, "Deep-Learning for Radar: A Survey," in IEEE Access, vol. 9, pp. 141800-141818, 2021, doi: 10.1109/ACCESS.2021.3119561.

[23] The MathWorks, Inc. "Statistics and Machine Learning Toolbox Documentation" mathworks.com. Accessed: February 07, 2023. [Online]. Available: https://www.mathworks.com/help/stats/index.html.

[24] T. Huynh-The, C. -H. Hua, V. -S. Doan, Q. -V. Pham, T. -V. Nguyen and D. -S. Kim, "Deep Learning for Constellation-based Modulation Classification under Multipath Fading Channels," 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2020, pp. 300-304, doi: 10.1109/ICTC49870.2020.9289413.

[25] Lay, Bunlong & Charlish, Alexander. (2020). Classifying LPI signals with transfer learning on CNN architectures. 10.1109/SSPD47486.2020.9272032.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 20,2024 at 07:13:10 UTC from IEEE Xplore. Restrictions apply.

Check for updates

Received: 2 February 2024 | Revised: 15 April 2024 | Accepted: 15 May 2024

DOI: 10.1002/aoc.7570

Applied Organometallic Chemistry WILEY

FULL PAPER

Development of $Ti_3C_2T_x$-based novel immunosensor for cancer biomarker detection

Sweety | **Devendra Kumar**

Department of Applied Chemistry, Delhi Technological University, Delhi, 110042, India

Correspondence
Devendra Kumar, Department of Applied Chemistry, Delhi Technological University, Delhi 110042, India.
Email: dkumar@dce.ac.in

Funding information
Council Scientific and Industrial Research (CSIR), Government of India, New Delhi, Grant/Award Number: 08/133 [0051]/2020-EMR-I

In this study, an ultrasensitive label-free electrochemical immunosensor based on $Ti_3C_2T_x$ (Mxene) with a 2D-layered morphology has been proposed for the detection of EpCAM antigen. A hydrothermal method is employed for the synthesis of $Ti_3C_2T_x$ by using a less toxic exfoliating reagent, $NaBF_4$. The bioelectrode, BSA/anti-EpCAM/$Ti_3C_2T_x$@ITO, has been fabricated by electrophoretic deposition of $Ti_3C_2T_x$ onto the ITO electrode, followed by an immobilization of EpCAM antibody. Electrochemical response studies reveal that the immunosensor shows a high sensitivity of 29.22 μA fg^{-1} ml cm^{-2} and a wide linear range from 0.1 fg/ml to 100 ng/ml, for EpCAM antigen detection. Further, the good stability of this fabricated immunosensor is an additional advantage for EpCAM antigen detection in serum samples.

KEYWORDS
electrophoretic deposition, epithelial cell adhesion molecules, immunosensor, $Ti_3C_2T_x$

1 | INTRODUCTION

In recent years, cancer has become the second leading reason of death worldwide. Early and sensitive detection of tumor markers is a great strategy for the diagnosis and prognosis of cancer patients.[1,2] Epithelial cell adhesion molecule (EpCAM) is a cell-to-cell adhesion homophilic glycoprotein antigen expressed on circulating tumor cells (CTCs).[3] An average concentration range of CTCs is identified as 1–100 CTCs per 10^7 white blood cells in 1 ml of blood.[4] EpCAM antigen contributes a significant role in the diagnosis and screening of various types of carcinomas arising from the head, neck, prostate, breast, stomach, colon, rectum, pancreas, biliary tract, and hepatic origin.[5] Mostly, immunoassay techniques are frequently used for the diagnosis of biomarkers, which are based on the specific biorecognition of antigens by antibodies.[6] Several techniques, like chemiluminescence immunoassay,[7] enzyme-linked immunosorption assay,[8] radioimmunoassay,[9] and sandwich-type immunosensor[10] have been used for the detection of tumor biomarkers. Although, these techniques give trustworthy results but have some serious limitations, such as requiring highly trained manpower, time-consuming, complex, and expensive.

Recently, biosensors have been considered the most accurate, fast, and sensitive analytical method for biomarker detection.[11] They identify cancer biomarkers by employing biorecognition elements like antibodies, DNA, RNA, etc., which transform biological signals into observable electrical signals.[12] Though numerous varieties of biosensors have been published for the identification of biomarkers.[13] Nonetheless, electrochemical biosensors have been widely used due to their ease of operation, high sensitivity, low cost, and fast response.[14–17] Electrochemical sensors that monitor the interaction of antigens and antibodies are referred to as electrochemical immunosensor. Certain biosensors have been developed in the last couple of years for the detection of epithelial cancer biomarkers.[18–20] Recently, Ying Ma et al. described an aptamer-modified quasi-ZIF-67@Au@methylene blue hybrid for highly sensitive detection of EpCAM.[21] Lina Zhu et al. proposed an electrochemical aptasensor based on CdSe/ZnS QDs for the successful detection of EpCAM.[22] In our previous work, we developed a

 © 2024 John Wiley & Sons Ltd.

biodegradable paper-based electrochemical immunosensor containing CuS/PEDOT:PSS for EpCAM antigen detection.[23] We feel that a modification with nano- or micro-materials is still required to the electrode surface that could offer better performances, such as lowering the detection threshold, increasing the signal, and better electron transport. Further, small sample volumes, fast test times, and tiny transducers are also critical components for the development of point-of-care devices (POCDs). Hence, the most important factor to improve the sensitivity of electrochemical immunosensors is based on the highly active transducer material having a large specific surface area, and numerous active sites.[23]

Different 2D-nanomaterials such as graphitic carbon nitride, graphene oxide, transition metal dichalcogenides etc., have been explored in the field of biosensing.[24,25] Recently, graphene-derived nanohybrids have been used for the fabrication of electrochemical biosensors, because of their high electrical conductivity and adsorption capacity, increased flexibility and mechanical strength, and superior immobilization of biomolecules due to the presence of oxygen-containing functional groups. However, some materials are associated with poor and unpredictable deposition onto the substrate owing to their low dispersibility, which further results in the lack of reproducibility of the biosensor.[26] Additionally, these nanomaterials have a very stable surface, which makes the modification with any functional group very difficult and possesses a major limitation of hydrophobicity.[27]

Recently, an emerging class of novel 2D materials like MXenes possessing extraordinary physiochemical properties has been explored to overcome the aforementioned limitations. MXenes have gained popularity owing to their 2D sheet-like morphology, excellent metal-like conductivity, large specific surface area, superior ability for incorporation of several functional groups, and outstanding ion intercalation behavior.[28] Also, MXenes have been chosen, because it is highly hydrophilic, biodegradable, and biocompatible; 2D-material with rich surface chemistry.[29] These materials belong to a group of transition metal carbide, nitride, or carbonitride; which are obtained by the selective etching of 'A' layers from the MAX phase.[30] Generally, the etching of 'A' layers from the MAX phase results in the surface terminations such as –OH, –F, –O, which provide a favorable microenvironment for the bioreceptor (DNA, enzymes, proteins, etc.) immobilization and help to retain their bioactivity and stability.[31] As of now, MXene-based biosensors have been developed for the detection of CEA, pesticides, phenol, nitrite, and H_2O_2.[30–36] These biosensors display promising results but limit their use to detect only a high concentration of analyte. Ti_3C_2 is one of the most extensively studied MXene, because of its unique electronic properties, excellent hydrophilicity, and high biocompatibility.[37] Therefore, it is considered a suitable candidate to detect very low concentrations of cancer biomarkers. Due to the presence of unpaired electrons on the surface of Ti_3C_2, it can be associated with different functional groups like –OH, –F, and –O, without affecting its metallic conductivity, and represented as $Ti_3C_2T_x$ (T = –OH, –F, –O).[38]

In this study, $Ti_3C_2T_x$ is being used to fabricate an immunosensor for EpCAM antigen detection. $Ti_3C_2T_x$, having robust electrochemical properties is synthesized by hydrothermal method by using a less toxic exfoliating reagent, $NaBF_4$. Electrophoretic deposition of the synthesized material is done onto the indium-tin-oxide (ITO) coated glass substrate. The 2D-layered sheets of $Ti_3C_2T_x$ could increase the active surface area of the modified electrode for the binding of biomolecules. Furthermore, the surface terminations (–OH, –F, –O) may provide biocompatible and aqueous environment for the immobilized antibody, while preserving their stability and bioactivity.[39] The constructed immunosensor, BSA/anti-EpCAM/$Ti_3C_2T_x$@ITO shows a good analytical performance with high sensitivity, quick response, and low detection limit with a broad range of EpCAM antigen detection. This biosensing approach also shows good repeatability, reproducibility, and long-term stability. To explore the applicability of this immunosensor, it is tested with artificial human serum, and data so acquired has been reported in this work.

2 | MATERIALS AND METHODS

2.1 | Materials

Epithelial cell adhesion molecule (EpCAM) antigen, N-(3-dimethyl aminopropyl)-N′-ethylcarbodiimide hydrochloride (EDC), bovine serum albumin factor-V, N-hydroxy succinimide (NHS), artificial human serum, MAX phase (Ti_3AlC_2), sodium tetrafluoro borate ($NaBF_4$), hydrochloric acid (HCl) and buffer salts of analytical grade have been procured from Sigma-Aldrich. Anti-EpCAM (antibody) was purchased from Atlas antibodies. All the solutions were prepared by using Millipore-purified deionized (DI) water.

2.2 | Synthesis of $Ti_3C_2T_x$

2D-$Ti_3C_2T_x$ was synthesized by two different methods viz. co-precipitation and hydrothermal method. In the co-precipitation method, 1 g of LiF was dissolved in 30 ml of 9 M HCl solution under stirring, and then 1 g of

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7570 by Delhi Technological University, Wiley Online Library on [04/06/2024]. 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

MAX phase was added slowly to avoid a violent reaction. This mixture was stirred for 24 h at 35 °C followed by washing with deionized water until the pH value of the supernatant reached higher than 6. Finally, it was filtered and dried under vacuum at 60 °C for 10 h.[40]

In the case of the hydrothermal method, the $Ti_3C_2T_x$ was prepared in a similar manner as reported in our previous work with a little modification.[41] Here, the reaction was carried at 160 °C for 12 h instead of a period of 16 h. It reveals that the results are good with lesser reaction time.

A comparison of both methods reveals that $Ti_3C_2T_x$ synthesized by hydrothermal method possesses a 2D-layered sheet-like structure with no traces of aluminum. This was confirmed by SEM and XRD characterization as clearly described in detail in the results and discussion (Sections 3.1 and 3.2). Hence, $Ti_3C_2T_x$ (hydrothermal) was used in all further studies in the present work.

2.3 | Electrophoretic deposition of $Ti_3C_2T_x$ onto ITO-coated glass substrate

Firstly, $Ti_3C_2T_x$ (50 μg/ml) was ultrasonicated in deionized water to make a colloidal suspension. A series of different potentials were used to optimize the smooth deposition of $Ti_3C_2T_x$ onto the pre-hydrolyzed ITO electrode. A DC potential of 10 V applied for 8 s resulted in the production of a uniform layer of $Ti_3C_2T_x$ on ITO substrate, whereas no film was obtained at a lower potential and a burning of the electrode was seen at a higher potential.

2.4 | Fabrication of $Ti_3C_2T_x$-based immunosensor

$Ti_3C_2T_x$@ITO electrode, as prepared in section 2.3, was incubated with 2 mM EDC (10 μl) and 5 mM NHS (10 μl) at room temperature for 40 min. Subsequently, 20 μl of anti-EpCAM (40 μg/ml) was dropcasted and allowed to immobilize for 12 h at 4 °C. The EDC:NHS, a crosslinker, activated the -COOH groups present in the structure of the antibody, invertibly. The activated carboxylic groups were then combined covalently with the hydroxyl groups available onto the surface of $Ti_3C_2T_x$, thereby producing an ester.[42] The anti-EpCAM/$Ti_3C_2T_x$/ITO electrodes were also treated with 1% BSA followed by washing with PBS to avoid any nonspecific binding with antigen. The fabrication of $Ti_3C_2T_x$-based immunosensor is represented schematically in Figure 1. These fabricated electrodes were stored at 4 °C prior to being used for immunosensing application.

2.5 | Design of the immunosensor

The fabrication of immunosensor is presented schematically in various steps in Figure 1. A novel 2D-layered sheet-like material, $Ti_3C_2T_x$ was prepared by hydrothermal method as depicted in step 1, Figure 1. The surface of $Ti_3C_2T_x$ is associated with different functional groups like –OH, –F, and –O providing a large electrochemically active hydrophilic platform, which further improves the analytical performance of the immunosensor. The

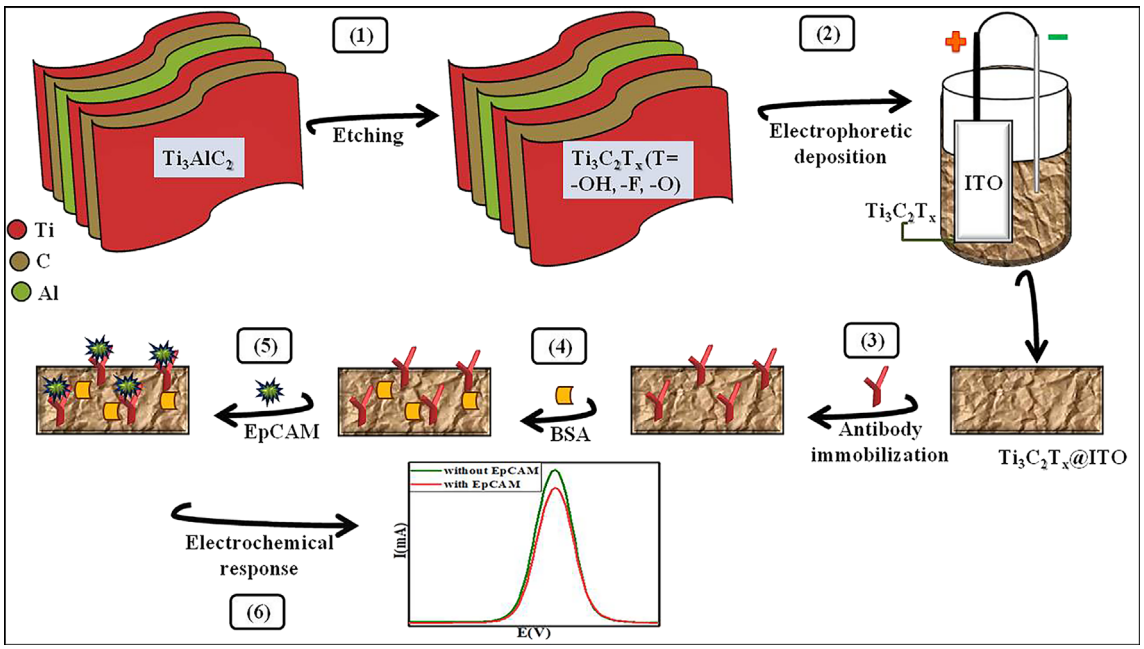


FIGURE 1 Schematic representation of $Ti_3C_2T_x$-based immunosensor for EpCAM antigen detection.

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7570 by Delhi Technological University, Wiley Online Library on [04/06/2024]. 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

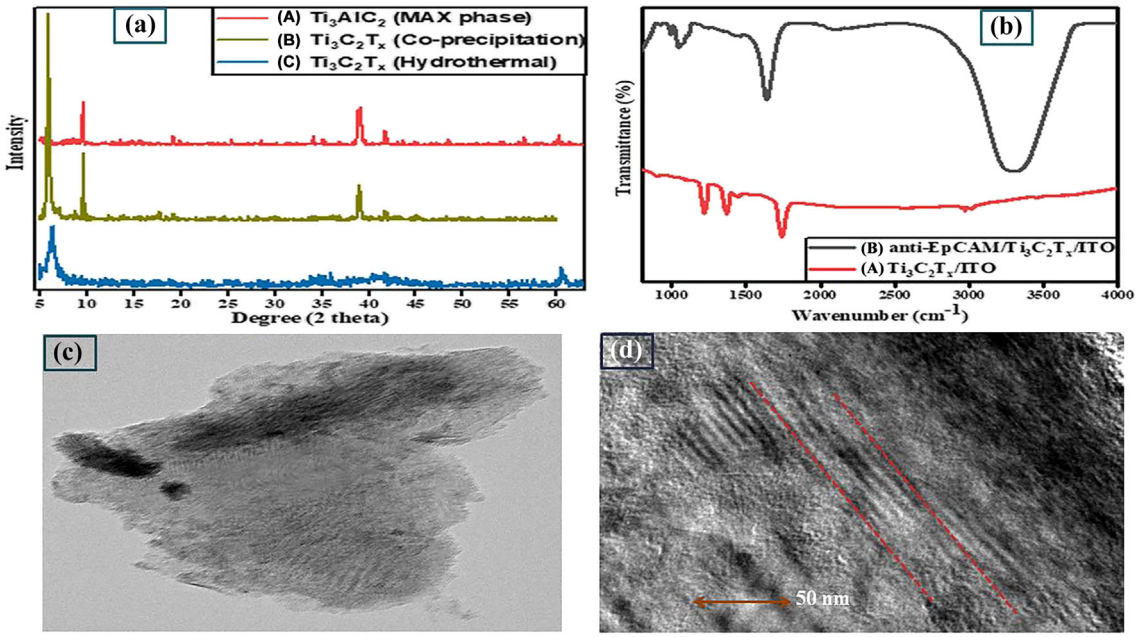


FIGURE 2 (a) XRD spectra, (b) FTIR spectra, (c) TEM image, and (d) cross-sectional image of $Ti_3C_2T_x$ (hydrothermal).

prepared $Ti_3C_2T_x$ was electrophoretically deposited onto the pre-hydrolyzed ITO electrode in the next step as shown in step 2, Figure 1. The functional groups of $Ti_3C_2T_x$ provide a compatible environment for the immobilization of biomolecules. Thus, the fabricated transducer surface was used to immobilize anti-EpCAM by using EDC-NHS as a cross-linker followed by the treatment with BSA, which is represented in steps 3 and 4 of Figure 1. It is worthwhile to mention that there is a formation of covalent bonding between the activated carboxylic groups of antibodies and the hydroxyl groups present on the surface of the transducer. Finally, the developed immunosensor, BSA/anti-EpCAM/$Ti_3C_2T_x$@ITO is utilized for the detection of EpCAM antigen, which is given in step 5, Figure 1. There is a formation of antigen–antibody complex on the surface of immunosensor, which blocks the effective surface area and results in the current decline, as shown in step 6 (Figure 1).

2.6 | Instruments

XRD data were collected by using Bruker D-8 advance monochromatic (Cu K_α) radiation X-ray diffractometer at wavelength 1.5406 A°.

Fourier transform infrared (FTIR) study has been performed by using Perkin Elmer spectrophotometer 2.

The morphological behavior of synthesized material was studied by ZEISS- scanning electron microscope (model- EVO 18) and CYRO- transmission electron microscope (model- TALOS).

The Autolab potentiostat/galvanostat (metrohm, Netherlands) has been employed for all the electrochemical and biosensing investigations. The operation of this instrument may refer to literature.[43]

3 | RESULTS AND DISCUSSION

3.1 | Chemical structure of MXene

XRD spectra of Ti_3AlC_2 (MAX phase), $Ti_3C_2T_x$ (co-precipitation), and $Ti_3C_2T_x$ (hydrothermal) are represented by curves (A), (B), and (C), respectively, in Figure 2a. The spectra of Ti_3AlC_2 (curve A) showed peaks at $2\theta = 9.63°$, 39.14°, 41.90°, and 60.29°. In curves (B) and (C), a dominant diffraction peak of Ti_3AlC_2 at $2\theta = 9.63°$ moved towards a lower angle at $2\theta = 6.36°$, showing that Ti–Al bond is interrupted and it is found in agreement with the literature report.[44] In curve (B), the presence of peaks at 9.63° and 39.14° indicate that the Al has not been etched completely from the Ti_3AlC_2 by the co-precipitation method. On the other hand, in the spectra of $Ti_3C_2T_x$ (hydrothermal), the main peak of aluminum at 39.14° is not observed (curve C), thereby suggesting that the Ti_3AlC_2 has been successfully exfoliated to $Ti_3C_2T_x$ by hydrothermal method. Furthermore, the interlayer spacing (d) is simply calculated by using Bragg's equation, i.e., $2d \sin\theta = n\lambda$. It is found that the interlayer spacing of $Ti_3C_2T_x$ (hydrothermal) increases to 13.88 Å (d-spacing of Ti_3AlC_2 is 9.18 Å), thereby confirming the etching of Al from Ti_3AlC_2.

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7570 by Delhi Technological University, Wiley Online Library on [04/06/2024]. 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

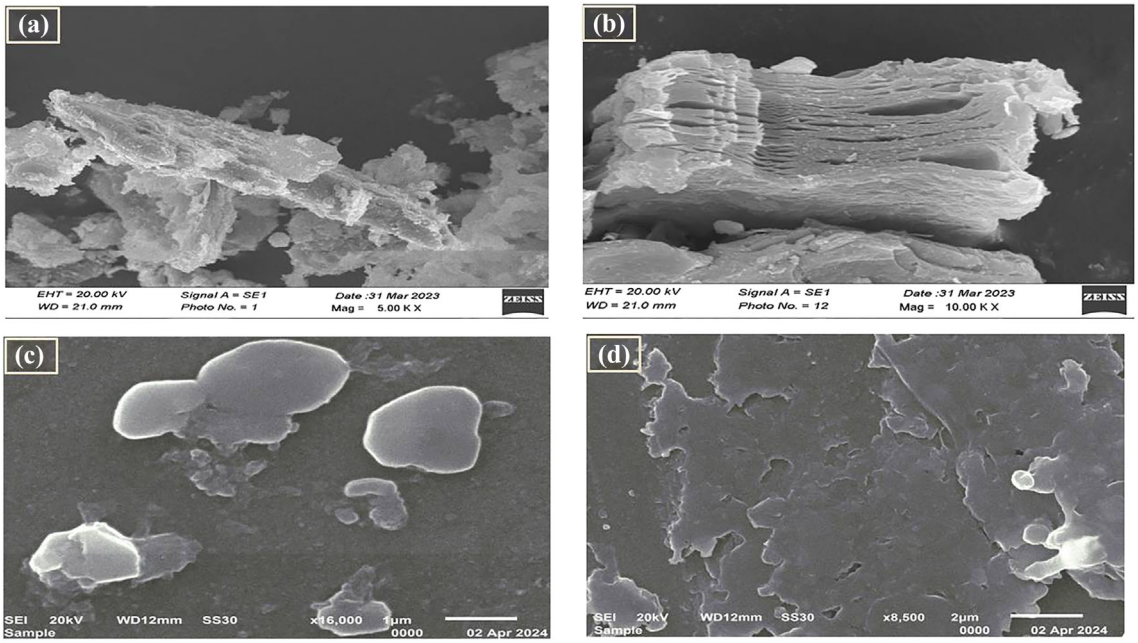


FIGURE 3 SEM images of (a) $Ti_3C_2T_x$ (co-precipitation), (b) $Ti_3C_2T_x$ (hydrothermal), (c) $Ti_3C_2T_x$/ITO electrode, and (d) anti-EpCAM/$Ti_3C_2T_x$/ITO electrode.

FTIR spectra of $Ti_3C_2T_x$/ITO (curve A) and anti-EpCAM/$Ti_3C_2T_x$/ITO (curve B) have been illustrated in Figure 2b. In curve A, the characteristic band at 1216 cm^{-1} is appeared due to the C–F bond stretching vibrations. The peaks at 1366 and 1738 cm^{-1} are attributed to the O–H and C=O bonds stretching vibrations, respectively.[45] On the other hand, FTIR spectra of anti-EpCAM/$Ti_3C_2T_x$/ITO (curve B), show that the characteristic band appeared at 1738 cm^{-1} due to the C=O bond stretching vibrations (curve A) has shifted to 1,638 cm^{-1}, due to formation of ester after immobilization of antibody. The characteristic peak at 1043 cm^{-1} is attributed to the C–O stretching vibrations. The broad peak observed at 3300 cm^{-1} is due to the O–H stretching vibrations.

3.2 | Morphological study of $Ti_3C_2T_x$

$Ti_3C_2T_x$ (hydrothermal) is characterized by transmission electron microscope (TEM) as shown in Figure 2c, where the micrograph represents the two-dimensional sheet-like structure of $Ti_3C_2T_x$. Figure 2d reveals a cross-sectional image of multilayered $Ti_3C_2T_x$. The red dotted lines indicate the individual layers of titanium and carbon atoms present in the structure of $Ti_3C_2T_x$.[46]

The morphology of $Ti_3C_2T_x$ prepared by two different methods was studied by scanning electron microscope (SEM). No layered sheet-like structure is visible in $Ti_3C_2T_x$ prepared by the co-precipitation method in Figure 3a. The layers appeared to be agglomerated during the synthesis process. Figure 3b presents a unique two-dimensional layered structure of $Ti_3C_2T_x$ sheet obtained by hydrothermal method. The loosely packed multilayers provide a large electroactive surface area for antibody immobilization.[47] It reveals that the $Ti_3C_2T_x$ prepared by hydrothermal method possesses better structural characteristics; hence it is used for all further characterization and electrochemical studies.

The SEM image of $Ti_3C_2T_x$/ITO electrode {Figure 3c} shows that 2D sheets of $Ti_3C_2T_x$ are well distributed onto the surface of the ITO electrode. After immobilization of anti-EpCAM onto the $Ti_3C_2T_x$/ITO electrode, the layered morphology of $Ti_3C_2T_x$ has been changed due to the covalent interactions, which further confirms that the $Ti_3C_2T_x$/ITO electrode has been covered by anti-EpCAM, as shown in Figure 3d.

The energy-dispersive X-ray (EDX) spectrum of $Ti_3C_2T_x$ is illustrated in Figure S1. The EDX peaks corresponding to Ti, C, O, F, and Cu also confirm the formation of $Ti_3C_2T_x$. However, an additional peak of Cu is due to the copper grid used during the TEM-EDX analysis.

3.3 | Electrochemical characterization of different electrodes

Electrochemical studies of fabricated electrodes were carried out by using the differential pulse voltammetry (DPV) technique in PBS (7.4 pH) having 5 mM

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7570 by Delhi Technological University, Wiley Online Library on [04/06/2024]. 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

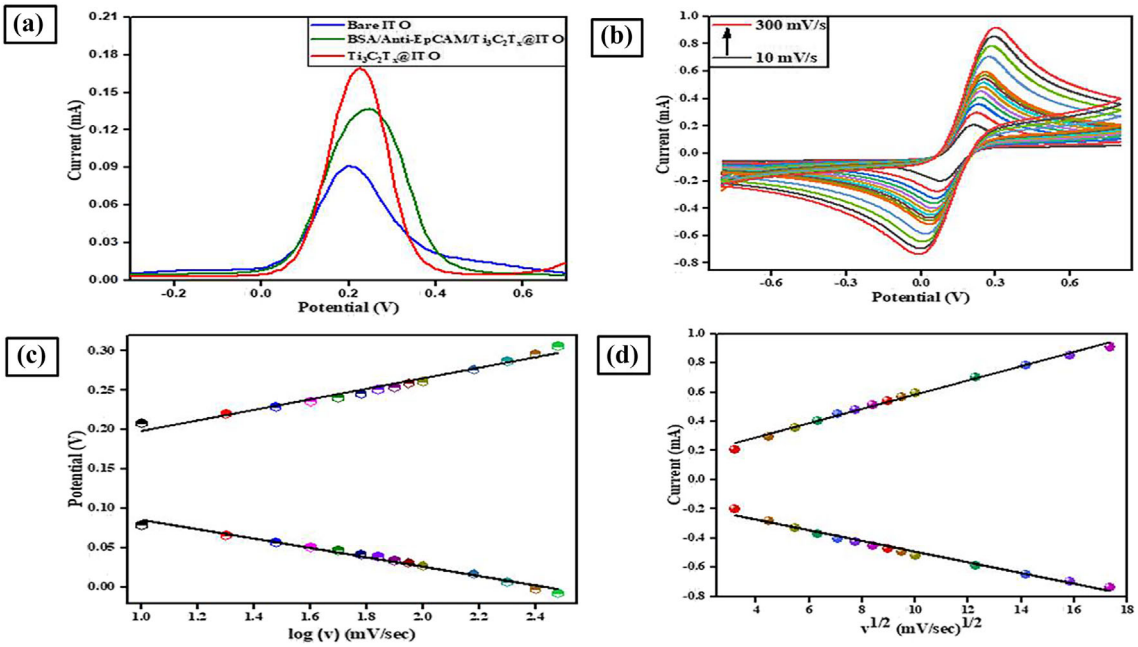


FIGURE 4 (a) DPV peaks of different fabricated electrodes. (b) CV analysis of $Ti_3C_2T_x$@ITO electrode at different scan rates (10–300 mV/sec). (c) Peak potential with log (scan rate) for $Ti_3C_2T_x$@ITO electrode. (d) Peak current with square root of scan rate for $Ti_3C_2T_x$@ITO electrode in PBS (0.2 M, pH = 7.4, NaCl = 0.9%) including $[Fe(CN)_6]^{3-/4-}$ (5 mM).

concentration of $[Fe(CN)_6]^{3-/4-}$. Figure 4a reveals that a bare hydrolyzed ITO electrode shows the least electrochemical current (0.074 mA), while the $Ti_3C_2T_x$@ITO electrode exhibits the maximum value of current (0.169 mA). This increase in current is due to the outstanding electrical conductivity of $Ti_3C_2T_x$. Further, the incorporation of $Ti_3C_2T_x$ increases the electrode's effective surface area, which facilitates fast electron transport, and hence redox peak current is increased significantly. The electrochemical current of the BSA/anti-EpCAM/$Ti_3C_2T_x$@ITO electrode decreases to 0.136 mA, due to the blockage of the conducting area of the electrode by antibody.

Figure 4b illustrates the cyclic voltammetry (CV) analysis of the $Ti_3C_2T_x$@ITO electrode by varying the scan rate (ν) from 10 to 300 mV/s. The CV plots of the ITO electrode and anti-EpCAM/$Ti_3C_2T_x$@ITO electrode are represented in Figures S2(A) and S2(B). The figures clearly show that with an increase in scan rate, the oxidation peak shifts towards a more positive value, while the reduction peak shifts towards a more negative value. It infers that there is a linear relationship between peak potentials and log (scan rate) as shown in Figure 4c, which is also denoted by the following equations (1) and (2):

$$E_{pa}\,[Ti_3C_2T_x@ITO] = 0.0668\,V\,\log(\nu) + 0.1313\,V; R^2 = 0.9720 \quad (1)$$

$$E_{pc}\,[Ti_3C_2T_x@ITO] = -0.05937\,V\,\log(\nu) + 0.1446\,V; R^2 = 0.9836 \quad (2)$$

By using the slope of the above equations, the charge transfer coefficient (α) for the $Ti_3C_2T_x$@ITO electrode is estimated to be 0.893. The heterogeneous electron transfer rate constant (K_s) is found to be 0.127 s^{-1} by using the Laviron equation.[48] Figure 4d represents a linear relationship between anodic and cathodic peak current densities against the square roots of respective scan rates. Hence, it is confirmed that the electron transfer from $Ti_3C_2T_x$@ITO electrode is a surface-controlled process as given in equations (3) and (4).

$$I_{pa}\,[Ti_3C_2T_x@ITO] = 48.90\,\mu A\,(s/mV) \times \nu^{1/2} + 934.48\,\mu A; R^2 = 0.9928 \quad (3)$$

$$I_{pc}\,[Ti_3C_2T_x@ITO] = -36.70\,\mu A\,(s/mV) \times \nu^{1/2} - 128.24\,\mu A; R^2 = 0.9836 \quad (4)$$

Furthermore, slopes of the above equations are used to calculate the effective surface area (A) by using the Randle-Sevick equation as mentioned below:

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7570 by Delhi Technological University, Wiley Online Library on [04/06/2024]. 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

$$I_p = 2.69 \times 10^5\, n^{3/2} A C D^{1/2} \nu^{1/2}$$

Where I_p is the peak current, n = no of electrons oxidized or reduced, C is the concentration of $[Fe(CN)_6]^{3-/4-}$ in mole/cm^3, D = diffusion coefficient in cm^2/s, ν =scan rate in V/s. The effective surface area (A) for ITO, $Ti_3C_2T_x$@ITO, and anti-EpCAM/$Ti_3C_2T_x$@ITO electrode is calculated as 0.22, 0.27, and 0.25 cm^2, respectively.

Also, the surface coverage (τ) of $Ti_3C_2T_x$@ITO electrode is determined as 3.57×10^{-8} mol/cm^2 with the help of equation- [$I_p = n^2F^2A\tau\nu/4RT$]; where F is the Faradays constant and R is the universal gas constant.

3.4 | Optimization of experimental parameters for efficient bioelectrode response

To check the performance of the developed immunosensor for EpCAM antigen detection, different biosensing parameters are optimized with the help of the DPV technique. The viability of the antigen and the corresponding antibody is greatly influenced by the pH of the electrolytic solution. So, the various pH values (6, 6.5, 7, 7.5, 8) were applied to the BSA/anti-EpCAM/$Ti_3C_2T_x$@ITO electrode and a resulting change in response current was measured. Figure S3(A) validates that a 6.5 pH electrolytic solution shows the highest current and hence is taken as an optimal pH for immunosensing studies. The incubation time of EpCAM antigen is another important parameter that impacts the immunosensor's performance. The value of peak current decreases as the time of incubation increases and it reaches a plateau after 8 min, signifying the end of the reaction as shown in Figure S3(B).

3.5 | Electrochemical biosensing studies

DPV is used to investigate the electrochemical performance of BSA/anti-EpCAM/$Ti_3C_2T_x$@ITO electrode for EpCAM antigen detection. The DPV response of bioelectrode after incubation of varying EpCAM antigen concentrations in PBS (6.5 pH) with 5 mM concentration of $[Fe(CN)_6]^{3-/4-}$ is shown in Figure 5a. It reveals that as the concentration of EpCAM antigen increases from 0.1 fg/ml to 100 ng/ml, the current response is found to decrease linearly. The decreasing trend in peak current occurs owing to the combination of antigen and antibody onto the bioelectrode surface that slows down the movement of electrons from the electrolytic solution to the bioelectrode. Figure 5b shows the calibration plot of the

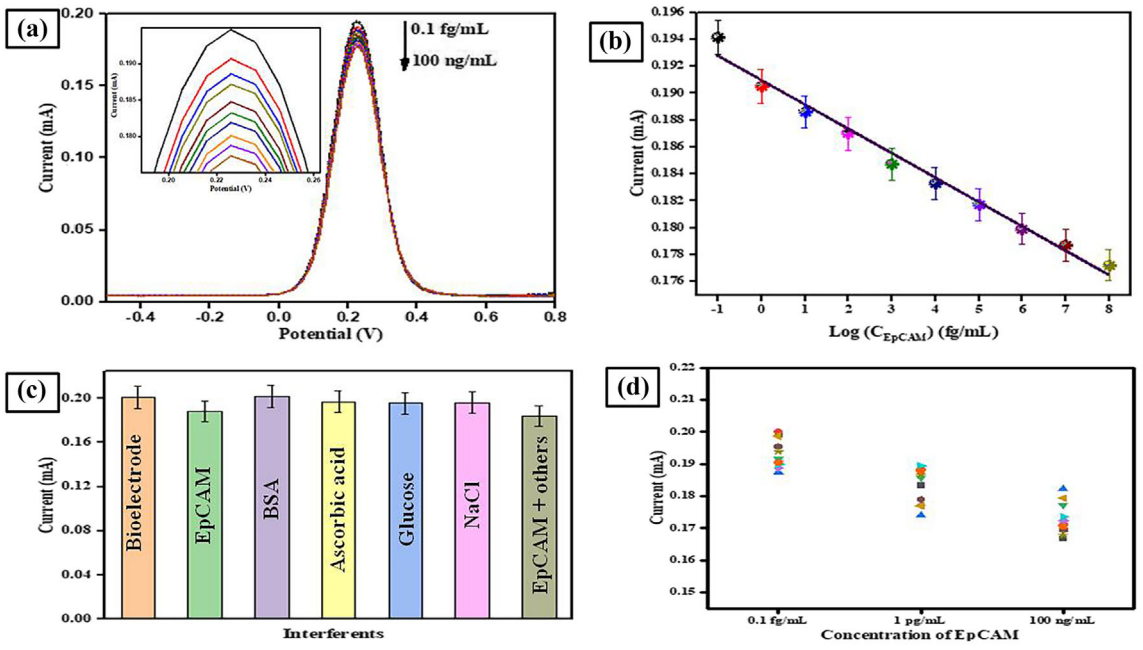


FIGURE 5 (a) Electrochemical response studies of BSA/anti-EpCAM/$Ti_3C_2T_x$@ITO electrode as a function of EpCAM antigen concentration (0.1 fg/ml to 100 ng/ml; top to bottom) in 6.5 pH PBS containing 5 mM $[Fe(CN)_6]^{3-/4-}$ using DPV technique; inset is the magnified view. (b) Plot showing linear relationship between peak current and logarithm of EpCAM antigen concentration. (c) Interference study of different analytes (1 pg/ml). (d) Repeatability of immunosensor.

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7570 by Delhi Technological University, Wiley Online Library on [04/06/2024]. 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 Comparison of recent biosensors with the proposed immunosensor for EpCAM detection.

S. No.	Materials	Detection method	Linear range of detection	Sensitivity	Limit of detection	Reference
1.	MCH, Hp1, AuE	SWV	0.1–20 ng/ml	–	20 pg/ml	[49]
2.	rGO, TiO_2	DPV	0.01–60 ng/ml	3.24 μA ml ng^{-1} cm^{-2}	6.5 pg/ml	[50]
3.	QZIF-67	SWV	0.005–100 ng/ml	–	3 pg/ml	[21]
4.	AgNPs, PVA	EM	2–2000 pg/ml	–	0.8 pg/ml	[51]
5.	CuS, PEDOT:PSS	Chronoamperometry	0.01 pg/ml to 1,000 ng/ml	104.31 μA pg^{-1} ml	0.01 pg/ml	[23]
6.	GRs–QDs	ASV	100 fg/ml – 7.5 μg/ml	–	100 fg/ml	[52]
7.	MoS_2, rGO	DPV	0.001–20 ng/ml	0.104 mA ml ng^{-1} cm^{-2}	44.22 fg/ml	[53]
8.	$Ti_3C_2T_x$	DPV	0.1 fg/ml – 100 ng/ml	29.22 μA fg^{-1} ml cm^{-2}	0.1 fg/ml	Present work

Abbreviations: AgNPs: silver nanoparticles, ASV: anodic stripping voltammetry, AuE: gold electrode, DPV: differential pulse voltammetry, EM: electrochemical microfluidic, GRs: graphene oxide nano-sheets, HP1: hairpin probe 1, MCH: 6-mercapto-1-hexanol, PVA: polyvinyl alcohol, QDs: quantum dots, QZIF-67: quasi zeolitic imidazolate framework-67, SWV: square wave voltammetry.

observed peak current against the logarithm of the concentration of EpCAM antigen (log C_{EpCAM}), which follows equation (5) as given below:

$$I_p = -1.813\,\mu A\,fg^{-1}\,ml \times \log C_{EpCAM} + 190.955\,\mu A; R^2 = 0.9809 \quad (5)$$

The aforementioned regression equation is used to calculate the sensitivity (S) and detection limit (LOD) of the BSA/anti-EpCAM/$Ti_3C_2T_x$@ITO immunosensor. The sensitivity (slope/effective surface area) of the immunosensor having LOD (0.1 fg/ml) is calculated as 29.22 μA fg^{-1} ml cm^{-2}. The recently developed EpCAM detection biosensors are being compared in the light of three parameters, namely, linear detection range, sensitivity, and limit of detection; and the data have been reported in Table 1. The immunosensor of the present work outperforms the other reported biosensors for the detection of EpCAM antigen in terms of all aforementioned parameters. The proposed immunosensor authenticates a very broad range for EpCAM antigen (0.1 fg/ml to 100 ng/ml) detection and significantly high sensitivity in comparison to other biosensors. This is mainly due to the excellent electrical conductivity, large surface area, and outstanding electrochemical properties of $Ti_3C_2T_x$.

3.6 | Immunosensor's selectivity, repeatability, reproducibility, and stability

The selectivity of the fabricated immunosensor is examined by analyzing its electrochemical response with other interfering agents like ascorbic acid, glucose, BSA, and NaCl. These analytes are the representative species of interfering organisms commonly seen in human plasma. It is observed that BSA/anti-EpCAM/$Ti_3C_2T_x$@ITO immunosensor showed a negligible change in current with an average RSD of 1.382%, even with 100 times more concentration of interfering species (1 pg/ml) as compared to EpCAM antigen (0.01 pg/ml). Furthermore, the specificity is examined by testing a mixture of all four analytes- containing EpCAM antigen, which shows a positive response for the detection of EpCAM as depicted in Figure 5c. Hence, it is inferred that the fabricated immunosensor is selective and specific for the detection of EpCAM antigen.

Reproducibility and repeatability are the crucial characteristics that demonstrate the immunosensor's analytical performance. In order to investigate repeatability, the DPV response of fabricated immunosensor has been recorded 10 times with each concentration of EpCAM antigen, viz., 0.1 fg/ml (low), 1 pg/ml (medium), and 100 ng/ml (high) as shown in Figure 5d. The RSD values for the anti-EpCAM/$Ti_3C_2T_x$@ITO immunosensor are calculated as 2.398%, 3.032%, and 2.909%, for 0.1 fg/ml, 1 pg/ml, and 100 ng/ml EpCAM antigen, respectively. The low values of relative standard deviations indicate that the fabricated immunosensor reveals good reproducibility and repeatability.

Finally, the long-term stability is measured with 10 fg/ml EpCAM antigen for 6 weeks as shown in Figure S4(A). The DPV response is recorded weekly and it is observed that the electrode retains 90% of its initial response after 5 weeks. After that, the current response remains at 77%, confirming the acceptable stability of the

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7570 by Delhi Technological University, Wiley Online Library on [04/06/2024]. 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 EpCAM antigen detection in human serum sample ($n = 3$).

S.No.	EpCAM concentration added (fg/ml)	EpCAM concentration found (fg/ml) (SD < 2.5%)	Recovery (%) (SD < 2%)	RSD (%)
1.	1	0.97	97.94	1.469
2.	10	9.91	99.16	0.597
3.	100	99.15	99.15	0.602
4.	1,000	999.40	99.94	0.038
5.	10,000	9989.00	99.89	0.077

immunosensor. It is believed that this significantly improved stability of the immunosensor is due to the excellent chemical stability of the $Ti_3C_2T_x$@ITO electrode.

3.7 | Human serum analysis

The clinical applicability of the fabricated immunosensor is explored by using artificial human serum. The human serum is diluted with PBS and spiked with EpCAM antigen (1, 10, 100, 1,000, 10,000 fg/ml). The DPV response of BSA/anti-EpCAM/$Ti_3C_2T_x$@ITO electrode incubated with different spiked concentrations is shown in Figure S4(B). The recoveries (97.93 to 99.94%) for the different serum sample analyses are reported in Table 2. Furthermore, RSD is found to be less than 1.5% suggesting that the developed immunosensor poses an excellent potential for EpCAM antigen detection.

4 | CONCLUSION

A $Ti_3C_2T_x$-based ultrasensitive electrochemical immunosensor capable of label-free detection of EpCAM antigen is reported in this study. The immunosensor shows remarkable results owing to a high specific area, superior electronic conductivity, good dispersion in the aqueous phase, and high biocompatibility of $Ti_3C_2T_x$. This immunosensor has a wide linear range from 0.1 fg/ml to 100 ng/ml and a high sensitivity of 29.22 μA fg^{-1} ml cm^{-2} for EpCAM antigen detection. Additionally, the accuracy and applicability of the immunosensor are validated by quick EpCAM detection in spiked human serum. Since the Mxene exhibits a great potential for effective immobilization of biomolecules, the suggested platform can be used for the detection of other biomarkers, which could revolutionize cancer detection.

AUTHOR CONTRIBUTIONS

Sweety: Data curation; validation; conceptualization; methodology; software; formal analysis; investigation; writing - original draft. **Devendra Kumar**: Supervision; writing - review and editing

ACKNOWLEDGMENTS

Sweety acknowledges the Council Scientific and Industrial Research (CSIR), Government of India, New Delhi, for their financial support (File no: 08/133 [0051]/2020-EMR-I).

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

Data can be made available on the basis of reasonable request.

FUNDING STATEMENT

This research did not receive any specific grant from funding agencies in public, commercial, or not-for-profit sectors.

ETHICS APPROVAL STATEMENT

Artificial human serum procured from Sigma Aldrich is used.

PATIENT CONSENT STATEMENT

N.A

ORCID

Sweety https://orcid.org/0009-0007-6498-6947
Devendra Kumar https://orcid.org/0000-0001-9118-2070

REFERENCES

[1] S. K. Arya, K. Y. Wang, C. C. Wong, A. R. A. Rahman, *Biosens. Bioelectron.* **2013**, *41*, 446.
[2] H. Amir, V. Subramanian, S. Sornambikai, N. Ponpandian, C. Viswanathan, *Bioelectrochemistry* **2024**, *155*, 108589.
[3] J. Macdonald, J. Henri, K. Roy, E. Hays, M. Bauer, R. N. Veedu, N. Pouliot, S. Shigdar, *Cancers* **2018**, *10*, 19.
[4] S. H. Hussain, C. S. Huertas, A. Mitchell, A.-L. Deman, E. Laurenceau, *Biosens. Bioelectron.* **2022**, *197*, 113770.

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7570 by Delhi Technological University, Wiley Online Library on [04/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[5] M. Hasan, M. Ahommed, M. Daizy, M. Bacchu, M. Ali, M. Al-Mamun, M. A. S. Aly, M. Khan, S. I. Hossain, *Biosens. Bioelectron.: X* **2021**, *8*, 100075.
[6] S. Jeong, M.-J. Park, W. Song, H.-S. Kim, *Clin. Biochem.* **2020**, *78*, 43.
[7] A. Liu, F. Zhao, Y. Zhao, L. Shangguan, S. Liu, *Biosens. Bioelectron.* **2016**, *81*, 97.
[8] S. Zhao, H. Yang, M. Zhang, D. Zhang, Y. Liu, Y. Liu, Y. Song, X. Zhang, H. Li, W. Ma, *Cell Biochem. Biophys.* **2013**, *65*, 263.
[9] W. Brummund, D. A. Aryan, M. T. Mennuti, N. A. Starkovsky, *Clin. Chim. Acta* **1980**, *105*, 25.
[10] G. Wang, Y. Qing, J. Shan, F. Jin, R. Yuan, D. Wang, *Microchim. Acta* **2013**, *180*, 651.
[11] A. Kaushal, N. Kaur, S. Sharma, A. K. Sharma, D. Kala, H. Prakash, S. Gupta, *Vaccines* **2022**, *10*, 2138.
[12] G. Yang, Z. Xiao, C. Tang, Y. Deng, H. Huang, Z. He, *Biosens. Bioelectron.* **2019**, *141*, 111416.
[13] V. Naresh, N. Lee, *Sensors* **2021**, *21*, 1109.
[14] S. Yadav, N. Kashaninejad, M. K. Masud, Y. Yamauchi, N.-T. Nguyen, M. J. Shiddiky, *Biosens. Bioelectron.* **2019**, *139*, 111315.
[15] F. S. Felix, L. Angnes, *Biosens. Bioelectron.* **2018**, *102*, 470.
[16] E. Sciurti, M. Signore, L. Velardi, R. Di Corato, L. Blasi, A. Campa, M. Martucci, P. Siciliano, L. Francioso, *Talanta* **2024**, *271*, 125726.
[17] S. Paneru, D. Kumar, *J. Appl. Electrochem.* **2024**, 1.
[18] J. Li, Q. Jiang, M. Chen, W. Zhang, R. Liu, J. Huang, Q. Xu, *Anal. Chim. Acta* **2024**, *1287*, 342055.
[19] F. Wang, Y. Zhang, D. Chen, Z. Zhang, Z. Li, *Sens. Actuators B* **2021**, *341*, 130034.
[20] J. Xu, X. Wang, C. Yan, W. Chen, *Sensors* **2019**, *19*, 1879.
[21] Y. Ma, X. Shi, Y. Ding, X. Zhang, J. Lu, D. Sun, *Chem. Eng. J.* **2024**, *483*, 149362.
[22] L. Zhu, B. Yang, K. Qian, L. Qiao, Y. Liu, B. Liu, *J. Electroanal. Chem.* **2020**, *856*, 113655.
[23] S. Paneru, D. Kumar, *Mater. Chem. Phys.* **2024**, *313*, 128687.
[24] X. Chen, X. Jia, J. Han, J. Ma, Z. Ma, *Biosens. Bioelectron.* **2013**, *50*, 356.
[25] S. Verma, C. M. Pandey, D. Kumar, *New J. Chem.* **2022**, *46*, 21190.
[26] S. Mukherjee, A. Mukherjee, Z. Bytesnikova, A. M. Ashrafi, L. Richtera, V. Adam, *Biosens. Bioelectron.* **2024**, *250*, 116050.
[27] H. Zhang, Z. Wang, F. Wang, Y. Zhang, H. Wang, Y. Liu, *Talanta* **2021**, *224*, 121879.
[28] X. Sheng, Y. Zhao, L. Zhang, X. Lu, *Compos. Sci. Technol.* **2019**, *181*, 107710.
[29] V. P. Gajdosova, L. Lorencova, P. Kasak, M. Jerigova, D. Velic, L. Orovcik, M. Barath, P. Farkas, J. Tkac, *Anal. Chim. Acta* **2022**, *1227*, 340310.
[30] S. Kumar, Y. Lei, N. H. Alshareef, M. Quevedo-Lopez, K. N. Salama, *Biosens. Bioelectron.* **2018**, *121*, 243.
[31] L. Wu, X. Lu, Z.-S. Wu, Y. Dong, X. Wang, S. Zheng, J. Chen, *Biosens. Bioelectron.* **2018**, *107*, 69.
[32] H. Liu, C. Duan, C. Yang, W. Shen, F. Wang, Z. Zhu, *Sens. Actuators B* **2015**, *218*, 60.
[33] F. Wang, C. Yang, M. Duan, Y. Tang, J. Zhu, *Biosens. Bioelectron.* **2015**, *74*, 1022.
[34] F. Shahzad, A. Iqbal, S. A. Zaidi, S.-W. Hwang, C. M. Koo, *J. Ind. Eng. Chem.* **2019**, *79*, 338.
[35] W. Xu, M. Sakran, J. Fei, X. Li, C. Weng, W. Yang, G. Zhu, W. Zhu, X. Zhou, *ACS Biomater Sci. Eng.* **2021**, *7*, 2767.
[36] F. Wang, C. Yang, C. Duan, D. Xiao, Y. Tang, J. Zhu, *J. Electrochem. Soc.* **2015**, *162*, B16. https://doi.org/10.1149/2.0371501jes
[37] L. Chen, X. Shi, N. Yu, X. Zhang, X. Du, J. Lin, *Materials* **2018**, *11*, 1701.
[38] R. Liu, W. Li, *ACS Omega* **2018**, *3*, 2609.
[39] X. Lu, Z. Wen, J. Li, *Biomaterials* **2006**, *27*, 5740.
[40] D. Zhang, Q. Mi, D. Wang, T. Li, *Sens. Actuators B* **2021**, *339*, 129923.
[41] D. Kumar, *J. Colloid Interface Sci.* **2023**, *652*, 549.
[42] G. T. Hermanson, *Bioconjugate techniques*, Academic press **2013**.
[43] S. Paneru, D. Sweety, *J. Appl. Electrochem.* **2023**, *53*, 2229.
[44] C. Yao, W. Zhang, L. Xu, M. Cheng, Y. Su, J. Xue, J. Liu, S. Hou, *Sep. Purif. Technol.* **2021**, *263*, 118365.
[45] J. Chen, X. Yuan, F. Lyu, Q. Zhong, H. Hu, Q. Pan, Q. Zhang, *J. Mater. Chem. A* **2019**, *7*, 1281.
[46] J. Choi, Y. J. Kim, S. Y. Cho, K. Park, H. Kang, S. J. Kim, H. T. Jung, *Adv. Funct. Mater.* **2020**, *30*, 2003998.
[47] D. P. Jena, S. Anwar, R. Parida, B. Parida, N. C. Nayak, *J. Mater. Sci. Mater. Electron.* **2021**, *32*, 8081.
[48] K. Li, Y. Li, L. Wang, L. Yang, B. Ye, *Arab. J. Chem.* **2019**, *12*, 2732.
[49] Q. Chen, W. Hu, B. Shang, J. Wei, L. Chen, X. Guo, F. Ran, W. Chen, X. Ding, Y. Xu, *Microchim. Acta* **2018**, *185*, 1.
[50] O. Jalil, C. M. Pandey, D. Kumar, *Microchim. Acta* **2020**, *187*, 1.
[51] K. Bravo, F. G. Ortega, G. A. Messina, M. I. Sanz, M. A. Fernández-Baldo, J. Raba, *Clin. Chim. Acta* **2017**, *464*, 64.
[52] M. J. Shiddiky, S. Rauf, P. H. Kithva, M. Trau, *Biosens. Bioelectron.* **2012**, *35*, 251.
[53] O. Jalil, C. M. Pandey, D. Kumar, *Bioelectrochemistry* **2021**, *138*, 107733.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Sweety, D. Kumar, *Appl Organomet Chem* **2024**, e7570. https://doi.org/10.1002/aoc.7570

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7570 by Delhi Technological University, Wiley Online Library on [04/06/2024]. 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

Environ Monit Assess (2024) 196:543
https://doi.org/10.1007/s10661-024-12681-4

RESEARCH

Diurnal rail noise measurement, analysis, and evaluation of associated health impacts on residents living in the proximity of rail track area

Kranti Kumar · Arun Bhartia · Rajeev Kumar Mishra · Ravi Pratap Singh Jadon · Jitendra Kumar

Received: 7 October 2023 / Accepted: 27 April 2024
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

Abstract In India, railway is the major transportation mode for carrying goods and people. The tracks for the movement of the rail were initially constructed in the city for the pre-eminence and expediency of the vantage of the people. Rapid modernization and increasing population in the city crammed the area around the railway tracks. Moving rail on the tracks passing through the city is not compatible, which is creating problems for the nearby residents. In the urban and suburban regions, the railway noise has become a major problem. This study was conducted to examine the perception of the physiological and psychological effects of railway noise in the nearby areas of railway stations in Delhi, India. For this purpose, 10 sites near the railway station were selected for the study. To assess the impact of railway noise pollution on the health of humans, a questionnaire survey was conducted. The data of 344 individuals were collected through the questionnaire survey and analyzed to get the perception towards railway noise. Noise level was monitored by a Sound Level Meter (SLM) and the equivalent noise level (L_{eq}) in dB(A) was used to compute the noise pollution in three shifts, i.e., morning, noon, and evening time. Results showed that 57.65% of female and 86.11% of male respondents in the survey reported the disturbance due to railway noise. The level of noise pollution was found higher in the evening time as compared to the noon and morning period, which exceeds the limit set by the Central Pollution Control Board (CPCB) at all the monitored locations. Findings of the study show that the primary cause of the health problems is railroad noise, which is negatively impacting the health of the residents, who are living in the proximity of the rail track region. The perception survey reported that headache, sleep disturbance, irritation, and stress are common health issues among the locals residing around the railway track proximity in Delhi.

Keywords Noise pollution · L_{eq} · Health effects · Questionnaire survey · Rail traffic noise

K. Kumar (✉) · A. Bhartia
School of Liberal Studies, Dr. B. R. Ambedkar University Delhi, Delhi 110006, India
e-mail: kranti31lu@gmail.com

K. Kumar
Department of Mathematics, Central University of Himachal Pradesh, Dharamshala 176206, India

R. K. Mishra · R. P. S. Jadon
Department of Environmental Engineering, Delhi Technological University, Delhi 110042, India

J. Kumar
Department of Mathematics, Central University of Haryana, Mahendragarh 123031, India

Introduction

Transportation plays a significant role in civilization, economic affluence, and interconnectivity (Mohammadi, 2009). Different transportation modes (rail, road,

Published online: 13 May 2024

air, etc.) have one consequence, i.e., generation of noise, which is influencing the living and working environment (Asdrubali & D'Alessandro, 2018; S. A. Stansfeld & Matheson, 2003). The noise generated from different kinds of transport medium, i.e., rail, road, and air, is considered a health hazard, which is majorly disturbing sleep and creating annoyance (Torija et al., 2011; Wrótny & Bohatkiewicz, 2021). Long-term exposure of the inhabitants to excessive noise is harmful to human health, having an adverse impact on numerous organs in addition to the auditory system (Petri et al., 2021; Wosniacki & Zannin, 2021). Different epidemiological and experimental studies show the determinantal effect of noise on health particularly regarding annoyance, blood pressure, and irritation (Licitra et al., 2016; Schreckenberg et al., 2012; R. Thompson et al., 2022). Few studies also have been reported in the literature showing evidence of metabolic disease and mental health hazards (Cerletti et al., 2020; Vincens & Persson Waye, 2022). Therefore, it is important to focus on lowering the impact of noise on the human population. Because each person's sensitivity to noise is unique (both subjectively and objectively), determining the detrimental effects of noise on people may be difficult. With the determination of appropriate criteria and their upper allowable values, generally recognized scientific techniques for its assessment were created (Lercher et al., 2017). In developing countries like India, the common noise sources are various construction activities, industries, and road traffic (Kumar, 2019). Due to fast infrastructure expansion, economic development, travel, and tourism, the city faces a problem of rail-induced noise pollution (Dupont, 2017). One of the greatest acceptable modes of transportation in India is the railroad. Continual, inescapable noise levels are putting a burden on the city's residents' auditory, non-auditory, and sensory systems (S. A. Stansfeld & Matheson, 2003). In recent years, due to sustainable alternative transport, rail traffic has drastically increased as a convenient way of transportation mode rather than the road and airway as it is inducing less environmental noise compared to other transport mechanisms (Perron et al., 2016). The high-speed trains are creating more noise, which may vary according to the frequency of passage (Poisson, 2015; Sarikavak & Boxall, 2019). The frequency and speed of the train were directly identified as major contributors to noise annoyance (Aasvang et al., 2008), which increases the risk of disease related to blood pressure and sleep disturbance (Zou et al., 2015). Numerous studies have demonstrated that exposure to different types of traffic noise in densely populated urban areas may adversely impact human health (Aasvang et al., 2008; Lercher et al., 2010; Maljaee et al., 2024; Sørensen et al., 2011). This environmental factor may lead to various health issues, including headache, sleep disturbance, psychiatric disorder, decreased concentration, hearing impairment, learning difficulties, stroke, hypertension, and, ultimately, a diminished overall quality of life (Guski, 2016; Lercher et al., 2008; Michaud et al., 2022; Moreno et al., 2023; Öhrström et al., 2007). Railway traffic noise constitutes a major environmental hazard, as mentioned in an extensive study conducted by Fields & Walker (1982). The study reported the assessment of the public's response towards railway noise in residential areas across Great Britain and arrived at the conclusion that "Noise is considered the most severe environmental disturbance attributable to railway operations." A study examining the community's reaction to high-speed railway noise was conducted in Tianjin, China. Residents living near various railway lines were interviewed about their annoyance and interference with daily activities due to the noise. Findings revealed that as the high-speed railway noise levels increased, so did the proportion of Chinese residents experiencing high noise annoyance and activity disruption. The length of a single noise event appeared to impact activity disturbance. Additionally, factors beyond just sound characteristics, such as noise sensitivity, satisfaction with the living environment, and housing satisfaction, significantly influenced the community's response to high-speed railway noise (Zhang & Ma, 2022). Research conducted in the railway work facility settings in Indonesia highlights the potential health and social issues that may arise from exposure to high levels of noise. Disorders, including nausea, insomnia, and masking effect, can negatively impact the well-being of individuals working or living near these areas (Damayanti et al., 2023). A field study in the UK involved interviews with participants and measurements of their exposure to vibrations from two sources of environmental vibration, specifically freight and passenger railway vibrations. By employing ordinal probit modeling with fixed thresholds to establish exposure–response relationships, it was found that individuals can differentiate between the two types of railway vibration. Moreover, the annoyance response to freight railway vibration was significantly higher than that to passenger railway vibration, even when considering

equal levels of exposure. The study also revealed that the population under investigation exhibited a community tolerance level 15dB higher for passenger railway vibration compared to freight railway vibration (Sharp et al., 2014). A questionnaire survey and on-site measurements were conducted in the railway station area of Bangladesh to assess the impact of railway noise pollution. The study discovered that various sources, including honking, trolley movement, announcements, audio speakers, and passengers shouting, contribute to the predominant noise alongside the primary noise. Additionally, the research highlights the severe health effects of noise exposure on staff, passengers, and employees in the railway station environment (Uddin et al., 2015).

A study was conducted to further understand the relationship between nocturnal railway noise exposure and its effects on cardiovascular reactivity in various age groups (Tassi et al., 2010). By comparing the cardiac responses to different types of trains, the study highlights the significance of assessing and mitigating railway noise pollution to safeguard public health and ensure a more comfortable living environment for those residing near railway stations. A socio-acoustic survey was conducted in Sweden's residential areas, examining the impact of railway and road traffic noise on human health. The study evaluated the effects of exposure to single and multiple noise sources on inhabitants' well-being. Results indicate that areas with multiple noise sources led to a significantly higher percentage of annoyed residents compared to those experiencing only one noise source at the same total sound exposure for 24 h. Consequently, researchers emphasized the importance of incorporating the combined effects of multiple noise sources in risk assessments and the development of mitigation strategies (Öhrström et al., 2007). In Tehran, the capital of Iran, a study conducted through a questionnaire survey on the residents living around the railway noise and vibrations in different districts shows that the distance from the railway affects both disaffection from noise and vibration (Maljaee et al., 2022). To aid decision-makers in assessing the financial aspects of installing new monitoring stations and evaluating existing ones, a study in Italy proposed a set of key performance indicators (KPIs) specifically tailored for noise monitoring. These KPIs were derived from those defined for general monitoring networks, ensuring compliance with current legislation. The study evaluated the performance of three noise monitoring networks in ports, offering valuable insights for improving the efficiency and effectiveness of such systems (Bolognese et al., 2023). An innovative exposure assessment strategy developed by the review of past and current cardiovascular noise effect research in Austria by a large epidemiological health survey was reported (Lercher et al., 2011). It presents the exposure–response curves, with and without interaction assessment, based on standardized re-analyses as the cardiovascular effects of noise rank second after annoyance in terms of disability-adjusted life year. A study conducted in Australia quoted that rolling contact between the rail and wheel induces vibration and consequent noise radiation along with railway noise. To promote railway transport and sustainability, it is necessary to reduce railway noise (Jiang et al., 2013).

Railway noise has generally been considered the least annoying transportation source, and, therefore, it has not achieved the same attention as road traffic and other types of transportation noise. However, recent studies suggest that nocturnal railway noise is as disturbing to sleep as road traffic noise and that railway noise has an important short-term impact on the health of the people living around railway stations and might increase the risk for different types of diseases (Lim et al., 2006; Pultznerova et al., 2018; Sørensen et al., 2011; Vincens & Persson Waye, 2022). The effects of noise pollution due to the different types of transportation medium such as road traffic and other sources are studied in India (Agarwal & Swami, 2011; Gilani & Mir, 2021; Khaiwal et al., 2016; Mishra et al., 2014) but no study has been conducted till this investigation which includes the railway transportation noise pollution as per the authors' knowledge. This makes the unique investigation to identify the effects of railway noise based on the questionnaire survey.

This study aims to bridge the gap in understanding the relationship between railway noise pollution and people's perception. By analyzing the impact of railway noise on residents residing near monitoring sites through a questionnaire survey, we aim to shed light on the pressing issue of noise pollution and its effect on overall well-being and quality of life. Given the dense population in India, the harm of noise pollution is particularly significant. As one of the most populous developing countries, India has witnessed a surge in noise pollution due to rapid urbanization (Banerjee, 2012; Firdaus & Ahmad,

2010; Jamir et al., 2014). Moreover, excessive noise pollution can disrupt daily routines and sleep patterns, leading to adverse health effects that negatively impact overall well-being (Alías & Alsina-Pagès, 2019).

The aim and objective of the proposed study are to identify and understand the railway noise pollution distribution pattern and the perception of the people residing near to rail track area. Noise distribution patterns were identified by monitoring the noise pollution in different time periods, i.e., morning, noon, and evening. It can help the higher authorities to make policies in that particular period to combat excessive noise levels. Face-to-face interviews of the participants through the questionnaire survey may give a constructive view of the various effects of railway noise on them, which may be an important factor in demitting the effect of noise. The physical and psychological consequences of the rail noise may be reduced with the dividing effective strategies, which may also bring the noise level within acceptable limits. This study may also help to make the plan for other places, which will improve the living environment of the people residing near the rail track region.

This study seeks to find a correlation between noise pollution and how it affects human mental and physical health, which has led to the following research questions: what is the railway noise distribution pattern in the morning, noon, and evening time? What is the perception of the people residing in the vicinity of the railway track? Is there any correlation between the perception of noise pollution and its various sources? Is there any relationship between the perception of noise pollution and health problems? Are there any physical and psychological consequences for the inhabitants living near monitoring sites?

To assess railway noise level patterns between morning, afternoon, and noon time period, hypothesis testing was performed and inferential analysis was carried out between (1) perception of noise pollution and various sources (railways, vehicles, industries, households, others) to know the correlation. (2) Perceived noise pollution and a number of issues (disturbance, irritation, headache, blood pressure, sleep disruption, and stress) to know the correlation. Figure 1 shows different steps of the study to achieve the above-mentioned objectives.

The rest of the article is organized as follows: "Literature survey" describes important studies conducted by various researchers related to the theme of the present paper. A description of measurement locations, noise level data collection and questionnaire survey, and study methodology adopted has been provided in "Fieldwork and methodology of the study." Detailed discussion about the results obtained from the noise level data and questionnaire survey data analysis has been presented in "Results and discussion." "Conclusion" concludes the findings of the study along with its major limitations and recommendations for possible future work.

Literature survey

Over the last 20 years, several academics have conducted extensive research on the impacts of noise pollution. An investigation carried out on Youngdeungpogu's transportation noise levels using noise prediction models in Seoul Metropolitan City, the Republic of Korea, found that many persons are susceptible to noise levels that would bother and interrupt their sleep and assessed the population exposed to vehicle noise levels. It was suggested to regulate and lower traffic noise levels in Seoul to safeguard locals from any potential negative health impacts. The European Commission has listed community noise as an environmental issue; it will become the biggest threat to the lifestyle of the people if not controlled (Ko et al., 2011). An assessment of the population exposed to noise pollution and acoustical planning mitigation analysis in a European capital city was carried out, which demonstrates that exposure exceeds the WHO-recommended guideline levels. The extent of low-frequency noise from transportation activities was examined, and it was discovered that there was a sizable low-frequency component, indicating potential health problems from exposure to noise from transportation (Douglas & Murphy, 2016; Murphy & King, 2010, 2014). Noise pollution has been referred to as a modern plague because highway, rail, and aviation traffic, which continue to be key sources of environmental noise, have grown steadily, making noise more intense and pervasive than ever before (Goines & Hagler, 2007). Possible health consequences of noise pollution are extensive, all-pervasive, long-lasting, and substantial from both a medical and societal standpoint. Health is harmed by the direct and cumulative negative

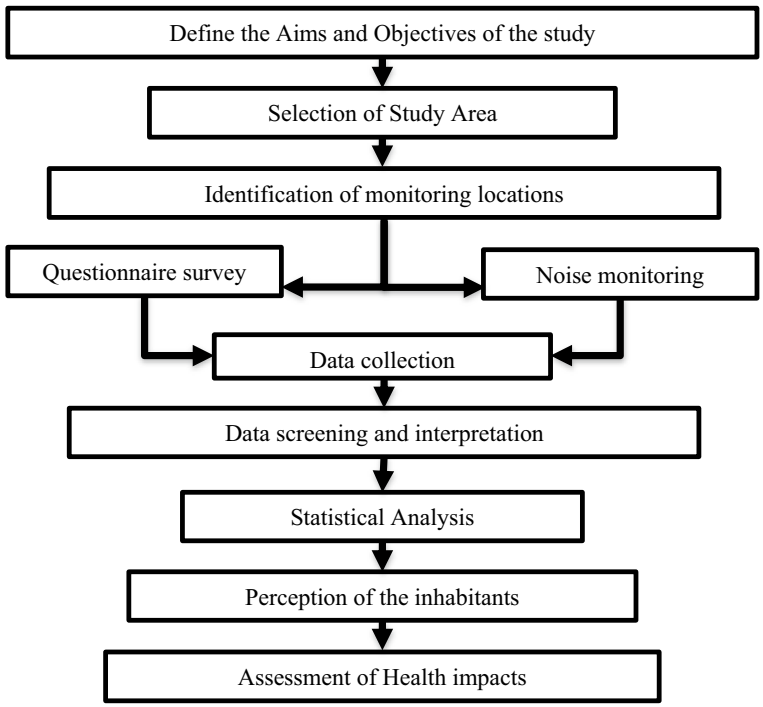


Fig. 1 Flowchart or framework for the study

impacts of noise, which also harm social, working, and learning settings. These consequences result in losses in both physical (economic) and intangible (well-being) assets. Large-scale epidemiological research has provided substantial proof that population exposure to ambient noise has a negative impact on health. As a result, environmental noise should be taken into account as both a source of annoyance and a worry for both public and environmental health (Themann & Masterson, 2019; World Health Organization, 2011). The main contributors to noise pollution in urban areas include transportation, construction equipment, manufacturing operations, and household activities. Despite the fact that this problem is continually becoming worse in emerging nations, it has not received the proper attention (Cheuk Fan, 2000; Miedema & vos, 1998) (Zannin et al., 2018). A study was conducted on a corridor used for both operational and under-construction public transportation, and discovered that while public transportation contributes less to overall noise pollution than private vehicles do, it still exceeds standards. This can be reduced by designing a noise barrier along the corridor. It is an issue that is not given enough consideration considering that individuals cannot sight, smell, or taste noise (Mishra et al., 2014). Railroads are often viewed as a practical mode of transportation, although those who live close to railroad terminals must endure high levels of train noise (Mejvald & Konopa, 2019). Train noise is issue with the environment. It is regarded as one of the primary sources of noise pollution in the vicinity of railroad stations. Negative impacts of noise on human existence will get worse in the next years and include decreased activity and performance, high blood pressure, cardiovascular disease, exhaustion, headaches, and disrupted sleep (Biswas, 2014; S. A. Stansfeld & Matheson, 2003), (Kumar, 2019). It is now well established that both wheel and rail are major radiators of sound, their exact balance depending on wheel and track design parameters. Sleepers are also seen to contribute significantly at low frequencies, and for very still rail pads may dominate the noise from the track (D. J. Thompson & Jones, 2000). Many existing techniques for acoustical properties and transportation noise still have problems with accuracy limitations. Different studies have presented methodologies and techniques for the ease of noise measurement in different kinds of acoustic environment which can help to locate the main noise sources during the pass-by of a car, but also their characterization in terms of source strength

(Ballesteros et al., 2015) such as adjusted controlled pass-by method for urban traffic noise assessment (Moreno et al., 2023), statistical pass-by method for unattended traffic noise measurement (Ascari et al., 2022), digital sensor for the measurement of outdoor noise (López et al., 2020), and deep learning, machine learning, and artificial intelligence–based algorithms for real time detection of environmental noise in different field conditions (Alías et al., 2018; Nourani et al., 2020; Renaud et al., 2023; Shah et al., 2020; Takenaka & Ozawa, 2022; Zhang & Kuehnelt, 2020). While certain literature demonstrates that gradient lifting technology models effectively predict specific targets, such as decision trees obtaining initial values for the fitting function through multiple regression (Yoon et al., 2018). This approach handles numerous input variables considered in the study. Nevertheless, calculated output values exhibit observed data information errors, which rely on loss functions like square-error, absolute-error, and negative binomial log-likelihood functions to minimize prediction discrepancies (Němec et al., 2021). The sensor network (Bolognese et al., 2023) and IoT (Liu et al., 2020) can be used for the depiction of the noise level in a specific environment over a time period in a more specific way.

A sociological study conducted in the city's neighborhoods revealed that noise irritation is a significant issue for many inhabitants. Number of persons who experience sleep disruption has been shown to be plainly unsatisfactory (Arana & García, 1998). A study conducted across four Indonesian cities, namely Bandung, Surabaya, Yogyakarta, and Semarang, revealed that noise levels at all measurement points within residential areas surpassed the prescribed noise threshold for such zones (Huboyo et al., 2022). This finding is crucial because excessive noise can negatively impact the comfort and well-being of people residing near railway tracks. Numerous studies confirm that noise exposure can harm health, leading to reduced quality of life. This is evident through its effects on hearing, sleep quality, concentration and memory, and cognitive and cardiovascular activities (Gallasch et al., 2016; Guski et al., 2016). Murphy et al. (2009) performed a study to evaluate how much road traffic noise is being exposed to locals and workers. It demonstrates that nighttime exposure to noise is more severe, also revealed that if the influence of traffic flow on residential populations is carefully considered, traffic management strategies have the potential to result in considerable reductions in the degree of noise exposure. According to Hong et al. (2011), road traffic noise masks the aero-plane (or railway) noise irritation, and the higher the level of road traffic noise, the lower the subject ranks the aviation (or railway) noise discomfort. Overall irritation of combined noise is greater than the maximum source-specific aggravation when the sound levels of the two sources are equivalent or equally bothersome, but it is equal to the maximum source-specific annoyance when one source's level is higher than the other. A study conducted by Zhang & Ma (2021) shows that the noise made by trains travelling on elevated lines regularly disturbs the daily lives of the surrounding residents. Investigations reveal that locals who live close to elevated railroads commonly complain about the noise. Another study conducted by da Paz et al. (2005) and Lakušić & Ahac (2012) shows that elevated sound levels are the primary causes of unhappiness, especially in metropolitan areas, and prolonged exposure to them can cause hearing loss due to noise. The main causes of railroad noise are loading and unloading, truck movements, braking, squeals, and whistles. As a result, at low levels of exposure to railroad noise, people are more disturbed by unusual noises, whereas at higher levels, people become accustomed to noise and pay less attention to unusual railway noise events (Licitra et al., 2016). Rylander (2004) analyzed environmental noise exposure and physiological stress reactions based on the physiological characteristics of the hearing system. Kuhnt et al. (2008) did a laboratory experiment to experimentally assess the real irritation brought on by various combinations of road and rail noise. They discovered that the 72 volunteers were subjected to several noise scenarios made up of various mixes of road and rail traffic noise. They were required to assess how irritated they felt after each loud situation. The study concluded that sleep quality decreased and reaction time increased gradually with noise levels. Disturbance in sleep due to noise may increase the risk of obesity (Oftedal et al., 2015). A survey conducted by Agarwal & Swami (2011) in Jaipur City, India, reported that 52% population was suffering from noise pollution. A study performed by Roswall et al., (2018a, b) explored the relationship between exposure to residential railway noise and the incidence of diabetes which came to no conclusion; however, it did

support the conclusion that there is a direct correlation between residential road traffic noise and diabetes incidence. Stansfeld et al. (2000) examined how noise pollution in metropolitan areas affects people's non-auditory health. In the lab, exposure to transport noise interferes with sleep, but in field trials, where adaptation takes place, it typically does not. Noise annoys people, makes it difficult to do complicated tasks, and alters social behavior. Aircraft traffic noise (1.6% threat increment per 10 dB raise in the 24-h uninterrupted level of noise; 95% CI 0.3–3.0%), road traffic sound (2.4% per 10 dB; 95% CI 1.6–3.2%), and railroad noise (3.1% per 10 dB; 95% CI 2.2–4.1%) all had statistically relevant regression exposure-risk relationships with cardiac arrest or hypertensive heart disease (Seidler et al., 2016). A study emphasized that the railway is an alternative model for the transportation of goods compared to road transportation, but due to the lack of appropriate space for loading and unloading, lack of parking space causes traffic congestion which is the primary source of noise pollution in the nearby areas of rail stations (Singh & Gupta, 2020). According to a study conducted at Chennai Central railway station by Kanakasabai et al. (2005), it was found that the noise pollution at the railway station was exceeding the permissible limits. A case study on railway noise pollution and its impacts in Bangladesh was carried out by Uddin et al. (2015) which showed that people living around railways are highly exposed to noise pollution. Bunn & Zannin (2016) executed the study to assess the noise levels generated by a railway that runs through a crowded area of a large city and to examine possible remedies that may be taken to reduce the noise levels. Three potential solutions that excluded the train horn, included acoustic barriers, and removed the railway tracks from the edge of the urban area were simulated as ways to reduce the noise pollution caused by rail traffic. Wosniacki & Zannin (2021) also presented a framework to regulate railway noise exposure in Brazil and discovered that more than 40% of the population is influenced by rail noise that is 5 dB(A) or louder than background noise, and that road traffic noise exposure levels are above WHO recommendations. Different researchers have developed numerous methods and technologies (Ascari et al., 2023; Fredianelli et al., 2022; Pallas et al., 2016) to mitigate environmental noise in a proper manner, which will be helpful in making effective noise mitigation action plans.

Fieldwork and methodology of the study

Study area and data collection

Delhi City was selected as a research area for the fulfillment of the aim and objectives of the study. Delhi is the capital of India and is surrounded by the states of Haryana and Uttar Pradesh, with an area of 1484 km^2. There are 47 railway stations in Delhi, five of which are prominent ones, which are New Delhi Railway Station, Old Delhi Railway Station, Hazrat Nizamuddin Railway Station, Anand Vihar Railway Station, and Delhi Sarai Rohilla Railway Station (Gupta & Anand, 2018). New Delhi railway station, one of the largest stations in India, is located between Ajmeri Gate and Paharganj (Government of India; Ministry of Railways; New Delhi, 2015). Continual inescapable noise levels are putting a burden on the auditory, non-auditory, and sensory systems of the residents of the city living in proximity to the railway station zone (S. A. Stansfeld & Matheson, 2003). The main objective of this study is to find out the effect of train noise, which occurs near the railway stations as shown in the framework of the research (Fig. 1). To comprehend the research locations and their surroundings/neighborhood regions better, field observations or field surveys were carried out. In this study, noise pollution monitoring was conducted only during the day in three shifts (morning, afternoon, and evening). Due to several constraints like battery backup and safety of the instrument, availability of data collection assistants, security clearance permission from Delhi police for night monitoring, and lack of sufficient funds, noise pollution levels at night could not be monitored. Therefore, in this study, health effects have been analyzed based on subjective responses collected through questionnaire surveys only. The basic elements of monitoring stations have been kept in mind during this study, considering the most popular along with different land use patterns, all the top 10 railway stations were selected out of 47 railway stations in Delhi to conduct this study as shown in Fig. 2 (A). Table 1 shows the salient features of the chosen locations. For the validation and significance, ten locations were selected

for the study keeping all the land use patterns. The noise level was monitored with the help of a class I Sound Level Meter (SLM), which had windscreen microphones confirming to international standards of IEC 61672–1:2002 (López et al., 2020). To reduce the impact of weather conditions, the devices were placed 1.5 m above ground level in open fields, close to noise sources (railway tracks), and with reflecting surfaces at a minimum distance of 3 m (Bhavan & Nagar, 2015; Wosniacki & Zannin, 2021). The primary data of noise pollution level was collected by sound level meter at each location and monitoring was conducted throughout the daytime under typical metrological conditions. For two consecutive days, at each railroad station, noise level data were recorded in three shifts (2 h/shift) of daytime at each identified location, whenever a train passed by. Figure 2 (B) shows the noise level measurement being carried out using the noise level meter at the Delhi Sarai Rohilla and Sabzi Mandi railway stations. In a similar way, noise level measurements were carried out at all the other railway stations and collected data were entered into an Excel sheet for further calculation and analysis purposes. The in situ face-to-face interview questionnaire survey was carried out with the noise level monitoring simultaneously at each selected location. In order to understand how inhabitants felt about noise pollution, a questionnaire-based study was used to measure the impact of noise on health in the neighborhood areas of the selected railway station of Delhi City. To understand the physiological and psychological effects on the inhabitants, questions were



Fig. 2 **A** Google Earth view of the study area along with the sampling locations in Delhi. **B** Monitoring locations at Delhi Sarai Rohilla (S2) and Sabzi Mandi (S8) showing noise level measurement

Table 1 Salient features of the selected locations

S. no	Location	Land pattern	Latitude	Longitude
1	Delhi Shahdara (S1)	Mixed	28.6734° N	77.2922° E
2	Delhi Sarai Rohilla (S2)	Residential	28.6628° N	77.1852° E
3	Delhi Cantt (S3)	Commercial	28.6119° N	77.1157° E
4	Vivek Vihar (S4)	Industrial	28.6683° N	77.3144° E
5	Dayabasti (S5)	Mixed	28.6693° N	77.1730° E
6	Anand Vihar (S6)	Mixed	28.6471° N	77.3158° E
7	New Delhi (S7)	Commercial	28.6429° N	77.2191° E
8	Sabzi Mandi (S8)	Residential	28.6685° N	77.2004° E
9	Old Delhi (S9)	Commercial	28.6606° N	77.2271° E
10	Hazarat Nizamuddin (S10)	Mixed	28.5889° N	77.2534° E

asked about sleep disturbance, early awakenings, and subjective responses from the residents. To comprehend people's perceptions pertaining to railway noise, a questionnaire of 344 people (almost 34 respondents at each chosen railway station) was conducted. Obtained questionnaire data were further analyzed to understand the attitude of respondents towards railway noise.

Descriptive statistics

In this part of the study, the descriptive statistical analysis (Brainard et al., 2004) has been conducted to validate and correlate the observed data with noise monitoring and a questionnaire survey. The Kalmogorov-Smirnov and Shapiro–Wilk test (Nassiri et al., 2007) was conducted to determine the normality of the noise monitoring data. The data was found normal, which was further statistically analyzed with a *t*-test, to see the distribution pattern of the data (Sahoo, 2014). The hypothesis was developed to determine the relation between different time periods, i.e., morning, noon, and evening noise levels at different sites. To see the significance of the correlation between different variables, inferential statistical analysis was conducted (Maljaee et al., 2022).

Test of normality

Before applying the *t*-test, a normality test was conducted on the collected noise level data set. Noise level data were divided into three sets for morning, noon, and evening. Table 2 presents the results of the normality test.

All three data sets are evidently normal based on the results of the normality test, permitting the *t*-test to be conducted.

t-test was applied to compare the means of two samples, i.e., between morning and noon noise levels, morning and evening noise levels, and noon and evening noise levels respectively.

Hypothesis testing between morning and noon noise levels To assess railway noise level patterns between the morning and noon time period, hypothesis testing was performed.

H_0: Morning noise level = Noon noise level
H_1: Morning noise level > Noon noise level

From Table 3, *p*-value < significance level; therefore, we reject the null hypothesis. Thus, railway noise in the morning hours is more than in the noon hours.

Table 2 Test of normality

	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistic	df	Sig	Statistic	df	Sig
VAR00004	0.114	20	0.200*	0.953	20	0.416
VAR00005	0.147	20	0.200*	0.934	20	0.181
VAR00006	0.174	20	0.114	0.934	20	0.183

*This is a lower bound of true significance

Hypothesis testing between morning and evening noise levels To assess railway noise level pattern between morning and evening time period, hypothesis testing was performed.

H_0: Morning noise level = Evening noise level
H_1: Morning noise level > Evening noise level

From Table 4, p-value < significance level; therefore, we reject the null hypothesis. Thus, railway noise in the evening hours is more than in the morning hours.

Hypothesis testing between noon and evening noise levels To assess railway noise level patterns between noon and evening time period, hypothesis testing was performed.

H_0: Noon noise level = Evening noise level
H_1: Noon noise level > Evening noise level

From Table 5, p-value < significance level; therefore, we reject the null hypothesis. Thus, railway noise in the evening hours is more than the noon hours.

Analysis of variance (ANOVA)

ANOVA was performed to compare the means of three groups (i.e., morning, noon, and evening noise levels). The equation for ANOVA is given by

Table 4 *T*-test findings for the noise levels in morning and evening

	Morning	Evening
Mean	68.035	71.03
Variance	0.485553	3.381158
Observations	20	20
Hypothesized mean difference	0	
Df	24	
t stat	−6.81147	
P(T <= t) one-tail	2.4E−07	
t critical one-tail	1.710882	
P(T <= t) two-tail	4.8E−07	
t critical two-tail	2.063899	

$$F = \frac{variance\ between\ treatments}{variance\ within\ treatments} \tag{1}$$

$$F = \frac{MS_{Treatments}}{MS_{Error}} = \frac{SS_{Treatments}/(I-1)}{SS_{Error}/(n_T-1)} \tag{2}$$

where

F ANOVA coefficient

M mean square

I number of groups

n_T number of observations

Table 3 *T*-test findings for the noise levels in morning and noon

	Morning	Noon
Mean	68.035	66.6
Variance	0.485553	3.705263
Observations	20	20
Hypothesized mean difference	0	
Df	24	
t stat	3.134856	
P(T <= t) one-tail	0.002247	
t critical one-tail	1.710882	
P(T <= t) two-tail	0.004495	
t critical two-tail	2.063899	

Table 5 *T*-test findings for the noise levels in noon and evening

	Noon	Evening
Mean	66.6	71.03
Variance	3.705263	3.381158
Observations	20	20
Hypothesized mean difference	0	
Df	38	
t stat	−7.442267	
P(T <= t) one-tail	3.125E−09	
t critical one-tail	1.68595	
P(T <= t) two-tail	6.25E−09	
t critical two-tail	2.02439	

$$F = \frac{(Sum\ of\ variance\ within\ groups)^2/(degree\ of\ freedom\ within\ groups)}{Total\ (variance)^2/(total\ degree\ of\ freedom)} \quad (3)$$

If $F > F_{critical}$, then the groups are significantly different and if $F < F_{critical,}$ then the groups are significantly similar.

Table 6 summarizes the noise levels of all three groups, i.e., morning, noon, and evening. ANOVA results (Table 7) demonstrate that $F_{critical} < F$; therefore, all the observations of the three groups (morning, noon, and evening noise levels) are not significantly similar. Hence ANOVA results confirm the results of hypothesis testing that the railway noise levels during the evening period are higher than those in the morning and noon

Results and discussion

This section provides the output of the analysis of the questionnaire survey and the findings of the statistical correlation assessment with the different variables, which includes the discussion on the various findings and association with the previous researchers' work.

Questionnaire survey analysis

For the analysis and assessment of railway noise, the questionnaire form was designed (Lim et al., 2006). The questionnaire form was used for the data collection, which was focused on knowing the perception of the individual about the railway noise impact on human health. The questionnaire form consists of three sections (Part A, Part B, and Part C). The first part (Part A) of the questionnaire is related to personal information like age, sex, contact number, educational background, annual income, and distance from the railroad. The second part (Part B) deals with sources of noise pollution (vehicles, railways, households, industry, others), time of stay in the current house, noise pollution understanding, residential area type, etc. In the third part (Part C), the questions are related to health problems such as disturbance, headache, sleep disturbance, irritation, blood pressure, and stress. All participants were asked about the train-induced noise level of irritation like extreme, very much, moderate, slight, or not at all during their stay at home. Respondents were also asked to rate their neighborhood noise on a scale of 0 to 10 depending on how much noise bothers, disturbs, and annoys them.

Random sampling method was adopted to direct the questionnaire on various age categories, sex, educational levels, and income levels of participants, which was also used in different studies previously (Goswami & Swain, 2017; Lim et al., 2008; Tanni et al., 2015). Population characteristics of the studied area are presented in Table 8. Male participants were found more interested in participating in the survey than female participants, as reported in the previous studies also (Kumar, 2019; Mimani & Singh, 2021). No female respondent was found at Delhi Shahdara and Delhi Sarai Rohilla railway station to answer the questions. Respondents aged 15–30 and 30–45 years combined were over 70% for all areas except Delhi Shahdara, Delhi Sarai Rohilla, and Delhi Cantt railway station. Age demographics are presented in Fig. 3. More than 45% of participants did not agree to share the data regarding their annual income except at Delhi Cantt, Dayabasti, and New Delhi railway stations. Except for Delhi Cantt and Sabzi Mandi railway station, more than 60% of participants were educated to 10 + 2. The survey observed that individuals living/working less than 5 years in their present house/business area were < 29%. From Table 8, it was found that

Table 6 Summary of the railway noise level in the daytime

Groups	Minimum noise level	Maximum noise level	Average noise level	Variance
Morning	66.6	69	68.04	0.485553
Noon	64	71.3	66.6	3.705263
Evening	68.3	74	71.03	3.381158

Table 7 Analysis of variance results

Source of variation	SS	df	MS	*F*	*P*-value	*F* crit
Between groups	204.361	2	102.1805	40.4837	1.14E − 11	3.15884272
Within groups	143.8675	57	2.523991			
Total	348.2285	59				

Table 8 Socio-demographic characteristics

Factors		Percentage (%) of different factors at selected locations									
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Sex	**Male**	100	100	97.37	70.59	94.29	84.38	94.44	77.42	96.97	75
	Female	0	0	2.63	29.41	5.71	15.63	5.56	22.58	3.03	25
Age	**15–30**	25.64	30.30	21.05	32.35	22.86	37.50	22.22	32.26	33.33	37.50
	30–45	35.90	36.36	21.05	41.18	65.71	50.00	52.78	45.16	48.48	56.25
	45–60	28.21	27.27	42.11	20.59	11.43	12.50	25.00	19.35	18.18	3.13
	> 60	10.26	6.06	15.79	5.88	0	0	0	3.23	0	3.13
Annual income (in million)	**< 0.2**	48.72	33.33	44.74	41.18	82.86	40.63	41.67	45.16	30.30	37.50
	0.2–0.4	5.13	0	36.84	5.88	0	6.26	25.00	9.68	3.03	0
	0.4–0.6	0	0	0	0	0	0	0	0	0	0
	Not revealed	46.15	66.67	18.42	52.94	17.14	53.13	33.33	45.16	66.67	62.50
Education	**10**	61.54	51.52	23.68	76.47	62.86	43.75	44.44	9.68	39.39	65.63
	10 + 2	10.26	24.24	23.68	14.71	17.14	28.13	25.00	29.03	48.48	12.50
	U. G	17.95	24.24	36.84	8.82	14.29	0	30.56	54.84	6.06	0
	P. G	5.13	0	5.26	0	0	0	0	6.45	0	0
	Other	5.13	0	10.53	0	5.71	28.13	0	0	6.06	21.88
Time period of stay in current house or work period in shop (years)	**0–5**	17.95	12.12	15.79	11.76	17.14	18.75	11.11	22.58	21.21	28.13
	5–10	15.38	12.12	7.89	23.53	20.00	31.25	11.11	19.35	30.30	21.88
	10–15	23.08	21.21	7.89	23.53	20.00	15.63	19.44	12.90	18.18	12.50
	15–20	17.95	12.12	7.89	14.71	8.57	15.63	16.67	9.68	9.09	12.50
	> 20	25.64	42.43	60.54	26.47	34.29	18.74	41.67	35.49	21.22	24.99

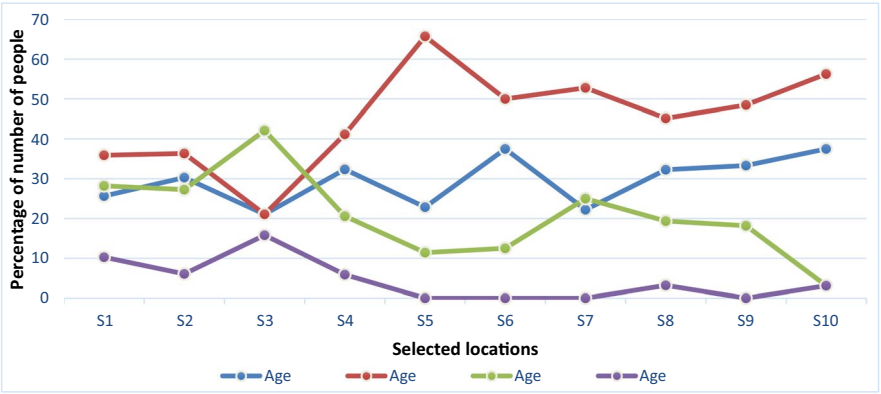


Fig. 3 Age demographics at the selected locations

the individuals living/working over 20 years in their current area are over 21% except for Anand Vihar railway station, which is of mixed type.

Analysis of the relationships between perception of noise pollution and its various sources

Outcomes obtained from the analysis of the correlations between the perception of noise pollution and various sources are shown in Table 9 to first identify the significance of the source. With regard to that, inferential analysis was carried out between the perception of noise pollution and various sources (railways, vehicles, industries, households, others) and it can be observed that only railways showed significant correlations with the perception related to noise. According to the results reported in previous studies, the railway is the prime source of noise pollution and significantly harms the health of the people living in proximity of the railway area (Roswall et al., 2018a,

Table 9 Correlation coefficient between perception of noise pollution and its sources

Correlations								
			Noise pollution	Railways	Vehicles	Industry	Household	Others
Spearman's rho	Noise pollution	Correlation coefficient	1.000	0.300**	0.034	0.074	0.008	0.089
		Sig. (2-tailed)		0.000	0.532	0.174	0.890	0.104
		N	337	337	337	337	337	332
	Vehicles	Correlation coefficient	0.300**	1.000	0.044	0.050	0.076	0.093
		Sig. (2-tailed)	0.000		0.421	0.360	0.161	0.087
		N	337	343	343	343	343	337
	Railways	Correlation coefficient	0.034	0.044	1.000	0.174**	−0.014	0.213**
		Sig. (2-tailed)	0.532	0.421		0.001	0.794	0.000
		N	337	343	343	343	343	337
	Industry	Correlation coefficient	0.074	0.050	0.174**	1.000	−0.041	0.562**
		Sig. (2-tailed)	0.174	0.360	0.001		0.448	0.000
		N	337	343	343	343	343	337
	Household	Correlation coefficient	0.008	0.076	−0.014	−0.041	1.000	−0.043
		Sig. (2-tailed)	0.890	0.161	0.794	0.448		0.428
		N	337	343	343	343	343	337
	Others	Correlation coefficient	0.089	0.093	0.213**	0.562**	−0.043	1.000
		Sig. (2-tailed)	0.104	0.087	0.000	0.000	0.428	
		N	332	337	337	337	337	337

**Correlation is significant at the 0.01 level (2-tailed)

b; Seidler et al., 2016; Sørensen et al., 2021). Perception of noise pollution analyzed here with respect to the other sources shows an insignificant positive correlation except for railways at all selected locations. The degree of association between the perception of noise pollution and its sources showed a degree of significance of less than 0.001. This indicates that, as expected, there is a close relationship between the sources of noise and the effects of noise.

Analysis of the relationships between perception of noise pollution and various problems

In order to first determine the significance of the difficulties caused by noise pollution on the inhabitants based on the questionnaire survey, results from the analysis of correlations between perceptions of noise pollution and various problems are provided in Table 10. In this regard, inferential analysis was conducted between perceived noise pollution and a number of issues (disturbance, irritation, headache, blood pressure, sleep disruption, and stress), and it was found that only sleep disruption significantly correlated with perceived noise pollution. According to findings from earlier studies, railroads are the main cause of noise pollution and severely interfere with the quality of sleep of adjacent inhabitants (Reiter & Kostek, 2008; Szwarc et al., 2011). Except for sleep disturbance (SD) at all sites surveyed, the impression of noise pollution as compared to the other issues reveals a negligible positive correlation. With a corresponding *p*-value of 0.001, the degree of connection between the impression of noise pollution and its sources was 0.141. This suggests that, as would be predicted, there is a close connection between noise sources and sleep disruption.

Noise annoyance

Noise from railway traffic is the main annoyance factor for those living close to railways (Licitra et al., 2016). Expanding rail route traffic in Delhi will prompt more noise and vibration issues in the future. Two questions were asked in the survey from respondents to assess the noise annoyance in the vicinity of railway stations. First, respondents were asked to react to annoyance due to noise exposure among four categories: Slightly, Moderately, Very, Extremely, or

Table 10 Correlation coefficient between perception of noise pollution and various problems

Correlations									
			NP	D	I	H	BP	SD	S
Spearman's rho	Noise pollution (NP)	Correlation coefficient	1.000	0.048	−0.032	0.003	−0.013	−0.056	0.141**
		Sig. (2-tailed)		0.380	0.557	0.957	0.812	0.304	0.010
		N	337	337	337	335	337	337	337
	Disturbance (D)	Correlation coefficient	0.048	1.000	−0.011	−0.010	0.022	0.016	−0.051
		Sig. (2-tailed)	0.380		0.845	0.853	0.684	0.766	0.342
		N	337	343	343	341	343	343	343
	Irritation (I)	Correlation coefficient	−0.032	−0.011	1.000	0.020	−0.116*	−0.042	−0.030
		Sig. (2-tailed)	0.557	0.845		0.708	0.032	0.435	0.574
		N	337	343	343	341	343	343	343
	Headache (H)	Correlation coefficient	0.003	−0.010	0.020	1.000	0.051	0.089	0.050
		Sig. (2-tailed)	0.957	0.853	0.708		0.345	0.100	0.355
		N	335	341	341	341	341	341	341
	Blood pressure (BP)	Correlation coefficient	−0.013	0.022	−0.116*	0.051	1.000	0.001	0.022
		Sig. (2-tailed)	0.812	0.684	0.032	0.345		0.983	0.688
		N	337	343	343	341	343	343	343
	Sleep disturbance (SD)	Correlation coefficient	−0.056	0.016	−0.042	0.089	0.001	1.000	−0.090
		Sig. (2-tailed)	0.304	0.766	0.435	0.100	0.983		0.097
		N	337	343	343	341	343	343	343
	Stress (S)	Correlation coefficient	0.141**	−0.051	−0.030	0.050	0.022	−0.090	1.000
		Sig. (2-tailed)	0.010	0.342	0.574	0.355	0.688	0.097	
		N	337	343	343	341	343	343	343

**Correlation is significant at the 0.01 level (2-tailed)

*Correlation is significant at the 0.05 level (2-tailed)

Not at all. Second, respondents were asked to give their opinion on a scale that ranges from zero to ten. If they were not annoyed, then respondents were asked to pick zero value and the value 10, in case they were extremely annoyed. In the first question, if a person chooses "Slightly" then it ranges from 1 to 3 on the opinion scale; if a person decides "Moderately," then it ranges from 4 to 7, and if a person selects "Very," then the value will be greater than 7 on opinion scale. Annoyance is a common issue faced by individuals for a longer period due to the high noise level. Figures 4 and 5 show the annoyance from noise exposure faced by respondents at the selected sites in the daytime and nighttime. It is evident from the figures that most of the people were annoyed due to train noise in the morning (08:00–09:00 h) and the evening (20:00–21:00) hours (Fig. 4 and 5).

Figure 6 indicates the significant sources of noise pollution in the vicinity of the railway stations. Clearly, it shows that transportation, i.e., railways, is a significant factor of noise pollution for the people living around the railroads. From the data analysis, it was found that more than 60% of noise comes from railway operations.

During the survey, participants were asked to react to the medical issues they faced because of noise pollution. They were asked to reply in three classifications: Yes, No, and Sometimes. Figure 7 shows the degree of disturbance faced by participants because of noise pollution at the selected locations. From the data, it was observed that participants were most affected by noise disturbance. The majority of respondents replied "Yes" at Delhi Shahdara, Delhi Sarai Rohilla, Delhi Cantt, Dayabasti, and Old Delhi railway stations whereas mixed responses were received at the other five locations.

Figure 8 shows the headaches from noise exposure at selected locations faced by participants. It was observed that more than 25% of people were facing headache problems near the New Delhi railway station due to railway operations and traffic. While more than 20% of people residing near Dayabasti and Anand Vihar railway stations were highly exposed to railway noise as both the stations operate an extensive number of passenger trains. Nearly 10% of people were affected near Hazarat Nizamuddin railway station. Nowadays, railway noise has become a serious

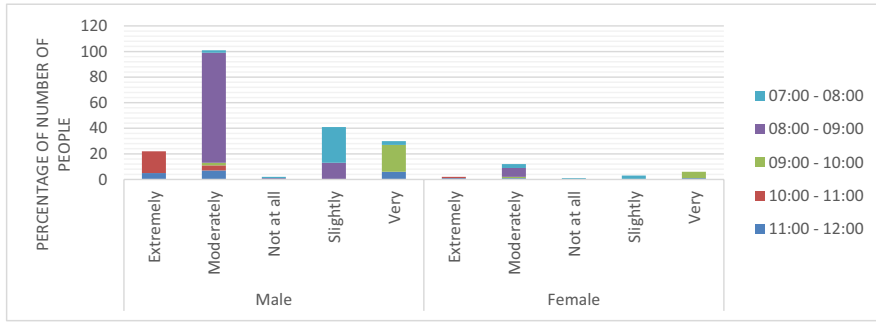


Fig. 4 People annoyed at daytime

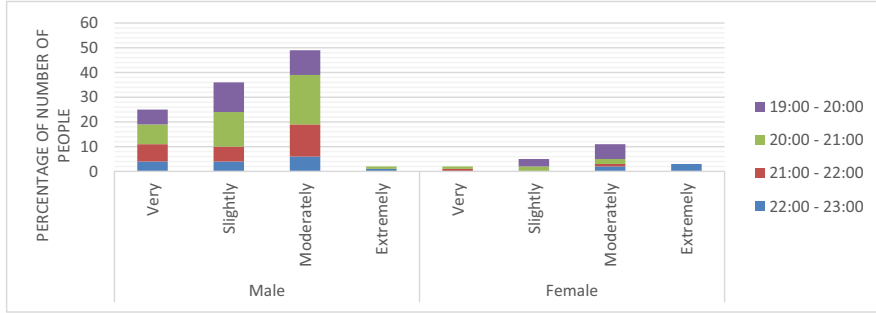


Fig. 5 People annoyed at nighttime

source of headache for people residing or working near stations.

Table 11 shows the health factors among participants due to noise exposure. The disturbance is found to be a common problem faced by participants. About 58% of male and 86% of female respondents reported disturbance issues due to noise exposure. It was found that 24%, 12%, 36%, and 16% of males answered "Yes" for Irritation, Blood Pressure, Sleep Disturbance, and Stress, respectively. Similarly, female participants replied "Yes" for Irritation, Blood Pressure, Sleep Disturbance, and Stress at 47%, 17%, 61%, and 47%, respectively. Male and female respondents replied "Yes" for headaches caused due to noise exposure at 44% and 58%, respectively. Female participants were found more affected compared to male participants.

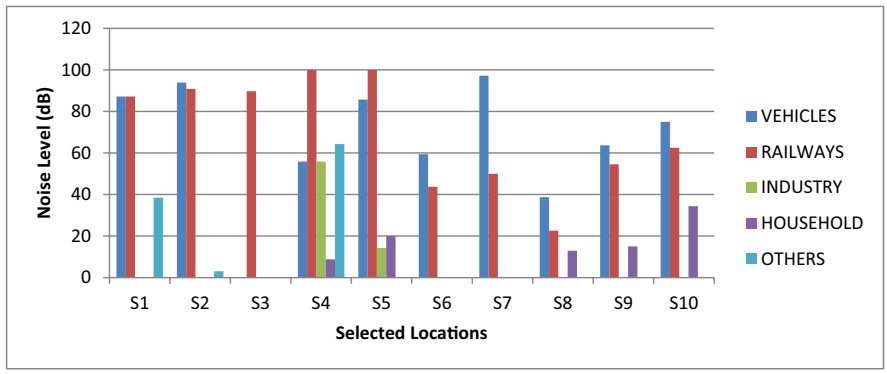


Fig. 6 Sources of noise pollution at selected sites

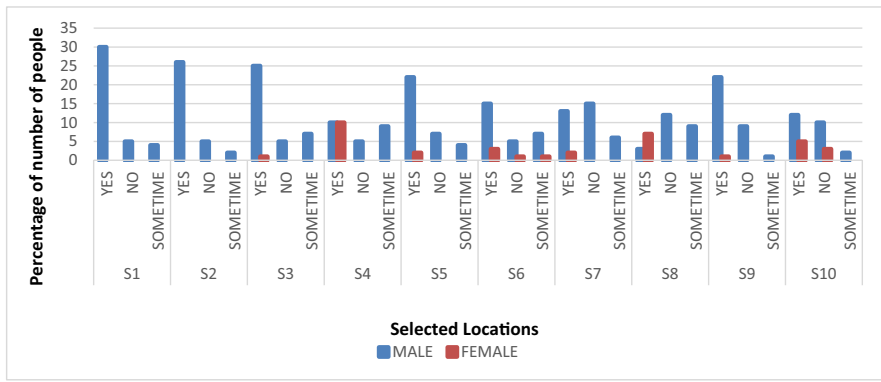


Fig. 7 Disturbance due to noise exposure at selected locations



Fig. 8 Headache due to noise exposure at selected locations

Table 11 Health factors due to noise exposure

Health factor	Male			Female		
	Yes	No	Sometime	Yes	No	Sometime
Disturbance	57.65	25.41	16.94	86.11	11.11	2.78
Irritation	24.10	61.24	14.66	47.22	19.44	33.33
Headache	43.65	24.43	31.92	58.33	11.11	30.56
Blood pressure	12.05	87.30	0.65	16.67	69.44	13.89
Sleep disturbance	36.48	52.12	11.40	61.11	13.89	25
Stress	15.64	72.96	11.40	47.22	19.44	33.33

Conclusion

The present study evaluated the railway-induced noise problem in Delhi. Several techniques were used to examine the noise levels during the train's passage. It was found that noise levels exceeded all the locations and violated the prescribed limits set by the Central Pollution Control Board. Railway noise level in the evening was found significantly higher than morning and noon time. General disturbance, headache, sleep disturbance, and irritation were the main health problems faced by the residents due to noise exposure. Results clearly show that female respondents faced railway noise–induced health problems more often as compared to male respondents. One of the possible causes of the more severe noise impact on female responders is that they spend more time at home than males do, which sensitizes them to longer periods of noise exposure. Nearly 60% of respondents agreed that railroad noise pollution is a serious problem for children. It is also clear from the self-reported questionnaire survey that perception of noise pollution shows a significant correlation with railway noise sources and majorly with sleep disturbance (SD). Face-to-face interviews were conducted with the people residing near the railway track to evaluate the level of annoyance due to railway noise. Residents faced more noise annoyance during the morning hour (08:00–09:00) and the evening hour (20:00–21:00). Residents of the selected areas suggested strict restrictions on honking and noise barrier installations along with the particular train corridor. Due to different constraints like battery back-up and safety of the instrument, availability of data collection assistants, security clearance permission from Delhi police for night monitoring, and lack of sufficient funds, noise data could not be collected during nighttime, which is one of the major limitations of this study. Another limitation of the study is the smaller number of female respondents for questionnaire survey/face-to-face interview. This is because most of the selected locations/railway stations were dominated by commercial/industrial or mixed land use pattern. Further recommendations to reduce train noise by the participants were strict control of horns blowing by trains, tree plantation, construction of under-pass, parking facility, and sound-proofing systems. Future studies may focus on the collection of more responses from more residents along with noise level data collection for a longer duration covering the entire day and nighttime so that a more precise and accurate picture of the harmful effects of noise pollution on residents can be reported.

Acknowledgements The authors would like to thank the Department of Environmental Engineering, Delhi Technological University, Delhi, for providing the noise level meter and other instruments to carry out this research work.

Author contributions Kranti Kumar: Conceptualization, Data analysis, Editing the original draft, Methodology, Supervision.

Arun Bhartia: Data collection, Data extraction, Writing the original draft, methodology and review writing.

Rajeev Kumar Mishra: Data analysis, Editing the original draft, Methodology.

Ravi Pratap Singh Jadon: Review writing, data collection, data analysis, methodology.

Jitendra Kumar: Editing the original draft, data analysis.

Data availability Data used in this study were collected by authors from the selected locations. If there are specific needs, data can be obtained from the corresponding author by sending an email to kranti31lu@gmail.com. Please state the purpose of data requirement.

Declarations

All authors have read, understood, and have complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors.

Ethics approval and consent to participate Not applicable.

Competing interests The authors declare no competing interests.

References

Aasvang, G. M., Moum, T., & Engdahl, B. (2008). Self-reported sleep disturbances due to railway noise: Exposure-response relationships for nighttime equivalent and maximum noise levels. *The Journal of the Acoustical Society of America, 124*(1), 257–268. https://doi.org/10.1121/1.2932074

Agarwal, S., & Swami, B. L. (2011). Road traffic noise, annoyance and community health survey - A case study for an Indian city. *Noise and Health, 13*(53), 272–276. https://doi.org/10.4103/1463-1741.82959

Alías, F., Alsina-Pagès, R. M., Orga, F., & Socoró, J. C. (2018). Detection of anomalous noise events for real-time road-traflc noise mapping: The DYNAMAP's project case study. *Noise Mapping, 5*(1), 71–85. https://doi.org/10.1515/noise-2018-0006

Alías, F., & Alsina-Pagès, R. M. (2019). Review of wireless acoustic sensor networks for environmental noise monitoring in smart cities. *Journal of Sensors, 2019,* 7634860. https://doi.org/10.1155/2019/7634860

Arana, M., & García, A. (1998). A social survey on the effects of environmental noise on the residents of Pamplona, Spain. *Applied Acoustics, 53*(4), 245–253. https://doi.org/10.1016/S0003-682X(97)00067-4

Ascari, E., Cerchiai, M., Fredianelli, L., & Licitra, G. (2022). Statistical pass-by for unattended road traffic noise measurement in an urban environment. *Sensors, 22*(22), 8767. https://doi.org/10.3390/s22228767

Ascari, E., Cerchiai, M., Fredianelli, L., Melluso, D., Rampino, F., & Licitra, G. (2023). Decision trees and labeling of low noise pavements as support for noise action plans. *Environmental Pollution*, *337*. https://doi.org/10.1016/j.envpol.2023.122487

Asdrubali, F., & D'Alessandro, F. (2018). Innovative approaches for noise management in smart cities: A review. *Current Pollution Reports, 4*(2), 143–153. https://doi.org/10.1007/s40726-018-0090-z

Ballesteros, J. A., Sarradj, E., Fernández, M. D., Geyer, T., & Ballesteros, M. J. (2015). Noise source identification with beamforming in the pass-by of a car. *Applied Acoustics, 93*, 106–119. https://doi.org/10.1016/j.apacoust.2015.01.019

Banerjee, D. (2012). Research on road traffic noise and human health in India: review of literature from 1991 to current. *Noise and Health, 14*(58), 113–118. https://doi.org/10.4103/1463-1741.97255

Bhavan, P., & Nagar, E. A. (2015). July 2015 Protocol for ambient level noise monitoring central pollution control board page 2

Biswas, S. (2014). Health hazards associated with noise in urban area. Octa Journal of Environmental Research April - June, 2014, International Peer-Reviewed Journal. *Journal of Environment Research, 2*(2), 107–111

Bolognese, M., Carpita, S., Fredianelli, L., & Licitra, G. (2023). Definition of key performance indicators for noise monitoring networks. *Environments - MDPI, 10*(4), 1–16. https://doi.org/10.3390/environments10040061

Brainard, J. S., Jones, A. P., Bateman, I. J., & Lovett, A. A. (2004). Exposure to environmental urban noise pollution in Birmingham, UK. *Urban Studies, 41*(13), 2581–2600. https://doi.org/10.1080/0042098042000294574

Bunn, F., & Zannin, P. H. T. (2016). Assessment of railway noise in an urban setting. *Applied Acoustics, 104*, 16–23. https://doi.org/10.1016/j.apacoust.2015.10.025

Cerletti, P., Eze, I. C., Schaffner, E., Foraster, M., Viennau, D., Cajochen, C., Wunderli, J. M., Röösli, M., Stolz, D., Pons, M., Imboden, M., & Probst-Hensch, N. (2020). The independent association of source-specific transportation noise exposure, noise annoyance and noise sensitivity with health-related quality of life. *Environment International*, *143*. https://doi.org/10.1016/j.envint.2020.105960

Cheuk Fan NG (2000). Effects of building construction noise on residents: a quasiexperimen. *Journal of Environmental Psychology, 20*, 375–385. https://doi.org/10.1006/jevp.2000.0177

da Paz, E. C., Ferreira, A. M. C., & Zannin, P. H. T. (2005). Comparative study of the perception of urban noise. *Revista de Saude Publica, 39*(3), 467–472. https://doi.org/10.1590/s0034-89102005000300019

Damayanti, A. T., Perkeretaapian, P., & Madiun, I. (2023). Analysis of noise intensity in the railway facility. *Jurnal Syntax Transformation, 4*(7), 10–20. https://doi.org/10.46799/jst.v4i7.756

Douglas, O., & Murphy, E. (2016). Source-based subjective responses to sleep disturbance from transportation noise. *Environment International, 92–93*, 450–456. https://doi.org/10.1016/j.envint.2016.04.030

Dupont, V. (2017). Urban development and population redistribution in Delhi: Implications for categorizing population. *New Forms of Urbanization: Beyond the Urban-Rural Dichotomy*, *2017*, 171–190. https://doi.org/10.4324/9781315248073-9

Fields, J. M., & Walker, J. G. (1982). The response to railway noise in residential areas in Great Britain. *Journal of Sound and Vibration, 85*(2), 177–255. https://doi.org/10.1016/0022-460X(82)90519-3

Firdaus, G., & Ahmad, A. (2010). Noise pollution and human health: A case study of municipal corporation of Delhi. *Indoor and Built Environment, 19*(6), 648–656. https://doi.org/10.1177/1420326X10370532

Fredianelli, L., Carpita, S., Bernardini, M., Del Pizzo, L. G., Brocchi, F., Bianco, F., & Licitra, G. (2022). Traffic flow detection using camera images and machine learning methods in ITS for noise map and action plan optimization. *Sensors*, *22*(5). https://doi.org/10.3390/s22051929

Gallasch, E., Raggam, R., Cik, M., Rabensteiner, J., Lackner, A., Piber, B., & Marth, E. (2016). Road and rail traffic noise induce comparable extra-aural effects as revealed during a short-term memory test. *Noise and Health, 18*(83), 206–213. https://doi.org/10.4103/1463-1741.189243

Gilani, T. A., & Mir, M. S. (2021). A study on the assessment of traffic noise induced annoyance and awareness levels

about the potential health effects among residents living around a noise-sensitive area. *Environmental Science and Pollution Research, 28*(44), 63045–63064. https://doi.org/10.1007/s11356-021-15208-3

Goines, L., & Hagler, L. (2007). Noise pollution: A modern plague. *Southern Medical Journal, 100*(3), 287–294. https://doi.org/10.1097/SMJ.0b013e3180318be5

Goswami, S., & Swain, B. K. (2017). Environmental noise in India : a Review. *Current Pollution Report, 3*, 220–229. https://doi.org/10.1007/s40726-017-0062-8

Gupta, S., & Anand, R. (2018). *Productivity of railway stations: Case Study-New Delhi Railway Station*. Transport Research Arena, pp. 1–10. https://doi.org/10.5281/zenodo.1491536

Guski, R. (2016). NORAH overview. *Proceedings of the INTER-NOISE 2016 - 45th international congress and exposition on noise control engineering: Towards a quieter future*, 7759–7763.

Guski, R., Klatte, M., Moehler, U., & Müller, U. (2016). Noise Assessment and Control: Paper ICA2016-157 NORAH (Noise Related Annoyance, Cognition, and Health): Questions, designs, and main results NORAH (Noise Related Annoyance, Cognition, and Health): Questions, designs, and main results. In *22nd International Congress on Acoustics studies Ica, 2016*, 0–6.

Hong, J., Kim, J., Kim, K., Cho, Y., & Lee, S. (2011). Quantitative model of combined annoyance caused by simultaneous exposure to outdoor traffic sounds. *Proceedings of the Institute of Acoustics, 33 2*(PART 3), 727–733.

Huboyo, H.S., Hardyanti, N., Meilasari, F. (2022). Noise level of railroad settlements JPL 05 Kaligawe Street, Semarang. *Jurnal Presipitasi: Media Komunikasi dan Pengembangan Teknik Lingkungan, 19*(3), 710-719. https://doi.org/10.14710/presipitasi.v19i3.710-719

Jamir, L., Nongkynrih, B., & Gupta, S. K. (2014). Community noise pollution in urban India: Need for public health action. *Indian Journal of Community Medicine, 39*(1), 8–12. https://doi.org/10.4103/0970-0218.126342

Jiang, S., Meehan, P. A., Thompson, D. J., & Jones, C. J. C. (2013). Railway rolling noise prediction: Field validation and sensitivity analysis. *International Journal of Rail Transportation, 1*(1), 109–127. https://doi.org/10.1080/23248378.2013.788359

Kanakasabai, V., Senthilkumar, G., & Chandrasekar, E. (2005). Study of noise level in railway stations. *Journal of Scientific and Industrial Research, 64*(12), 957–959.

Khaiwal, R., Singh, T., Tripathy, J. P., Mor, S., Munjal, S., Patro, B., & Panda, N. (2016). Assessment of noise pollution in and around a sensitive zone in North India and its non-auditory impacts. *Science of the Total Environment, 566–567*, 981–987. https://doi.org/10.1016/j.scitotenv.2016.05.070

Ko, J. H., Chang, S. I., Kim, M., Holt, J. B., & Seong, J. C. (2011). Transportation noise and exposed population of an urban area in the Republic of Korea. *Environment International, 37*(2), 328–334. https://doi.org/10.1016/j.envint.2010.10.001

Kuhnt, S., Schürmann, C., Schütte, M., Wenning, E., Griefahn, B., Vormann, M., & Hellbrück, J. (2008). Modelling annoyance from combined traffic noises: An experimental study. *Acta Acustica United with Acustica, 94*(3), 393–400. https://doi.org/10.3813/AAA.918047

Kumar, K., (2019). Study on the effects of traffic noise on human health. In *INTER-NOISE and NOISECON Congress and Conference Proceedings, 259*(3), 6783–6794.

Lakušić, S., & Ahac, M. (2012). Rail traffic noise and vibration mitigation measures in urban areas. *Tehnicki Vjesnik, 19*(2), 427–435.

Lercher, P., Brink, M., Rüdisser, J., Van Renterghem, T., Botteldooren, D., Baulac, M., & Defrance, J. (2010). The effects of railway noise on sleep medication intake: Results from the ALPNAP-study. *Noise and Health, 12*(47), 110–119. https://doi.org/10.4103/1463-1741.63211

Lercher, P., Botteldooren, D., Widmann, U., Uhrner, U., & Kammeringer, E. (2011). Cardiovascular effects of environmental noise: Research in Austria. *Noise & Health, 13*(52), 234–250. https://doi.org/10.4103/1463-1741.80160

Lercher, P., de Greve, B., Botteldooren, D., Dekoninck, L., Oettl, D., Uhrner, U., & Rüdisser, J. (2008). Health effects and major co-determinants associated with rail and road noise exposure along transalpine traffic corridors. In *Proceedings of the 9th International Congress on Noise as a Public Health Problem (ICBEN 2008), 2015*, 322–332.

Lercher, P., De Coensel, B., Dekonink, L., & Botteldooren, D. (2017). Community response to multiple sound sources: Integrating acoustic and contextual approaches in the analysis. In *International Journal of Environmental Research and Public Health*, *14*(6). https://doi.org/10.3390/ijerph14060663

Licitra, G., Fredianelli, L., Petri, D., & Vigotti, M. A. (2016). Annoyance evaluation due to overall railway noise and vibration in Pisa urban areas. *Science of the Total Environment, 568*, 1315–1325. https://doi.org/10.1016/j.scitotenv.2015.11.071

Lim, C., Kim, J., Hong, J., & Lee, S. (2006). The relationship between railway noise and community annoyance in Korea. *The Journal of the Acoustical Society of America, 120*(4), 2037–2042. https://doi.org/10.1121/1.2266539

Lim, C., Kim, J., Hong, J., & Lee, S. (2008). Effect of background noise levels on community annoyance from aircraft noise. *The Journal of the Acoustical Society of America, 123*(2), 766–771. https://doi.org/10.1121/1.2821985

Liu, Y., Ma, X., Shu, L., Yang, Q., Zhang, Y., Huo, Z., & Zhou, Z. (2020). Internet of things for noise mapping in smart cities: State of the art and future directions. *IEEE Network, 34*(4), 112–118. https://doi.org/10.1109/MNET.011.1900634

López, J. M., Alonso, J., Asensio, C., Pavón, I., Gascó, L., & de Arcas, G. (2020). A digital signal processor based acoustic sensor for outdoor noise monitoring in smart cities. *Sensors (Switzerland), 20*(3), 1–17. https://doi.org/10.3390/s20030605

Maljaee, S. S., Khadem Sameni, M., & Ahmadi, M. (2022). Effects of railway noise and vibrations on dissatisfaction of residents: Case study of Iran. In *Environment, Development and Sustainability* (Issue 0123456789). Springer Netherlands. https://doi.org/10.1007/s10668-022-02718-4

Maljaee, S. S., Khadem Sameni, M., & Ahmadi, M. (2024). Effects of railway noise and vibrations on dissatisfaction of residents: Case study of Iran. In *Environment,*

Development and Sustainability, *26*(1). Springer Netherlands. https://doi.org/10.1007/s10668-022-02718-4

Mejvald, P., & Konopa, O. (2019). Continuous acoustic monitoring of railroad network in the Czech Republic using smart city sensors. *ICTEP 2019 - International Council of Environmental Engineering Education - & Technologies of Environmental Protection & - Proceedings*, pp. 181–186. https://doi.org/10.1109/ICTEP48662.2019.8969003

Michaud, D. S., Marro, L., Denning, A., Shackleton, S., Toutant, N., & McNamee, J. P. (2022). A comparison of self-reported health status and perceptual responses toward environmental noise in rural, suburban, and urban regions in Canada. *The Journal of the Acoustical Society of America, 151*(3), 1532–1544. https://doi.org/10.1121/10.0009749

Miedema, H. M. E., & Vos, H. (1998). Exposure-response relationships for transportation noise. *The Journal of the Acoustical Society of America, 104*(6), 3432–3445. https://doi.org/10.1121/1.423927

Mimani, A., & Singh, R. (2021). Anthropogenic noise variation in Indian cities due to the COVID-19 lockdown during March-to-May 2020. *The Journal of the Acoustical Society of America, 150*(5), 3216–3227. https://doi.org/10.1121/10.0006966

Mishra, R. K., Shukla, A., Parida, M., & Rangnekar, S. (2014). Eia based comparative urban traffic noise analysis between operational and under construction phase public transport corridor. *International Journal for Traffic and Transport Engineering, 4*(3), 352–362. https://doi.org/10.7708/ijtte.2014.4(3).08

Mohammadi, G. (2009). An investigation of community response to urban traffic noise. In *Global Perspective for Competitive Enterprise, Economy and Ecology - Proceedings of the 16th ISPE International Conference on Concurrent Engineering*, pp. 673–680. https://doi.org/10.1007/978-1-84882-762-2_64

Moreno, R., Bianco, F., Carpita, S., Monticelli, A., Fredianelli, L., & Licitra, G. (2023). Adjusted controlled pass-by (CPB) method for urban road traffic noise assessment. *Sustainability (Switzerland), 15*(6), 1–17. https://doi.org/10.3390/su15065340

Murphy, E., & King, E. A. (2010). Strategic noise mapping, population exposure estimates and "acoustical planning" mitigation analysis in a european capital city. *39th International Congress on Noise Control Engineering 2010, INTER-NOISE 2010, 4*, 2535–2542.

Murphy, E., & King, E. A. (2014). An assessment of residential exposure to environmental noise at a shipping port. *Environment International, 63*(2014), 207–215. https://doi.org/10.1016/j.envint.2013.11.001

Murphy, E., King, E. A., & Rice, H. J. (2009). Estimating human exposure to transport noise in central Dublin, Ireland. *Environment International, 35*(2), 298–302. https://doi.org/10.1016/j.envint.2008.07.026

Nassiri, P., Abbaspour, M., Mahmoodi, M., & Givargis, S. (2007). A rail noise prediction model for the Tehran-Karaj commuter train. *Applied Acoustics, 68*(3), 326–333. https://doi.org/10.1016/j.apacoust.2006.02.003

Němec, M., Gergel', T., Gejdoš, M., Danihelová, A., & Ondrejka, V. (2021). Selected approaches to the assessment of environmental noise from railways in urban areas. *International Journal of Environmental Research and Public Health*, *18*(13). https://doi.org/10.3390/ijerph18137086

Nourani, V., Gökçekuş, H., & Umar, I. K. (2020). Artificial intelligence based ensemble model for prediction of vehicular traffic noise. *Environmental Research, 180*, 108852. https://doi.org/10.1016/j.envres.2019.108852

Oftedal, B., Krog, N. H., Pyko, A., Eriksson, C., Graff-Iversen, S., Haugen, M., Schwarze, P., Pershagen, G., & Aasvang, G. M. (2015). Road traffic noise and markers of obesity – A population-based study. *Environmental Research, 138*, 144–153. https://doi.org/10.1016/j.envres.2015.01.011

Öhrström, E., Barregård, L., Andersson, E., Skånberg, A., Svensson, H., & Ängerheim, P. (2007). Annoyance due to single and combined sound exposure from railway and road traffic. *The Journal of the Acoustical Society of America, 122*(5), 2642. https://doi.org/10.1121/1.2785809

Pallas, M. A., Bérengier, M., Chatagnon, R., Czuka, M., Conter, M., & Muirhead, M. (2016). Towards a model for electric vehicle noise emission in the European prediction method CNOSSOS-EU. *Applied Acoustics, 113*, 89–101. https://doi.org/10.1016/j.apacoust.2016.06.012

Perron, S., Plante, C., Ragettli, M. S., Kaiser, D. J., Goudreau, S., & Smargiassi, A. (2016). Sleep disturbance from road traffic, railways, airplanes and from total environmental noise levels in montreal. *International Journal of Environmental Research and Public Health*, *13*(8). https://doi.org/10.3390/ijerph13080809

Petri, D., Licitra, G., Vigotti, M. A., & Fredianelli, L. (2021). Effects of exposure to road, railway, airport and recreational noise on blood pressure and hypertension. *International Journal of Environmental Research and Public Health*, *18*(17). https://doi.org/10.3390/ijerph18179145

Poisson, F. (2015). Railway noise generated by high-speed trains. *Notes on Numerical Fluid Mechanics and Multidisciplinary Design, 126*, 457–480. https://doi.org/10.1007/978-3-662-44832-8_55

Pultznerova, A., Eva, P., Kucharova, D., & Argalasova, L. (2018). Railway noise annoyance on the railway track in northwest Slovakia. *Noise and Health, 20*(94), 90–100. https://doi.org/10.4103/nah.NAH_59_17

Reiter, M., & Kostek, B. (2008). Comparing railway noise prediction results for passenger trains using various models. *Proceedings - European Conference on Noise Control*, pp. 2637–2642. https://doi.org/10.1121/1.2933582

Renaud, J., Karam, R., Salomon, M., & Couturier, R. (2023). Deep learning and gradient boosting for urban environmental noise monitoring in smart cities. *Expert Systems with Applications, 218*(October 2022), 119568. https://doi.org/10.1016/j.eswa.2023.119568

Roswall, N., Raaschou-Nielsen, O., Jensen, S. S., Tjønneland, A., & Sørensen, M. (2018a). Long-term exposure to residential railway and road traffic noise and risk for diabetes in a Danish cohort. *Environmental Research, 160*(August 2017), 292–297. https://doi.org/10.1016/j.envres.2017.10.008

Roswall, N., Raaschou-nielsen, O., Solvang, S., & Tjønneland, A. (2018b). Long-term exposure to residential railway and road tra ffi c noise and risk for diabetes in a Danish cohort. *Environmental Research, 160*(August 2017), 292–297. https://doi.org/10.1016/j.envres.2017.10.008

Rylander, R. (2004). Physiological aspects of noise-induced stress and annoyance. *Journal of Sound and Vibration, 277*(3), 471–478. https://doi.org/10.1016/j.jsv.2004.03.008

Sahoo, S. (2014). Hypothesis Testing. (2014). In *Analysis of traffic noise* (pp. 25–30). http://ethesis.nitrkl.ac.in/5590/1/E-49.pdf

Sarikavak, Y., & Boxall, A. (2019). The impacts of pollution for new high-speed railways: The case of noise in Turkey. *Acoustics Australia, 47*(2), 141–151. https://doi.org/10.1007/s40857-019-00154-5

Schreckenberg, D., Eikmann, T., Guski, R., Klatte, M., Müller, U., Peschel, C., Schmidt, J., Seidler, A., & Möhler, U. (2012). NORAH (Noise-Related Annoyance, Cognition, and Health) - Konzept einer studie zur wirkung von verkehrslärm bei anwohnern von flughäfen. *Larmbekampfung, 7*(3), 107–114.

Seidler, A., Wagner, M., Schubert, M., Dröge, P., Römer, K., Pons-Kühnemann, J., Swart, E., Zeeb, H., & Hegewald, J. (2016). Aircraft, road and railway traffic noise as risk factors for heart failure and hypertensive heart disease—A case-control study based on secondary data. *International Journal of Hygiene and Environmental Health, 219*(8), 749–758. https://doi.org/10.1016/j.ijheh.2016.09.012

Shah, S. K., Tariq, Z., Lee, J., & Lee, Y. (2020). Real-time machine learning for air quality and environmental noise detection. In *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, 3506–3515. https://doi.org/10.1109/BigData50022.2020.9377939

Sharp, C., Woodcock, J., Sica, G., Peris, E., Moorhouse, A. T., & Waddington, D. C. (2014). Exposure-response relationships for annoyance due to freight and passenger railway vibration exposure in residential environments. *The Journal of the Acoustical Society of America, 135*(1), 205–212. https://doi.org/10.1121/1.4836115

Singh, M., & Gupta, S. (2020). Urban rail system for freight distribution in a mega city: Case study of Delhi, India. *Transportation Research Procedia, 48*(2019), 452–466. https://doi.org/10.1016/j.trpro.2020.08.052

Sørensen, M., Hvidberg, M., Hoffmann, B., Andersen, Z. J., Nordsborg, R. B., Lillelund, K. G., Jakobsen, J., Tjønneland, A., Overvad, K., & Raaschou-Nielsen, O. (2011). Exposure to road traffic and railway noise and associations with blood pressure and self-reported hypertension: A cohort study. *Environmental Health: A Global Access Science Source, 10*(1), 92. https://doi.org/10.1186/1476-069X-10-92

Sørensen, M., Poulsen, A. H., Kroman, N., Hvidtfeldt, U. A., Thacher, J. D., Roswall, N., Brandt, J., Frohn, L. M., Jensen, S. S., Levin, G., & Raaschou-Nielsen, O. (2021). Road and railway noise and risk for breast cancer: A nationwide study covering Denmark. *Environmental Research, 195*. https://doi.org/10.1016/j.envres.2021.110739

Stansfeld, S., Haines, M., & Brown, B. (2000). Noise and health in the urban environment. *Reviews on Environmental Health, 15*(1–2), 43–82. https://doi.org/10.1515/reveh.2000.15.1-2.43

Stansfeld, S. A., & Matheson, M. P. (2003). Noise pollution: Non-auditory effects on health. *British Medical Bulletin, 68*, 243–257. https://doi.org/10.1093/bmb/ldg033

Szwarc, M., Kostek, B., Kotus, J., Szczodrak, M., & Czyżewski, A. (2011). Problems of railway noise—a case study. *International Journal of Occupational Safety and Ergonomics, 17*(3), 309–325. https://doi.org/10.1080/10803548.2011.11076897

Takenaka, K., & Ozawa, K. (2022). Noise suppression system using deep learning for smart devices. In *LifeTech 2022 - 2022 IEEE 4th Global Conference on Life Sciences and Technologies*, 559–560. https://doi.org/10.1109/LifeTech53646.2022.9754759

Tanni, T., Hasan, M., Azad, A., & Bakali, B. (2015). State of the environment in slum area: A case study on Khora Slum, Khulna. *Journal of Environmental Science and Natural Resources, 7*(1), 295–304. https://doi.org/10.3329/jesnr.v7i1.22187

Tassi, P., Saremi, M., Schimchowitsch, S., Eschenlauer, A., Rohmer, O., & Muzet, A. (2010). Cardiovascular responses to railway noise during sleep in young and middle-aged adults. *European Journal of Applied Physiology, 108*(4), 671–680. https://doi.org/10.1007/s00421-009-1270-8

Themann, C. L., & Masterson, E. A. (2019). Occupational noise exposure: A review of its effects, epidemiology, and impact with recommendations for reducing its burden. *The Journal of the Acoustical Society of America, 146*(5), 3879–3905. https://doi.org/10.1121/1.5134465

Thompson, D. J., & Jones, C. J. C. (2000). Review of the modelling of wheel/rail noise generation. *Journal of Sound and Vibration, 231*(3), 519–536. https://doi.org/10.1006/jsvi.1999.2542

Thompson, R., Smith, R. B., Bou Karim, Y., Shen, C., Drummond, K., Teng, C., & Toledano, M. B. (2022). Noise pollution and human cognition: An updated systematic review and meta-analysis of recent evidence. *Environment International, 158*(June 2021), 106905. https://doi.org/10.1016/j.envint.2021.106905

Torija, A. J., Ruiz, D. P., Coensel, B. D., Botteldooren, D., Berglund, B., & Ramos-Ridao, Á. (2011). Relationship between road and railway noise annoyance and overall indoor sound exposure. *Transportation Research Part*

d: Transport and Environment, 16(1), 15–22. https://doi.org/10.1016/j.trd.2010.07.012

Uddin, M. A., Uddin Miah, S. M. S., & Hossan, M. A. (2015). Noise pollution and its impacts case study of selected rail stations of Bangladesh.pdf. *Stamford Journal of Environment and Human Habitat, 4*(May), 15–23.

Vincens, N., & PerssonWaye, K. (2022). Railway noise and diabetes among residents living close to the railways in Västra Götaland, Sweden: Cross-sectional mediation analysis on obesity indicators. *Environmental Research, 212*(PB), 113477. https://doi.org/10.1016/j.envres.2022.113477

World Health Organization, Regional Office for Europe, (2011). Burden of disease from environmental noise: Quantification of healthy life years lost in Europe. World Health Organization. Regional Office for Europe. https://iris.who.int/handle/10665/326424

Wosniacki, G. G., & Zannin, P. H. T. (2021). Framework to manage railway noise exposure in Brazil based on field measurements and strategic noise mapping at the local level. *Science of the Total Environment, 757*(xxxx), 143721. https://doi.org/10.1016/j.scitotenv.2020.143721

Wrótny, M., & Bohatkiewicz, J. (2021). Traffic noise and inhabitant health—A comparison of road and rail noise. *Sustainability, 13*, 7340. https://doi.org/10.3390/su13137340

Yoon, K., Gwak, D. Y., Chun, C., Seong, Y., Hong, J., & Lee, S. (2018). Analysis of frequency dependence on short-term annoyance of conventional railway noise using sound quality metrics in a laboratory context. *Applied Acoustics, 138*(April 2017), 121–132. https://doi.org/10.1016/j.apacoust.2018.03.024

Zannin, P. H. T., Quadros, F., De Oliveira, F. L., & Nascimento, E. ODo. (2018). Evaluation of environmental noise generated by household waste collection trucks. *Journal of Environmental Assessment Policy and Management, 20*(4), 1–17. https://doi.org/10.1142/S1464333218500102

Zhang X., Kuehnelt H., D R, W. (2020). Pass-by Noise Modelling Applying Machine Learning. Forum Acusticum, Dec 2020, Lyon, France. pp. 2251-2258, https://doi.org/10.48465/fa.2020.0288.hal03235439

Zhang, L., & Ma, H. (2021). Investigation of Chinese residents' community response to high-speed railway noise. *Applied Acoustics, 172*, 107615. https://doi.org/10.1016/j.apacoust.2020.107615

Zhang, L., & Ma, H. (2022). The effects of environmental noise on children's cognitive performance and annoyance. *Applied Acoustics, 198*, 108995. https://doi.org/10.1016/j.apacoust.2022.108995

Zou, C., Wang, Y., Wang, P., & Guo, J. (2015). Measurement of ground and nearby building vibration and noise induced by trains in a metro depot. *Science of the Total Environment, 536*, 761–773. https://doi.org/10.1016/j.scitotenv.2015.07.123

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.

EEG-Based Emotion Recognition Using Optimized Deep-Learning Techniques

2024 11th International Conference on Signal Processing and Integrated Networks (SPIN) | 979-8-3503-0843-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/SPIN60856.2024.10512074

Amit Kumar Dwivedi
Department of Electronics and Communication
Delhi Technological University
Rohini, 110042, New Delhi, India.
amitkumardwivedi_2k20phdec512@dtu.ac.in

Om Prakash Verma
Department of Electronics and Communication
Delhi Technological University
Rohini, 110042, New Delhi, India.
opverma@dce.ac.in

Sachin Taran
Department of Electronics and Communication
Delhi Technological University
Rohini, 110042, New Delhi, India.
taransachin2@gmail.com

***Abstract*— Emotions are crucial in identifying the current mental state. Electroencephalography (EEG) signals can accurately determine the current mental state. The automatic identification and analysis of human emotions using EEG data is vital in the treatment of psychiatric diseases due to the influence that emotions have on interactions, interpretations, and decisions. However, one major problem with EEG recorders is their low spatial resolution. EEG signals are nonlinear and complicated. Advanced signal processing techniques are required to analyze and extract useful characteristics. Emotion recognition using EEG can improve human-machine interactions. A convolutional neural network (CNN) is suggested in this article for automatic feature extraction and categorization of emotions. First, EEG signals are converted into pictures by applying time-frequency representation techniques. Following this, these pictures are fed into CNN and optimized CNN models for training. The optimized CNN model produces better accuracy for classifying emotions. Deep learning (DL) techniques are popular for determining complex patterns. Deep learning methods have shown good results in EEG-based emotion identification. It can automatically extract high-level information. DL's performance depends on hyperparameters. However, optimizing the performance of a deep-learning model requires accurate hyperparameter tuning. The manual hyperparameter tuning method is tiresome, costly, computationally expensive, and time-consuming. As a result, an automated method is required to determine the best hyperparameters to maximize DL's efficacy. To automate the hyperparameter optimization process, the proposed work presents a novel framework based on Bayesian optimization. Bayesian optimization (BO) is used to automate hyperparameter selection to predict emotional states.**

Keywords—EEG, Emotion Recognition, Deep-Learning Methods, Bayesian Optimization.

I. INTRODUCTION

Emotions are physiological phenomena often resulting in motivation, personality, temperament, behavior, and feelings. The study of the link between emotions and physiological functions is becoming increasingly popular in research. Emotions have a big effect on daily activities such as work, play, and interpersonal relations. Emotions are broadly classified into two types: positive and negative, positive emotions result in good mental health[1]. Whereas, negative emotions result in poor mental and physical health and can also produce psychiatric issues.

EEG signals are utilized to identify an individual's emotional state. Human-machine interactions can be enhanced by brain-computer interfaces (BCIs) by utilizing these signals [2]-[3]. Examining useful EEG characteristics is essential for identifying emotions. Even though EEG tests usually require a large array of electrodes on the scalp (sometimes more than 30), it is rarely addressed to identify the spatial characteristics that best reflect the functional interactions among different EEG channels. Innovative research that uses multichannel EEG recordings to investigate these spatial aspects has recently come to light [4]. Investigating the complex interplay between external and psychological elements. Emotion recognition is a rapidly expanding interdisciplinary research field spanning computer engineering and neuroscience[5].

A study that combined machine learning with EEG signals discovered a strong relationship between emotional stimuli and EEG responses in a variety of models [6]. Deep-learning (DL) techniques have seen tremendous and quick progress in the last several years, attracting a lot of interest from a variety of study fields. Recurrent neural networks (RNNs) [7], CNNs [8], and deep belief networks (DBNs) [9] are a few network architectures that have been proposed for emotion recognition. These models have had a major influence on areas including speech recognition [10], image classification [11], and time-series prediction [12], [13], because of their excellent performance and computational economy.

Dynamical graph convolutional neural networks (DGCNN) trained from EEG channel relationships result in improved feature extraction and produce precise emotion classification [14]. ERACNN (emotion recognition convolution neural network) [15] is a three-layer CNN consisting of a fully connected layer, a pool, and a convolutional layer that provides accurate emotion classification results.

A two-dimensional CNN, which consists of four convolutional layers and seven fully connected layers, has an accuracy of 76.56% for arousal emotions and 80.46% for valence feelings [16]. In [17], use a CNN framework with a residual block, convolutional layer, max pooling layer, fully connected layer, and softmax layer to identify emotions. The CNN uses three convolutional layers in the residual block for EEG feature extraction, with the third layer acting as a shortcut connection. It attains a mean accuracy of 82.84% on the DEAP dataset. A CNN technique for multichannel EEG-based emotion identification is presented in [18]. The method consists of generating a 3-D feature matrix, applying three 1×1 convolutional layers as a univariate convolutional layer, adding a four-way convolution to a multivariate convolutional layer, and finishing with a fully-connected layer. On the DEAP dataset, the approach obtains an accuracy of 85.53% for valence and 85.88% for arousal.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 06:08:25 UTC from IEEE Xplore. Restrictions apply.

In [19], the regional-asymmetric CNN (RACNN) has been used for EEG-based emotion identification, employing three feature extractors and achieving 95% recognition accuracy on DEAP and DREAMER datasets. Deep learning approaches have shown good accuracy, but their practical implementation is limited by the time-consuming and specialized knowledge. Thus, an automated approach is needed to replace laborious manual experimentation and expedite the creation of appropriate and effective deep-learning models. This paper suggests a method for emotion recognition using convolutional neural networks and time-frequency images. Time-domain EEG signals are converted into representations that include frequency, amplitude, and time via a scalogram, spectrogram, and smoothed pseudo-Wigner–Ville distribution. These representations serve as inputs for CNN architectures and are shown as time-frequency images. The classification task utilizes CNN and along with an optimized CNN. The optimized CNN shows better accuracy. The structure of this paper is as follows: In Section 2, explains the details of dataset and different time-frequency methods used in the work. Section 3 outlines the research methodology, detailing the processes of applying deep learning, feature extraction, and classification. Section 4, shows concluding remark.

II. THE PROPOSED METHOD

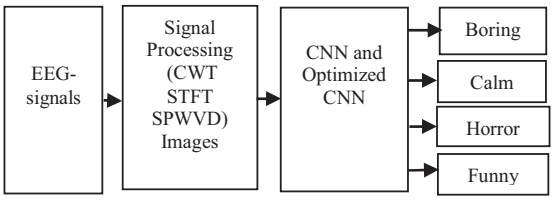


Figure 1. Block diagram of proposed work.

A. Dataset

The GAMEEMO dataset that has been utilized in this study contains recorded EEG signal data from 28 people between the ages of 20 and 27. The data were captured using a portable 14-channel EMOTIV EPOC + mobile EEG device[20]. Each participant played the games for five minutes, for a total of twenty minutes. The EEG data from four computer games: 'calm' (G2), 'boring' (G1), 'funny' (G4), or 'horror' (G3) has been used. The EEG signals were sampled at 128 Hz and subjected to band-pass filtering (0.16 Hz minimum, 43 Hz maximum). The dataset contains 1,568 EEG signals, with 38,252 samples per game (4 games × 14 channels × 28 participants). Figure 2 displays the 14-channel EEG data for the four games that were recorded from a single subject.

B. Time-Frequency Representation Continuous Wavelet Transform:

Time-Frequency Representation (TFR) is a useful technique for displaying the spectral variations of a signal over time. This article's methodology in Figure 1 makes use of the Short-Time Fourier Transform (STFT), the Continuous Wavelet Transform (CWT), and the SPWVD. It is simpler to convert EEG signals into their time-frequency representations when these transformation methods are coupled. The unique insights that each technique provides into the behavior of the signal allow for a more thorough and sophisticated study of the dynamic frequency components of the signal over time.

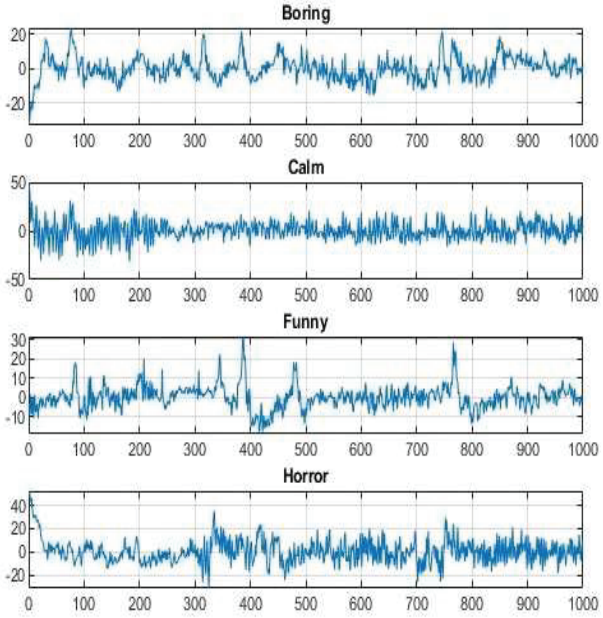


Figure 2. EEG Signals of four emotions.

C. Continuous Wavelet Transform (CWT):

A band-pass filter with finite energy and a zero mean function that oscillates within a particular frequency range is called a wavelet. There are two ways to accomplish the wavelet transform: the CWT and the discrete wavelet transform (DWT). The signal x(t)'s CWT is represented by [21].

$$W_{c,d}[x(t)] = \frac{1}{\sqrt{c}} \int_{-\infty}^{\infty} x(t) \Psi^{*}\left(\frac{t-d}{c}\right) \mathrm{dt} \quad (1)$$

Where the scaling parameter c regulates the function's spread and d represents the time shift. The scaled and shifted form of the mother wavelet ψ(t) is ψ((t − d)/c). CWT has the benefit of having adjustable window sizes. Thus, low-frequency components can be analyzed with bigger windows, and high-frequency components can be examined with smaller windows. Figure 3 shows the scalogram of EEG data during the boring game from a single participant.

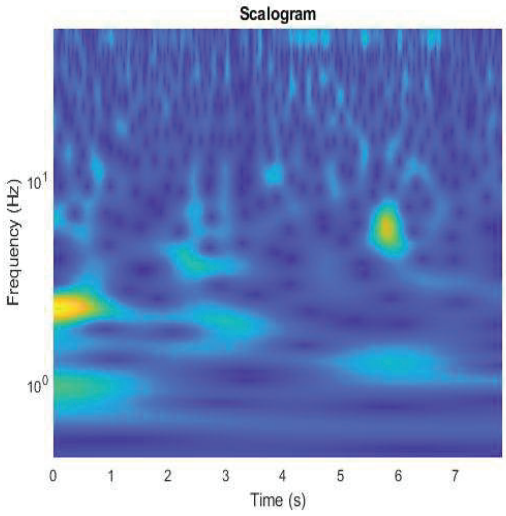


Figure 3.Scalogram of EEG data.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 06:08:25 UTC from IEEE Xplore. Restrictions apply.

D. Smoothed Pseudo Wigner-Ville Distribution

SPWVD is a robust transformation method that gives complete details about the time-frequency localization of a signal's energy. SPWVD gives information on the time-domain localization of the signal and makes it easier to examine higher harmonics. Cross-term interference is efficiently minimized in SPWVD by utilizing both time and frequency windows. The formula for SPWVD is defined as [22]:

$$SPWVD(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m\left(\frac{\tau}{2}\right) m^*\left(-\frac{\tau}{2}\right) \quad (2)$$

$$\text{n(t-t')}\ldots\ldots\ldots\ldots\times x(t'+\frac{\tau}{2}) x^*\left(t'-\frac{\tau}{2}\right) exp^{-j2\pi f t} dt'\, d\tau \quad (3)$$

SPWVD utilizes windows in the time (m (t)) and frequency (n (t)) domains to minimize interference. These windows are one by one adjusted for precise control over time and frequency details. The optimum between interference and time-frequency precision in conventional techniques is overcome by SPWVD, in contrast to STFT and CWT. Figure 4 shows the SPWVD of EEG data.



Figure 4. SPWVD of EEG Data.

E. Short Time Fourier Transform (STFT)

The Fourier transform in its enhanced form is known as the Short-Time Fourier transform (STFT). It evaluates the frequency variations of a signal over time. The time-domain signal is split into discrete blocks, and each block is then subjected to the Fourier transform. STFT has found several uses with EEG signals because of its ability to divide a signal into blocks [23], [24], and [25]. Several features are extracted from each time-domain signal segment by applying the Fourier transform using a window function. In symmetric band pass filtering, the frequency bands are evenly spaced in STFT. The signal x(t)'s STFT is calculated as [26]

$$X_\omega(t,f) = \int_0^T x(\tau)\omega(t-\tau) exp^{-j2\pi f\tau} \mathrm{d}\tau \quad (4)$$

The squared magnitude of the STFT defines the spectrogram. The spectrogram can be expressed mathematically as $S_\omega(t,f) = |X_\omega(t,f)|^2$, where the window function is denoted by ω(t). There must be uniformity in the length of the windows in every block. During the window's duration, a signal x(t) is taken to be almost stationary. Figure 5 shows STFT of the EEG data.

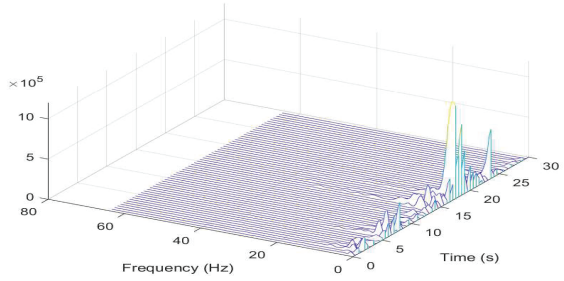


Figure 5.Spectrogram of EEG Signal.

F. Deep-Learning Hyperparameters Tuning using Bayesian optimization

Bayesian optimization is a process that combines deep learning models with optimization methods to find the best hyperparameters. In deep learning, hyperparameters are configuration parameters that must be defined before the model can be trained. These parameters were not learned from the training data. These may include batch sizes, learning rates, dropout rates, and other factors. Bayesian optimization is a probabilistic model-based optimization. It uses iterative exploration and exploitation of the hyperparameter space to determine the best set of hyperparameters. It builds a probabilistic model of the objective function of the deep learning model's performance and uses this model to guide the search for the next set of hyperparameters to evaluate. Grid search, random search, evolutionary algorithms, and gradient-based optimization techniques are some of the optimization strategies available for hyperparameter optimization. Grid search is computationally expensive. Random search is computationally efficient, but finding an optimal solution is difficult; and after every iteration, there is no guarantee of better model performance. Bayesian optimization is preferred over other methods due its advantages. Bayesian optimization requires fewer evaluations of objective functions. Bayesian optimization dynamically adjusts the sampling approach based on previous evaluations. Bayesian optimization can deal with noisy or stochastic objective functions using the surrogate model. Probabilistic surrogate models include some uncertainty, which helps with decision-making when choosing the subsequent set of evaluation hyperparameters.

III. RESULTS & DISCUSSION

In this work, the GAMEEMO dataset has been used, which has 14 channels and preprocessed EEG signals and contains 38252 samples from 28 subjects playing four distinct games. Nine segments, each measuring 3840 samples, are created from each EEG signal. To analyze the data channel-wise, firstly, the data from all 28 subjects is combined for a single epoch of the same channel, yielding a dimension for each epoch of (4×28) ×3840. Finally, signals from all nine epochs are combined, yielding a total signal dimension for each channel total signal of ((4×28×9) ×3840) or 1008×3840. A sequential transformation is required to transform the EEG signals into EEG images. The 14-channel EEG readings are analyzed individually.

The window length is automatically calculated based on time-frequency variations by applying the continuous wavelet transform (CWT) to the chosen EEG channel signals. In this work, another technique is employed to transform the 1-D EEG signal into a 2-D signal, called SPWVD. The Kaiser window is used to reduce cross-terms in the frequency and temporal domains. Selecting the right window size is important since a large window might result in a considerable increase in the size of the image, while a very small window may yield inferior resolution. As a result, a medium-sized window measuring 31 is chosen using empirical means.

To guarantee computational efficiency, the window size is kept constant at 2^n - 1, where n is the bit count. The third technique used for time-frequency analysis is STFT. It is a

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 06:08:25 UTC from IEEE Xplore. Restrictions apply.

method for analyzing the temporal evolution of frequency components in a nonstationary signal. The spectrogram, or squared magnitude of the STFT, shows the spectral characteristics of the signal in a time-frequency format. A Hann window is used to window each of the 128 sample segments with an overlap length of 64. Spectrogram images are produced using a water plot. After that, the time-frequency representation is transformed into a grayscale image with dimensions of 1008 x 3840 after accounting for a scaling factor. After that, a jet128 color map is used to color the grayscale image's pixel brightness and contrast values, creating an EEG color image that is 1024 × 1024×3. The colorful photos are downsized to make them compatible with the Google Net model, which demands an input size of 224 × 224 × 3. The colorful photos are downsized to make them compatible with the Google Net model, which demands an input size of 224 × 224×3. The result of this multi-step procedure is the encoding of EEG signals as images that may be used as input into deep learning models for applications like gaming emotion recognition related to cognitive state assessments.

To train the deep-learning network, import the scalogram, spectrogram, and SPWVD images separately to create the image data store. The pictures are sorted and labeled according to their folder names using the image data store function. It reads batches of images quickly, which helps with CNN training. The images of three types of time-frequency methods are randomly split into three types: 70% of images are allocated for training, 10% of images are allocated for validation, and the rest of 10% are allocated for testing.

In deep learning, every layer in the network design is a filter that picks up basic image characteristics at the lower levels, like blobs, edges, and colors. The focus changes to more complex characteristics to differentiate between categories as one moves through the layers. A pre-trained network gives good results in the classification of images. GoogleNet is one of the most popularly used pre-trained networks. It is necessary to retrain GoogLeNet to tackle the particular EEG classification challenge, as it was initially trained to identify 1000 object categories in pictures. The network's convolutional layers extract image features, which are subsequently used to categorize the input image by the final classification layer and the final learnable layer. These two GoogLeNet layers—loss3-classifier and output—contain vital details about how to combine the extracted features into projected labels, a loss value, and class probabilities. To modify GoogLeNet for RGB picture categorization, these two layers must be replaced with new layers that are customized for the unique properties of the input.

Table 1 First seven channel validation accuracy

Deep-Learning Model	Time-Frequency Technique	First Seven Channel's validation Accuracy for classification of four gaming emotions (%)						
		AF3	AF4	F3	F4	F7	F8	FC5
Google-Net	Scalogram	60	55.4	63	64	48	60	65.5
	Spectrogram	42.7	38.4	56	40	37	37	38
	SPWVD	70	64	74	78	60	70	76
Optimized Google Net	Scalogram	71	64	74	73	59	71	76
	Spectrogram	53	47	67	52	48	48	50
	SPWVD	81	85	87	92	87	81	95

Table 2 Next eight channel validation accuracy

Deep-Learning Model	Time-Frequency Technique	Next eight Channel's validation Accuracy for classification of four gaming emotions (%)						
		FC 6	T7	T8	P7	P8	O1	O2
Google-Net	Scalogram	63	59	63	61	57	66	62
	Spectrogram	40	37	47	39	38	38	39
	SPWVD	74	71	74	73	68	75	71
Optimized Google Net	Scalogram	76	71	75	72	68	74	73
	Spectrogram	51	48	56	48	49	49	51
	SPWVD	83	81	88	82	78	87	83

Table 1 and Table 2 display the validation accuracy for the classification of four different emotions. It shows that the Optimized version of GoogleNet provides better accuracy than the normal version. It can be inferred from the tables that SPWVD images give better results in terms of validation accuracy than the scalogram and STFT time-frequency images. EEG channels AF3, F3, F4, F8, O1, T8, P7, FC5, and O2 are showing better accuracy than the other channels. Table 3 displays the empirical values of hyperparameters used to train GoogleNet using time-frequency images.

Table 3 Training parameter for GoogleNet

Parameters	Value
Minimization of Loss Function algorithm	'sgdm'
Max-Epochs	20
Minimum Batch Size	15
Initial-Learn-Rate	$1\times e^{-4}$
Validation Data	imgsValidation
Validation Frequency	10
Momentum	0.9
L2Regularization	0.0001

The performance of GoogleNet can be improved by using optimization. Bayesian optimization algorithms have been used to choose the optimal values of the hyperparameters. The hyperparameters which has optimized is section depth, initial learning rate, momentum, regularization, and loss function. Section depth shows the number of layers within a module of neural network, adjusting section depth allows to find a balance between model complexity and efficiency. Initial Learning Rate is the parameter that determines how quickly the model learns from the data during training, variations in the initial learning rate have an impact on the stability and speed of model training. Momentum optimization expedites convergence. Increases in momentum lead to smoother and more consistent updates since the optimization method gives the previous gradients more weight. Lower momentum levels, on the other hand, lessen the impact of previous gradients, which may assist the algorithm in escaping local minima more readily but may also lead to more uncertain updates. The influence of previous gradients changes with momentum, affecting convergence stability and speed. To avoid overfitting, L2 regularization penalizes large weights. Modifying it affects how well training data fit and generalization are balanced. By calculating the difference between expected and actual values, the loss function algorithm evaluates the effectiveness of the model. Modifying the loss function has an impact on the model's learning process and how it prioritizes different training set elements, which could change the model's behavior and performance. Table 4 displays the ranges of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 06:08:25 UTC from IEEE Xplore. Restrictions apply.

values of the hyperparameters that are to be optimized and their optimized values.

Table 4 Range of hyperparameters used in Bayesian optimization

Hyperparameters	Range	Data Type	Transform
Section Depth	[1 5]	Integer	None
Initial Learning Rate	[1e-4 1]	Real	log
Momentum	[0.5 0.95]	Real	None
L2Regularization	[1e-12	Real	log
Loss Function algorithm	[‘Adam’, ‘sgdm’]	Categorical	None
Optimized values of hyperparameters			
Section Depth	3		
Initial Learning Rate	0.01327		
Momentum	0.65334		

Table 5 displays the final set of hyperparameters used for training the GoogleNet.

Table 5 Training Parameters for Optimized Google Net

Parameter	Value
Minimization of Loss Function algorithm	‘Adam’
Max-Epochs	20
Minimum Batch Size	15
Initial-Learn-Rate	0.01327
Validation Data	imgsValidation
Validation Frequency	10
Momentum	0.65334
L2Regularization	1.7732×e-7
Section Depth	3

Optimized Google is trained using the combined image data of a scalogram, spectrogram, and SPWVD from all channels. The optimized GoogleNet extracts the characteristics from the images, which are used as input to SVM classifiers for training. The trained SVM classifier tests the image data that was not used in training and testing. The classification performance is displayed using an ROC curve and confusion matrix in Figure 6 and Figure **7**. The fact that Figure 6 demonstrates that the true class and predicted class accuracy are higher than 82% indicates the dependability of the suggested model. Figure **7** shows that the AUC for the four classes is 0.9712, 0.9596, 0.9597, and 0.9625, demonstrating the clear distinction among the four classes.

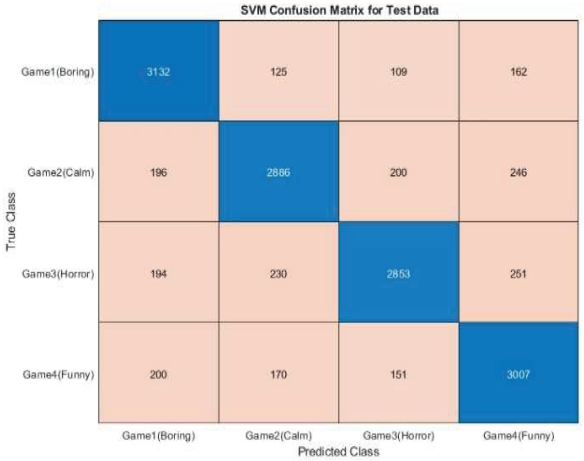


Figure 6. Confusion Matrix for Test Data.

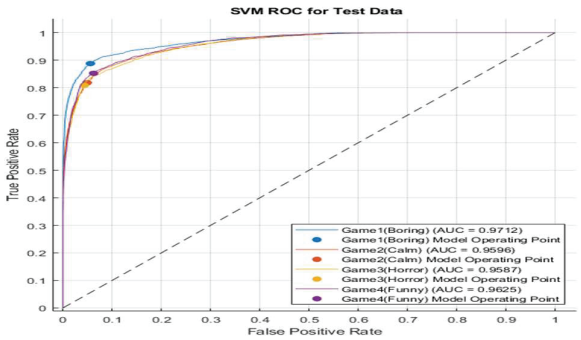


Figure 7.ROC Curve for Test Data.

Table 6 displays the results of this previous work, along with pertinent findings from similar studies that used similar statements and datasets. A multi-layer perceptron neural network (MLPNN) was used in the study of Alakus et al. [27] to classify data into four groups, with a maximum accuracy of 73.21% and 42.85% and 55% accuracy, respectively. The kNN and SVM classifiers showed relatively poorer performance. According to Alakus et al., regarding 2-class classifications, the accuracy was 81.3% with MLPNN, 72.5% with SVM, and 66% with kNN. Calculating the spectral entropy values of EEG signals across all channels was a different strategy by Alakus and Turkoglu et al. [28]. Next, a bidirectional long-short-term memory (LSTM) architecture was applied to the classification task, yielding 76.91% accuracy. For 2-class classification using SVM, kNN, and ELM classifiers, Aslam et al. [29] found high accuracies of 98.78%, 98.53%, and 98.41%, respectively. Tuncer et al.[30] used a cubic SVM classifier to achieve a channel-wise 2-class classification accuracy of 99.29%. With the use of projection dictionary learning, Su et al.[31] recorded an accuracy of 64.34% for two classes and 49.01% for four.

Table 6. Comparative study with previous work.

Author	Features used	Classification algorithm	Number of Classes	Performance Measure (%)
Alakus [27] et al., 2020	Hjorth Parameter, Entropy, Standard deviation, Variance and zero crossing	SVM, kNN and MLPNN	4-Class	k-NN-42.85 SVM-55 MLPNN-73.21
Alakus [27] et al., 2020	Hjorth Paramers, Entropy feature, Standard deviation, Variance and zero crossing	SVM, k-NN and MLPNN	2- Class	k-NN-66 SVM-72.5 MLPP-81.3
Alakus and Turkoglu [28] et al., 2020	Spectral Entropy	Bi-LSTM	2- Class	Bi-LSTM 76.93(%)
Aslam [29] et al., 2021	EEG images (Scalogram) features extracted by pertained Google Net	SVM, k-NN and ELM	2-Class	k-NN-98.53 SVM-98.87 ELM-98.41
Tuncer [30] et al., 2021	Fused feature is generated by LEDPatNet19	Cubic SVM	2- Class	LEDPatNet19 Cubic SVM-99
Su[31] et al., 2023	Genetic Algorithm-projection dictionary pair learning (GA-PDPL)	GA-PDPL	4-Class 2-Class	49.01 64.34
Proposed Method	Optimized GoogleNet	SVM	4-Class	84.2

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 06:08:25 UTC from IEEE Xplore. Restrictions apply.

IV. CONCLUSION

This study investigates how well CNN-based models and various time-frequency techniques distinguish between various gaming emotions. Three conditions were investigated one showed better classification than the other two conditions. Compared to CWT and STFT methods, the SPWVD method yielded higher accuracy. This suggests that the time-frequency representation of SPWVD offers more comprehensive information than spectrogram and scalogram plots. Among other CNN-based networks, the suggested optimized CNN model demonstrated the fastest and most efficient computation but also the highest accuracy. Integration of EEG signals with additional wearable sensors is planned for future work. Better results are expected as a result of this integration, which attempts to improve classification accuracy.

REFERENCES

[1] A. Baudry, D. Grynberg, C. Dassonneville, S. Lelorain, and V. Christophe, "Sub-dimensions of trait emotional intelligence and health: A critical and systematic review of the literature," *Scandinavian J Psychology*, vol. 59, no. 2, pp. 206–222, Apr. 2018, doi: 10.1111/sjop.12424.

[2] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 10, pp. 1175–1191, Oct. 2001, doi: 10.1109/34.954607.

[3] S. Koelstra *et al.*, "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012, doi: 10.1109/T-AFFC.2011.15.

[4] X. Li, D. Song, P. Zhang, G. Yu, Y. Hou, and B. Hu, "Emotion recognition from multi-channel EEG data through Convolutional Recurrent Neural Network," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Shenzhen: IEEE, Dec. 2016, pp. 352–359. doi: 10.1109/BIBM.2016.7822545.

[5] Zhihong Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009, doi: 10.1109/TPAMI.2008.52.

[6] R. Alhalaseh and S. Alasasfeh, "Machine-Learning-Based Emotion Recognition System Using EEG Signals," *Computers*, vol. 9, no. 4, p. 95, Nov. 2020, doi: 10.3390/computers9040095.

[7] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[8] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA: IEEE, Jun. 2015, pp. 1–9. doi: 10.1109/CVPR.2015.7298594.

[9] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006, doi: 10.1162/neco.2006.18.7.1527.

[10] O. Abdel-Hamid, A. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional Neural Networks for Speech Recognition," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 10, pp. 1533–1545, Oct. 2014, doi: 10.1109/TASLP.2014.2339736.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[12] W. Yu *et al.*, "Integrating Spatio-Temporal and Generative Adversarial Networks for Enhanced Nowcasting Performance," *Remote Sensing*, vol. 15, no. 15, p. 3720, Jul. 2023, doi: 10.3390/rs15153720.

[13] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human Trajectory Prediction in Crowded Spaces," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 961–971. doi: 10.1109/CVPR.2016.110.

[14] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks," *IEEE Trans. Affective Comput.*, vol. 11, no. 3, Art. no. 3, Jul. 2020, doi: 10.1109/TAFFC.2018.2817622.

[15] C. Cheng, X. Wei, and Z. Jian, "Emotion recognition algorithm based on convolution neural network," in *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, Nanjing: IEEE, Nov. 2017, pp. 1–5. doi: 10.1109/ISKE.2017.8258786.

[16] Y.-H. Kwon, S.-B. Shin, and S.-D. Kim, "Electroencephalography Based Fusion Two-Dimensional (2D)-Convolution Neural Networks (CNN) Model for Emotion Recognition System," *Sensors*, vol. 18, no. 5, p. 1383, Apr. 2018, doi: 10.3390/s18051383.

[17] F. Wang *et al.*, "Emotion recognition with convolutional neural network and EEG-based EFDMs," *Neuropsychologia*, vol. 146, p. 107506, Sep. 2020, doi: 10.1016/j.neuropsychologia.2020.107506.

[18] H. Chao and L. Dong, "Emotion Recognition Using Three-Dimensional Feature and Convolutional Neural Network from Multichannel EEG Signals," *IEEE Sensors J.*, vol. 21, no. 2, pp. 2024–2034, Jan. 2021, doi: 10.1109/JSEN.2020.3020828.

[19] H. Cui, A. Liu, X. Zhang, X. Chen, K. Wang, and X. Chen, "EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network," *Knowledge-Based Systems*, vol. 205, p. 106243, Oct. 2020, doi: 10.1016/j.knosys.2020.106243.

[20] T. B. Alakus, M. Gonen, and I. Turkoglu, "Database for an emotion recognition system based on EEG signals and various computer games – GAMEEMO," *Biomedical Signal Processing and Control*, vol. 60, p. 101951, Jul. 2020, doi: 10.1016/j.bspc.2020.101951.

[21] J.-M. Combes, Ed., *Wavelets: time-frequency methods and phase space ; proceedings of the International Conference, Marseille, France, December 14 - 18, 1987*, 2. ed. in Inverse problems and theoretical imaging. Berlin Heidelberg New York London Paris Tokyo Hong Kong Barcelona: Springer, 1990.

[22] E. Pereira De Souza Neto, M.-A. Custaud, J. Frutoso, L. Somody, C. Gharib, and J.-O. Fortrat, "Smoothed pseudo Wigner–Ville distribution as an alternative to Fourier transform in rats," *Autonomic Neuroscience*, vol. 87, no. 2–3, pp. 258–267, Mar. 2001, doi: 10.1016/S1566-0702(00)00211-3.

[23] A. Bachiller *et al.*, "Decreased entropy modulation of EEG response to novelty and relevance in schizophrenia during a P300 task," *Eur Arch Psychiatry Clin Neurosci*, vol. 265, no. 6, pp. 525–535, Sep. 2015, doi: 10.1007/s00406-014-0525-5.

[24] U. Budak, V. Bajaj, Y. Akbulut, O. Atila, and A. Sengur, "An Effective Hybrid Model for EEG-Based Drowsiness Detection," *IEEE Sensors J.*, vol. 19, no. 17, pp. 7624–7631, Sep. 2019, doi: 10.1109/JSEN.2019.2917850.

[25] S. Chaudhary, S. Taran, V. Bajaj, and A. Sengur, "Convolutional Neural Network Based Approach Towards Motor Imagery Tasks EEG Signals Classification," *IEEE Sensors J.*, vol. 19, no. 12, pp. 4494–4500, Jun. 2019, doi: 10.1109/JSEN.2019.2899645.

[26] S. Nawab, T. Quatieri, and Jae Lim, "Signal reconstruction from short-time Fourier transform magnitude," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, no. 4, pp. 986–998, Aug. 1983, doi: 10.1109/TASSP.1983.1164162.

[27] T. B. Alakus, M. Gonen, and I. Turkoglu, "Database for an emotion recognition system based on EEG signals and various computer games – GAMEEMO," *Biomedical Signal Processing and Control*, vol. 60, p. 101951, Jul. 2020, doi: 10.1016/j.bspc.2020.101951.

[28] T. B. Alakus and I. Turkoglu, "Emotion recognition with deep learning using GAMEEMO data set," *Electron. lett.*, vol. 56, no. 25, pp. 1364–1367, Dec. 2020, doi: 10.1049/el.2020.2460.

[29] M. Aslan, "CNN based efficient approach for emotion recognition," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 9, pp. 7335–7346, Oct. 2022, doi: 10.1016/j.jksuci.2021.08.021.

[30] T. Tuncer, S. Dogan, and A. Subasi, "LEDPatNet19: Automated Emotion Recognition Model based on Nonlinear LED Pattern Feature Extraction Function using EEG Signals," *Cogn Neurodyn*, vol. 16, no. 4, pp. 779–790, Aug. 2022, doi: 10.1007/s11571-021-09748-0.

[31] J. Su, J. Zhu, T. Song, and H. Chang, "Subject-Independent EEG Emotion Recognition Based on Genetically Optimized Projection Dictionary Pair Learning," *Brain Sciences*, vol. 13, no. 7, p. 977, Jun. 2023, doi: 10.3390/brainsci13070977.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 06:08:25 UTC from IEEE Xplore. Restrictions apply.

Emotion Recognition in Speech Using MFCC and Energy Based Ratio Features

Ravi
Department of electronics and communication engineering
Delhi Technological University (DTU),
Shahbad Daulatpur, New Delhi,110042
ravi_2k20phdec12@dtu.ac.in

Sachin Taran
Department of electronics and communication engineering
Delhi Technological University (DTU),
Shahbad Daulatpur, New Delhi,110042
sachintaran@dtu.ac.in

2024 11th International Conference on Signal Processing and Integrated Networks (SPIN) | 979-8-3503-0843-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/SPIN60856.2024.10511355

***Abstract*— Emotion recognition from speech signal has emerged as a prominent research focus within the realm of human computer interaction in the last decade. Nonetheless, the existing recognition accuracy requires enhancement. To improve the recognition accuracy the proposed framework adopts a global feature extraction approach, decomposing the input speech signal into Intrinsic Mode Functions (IMFs). After that, the Mel Frequency Cepstral Coefficients (MFCC), and energy-based ratio features are extracted from each IMF. The ReliefF algorithm is employed to select pertinent features, which are subsequently tested on an optimizable ensemble classifier for evaluation. The proposed method provides 95.3% emotion recognition accuracy.**

Keywords*— Speech signal processing, Emotion recognition, Empirical mode decomposition Ratio feature, MFCC.

I. INTRODUCTION

Speech Emotion Recognition (SER) has a significant place in enhancing human-computer interaction (HCI) systems. Its contributions are found in many areas, including as audio surveillance, computer games, robotic interfaces, e-learning, and clinical research. Notably, SER can support educators by providing insights into a learner's emotional state, enabling the establishment of a classroom atmosphere that is favorable to learning, as described in [1].

Speech signal non-stationarity makes emotion recognition in speech signals a difficult task even today. For example, human listeners only successfully identify an anonymous speaker's emotions 60% of the time [1]. Additionally, the scarcity of available training data poses a significant obstacle, with most SER methods reporting accuracies ranging from 70% to 90% [2-14]. Another hurdle in SER is the development of language-independent systems, as accuracy is affected by factors such as the number of emotions, the size of the training dataset, the gender, the language, the selected feature set, and the emotion class. SER models can be optimized by selecting carefully features that efficiently capture every facet of the input signal [2]. However, the local and global features are utilized for addressing the non-stationarity of speech signals. Local features represent temporal dynamics, while global features capture statistical measures such as maximum value, standard deviation, and mean value. Further, different features based on speech quality, spectral variation, and prosodic nature of speech signal is employed in SER systems based on their applications [3]. Prosodic features analyze the rhythm and intonation of speech signals, capturing human perception through estimations of energy, length, and fundamental frequency. Meanwhile, spectral features focus on vocal cord characteristics.

For emotion classification other methods are proposed as long-term temporal contexts using recurrent Neural Networks [4], Deep Learning [5], and speech spectrogram using Neural Networks [6]. Researchers in [7-14] employ various techniques for emotion classification across various datasets. In [7], a convolutional neural network (CNN) trained on BP4D+ images and Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) speech signals, achieving accuracy of 77.6% and 56.2% on FAU-Aibo and RAVDESS database, respectively. Addressing challenges in speech emotion recognition (SER), researcher proposed temporal modeling framework combines bidirectional long short-term memory networks, CNN, and Capsule networks, outperforming existing systems [8]. Utilizing rational dilation wavelet transform decomposition, [9] attains a maximum accuracy of 83.3% in the categorization of two classes. Use distinct datasets for four-class emotion classification [11–13]. In [10] SER system utilizing K Nearest Neighbor (KNN) achieves an impressive 86.02% classification accuracy based on Berlin emotional database features. Using six binary classifiers, [11] suggests a multi-task learning method that yields an accuracy of 54.76%. A significant SER accuracy of 77.5% is achieved in [12] using a radial basis kernel function. Traditional approaches are also employed in [13] and [14] for developing SER models for robotic and multi-language applications, though both models exhibit unsatisfactory recognition rates.

Nevertheless, the natural non-stationarity of speech signals makes it more practical to employ nonlinear signal processing methods. Recognizing speech-based emotions accurately is still difficult due to the variability, complexity, and subtle changes in the nonlinear features of speech signal. In order to address this difficulty, the Empirical Mode Decomposition (EMD) technique was presented by Huang et al. [15], to analyze and process the input speech signal, which has the potential to display nonlinearity and non-stationarity while maintaining its inherent characteristics. SER has demonstrated success with EMD approach [16], [17],[18]. Through the use of numerous components that each capture the maximum local frequency, the EMD approach provides intrinsic mode functions (IMFs) from speech signal or time series data. These IMFs can be efficiently utilized for further processing of speech signal to identify the emotion.

In the proposed framework, the global feature extraction approach is followed. The input speech signal is decomposed into IMFs and features are computed from each IMF. In this approach two types of features are extracted, one is energy-based ratio feature and another is based on Mel frequency cepstral coefficient (MFCC). Features are selected using the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:59:14 UTC from IEEE Xplore. Restrictions apply.

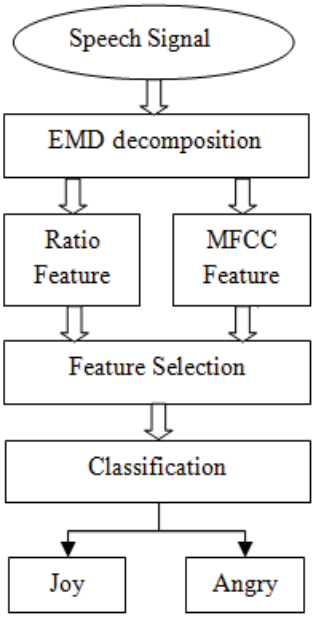


Fig. 1: Process flow of proposed SER framework

ReliefF algorithm and finally, the selected features are tested on the optimizable ensemble (OE) classifier. The obtained result is compared with the existing SER architectures in result and discussion section..

II. METHODOLOGY

The process flow of proposed SER framework is presented in Fig. 1. The detail description of each stage of Fig. 1 is discussed in subsequent sections..

A. Dataset

In the proposed framework publically available EMOVO dataset is used. This dataset, captured by three male and three female orators [19], comprises 14 sentences expressing seven distinct emotions. The dataset encompasses a total of 588 utterances. The recordings maintain a sampling frequency of 48KHz. In this experiment two emotions joy and sad are used for testing the proposed method.

B. EMD Decomposition

Before The input speech signal initially decomposed using EMD, which dissects the nonstationary signal into a amplitude and frequency modulated (AM-FM) oscillating components named as IMFs [15]. An IMF is a function that adheres to the following criteria:

1. Either the number of extrema is the same or it changes by a maximum of one.
2. Based on local extrema, the mean value of envelopes is zero.

The EMD procedure for extracting IMFs from a time series signal can be delineated through the underneath steps:

1. Determine each signal's local maximum and local minimum.
2. To obtain the envelopes $h_{max}(t)$ and $h_{min}(t)$, connect all the maxima and minima individually.
3. Compute the average value of the envelopes using (1) [15]:

$$m = \frac{h_{max}(t) + h_{min}(t)}{2} \quad (1)$$

4. Deduct m(t) from the as follows [15]:

$$q_1(t) = x(t) - m(t) \quad (2)$$

5. Verify whether $q_1(t)$ complies with the specified Intrinsic Mode Function (IMF) conditions mentioned earlier.
6. Continue from steps 2 through 5 until an IMF is obtained.

Once the initial Intrinsic Mode Function (IMF) is acquired, define $I_1(t)=q_1(t)$, representing the smallest temporal scale in $x(t)$. The subsequent IMF can be obtained by generating a residue $r_1(t) = x(t) - I_1(t)$, which is then used as the new signal in the aforementioned algorithm. The obtained residue is iterated until it becomes a monotonic function, signifying that no further IMFs can be produced. A set of narrow-band symmetric waveforms is formed by the resulting IMFs. The Fig. 2 shows the speech signal and their IMFs after decomposition. Following the decomposition, the signal $x(t)$ can be defined as [15]:

$$x(t) = \sum_{k=1}^{K} I_k(t) + r_K(t) \quad (3)$$

Where $I_k(t)$ is k_{th} IMF, K represents total number of IMF and $r_K(t)$ is the final residue.

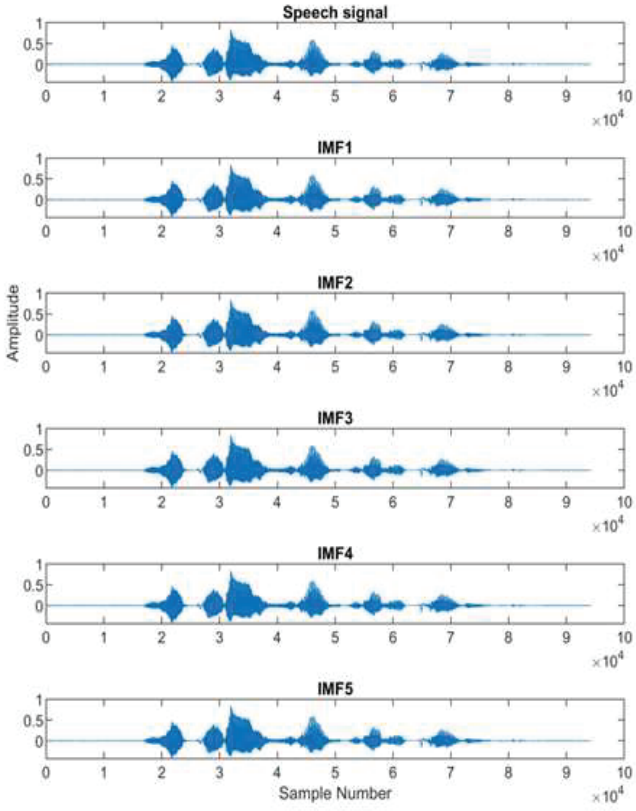


Fig. 2: Speech signal and 5 IMFs after EMD decomposition.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:59:14 UTC from IEEE Xplore. Restrictions apply.

C. Feature Extraction

In the proposed framework ratio feature based on energy and statistical measures are calculated from MFCC coefficients. The detail descriptions are available in following sections.

a) Ratio Feature: After decomposition the energy of each IMF is calculated using (4) [21].

$$E_k = \sum_{n=1}^{N} |I_k(t)|^2 \quad (4)$$

where N is preprocessed signal length.

To obtain the ratio feature each IMF, feature energy is divided by the energy of remaining IMFs. The ratio features from IMF is calculated using (5)

$$f = {}^{E_k}/_{E_l} \qquad \text{l} = 1,2,3 \ldots \ldots \text{except k} \quad (5)$$

b) MFCC Feature: The MFCC serves as a prominent cepstral feature in the development of SER systems. Its primary purpose is to accurately depict the short-time power spectrum of a speech signal. Each preprocessed signal undergoes MFCC extraction, as outlined in Fig. 3.

D. Feature Selection and Classification

The feature selection method employs the ReliefF algorithm, a rank-based approach, to choose relevant features [22]. This algorithm randomly selects instances from both the same class and different classes. When R data instances are randomly chosen from the total n instances, the ReliefF algorithm computes the feature score for each original feature as discussed in [22].

In proposed framework optimizable ensemble classifier is used for the emotion identification. It combines multiple individual models, such as decision trees or neural networks, to enhance predictive performance by leveraging diverse perspectives [23]. Through optimization techniques like boosting or bagging, it strategically weighs and combines the strengths of each constituent model, effectively mitigating individual weaknesses. This approach promotes robustness, generalization, and improved accuracy in complex classification tasks, making it a powerful tool for heightened performance and adaptability across dataset.

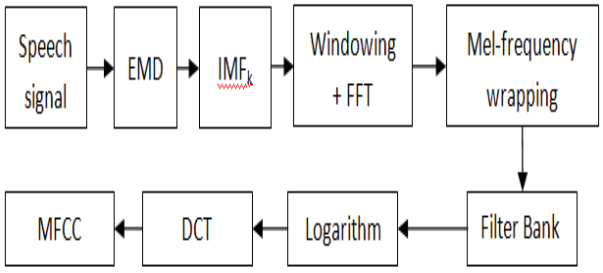


Fig 3: Schema of MFCC feature extraction.

III. RESULTS AND DISCUSSION

In this experiment publically available EMOVO database is used to train the classifier. Initially, the speech signal is decomposed into five IMFs using EMD method. After that energy based ratio feature and MFCC based statistical features are calculated. For the energy based ratio features, the energy of each IMF is calculated and obtained energy of each IMF is divided with remaining IMF energy as explained in section 2. In MFCC based features mean and variance are calculated using the first three coefficients because most of the information is lies in this region. The extracted features are optimized using ReliefF algorithm. Initially a total of 50 features are extracted. After that, top 25 features having high predictor importance weight are considered as significant feature and remaining are rejected. The predictor rank bar is shown in Fig. 4.

These selected features are utilized to train the classifier. The classifier is trained using 10 cross validation in which complete dataset is divided in 10 subsets. In which 9 subsets is utilized for the training and remaining 1 is used for testing. This process is continuing until all subsets have been utilized as the test dataset. The OE classifier is used for emotion classification and achieves an impressive 95.3% recognition accuracy. Additionally, Table 1 presents performance indicators for each emotion, such as recall, precision, and F1 score. Notably, the F1 scores show that the proposed framework exhibits a similar degree of efficacy for both emotions. Further, table 1 show how robust the framework's design is, since it constantly produces reliable outcomes for both emotional states. The precision and recall demonstrate how effectively the framework can recognize and categorize emotions. Finally, the proposed framework's effectiveness is confirmed by the high recognition accuracy and balanced F1 scores across emotions, shows its potential for use in a variety of situations where emotion classification is required. This work advances the field by establishing the framework as a reliable and strong solution for emotion recognition applications. Fig. 5 shows the region of convergence of proposed framework.

The comparison between the suggested framework and the current SER models is presented in Table 2. Earlier research showed that M. A. Jalal et al. [7] obtained 70.14% classification accuracy, R. Jannat et al. [6] obtained 66.41% and ravi et al. [9] achieves 83.3% accuracy. In contrast, the pr

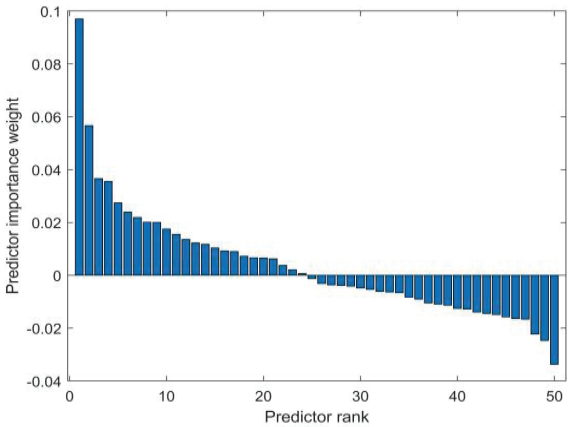


Fig. 4. ReliefF algorithm predictor rank bar for feature selection

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:59:14 UTC from IEEE Xplore. Restrictions apply.

Table:1 Precision, recall and F1 score of proposed framework.

Emotions	Precision %	Recall %	F1 score %
Joy	94.25	96.47	95.34
Sad	96.38	94.11	95.23

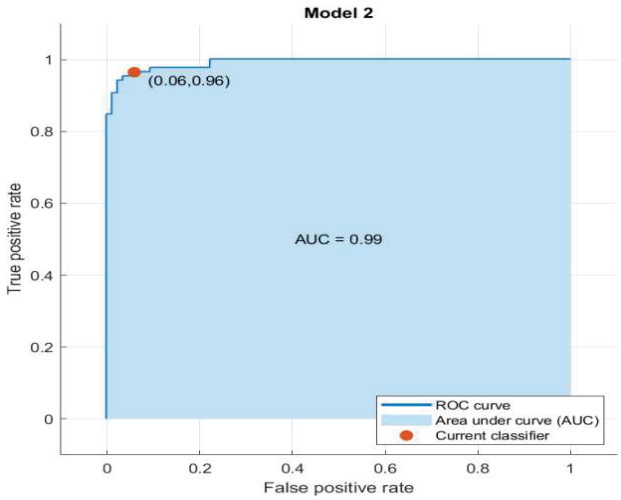


Fig. 5. Region of convergence of proposed framework.

Table: 2 Comparison of proposed framework with existing SER model.

Authors	Classified emotions	Method	Accuracy (%)
R. Jannat et al [6]	Happy, Sad	Ubiquitous computing with CNN	66.41
M. A. Jalal et al [7]	Happy, Sad	Used BLSTM, CNN and Capsule networks	70.14
Ravi et al [9]	Happy, Sad	RDWT with OE classifier	83.3
Proposed Method	Joy, Sad	EMD with OE classifier	95.3

proposed method in this study attains a significantly higher classification accuracy of 95.3%. Therefore, this research introduces a more precise approach for emotion recognition utilizing speech signals, surpassing the accuracies achieved by previous methodologies. The substantial improvement in classification accuracy underscores the effectiveness and advancement offered by the proposed method in accurately discerning emotional states from speech signals. This enhancement in performance positions the proposed method as a noteworthy contribution to the field of emotion recognition, showcasing its potential to outperform existing approaches and enhance the reliability of emotion classification systems.

IV. CONCLUSION

This study introduces features based on the EMD for analyzing speech signals associated with various emotions. Initially, EMD is used to decompose the input speech signals into IMFs. Subsequently, ratio and MFCC based features are computed from each IMF. The statistical significance of these features is evaluated using ReliefF algorithm and high rank features are selected for further processing. The energy based ratio feature shows high potential for the classification. These extracted features are employed for training and testing in OE classifiers. The classifier trained using 10 fold cross validation and achieves 95.3% accuracy for binary class. Finally the accuracy of proposed method is compared with the existing state of art. The proposed framework proved more efficient and robust compare to the other existing methods.

However, there is room for improvement in terms of SER system performance. For example, major research requires in SER system to develop a gender independent and language independent model. This enhancement can be achieved by combining several machine learning and deep learning approaches with an exploration of a novel and robust feature collection.

REFERENCES

[1] Ravi,Taran, Sachin. "A nonlinear feature extraction approach for speech emotion recognition using VMD and TKEO." Applied Acoustics 214 (2023): 109667.

[2] Yang, Ningning, et al. "Recognize basic emotional statesin speech by machine learning techniques using mel-frequency cepstral coefficient features." Journal of Intelligent & Fuzzy Systems 39.2 (2020): 1925-1936.

[3] Koolagudi, Shashidhar G., and K. Sreenivasa Rao. "Emotion recognition from speech: a review." International journal of speech technology 15.2 (2012): 99-117.

[4] Huang, Jian, Bin Liu, and Jianhua Tao. "Learning long-term temporal contexts using skip RNN for continuous emotion recognition." *Virtual Reality & Intelligent Hardware* 3, no. 1 (2021): 55-64.

[5] Z. Han, J. Zhao, H. Leung, K. F. Ma and W. Wang, "A Review of Deep Learning Models for Time Series Prediction," in IEEE Sensors Journal, vol. 21, no. 6, pp. 7833-7848, 15 March15, 2021.

[6] Prasomphan, Sathit. "Improvement of speech emotion recognition with neural network classifier by using speech spectrogram." 2015 International Conference on Systems, Signals and Image Processing (IWSSIP). IEEE, 2015.

[7] R. Jannat, I. Tynes, L. L. Lime, J. Adorno, and S. Canavan, ``Ubiquitous emotion recognition using audio and video data," in Proc. ACM Int. Joint Conf. Int. Symp. Pervas. Ubiquitous Comput. Wearable Comput.,Oct. 2018, pp. 956-959.

[8] M. A. Jalal, E. Loweimi, R. K. Moore, and T. Hain, ``Learning temporal clusters using capsule routing for speech emotion recognition," in Proc. Interspeech, Graz, Austria, Sep. 2019, pp. 17011705.

[9] Ravi and S. Taran, "Emotion Recognition Using Rational Dilation Wavelet Transform For Speech Signal," 2021 7th International Conference on Signal Processing and Communication (ICSC), 2021, pp. 156-160.

[10] A. Bombatkar, G. Bhoyar, K. Morjani, S. Gautam, and V. Gupta, ``Emotion recognition using speech processing using k-nearest neighbor algorithm," Int. J. Eng. Res. Appl., vol. 4, pp. 22489622, Apr. 2014.

[11] B. Zhang, E. M. Provost, and G. Essi, ``Cross-corpus acoustic emotion recognition from singing and speaking:Amulti-task learning approach," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Shanghai, China, Mar. 2016, pp. 5805_5809.

[12] T. Danisman and A. Alpkocak, ``Emotion classication of audio signals using ensemble of support vector machines," in Proc. Int. Tutorial Res. Workshop Perception Interact. Technol. Speech-Based Syst. Berlin, Germany: Springer, 2008, pp. 205216.

[13] Chen, Luefeng, Wanjuan Su, Yu Feng, Min Wu, Jinhua She, and Kaoru Hirota. "Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction." Information Sciences 509 (2020): 150-163.

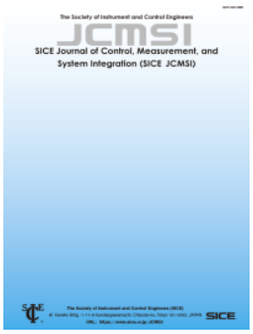
[14] Lalitha, S., Deepa Gupta, Mohammed Zakariah, and Yousef Ajami Alotaibi. "Investigation of multilingual and mixed-lingual emotion recognition using enhanced cues with data augmentation." Applied Acoustics 170 (2020): 107519.

[15] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, H. H. Liu,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:59:14 UTC from IEEE Xplore. Restrictions apply.

The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time
series analysis, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering
Sciences 454 (1998) 903–995.
[16] X. Li, X. Li, X. Zheng, D. Zhang, Emd-teo based speech emotion recognition, Life System Modeling
and Intelligent Computing (2010) 180–189.
[17] C. Shahnaz, S. Sultana, S. A. Fattah, R. M. Rafi, I. Ahmmed, W.-P. Zhu, M. O. Ahmad, Emotion
recognition based on emd-wavelet analysis of speech signals, in: Digital Signal Processing (DSP), 2015
IEEE International Conference on, IEEE, 2015, pp. 307–310.
[18] Tuncer, Turker & Dogan, Sengul & Acharya, U Rajendra. (2020). "Automated accurate speech emotion recognition system using twine shuffle pattern and iterative neighborhood component analysis techniques." Knowledge-Based Systems. 211. 10.1016/j.knosys.2020.106547.
[19] G. Costantini, I. Iaderola, A. Paoloni, and M. Todisco, ''EMOVO corpus: An Italian emotional speech database,'' in Proc. LREC, Reykjavik, Iceland, 2014, pp. 3501–3504.
[20] M. S. Likitha, S. R. R. Gupta, K. Hasitha and A. U. Raju, "Speech based human emotion recognition using MFCC," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 2017, pp. 2257-2260, doi: 10.1109/WiSPNET.2017.8300161.
[21] Li, Dongdong, et al. "Exploiting the potentialities of features for speech emotion recognition." Information Sciences 548 (2021): 328-343.
[22] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM computing surveys (CSUR)*, *50*(6), 1-45.
[23] Pant, Himanshu, Hitesh Kumar Dhanda, and Sachin Taran. "Sleep apnea detection using electrocardiogram signal input to FAWT and optimize ensemble classifier." *Measurement* 189 (2022): 110485.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:59:14 UTC from IEEE Xplore. Restrictions apply.



SICE Journal of Control, Measurement, and System Integration

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tmsi20

End-to-end handwritten Ge'ez multiple numerals recognition using deep learning

Ruchika Malhotra & Maru Tesfaye Addis

To cite this article: Ruchika Malhotra & Maru Tesfaye Addis (2024) End-to-end handwritten Ge'ez multiple numerals recognition using deep learning, SICE Journal of Control, Measurement, and System Integration, 17:1, 122-134, DOI: 10.1080/18824889.2024.2336682

To link to this article: https://doi.org/10.1080/18824889.2024.2336682

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 23 Apr 2024.

Submit your article to this journal

Article views: 60

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tmsi20

SICE JOURNAL OF CONTROL, MEASUREMENT, AND SYSTEM INTEGRATION
2024, VOL. 17, NO. 1, 122–134
https://doi.org/10.1080/18824889.2024.2336682

OPEN ACCESS Check for updates

End-to-end handwritten Ge'ez multiple numerals recognition using deep learning

Ruchika Malhotra [a] and Maru Tesfaye Addis [b,c]

[a]Department of Software Engineering, Delhi Technological University, New Delhi, India; [b]Department of Computer Science, Debre Tabor University, Debre Tabor, Ethiopia; [c]Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India

ABSTRACT

Ge'ez has been used in Ethiopian churches for centuries to read and interpret the Bible. As a part of the liturgy and religious ceremonies, it is also utilized in prayer and chanting. Since the language is ancient, plenty of handwritten physical documents are generated. Recognizing handwritten Ge'ez numerals poses significant challenges due to variations in handwriting styles, incomplete or overlapping strokes, noise, and distortion. This study introduces an end-to-end approach for handwritten Ge'ez multiple numeral recognition employing deep neural networks. The proposed method streamlines recognition by eliminating manual feature extraction stages. To enable end-to-end training without explicit alignment, the model uses attention mechanisms and a connectionist temporal classification-based loss function. The proposed model is evaluated on 120,000 handwritten Ge'ez multiple numeral images, which is a syntactically generated dataset. The developed recognition model achieved a character error rate (CER) of 2.81% and a word error rate (WER) of 18.13%. These results highlight the effectiveness of the approach in accurately and reliably recognizing handwritten Ge'ez multiple numerals.

ARTICLE HISTORY
Received 13 July 2023
Revised 29 February 2024
Accepted 26 March 2024

KEYWORDS
Deep learning; digit recognition; end-to-end learning; handwritten Ge'ez numerals; pattern recognition

Introduction

The Ge'ez language has a significant role in Ethiopian culture, functioning as an ancient Semitic language primarily used within Ethiopian churches [1]. It holds a sacred status and has been historically employed for writing Ethiopian historical accounts and religious texts, resulting in a substantial collection of Ge'ez documents found in Ethiopia. For centuries, Ge'ez has been used in Ethiopian churches for reading and interpreting the Bible. It is an integral part of liturgical practices, religious ceremonies, prayers, and chants. Due to its ancient nature, a large number of handwritten physical documents are available in the Ge'ez language. Ge'ez utilizes the Ethiopic script, which is its native alphabet [2]. The script follows a left-to-right reading and writing system similar to the Latin script.

Derived from the ancient South Arabian alphabet [1], the Ethiopic script has been used to write Ge'ez for over two thousand years and remains one of the few writing systems still in use today. Notably, Ge'ez incorporates a unique numbering system known as the Ge'ez numbering system, specifically designed for representing numerical data. Comprising a total of 20 distinct symbols as shown in Table 1, the Ge'ez numbering system employs complex shapes for each symbol, accompanied by two horizontal lines positioned above and below it. It is also important to note that the Ge'ez numeral system does not include a symbol that represents zero.

Table 1. Ge'ez numbering and English equivalent numeric symbols.

Digits*	English symbol	Numbers*	English symbol
፩	1	፲	10
፪	2	፳	20
፫	3	፴	30
፬	4	፵	40
፭	5	፶	50
፮	6	፷	60
፯	7	፸	70
፰	8	፹	80
፱	9	፺	90
		፻	100
		፼	10, 000

The Ge'ez numbering system consists of two main categories: digits and numbers. The digit's category includes symbols representing the numbers from one to nine: "፩," "፪," "፫," "፬," "፭," "፮," "፯," "፰," and "፱," where each symbol corresponds to its respective numerical value. The numbers category encompasses symbols representing multiples of ten and larger numbers. These symbols include ፲ (representing 10), ፳ (20), ፴ (30), ፵ (40), ፶ (50), ፷ (60), ፸ (70), ፹ (80), ፺ (90), ፻ (100), and ፼ (10,000). In this category, each symbol represents a specific value, and the English equivalent number appears inside the brackets. The Ge'ez numbering system can represent a wide range of numerical values by combining the 20 primary symbols. For example, the

CONTACT Maru Tesfaye Addis marutesfayeaddis_2k18phdco25@dtu.ac.in

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

symbol ፩ combined with other 9-digit symbols creates numbers such as ፳፩ (representing 21), ፳፪ (22), ፳፫ (23), and so on, up to ፳፱ (29).

Handwritten numeral or text recognition is a complex process that involves automatically identifying various handwritten numerals or text information [3]. While computer algorithms are employed to analyze and interpret numeral features, recognizing handwritten digits presents unique challenges in optical character recognition (OCR) and pattern recognition.

One of the main difficulties lies in the significant variations in handwritten digits exhibited in shape, size, and style [4]. These variations make it particularly challenging to develop an algorithm that can accurately recognize all possible variations. Furthermore, distinguishing between different digits can be problematic due to overlapping or incomplete strokes.

Handwritten digits are also susceptible to distortions and noise caused by factors like poor handwriting, smudges, or low-quality image documents [5–7]. So, the ability to generalize to new data makes handwritten numeral recognition more demanding than recognizing printed numerals. A variety of techniques have been explored to overcome these challenges, including pre-processing, feature extraction, and deep-learning approaches [8]. While considerable research has been conducted on pattern recognition for widely used scripts like Latin, Arabic, Urdu, Persian, and Mandarin, Ge'ez numeral recognition, which pertains to an ancient script, has not received the same level of attention in OCR research [9]. Previous works on Ge'ez script recognition have primarily been based on printed texts [10–13], and the recognition of handwritten Ge'ez scripts has remained largely unexplored due to the scarcity of datasets [3]. Although deep-learning algorithms typically require a large amount of labelled data for training [6], publicly available Ge'ez digit datasets are scarce, with the dataset prepared by Nur et al. being an exception [3]. Consequently, recognizing Ge'ez numerals poses a unique challenge in OCR.

Deep learning (DL) approaches have revolutionized pattern recognition, surpassing traditional machine-learning methods in terms of effectiveness [14]. In particular, convolutional neural networks (CNNs) have played a crucial role in image processing and pattern recognition. Building upon previous studies on Ge'ez digit recognition and handwritten text recognition [3,15], this study focused on the underexplored area of handwritten Ge'ez numeral recognition. To the best of our knowledge, this study will be the first of its kind in handwritten Ge'ez multiple-digit recognition. Our study makes the following significant contributions:

- Preparation of the dataset: To enhance the training process, we prepare synthetically generated image datasets, augmenting the available data to facilitate robust learning and recognition.
- Development of comprehensive end-to-end (E2E) learning model: The proposed model integrates automatic feature extraction, attention mechanism, bidirectional LSTM (BLSTM) based sequencing, and CTC loss functions. This comprehensive framework enables the model to automatically extract relevant features from Ge'ez numeral images, selectively focus on important regions of the input sequence, and gain a better understanding of temporal information. Additionally, it enables E2E training without the need for explicit alignment between images and labels.
- Experimental Evaluation: Leveraging the prepared dataset, we conduct different experiments, starting with hyperparameter selection, to evaluate our model's performance. The preliminary results demonstrate promising outcomes in Ge'ez handwritten numeral recognition.

Through these contributions, our study aims to advance Ge'ez handwritten numeral recognition by leveraging a comprehensive DL framework that combines CNNs, BLSTM, attention mechanisms, and CTC. This approach facilitates effective training and recognition, paving the way for improved performance in Ge'ez numeral recognition tasks.

Literature reviews

Numeral recognition has been extensively studied since the 1980s, employing various approaches ranging from traditional machine-learning techniques to deep-learning approaches [16]. Numerous published studies have explored different strategies for digit recognition, focusing on pattern recognition and feature extraction techniques. In this section, we provide an overview of the research conducted in the field of digit recognition. Some studies use traditional machine-learning approaches, while recent studies apply deep-learning and ensemble approaches.

Abdulrazzaq and Saeed compared three machine-learning recognition algorithms: Naive Bayes (NB), multilayer perceptron (MLP), and the K* algorithm [17]. They used a handwritten dataset from NIST, consisting of 48,060 samples with 128 × 128-pixel images. Pre-processing was performed, resulting in 16 × 16 feature images. The authors employed the "correlated feature selection" method and obtained 37 features for training. Their main objective was to identify the best classifier among the three algorithms, achieving high accuracy with minimal features. The K* algorithm performed better than NB and MLP, with a recognition rate of 82.36%. However, image resizing and random sample selection negatively affected recognition accuracy, leading to more misrecognized digits. The NIST

dataset presented challenges due to numerous confusing images, making it difficult to find the optimal classifier through random sampling. In a study by Boukharouba and Bennia [18] a novel feature extraction technique was proposed for recognizing Persian handwritten digits. They utilized a Freeman chain code that incorporated vertical and horizontal information. The advantage of their feature extraction approach is that it eliminates the need for normalization. They employed a support vector machine (SVM) as the classification algorithm and evaluated their model using 80,000 Persian numerals, achieving an accuracy of 98.55%.

Liu et al. [19] compared the recognition time and rate of neural networks with backpropagation (BP), support vector machines (SVM), k-nearest neighbour (KNN), and CNN. They used the MNIST handwritten dataset, which consists of 70,000 28 × 28 images for training and testing. Randomly selecting 5000 samples for training and 1000 samples for testing, the authors followed specific procedures for training the classifier, including input, pre-processing, feature selection, classifier design, and classification decision steps. For the testing phase, they repeated the input, preprocessing, feature selection, and classification decision steps to obtain recognition results. The advantages and disadvantages of the four algorithms were analyzed and reported. The simulation test results revealed that the CNN algorithm achieved the highest recognition accuracy. Abu Ghosh and Maghari [20] compared three popular neural network approaches: CNN, DBN, and DNN. Their system consisted of six steps: preprocessing, segmentation, feature extraction, classification, training, and recognition. They used the MNIST handwritten dataset and a self-created random dataset to verify their results. The random dataset contained 85 different handwritten digits from various sources. The authors evaluated the three algorithms based on accuracy and performance, considering execution time as an additional criterion. DNN outperformed the other algorithms, achieving an accuracy of 98.08%.

Ahlawat et al. [21] proposed an improved digit recognition system using the MNIST dataset and a CNN algorithm. They developed two CNN architectures: CNN_3L with three layers and CNN_4L with four layers. Their main objective was to explore various parameters of the CNN architecture using the MNIST handwritten dataset. They employed optimization algorithms, including stochastic gradient descent with momentum, Adam, Adagrad, and Adadelta, to enhance the performance and speed of the neural networks. Among these optimization parameters, the Adam optimizer achieved a detection rate of 99.89% for the CNN model with three layers, surpassing previously reported results in this field.

The authors [22] proposed recognizing Arabic handwritten characters using ensemble convolutional neural networks. They utilized adaptive gradient descent in the initial epoch for convergence and regular stochastic descent in the final epoch. The study employed K-fold and Monte Carlo cross-validation methods. With the MADbase dataset, they achieved a test accuracy of 99.47% and a classification accuracy of 99.74%. To enhance Arabic handwritten numeral recognition, Ahamed et al. [8] proposed a convolutional neural network architecture. Their CNN architecture comprised two convolutional networks with 32 filters and 5 × 5 kernels, two max-pooling networks with a 2 × 2 kernel, two dropout layers, and two fully connected layers. They conducted two activities in their study: improving a previous research model by Ashiquzzaman et al. [23] and introducing a newly proposed method. Using augmentation techniques, they expanded the initial dataset from 3000 to 72,000 images. The improved model achieved 98.95% accuracy, while the newly proposed model achieved 99.96% accuracy.

Ahlawat and Choudhary [24] proposed an automatic recognition model for handwritten digits utilizing CNN for feature extraction and SVM for output prediction. By combining CNN and SVM, they achieved a classification accuracy of 99.28% on the MNIST dataset of handwritten digits. Ali et al. [25] developed an improved method for recognizing handwritten digits using a CNN with a DeepLearning4J framework. The CNN model, consisting of three convolutional layers, one max-pooling layer, and two fully connected layers, is trained and tested on the MNIST handwritten dataset. The model achieved an overall accuracy of 99.21% when tested on a dataset of 5,130 images, with only 42 misidentifications. The computation time for training and testing was also improved. In their work [26], Yu et al. presented an improved LeNet5 digit recognition algorithm. They replaced the last two layers of LeNet5 with an SVM, utilizing SVM as the classifier and LeNet5 as the feature extractor. Yoa et al. [27] proposed a CapsNet-based recognition and separation method for fully overlapping digits. They introduced three main modifications to the original capsule network [28]: the use of a small convolution kernel, the expansion of capsule dimensions to express extracted features, and the application of dual dynamic routing. Their proposed model, called "FOD_DCNet," achieved an accuracy of 93.53%, surpassing the performance of CapsNet by 5.43%.

End-to-end learning

E2E learning is a DL or machine-learning approach that aims to simplify the process of solving a task by combining all the stages into a single model [29]. Traditionally, machine-learning pipelines involve multiple steps such as pre-processing, feature extraction, and classification. E2E learning eliminates the need for handcrafted features or intermediate representations and allows the

model to learn directly from raw input data to produce the desired output.

This approach has gained popularity with the advancement of deep-learning and neural networks [30]. DL models, such as CNN and recurrent neural networks (RNN), have demonstrated exceptional performance in various domains like computer vision, natural language processing, speech recognition, and pattern recognition.

The main advantage of E2E learning is its simplicity. By removing the need for complex feature engineering, the overall system design becomes more straightforward. Additionally, end-to-end learning models are flexible and can adapt to different domains and tasks. They automatically learn relevant features and representations from the raw data, capturing intricate patterns and dependencies [4].

However, there are some limitations to consider. E2E learning models typically require large amounts of labelled training data to generalize well, which can be time-consuming and costly to acquire. The complex representations learned by these models may also lack interpretability, making it challenging to understand the reasons behind their decisions. Furthermore, the integrated nature of E2E models can make it difficult to reuse or modify specific parts for different tasks [31].

So, we can generalize E2E learning as a powerful approach that allows a model to learn directly from raw input data to produce the desired output. It simplifies the machine-learning process, and achieves state-of-the-art results in various domains, but also requires substantial amounts of labelled data and may lack interpretability and modularity.

Studies done on handwritten Ge'ez numerals

In the realm of Ge'ez numeral recognition, limited prior studies have been conducted. Agegnehu et al. [32] applied deep learning to recognize Amharic punctuation marks and offline handwritten digits using a CNN architecture. Their dataset consisted of 5800 images collected from 100 different handwriting samples, achieving a testing accuracy of 70.04%. Furthermore, studies have also explored Ge'ez digits in contexts beyond printed and handwritten forms. Tamiru et al. [33] proposed a study on recognizing Amharic alphabetic sign language, benefiting individuals with hearing loss. Their model consisted of pre-processing, regimentation, feature extraction, and recognition stages, utilizing digital image processing and machine-learning algorithms such as SVM (98.06%) and ANN (80.86%) for classification. Abeje et al. [34] focused on Ethiopian sign language recognition using a deep convolutional neural network. Their system encompassed preprocessing, feature extraction, and classification stages, employing digital image processing techniques. The dataset used in their study was collected from hearing-impaired students.

While the majority of researchers have predominantly focused on the recognition of English numerals, the absence of publicly available Ge'ez digit datasets has deterred extensive research in this area. However, Nur et al. [3] advanced the field of Ge'ez digit recognition by developing a recognition model with improved accuracy and an experimental dataset comprising 51,952-digit images written by 524 individuals. Their proposed CNN architecture achieved a recognition accuracy of 96.21%. The authors recommended exploring multi-digit recognition using diverse deep-learning approaches for future research. Therefore, our motivation started from their recommendations.

Methodology

In this section, we present the details of our E2E handwritten Ge'ez multiple-number recognition system. The purpose of this study is to develop a recognition model that can accurately recognize handwritten multiple digits in Ge'ez numerals. The architecture of the model and the dataset used to train and evaluate the model are explained.

The dataset preparation

To address the issue of limited datasets in deep-learning studies, we recognize the significance of having a sufficiently large and diverse dataset for model training and testing. In this research, we made concerted efforts to overcome this challenge by developing an extensive dataset specifically tailored to our DL problem.

The dataset used in this study is derived from the initial handwritten dataset introduced by Nur et al. [3]. The original dataset comprises 51,952 sample preprocessed greyscale images of single digits with a dimension of 32 by 32 pixels (see Figure 1). To extend the scope of the dataset and sequentially accommodate multiple digits, we synthetically generated a new dataset having a multi-digit image from the original dataset.

Asthe dataset is synthetic, we utilize an automated multi-digit image creator algorithm to create sequential data. In achieving this, we have developed a program employing the OpenCV image creator tool, specifically using the concatenation method. The selection of digits is performed randomly. The algorithm is depicted in Table 2.

The newly generated dataset consists of 120,000 image samples with a dimension of 32 by 128 pixels (see Figure 2). Each image in this dataset contains a variable number of sequential digits. This expansion allows us to explore and evaluate the performance of the OCR model in scenarios involving multiple numerals within a single image.

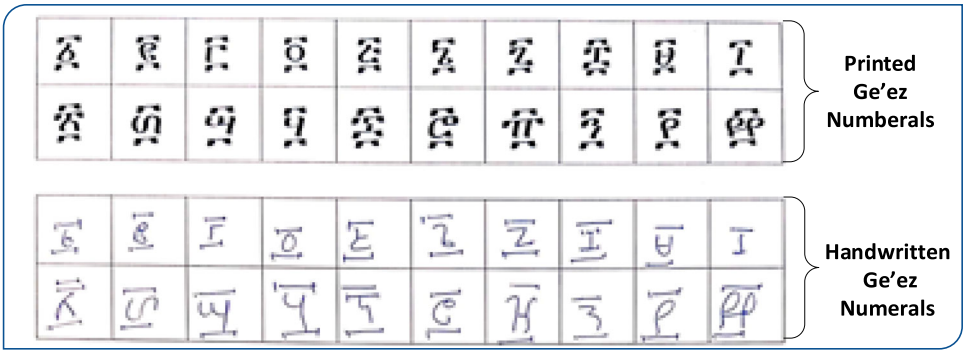


Figure 1. Sample data from the initial handwritten Ge'ez digit dataset.

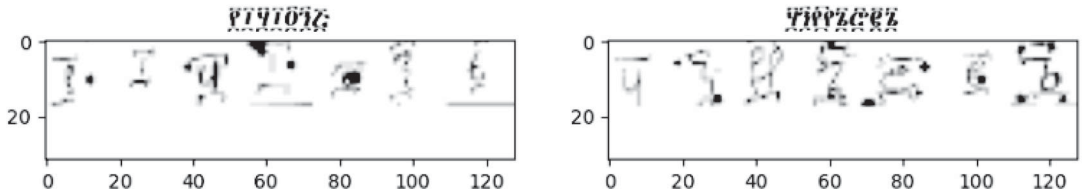


Figure 2. Sample images from synthetically generated 32 by 128 multiple numerals dataset.

Table 2. Synthetic sequential multi-digit creator algorithm.

Parameters:
n_items: Number of items per sequence
n_seqs: Total number of sequences to generate
Data:
items: Collection of available items
labels: Labels corresponding to the items
Result:
seq_labels: Array of labels for generated sequences
seq_data: Array of generated sequences
Algorithm:
1. Initialize seq_data ***as*** an ***empty list***
2. Initialize seq_labels ***as*** an ***empty list***
3. ***For*** i ***from*** 1 ***to*** n_seqs:
   a. Generate random indices ***using*** np.random.randint
   b. Retrieve random_items ***from*** items based ***on*** indices
   c. Transform ***each*** item ***in*** random_items ***as*** needed
   d. ***Convert*** transformed random_items ***to*** a numpy ***array and*** assign ***to*** seq
   e. Retrieve random_labels ***from*** labels based ***on*** random indices
   f. Append seq ***to*** seq_data
   g. Append random_labels ***to*** seq_labels
4. ***Convert*** seq_labels ***and*** seq_data ***to*** numpy arrays
5. Reshape seq_data ***to*** the desired shape
6. ***Return*** seq_labels and seq_data

By leveraging the initial dataset provided by Nur et al. [3], we were able to build upon their work and extend it to address the specific requirements of our study. The synthetic generation of the dataset provides a more diverse and challenging set of images, enabling a comprehensive evaluation of the OCR model's capabilities in handling sequential multiple digits.

Generating the sequential synthetic data follows the following algorithmic steps and the flowcharts of the steps are depicted in Figure 3.

(1) Original datasets from [3]: Random selection of images from the original dataset which has a single digit in each image.
(2) Data augmentation: Add noise, erosion, and dilatation.
(3) Pre-processing: Resize, transpose, and normalize the image
(4) Finally, generated 32 by 128-pixel image which consists of multiple handwritten Ge'ez numerals.

The proposed network architecture

The proposed recognition model consists of two main steps: feature extraction and the recognition process. These steps work together to achieve improved recognition accuracy.

Feature extraction process

In the feature extraction part, we employ seven CNN layer blocks, which automatically extract relevant features from the input image, and a SoftMax-based attention mechanism. Each CNN layer block consists of convolutional layers, max-pooling layers, and batch normalization. These components play specific roles in the feature extraction process.

Convolutional layers perform convolution operations on the input image. The convolution operation between an input image (I) and a filter/kernel (K) can be computed in Equation (1).

$$(I,K)[i,j] = \sum\sum (I[x,y] * K[i-x, j-y]) \quad (1)$$

In Equation (1), *(I,K)[i,j]* represents the value at position *[i,j]* in the resulting feature map. The double summation ($\sum\sum$) computes the element-wise multiplication and summation over the spatial dimensions. [x,y] and [i−x,−y] represent the corresponding values in the input image and filter/kernel, respectively.

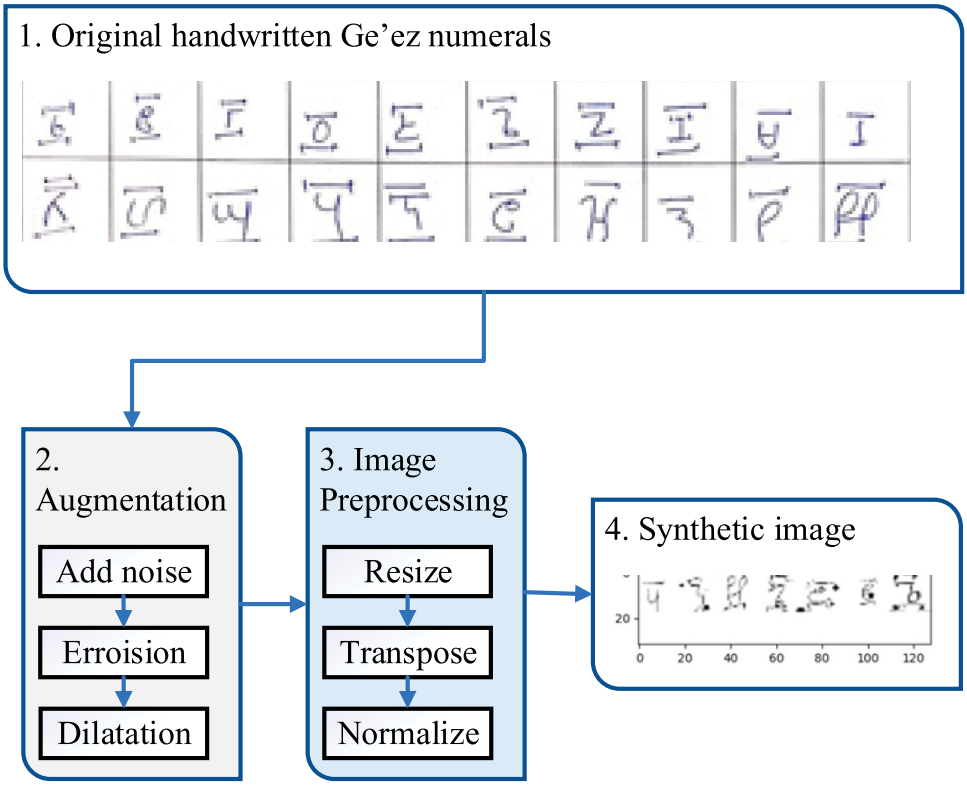


Figure 3. Experimental synthetic dataset preparation steps.

The convolutional layers are configured with parameters such as feature maps, kernel size, stride, and padding, which determine how the convolutions are applied to the input image. By applying these layers, the model can capture important patterns and structures within the image.

The max-pooling operation reduces the dimensionality of the feature maps by selecting the maximum value within each pooling region. Given a feature map (F), the max-pooling operation can be expressed in Equation (2).

$$MaxP(F)[i,j] = \max(F[x,y]) \tag{2}$$

Here, $MaxP(F)[i,j]$ represents the value at the position $[i,j]$ in the down-sampled feature map. The function $\max(F[x,y])$ returns the maximum value within the pooling region defined by coordinates $[x,y]$. This pooling operation helps to retain the most salient features while discarding unnecessary information, leading to more efficient processing and improved generalization.

Batch normalization helps normalize the feature maps by subtracting the mean and dividing by the standard deviation. Given a feature map the batch normalization operation can be represented in Equation (3).

$$BN(F)[i,j] = \frac{F[i,j] - mean(F)}{\sqrt{variance(F) + epsilon}} \tag{3}$$

where, $BN(F)[i,j]$ represents the normalized value at position (i,j) in the feature map. $mean(F)$ represents the mean of the feature map, $variance(F)$ denotes the variance, and is a small constant to avoid division by zero.

Table 3. The proposed model convolutional layer blocks along with the parameter values.

Layer blocks	Layers	Feature maps	Kernel size	Activation	Stride size
Block1	Conv1	32	(3,3)	ReLU	(1,1)
	pooling1	–	(2,2)		(2,2)
Block2	Conv2	64	(3,3)	ReLU	(1,1)
	pooling2	–	(2,2)		(2,2)
Block3	Conv3	128	(3,3)	ReLU	(1,1)
Block4	Conv4	256	(3,3)	ReLU	(1,)
	pooling4	–	(1,2)		–
Block5	Conv5	256	(3,3)	ReLU	–
	Normalization1	–	–		–
Block6	Conv6	512	(3,3)	ReLU	(1,1)
	pooling5	–	(1,2)		(1,1)
Block7	Conv7	512	(2,2)	ReLU	(1,1)

The subtraction and division operations ensure that the mean is zero and the standard deviation is one. This normalization step ensures that the recognition process is not affected by outliers or irregularities in the data. By maintaining consistent scale and distribution of the features, the model becomes more robust and reliable.

Through the combination of these CNN layer blocks, the proposed model can effectively extract meaningful features from the input image, which serves as a crucial foundation for accurate recognition. The feature extraction steps and the attribute value of each are outlined in Table 3. The attention mechanism selectively focuses on relevant parts of the feature maps. Given a set of feature maps (F) and their corresponding attention weights (W), the attention mechanism can be

computed using Equation (4).

$$A(F, W) = \sum (F[t, i] * W[t, i]) \tag{4}$$

where, $A(F, W)$ represents the attended feature vector. The summation ($\sum$) computes the element-wise multiplication and summation over the temporal (t) and spatial (i) dimensions of the feature maps. The attention weights (W) assign importance or relevance to each element in the feature maps, allowing the model to concentrate on the most informative regions. By assigning different weights to different regions, the model can allocate its attention to the most informative areas, improving recognition accuracy. This mechanism helps the model handle variations in scale, orientation, and appearance within the input image.

Recognition process

In the recognition process, two main techniques are utilized in addition to the SoftMax dense layer to further enhance the model's performance. These techniques include bidirectional RNNs and CTC.

Bidirectional RNNs are pivotal for capturing dependencies in both the forward and backward directions within a given input sequence X. This bidirectional processing occurs in two passes, and the outputs of the forward and backward RNNs at each time step are concatenated to form the final hidden state representation, as expressed by the mathematical representation in Equation (5):

$$h[t] = Con(h_f[t], h_b[t]) \tag{5}$$

where, $h[t]$ represents the hidden state at the time step t. h_f and h_b represents the outputs of the forward and backward RNNs, respectively.

The Concatenate operation fuses these two representations, enabling the model to capture contextual information from both past and future time steps. This bidirectional approach ensures a comprehensive understanding of the entire input sequence, enhancing the model's ability to discern sequential patterns and structures.

Moreover, recurrent neural networks (RNNs) have become standard in processing sequential data. In our case, RNNs maintain previous inputs within the internal network state, allowing the model to leverage past context for improved recognition accuracy.

Long Short-Term Memory (LSTM) networks, a specific type of RNN, are employed to address the challenge of learning long-term dependencies. Hochreiter and Schmidhuber [35] proposed the LSTM cell to overcome the vanishing gradient problem associated with traditional RNNs. Comprising three sigmoid activation function gates – the Input Gate, Forget Gate, and Output Gate – LSTM networks facilitate effective learning of intricate temporal dependencies. The Equations (6) to (11) outline the mathematical expressions representing the LSTM gates:

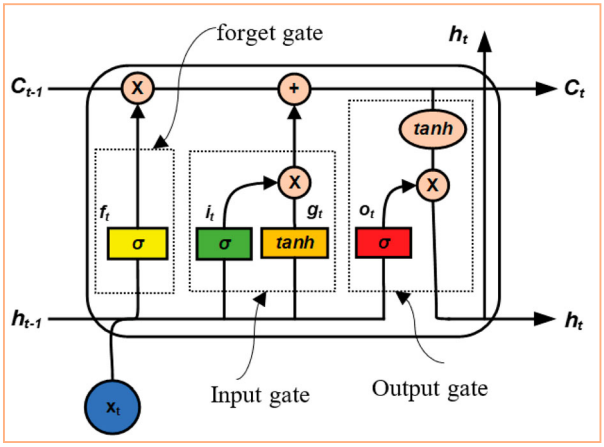


Figure 4. The LSTM gates architectures at timestamp ∼ t [12].

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{6}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{7}$$

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{8}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{9}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{10}$$

$$h_t = o_t \cdot tanh(C_t) \tag{11}$$

Figure 4 providesa visual representation of the LSTM gates' architecture at a specific timestamp, enhancing the conceptual understanding of the LSTM's inner workings.

CTC loss is utilized as a loss function during training to align the predicted sequence (PS) with the ground truth (GT). Given a predicted sequence (Y_{pred}) and the GT sequence (Y_{true}), the CTC loss can be calculated here in Equation (12). Here, $p(Y_{true}|Y_{pred})$ denotes the probability of obtaining the GT sequence given the PS.

$$CTCLoss(Y_{pred}, Y_{true}) = -\log(p(Y_{true}|Y_{pred})), \tag{12}$$

The CTC loss is typically computed using dynamic programming algorithms, such as the forward–backward algorithm, to sum over all possible alignments. The negative logarithm is applied to penalize the deviation between the predicted and GT sequences. CTC enables the model to handle input sequences of variable lengths and perform E2E training. It automatically aligns the PS with the GT, accounting for possible repetitions, deletions, and insertions. This approach eliminates the need for explicit alignment information and improves the model's ability to recognize sequences accurately.

By incorporating these techniques into the feature extraction process above, the proposed model achieves superior performance in recognizing complex patterns and sequences, leading to enhanced recognition results overall. In light of the aforementioned concepts, we

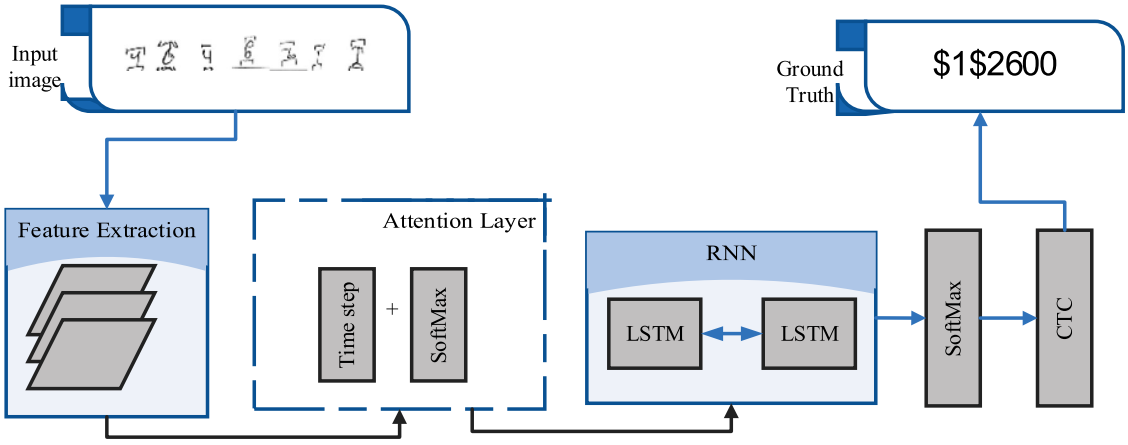


Figure 5. The proposed E2E handwritten Ge'ez multiple numerals recognition network architecture.

Table 4. The proposed model recurrent neural networks along with the parameter values.

Layer Blocks	Layers	Units	Dropouts	Activation
Bidirectional	LSTM1	128	0.25	Tanh
	LSTM2	128	0.25	Tanh
	Fully Connected	21	–	SoftMax

propose an E2E learning recognition model, as illustrated in Figure 5, specifically designed for recognizing handwritten multiple digits in Ge'ez numerals. The model architecture includes convolutional layer blocks with specific parameter values as mentioned in Table 3 to extract features from the input. Additionally, as depicted in Table 4, two layers of bidirectional LSTM cells, each with 128 hidden units, are utilized to capture long-term dependencies in sequential digit datasets. Dropout with a rate of 25% is applied to regularize the model and prevent overfitting.

Therefore, this model directly learns the mapping from input to output by optimizing the parameters of the entire model. During training, the model implicitly learns to extract relevant features or representations from the input, eliminating the need for manual feature engineering or intermediate stages.

Our proposed model's performance is evaluated using two metrics: CER and WER. CER is calculated by Equation (7).

$$CER = \left(\frac{(I + D + S)}{GT}\right) * 100, \tag{7}$$

Here, I, D, and S represent the number of character insertions, D deletions, and substitutions respectively. GT is the total number of characters in the GT text.

Similarly, WER is calculated by considering word-level errors. The formula WER can be expressed in Equation (8).

$$WER = \left(\frac{(I_w + D_w + S_w)}{N_w}\right) * 100, \tag{8}$$

where, I_w, D_w, S_w represents the number of words insertions, deletions, and substitutions respectively, and N_w is the total number of words in the ground truth.

Both CER and WER provide valuable insights into the error rates of the recognition system. By evaluating these metrics, we can measure the effectiveness of our proposed model in accurately transcribing characters and words, and identify areas where improvements may be required.

In addition to our primary evaluation metrics, CER and WER, we employed secondary evaluation metrics to provide a comprehensive analysis of the model's performance. These metrics, namely precision, recall, and F1-score, focus on the classification aspect of the model's predictions, specifically in terms of correctly predicted characters and their positions within the sequence.

To better interpret these metrics, let's define the parameters used in generating the classification report. TP (true positive) signifies the number of positive instances correctly predicted, whereas FP (false positive) denotes the number of positive instances predicted incorrectly. On the other hand, TN (true negative) indicates the number of negative instances correctly predicted, and FN (false negative) represents the number of negative instances predicted incorrectly. Equations (9), (10), and (11) are used to compute the precision (P), recall (R), and F1 score respectively.

$$P = \frac{TP}{(TP + FP)}, \tag{9}$$

$$R = \frac{TP}{(TP + FN)}, \tag{10}$$

$$F - score = \frac{2PR}{R + P}, \tag{11}$$

Precisionis also known as the positive predictive value (PPV) and it reveals the proportion of correctly identified positive instances out of all instances predicted as positive. It provides insights into the accuracy of the positive predictions made by the model. The recall represents the proportion of correctly identified

positive instances out of all actual positive instances. It measures the model's ability to capture positive instances correctly within the dataset. F1-score combines precision and recall into a single metric. It offers a balanced measure of the model's overall performance by considering both precision and recall. The F1 score allows us to assess the model's ability to achieve accurate positive predictions while capturing a sufficient number of positive instances.

By analysing these secondary evaluation metrics, including precision, recall, and F1-score, we gain a deeper understanding of the model's classification performance. These metrics provide valuable insights into the model's ability to accurately predict characters within the sequence and make informed assessments of its performance.

Results and discussion

The proposed model was implemented in Python using the Kera's framework with TensorFlow as the backend. To expedite training, the use of powerful GPUs was made accessible through Google Collaboratory, eliminating the need for expensive hardware. Collaboratory, a cloud-based service, enables remote code execution from anywhere with an internet connection. Its user-friendly interface and support for popular data science libraries make it an excellent choice for deep-learning projects.

A total of 120,000 images were used for the experiment. The dataset is first divided into three categories: training, validation, and testing. The training consists of 96,000 (80%) images, the validation comprises 12,000 (10%) images, and the testing comprises 12,000 (10%) images.

During the experimentation phase, we carefully considered and fine-tuned various network parameters to ensure optimal performance. The results reported in this study were obtained using an Adam optimizer in conjunction with a CNN that progressively increased the feature map size from 32 to 512. Additionally, a BLSTM network with two hidden layers, each consisting of 128 units, was utilized. The initial learning rate chosen for training the model was set to 0.001.

Through systematic parameter selection and experimentation, this configuration yielded favourable outcomes, which are discussed in this study. The combination of the Adam optimizer, CNN with increasing feature maps and BLSTM with two hidden layers proved effective in capturing relevant patterns within the data. Moreover, the learning rate of 0.001 facilitated a suitable balance between convergence speed and optimization accuracy.

Results

Discussions

In this experiment, the evaluation of the proposed recognition model was based on two key metrics (CER and WER) as it was described in section three of this study. The results demonstrate the model's impressive performance, with an average CER of 2.81% and a WER of 18.13% at empirically selected epoch size 30 and batch size 128. Moreover, the model showcased its capability by accurately predicting 97.18% of the digits in their entirety, while achieving coverage of 81.87% for the full-digit sequences. These outcomes are highly promising, especially considering that this study represents the first attempt at multiple Ge'ez digit recognition.

Errorsare a common occurrence in deep-learning models, and our recognition model is no exception. In Figure 6, we showcase some examples of errors in the results. In the first image sequence, the digit "፰" is mistakenly classified as "፱," while in the second sequence, the image of "፰" is incorrectly identified as "፬." The digit enclosed within the red box represents the digit that has been misrecognized. At the label parts blue represents the ground truth or correct

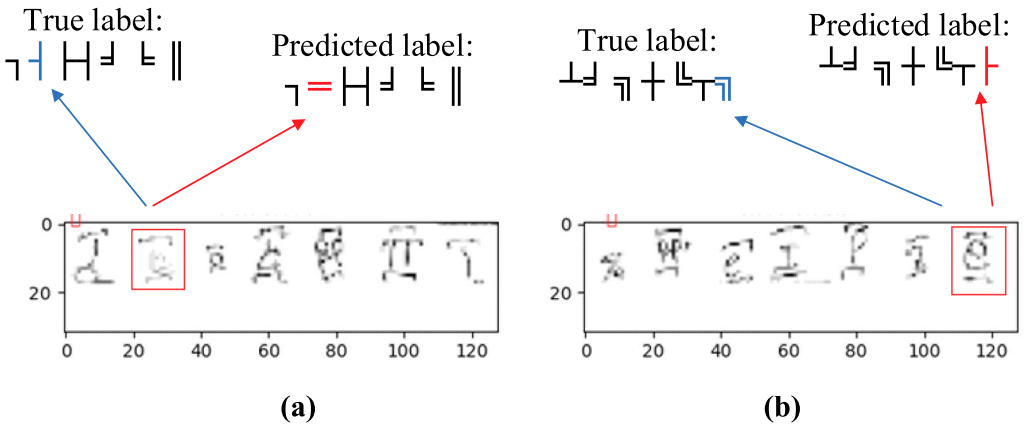


Figure 6. Sample misrecognized Ge'ez digits (a) The error that happened between the beginning and the end of the sequence. (b) The error that happens at the end of the sequence.

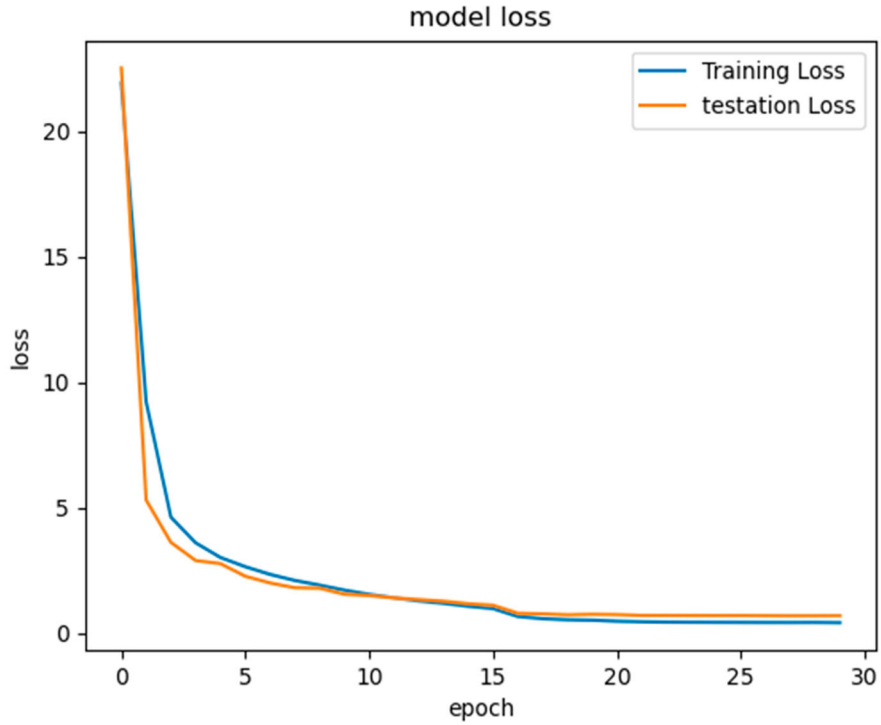


Figure 7. Training and validation loss versus epoch of the proposed multiple handwritten Ge'ez numerals model.

label for the corresponding character image and the red colour highlights the images that have been misrecognized based on the predicted label. These errors can be attributed to the close resemblance between the shapes of these digits, particularly their curved nature. The model may struggle to distinguish between them due to their visual similarities.

The training and validation loss of the proposed multiple handwritten Ge'ez numerals end-to-end learning model is also depicted in Figure 7. From this, the training and the validation graph converge on the proposed model. This convergence suggests that the model's performance on unseen data (validation set) is consistent with its performance on the training data. It is a positive sign that the model is learning effectively and is likely to perform well on new, unseen data.

Table 5. Precision, recall, f1-score, and support results of the proposed recognition model.

Digits & Numbers	precision	recall	f1 - score	support
፩	0.96	0.95	0.96	5249
፪	0.96	0.97	0.96	5273
፫	0.98	0.98	0.98	5286
፬	0.97	0.98	0.97	5075
፭	0.97	0.97	0.97	5239
፮	0.94	0.97	0.95	5291
፯	0.96	0.96	0.96	5287
፰	0.97	0.96	0.97	5267
፱	0.98	0.99	0.98	5151
፲	0.98	0.97	0.97	5139
፳	0.98	0.97	0.98	5347
፴	0.98	0.99	0.99	5390
፵	0.98	0.98	0.98	5271
፶	0.98	0.97	0.97	5273
፷	0.97	0.96	0.97	5248
፸	0.97	0.97	0.97	5213
፹	0.98	0.98	0.98	5213
፺	0.95	0.97	0.96	5186
፻	0.98	0.97	0.97	5303
፼	0.99	0.99	0.99	5299

The performance evaluation of the recognition model in recognizing Ge'ez numerals reveals high accuracy and effectiveness. Precision, recall, and F1 scores were used to assess the model's performance for each Ge'ez alphabet (see Table 5) The precision values ranged from 0.94 to 0.99, indicating the model's ability to accurately classify and identify the Ge'ez characters. Most of the digits achieved precision scores above 0.95, highlighting the model's accuracy in recognizing the Ge'ez numerals. Similarly, the recall values were consistently high, ranging from 0.95 to 0.99. This indicates the model's ability to effectively retrieve a significant portion of the true positive instances for each alphabet, reflecting its proficiency in capturing relevant patterns and features of the Ge'ez characters. The F1 scores, which consider both precision and recall, ranged from 0.95 to 0.99 for the different Ge'ez alphabets. These high F1 scores demonstrate the model's strong capability to achieve a balance between accurate classification and comprehensive retrieval of the Ge'ez numerals. The support column represents the number of instances for each Ge'ez alphabet in the evaluation dataset, which was well-balanced, with a similar number of instances for most characters, ranging from 5075 to 5390. Overall, the results indicate that the recognition model performed exceptionally well in recognizing the Ge'ez numerals, exhibiting high

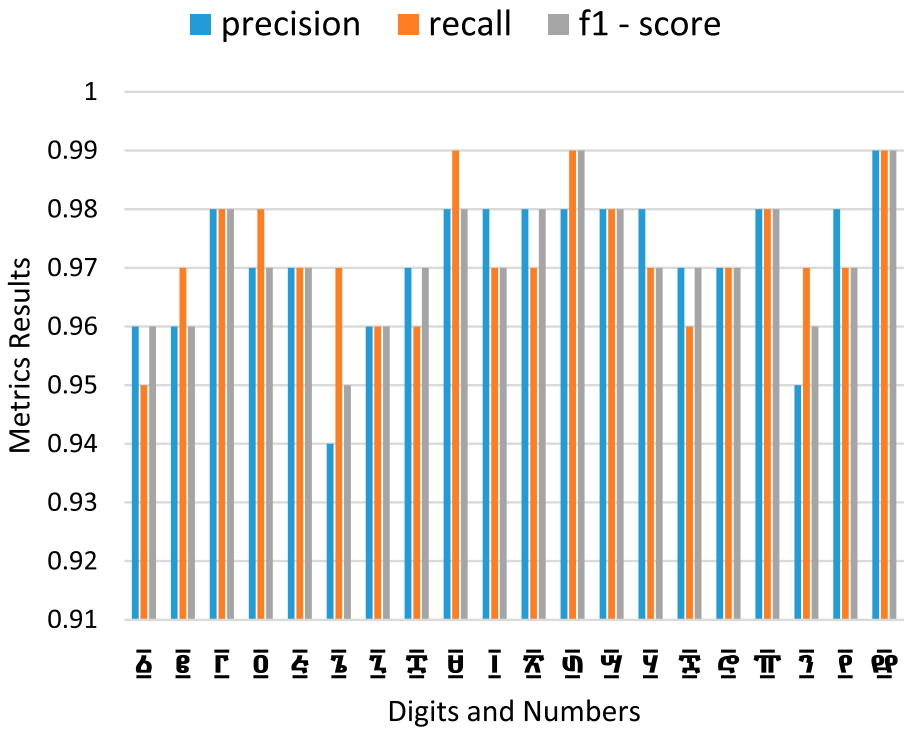


Figure 8. Precision, recall, f1-score, and support results of the proposed recognition model.

precision, recall, and F1 scores. This showcases the effectiveness of the model and its potential for various applications involving Ge'ez numeral recognition.

In terms of individual digit performance, as directed in Figure 8, the Ge'ez digit ፼ achieved the highest precision, recall, and F1-score, demonstrating exceptional accuracy and completeness in recognition. The digits ፱ and ፸ also exhibited strong performance, closely following the top-performing digit with high precision, recall, and F1 scores of 0.98. Overall, the Ge'ez digits ፼, ፱, and ፸ performed exceptionally well, exhibiting the highest precision, recall, and F1 scores. These digits consistently achieved high accuracy and retrieval rates, making them the most reliably recognized digits by the model.

Limitations. During the experiment, the proposed multiple handwritten Ge'ez numeral recognition method is not tested with real-life datasets. It is important to note that the datasets used in this experiment consisted solely of synthetic handwritten samples. As a result, the model's performance may differ when faced with real-life handwritten data, which often exhibits greater variability and complexity. Additionally, the recognition system faced challenges due to the visual similarities of the numerals. These similarities can lead to confusion and recognition errors, especially when the model is trained on a limited number of samples for certain Ge'ez numerals. The scarcity of training data for specific characters can hinder the system's ability to accurately recognize them. To address these limitations and enhance the recognition performance of the system, it is crucial to expand the training dataset by including a more diverse and representative collection of real-life handwritten samples. Increasing the number of training samples for each numeral would enable the model to learn and generalize better, improving its ability to handle the intricacies and variations present in real-life data.

Conclusion

In conclusion, this study introduces end-to-end learning model specifically designed for recognizing handwritten Ge'ez multiple numerals. By leveraging deep-learning techniques, our model automatically extracts relevant features from input images and effectively captures the sequential nature of the numerals. The integration of attention mechanisms and a CTC-based loss function allows the model to focus on crucial regions of the input, enabling end-to-end training without the need for explicit alignment between images and labels. The experimental results demonstrate a CER of 2.81% and a WER of 18.13% for the handwritten Ge'ez numeral recognition task, validating the effectiveness of our proposed approach.

Our model not only contributes to the underexplored area of OCR research for the Ge'ez script but also opens up new possibilities for Ge'ez numeral recognition. Future work can be improved by expanding the dataset to include other characters in the Ge'ez language, exploring alternate deep-learning architectures, and investigating the applicability of the approach to other languages and scripts. Continuous research and development efforts hold the potential to further improve the recognition of handwritten Ge'ez multiple numerals, benefiting various applications in pattern recognition and document processing.

To enhance the performance and reliability of our proposed method in real-life scenarios, future iterations of the recognition system should incorporate a larger and more diverse dataset. This step will address the challenges posed by real-life handwritten data and help mitigate recognition errors, ultimately improving the practical applicability of the recognition system. By continually refining and advancing our approach, we can achieve more accurate and robust recognition of handwritten Ge'ez multiple numerals, paving the way for broader adoption and utilization in various fields.

Acknowledgment

The authors would like to express our sincere gratitude to the faculty members, staff, and students of Delhi Technological University for their invaluable guidance and support throughout the completion of this manuscript. Their continuous assistance and expertise have played a crucial role in the success of this research project. The authors are also grateful for their unwavering commitment to fostering academic excellence and their contributions to our growth as researchers.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Ruchika Malhotra received a master's and Ph.D. insoftware engineering from the University School of Information Technology, Guru Gobind Singh Indraprastha University, Delhi, India. She is a Professor at the Department of Software Engineering, Delhi Technological University, Delhi. She has published more than 200 research papers in international journals and conferences. Her research interests include software testing, improving software quality, statistical and adaptive prediction models, software metrics, neural net modeling, and the definition and validation of software metrics.

Maru Tesfaye Addis received a bachelor's degree in computer science and information technology from Arba Minch University, Ethiopia, in 2010, and a master's degree in computer science from Bahir Dar University, Ethiopia, in 2017. He is currently pursuing a Ph.D. degree in computer science and engineering at Delhi Technological University, India. He is also a Lecturer with the Department of Computer Science, at Debre Tabor University. His research interests include pattern recognition, deep learning, artificial intelligence, and image processing. He is also dedicated to advancing knowledge and contributing to the development of innovative solutions in these areas.

ORCID

Ruchika Malhotra http://orcid.org/0000-0003-3872-6213
Maru Tesfaye Addis http://orcid.org/0000-0002-9416-9063

References

[1] Meyer R. The ethiopic script: linguistic features and socio-cultural connotations. Multiling Ethiopia: Ling Challen Capacity Build Efforts. 2016;8(1):137–172.
[2] Tadesse DA, Liu CM, Ta VD. Unconstrained bilingual scene text reading using octave as a feature extractor. Appl Sci. 2020;10(4474):1–14.
[3] Nur MA, Abebe M, Rajendran RS. Handwritten Geez digit recognition using deep learning [open access]. Appl Comput Intell Soft Comput. 2022;2022:1–12.
[4] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proc IEE: IEEE. 1998;86:2278–2324. doi:10.1109/5.726791
[5] Fateh A, Fateh M, Abolghasemi V. Multilingual handwritten numeral recognition using a robust deep network joint with transfer learning. Inform Sci. 2021;581:479–494. doi:10.1016/j.ins.2021.09.051
[6] Goel P, Garantra A. Handwritten Gujarati numerals classification based on deep convolution neural networks using transfer learning scenarios. IEEE Access. 2023;11:20202–20215. doi:10.1109/ACCESS.2023.3249787
[7] Qiao J, Wang G, Wenjing Li W, et al. An adaptive deep Q-learning strategy for handwritten digit recognition. Neural Netw. 2018;107:61–71. doi:10.1016/j.neunet.2018.02.010
[8] Ahamed P, Kundu S, Khan T, et al. Handwritten Arabic numerals recognition using convolutional neural network. J Amb Intel Hum Comp. 2020;11:5445–5457. doi:10.1007/s12652-020-01901-7
[9] Obsie EY, Qu H, Huang Q, editors. Amharic character recognition based on features extracted by CNN and auto-encoder models. The 13th International Conference on Computer Modeling and Simulation; Melbourne VIC, Australia: ACM; 2021 June 25–27.
[10] Meshesha M, Jawahar CV. Recognition of printed Amharic documents. Eighth International Conference on Document Analysis and Recognition (ICDAR'05); 31 Aug.-1 Sept.; Seoul, Korea (South): IEEE; 2005. p. 784–788.
[11] Belay BH, Habtegebrial TA, Stricker D. Amharic character image recognition. 18th IEEE International Conference on Communication Technology; 08-11 October Chongqing, China: IEEE; 2018. p. 1179–1182.
[12] Malhotra R, Addis MT. Ethiopic base characters image recognition using LSTM. 2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST); Mohali, India: IEEE; 2021. p. 94–98.
[13] Addis D, Liu CD, Ta VD. Printed Ethiopic script recognition by using LSTM networks. 2018 International Conference on System Science and Engineering (ICSSE); 28-30 June; New Taipei, Taiwan: IEEE; 2018. p. 1–6.
[14] Aly S, Mohamed A. Unknown-length handwritten numeral string recognition using cascade of PCA-SVMNet classifiers. IEEE Access. 2019;7:52024–52034. doi:10.1109/ACCESS.2019.2911851
[15] Belay B, Habtegebrial T, Meshesha M, et al. Amharic OCR: an end-to-end learning [open access]. Appl Sci-Basel. 2020 Feb;10(1117):1–13.
[16] Azawi N. Handwritten digits recognition using transfer learning. Comput Electr Eng. 2023;106:1–10. doi:10.1016/j.compeleceng.2023.108604
[17] Abdulrazzaq MB, Saeed JN. A comparison of three classification algorithms for handwritten digit recognition. 2019 International Conference on Advanced Science and Engineering (ICOASE); 02-04 April; Zakho - Duhok, Iraq: IEEE; 2019. p. 58–63.

[18] Boukharouba A, Bennia A. Novel feature extraction technique for the recognition of handwritten digits. Appl Comput Inform. 2017;13(1):19–26. doi:10.1016/j.aci.2015.05.001

[19] Liu W, Wei J, Meng Q. Comparisons on KNN, SVM, BP and the CNN for handwritten digit recognition. 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA); Nanjing, China: IEEE; 2020. p. 587–590.

[20] Abu Ghosh MM, Maghari AY. A comparative study on handwriting digit recognition using neural networks recognition using neural networks. 2017 International Conference on Promising Electronic Technologies; Palestine: IEEE; 2017. p. 77–81.

[21] Ahlawat S, Choudhary A, Nayyar A, et al. Improved handwritten digit recognition using convolutional neural networks (CNN). Sensors (Basel). 2020 Jun 12;20(12):1–8. doi:10.3390/s20123344

[22] de Sousa IP. Convolutional ensembles for Arabic handwritten character and digit recognition. Peerj Comput Sci. 2018;167:1–13.

[23] Ashiquzzaman A, Tushar AK, Rahman A, et al. An efficient recognition method for handwritten Arabic numerals using CNN with data augmentation and dropout. Data Management, Analytics and Innovation. Singapore: Springer Singapore; 2019. p. 299–309.

[24] Ahlawat S, Choudhary A. Hybrid CNN-SVM classifier for handwritten digit recognition. International Conference on Computational Intelligence and Data Science (ICCIDS 2019); New Delhi, India: Elsevier B.V.; 2019. p. 2554–2560.

[25] Ali S, Shaukat Z, Azeem M, et al. An efficient and improved scheme for handwritten digit recognition based on convolutional neural network. Sn Appl Sci. 2019 Sep;1:1–9.

[26] Yu N, Jiao P, Zheng Y. Handwritten digits recognition base on improved LeNet5. The 27th Chinese Control and Decision Conference (2015 CCDC); Qingdao, China: IEEE; 2015. p. 4871–4875.

[27] Yao H, Tan T, Xu C, et al. Deep capsule network for recognition and separation of fully overlapping handwritten digits. Comput Electr Eng. 2021;91(2021):1–12.

[28] Sabour S, Frosst N, Hinton GE. Dynamic routing between capsules. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems: ACM Digital Library; 2017. p. 3859–3869.

[29] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–444. doi:10.1038/nature14539

[30] Goodfellow I, Bengio Y, Courville A. Deep learning cambrege. Masssachusetts London: England MIT press; 2016.

[31] Sermanet P, Eigen D, Zhang X, et al. Overfeat: integrated recognition, localization and detection using convolutional networks. 2nd International Conference on Learning Representations, ICLR 2014; 14-16 April Banff, Canada: NYU; 2014. p. 1–16.

[32] Agegnehu M, Tigistu G, Samuel M. Offline handwritten amharic digit and punctuation mark script recognition using deep learning. 2nd Deep Learning Indaba-X Ethiopia Conference 2021; January 27–29; Adama, Ethiopia: Adama Science and Technology University; 2022. p. 53–61.

[33] Tamiru NK, Tekeba M, Salau AO. Recognition of amharic sign language with amharic alphabet signs using ANN and SVM. Vis Comput. 2022;38:1703–1718. doi:10.1007/s00371-021-02099-1

[34] Abeje BT, Salau AO, Mengistu AD, et al. Ethiopian sign language recognition using deep convolutional neural network. Multimed Tools Appl. 2021;81:29027–29043.

[35] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997 Nov 15;9(8):1735–1780. doi:10.1162/neco.1997.9.8.1735

Groundwater for Sustainable Development 26 (2024) 101207

Contents lists available at ScienceDirect

Groundwater for Sustainable Development

journal homepage: www.elsevier.com/locate/gsd

Research paper

Energetic, exergetic, exergoeconomic, and enviroeconomic analysis of solar dish-integrated water desalination system with various nanofluids

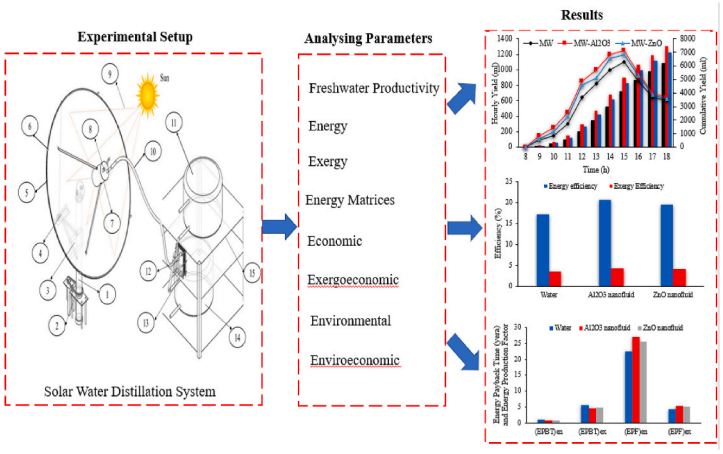
Ravi Kant [a], Anil Kumar [b,c,*], Mukul Kumar Dubey [a]

[a] Solar Energy Division, Sardar Patel Renewable Energy Research Institute, Vallabh Vidyanagar, Gujrat, 388120, India
[b] Department of Mechanical Engineering, Delhi Technological University, Delhi, 110 042, India
[c] Centre for Energy and Environment, Delhi Technological University, Delhi, 110 042, India

HIGHLIGHTS

- Solar dish collector's performance in water desalination is analyzed.
- Annual freshwater yield for SWDS is enhanced by 20.13% (Al_2O_3) and 12.56% (ZnO).
- Energy efficiency of SWDS is enhanced by 20.06 (Al_2O_3) and 12.55% (ZnO).
- Lowest cost per liter of US$0.050 is obtained for SWDS with Al_2O_3 nanofluids.
- Highest amount of net CO_2 mitigation is 72.54 tons of CO_2 for Al_2O_3 nanofluid-based SWDS.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Solar
Desalination
Nanofluids
Exergoeconomic
Enviroeconomic

ABSTRACT

Global demand for freshwater is increasing due to population growth and water scarcity. Traditional desalination systems are characterized by their high energy consumption, resulting in low efficiency and elevated costs. It is crucial to explore sustainable and eco-friendly desalination system to address these problems. Present study explores the impact of Al_2O_3 and ZnO nanofluids on a solar dish integrated water desalination system (SWDS). Performance evaluation includes energy, exergy, economic, exergoeconomic, and enviroeconomic perspectives. Compared to SWDS with water, annual freshwater productivity increased by 20.13% (Al_2O_3) and 12.56% (ZnO). SWDS with Al_2O_3 nanofluid exhibited the highest energy and exergy efficiency of 20.76% and 4.32%, respectively. Energetic parameters revealed a minimum energy payback time of 0.91 years and a maximum energy production factor of 27.18 for SWDS with Al_2O_3 nanofluid. The lowest cost per liter (CPL) was US$0.050 for Al_2O_3 nanofluids, while the highest CPL was US$0.0547 for ZnO nanofluid. The maximum exergoeconomic parameter was 2.50 kWh/US$ for SWDS with Al_2O_3 nanofluid. SWDS with Al_2O_3 nanofluid demonstrated the highest net CO_2 mitigation of 72.54 tonne, with corresponding carbon credits of US$2124.08 (energy) and US

* Corresponding author. Department of Mechanical Engineering, Delhi Technological University, Delhi, 110 042, India.
E-mail address: anilkumar76@dtu.ac.in (A. Kumar).

https://doi.org/10.1016/j.gsd.2024.101207
Received 17 February 2024; Received in revised form 12 May 2024; Accepted 12 May 2024
Available online 14 May 2024
2352-801X/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

$423.92 (exergy). The Al_2O_3 nanofluid-based SWDS demonstrated superior performance in energy and exergy perspective. Furthermore, Al_2O_3 nanofluid-based SWDS showed enhanced cost-effectiveness and environmental sustainability when compared to water-based SWDS.

1. Introduction

The rapid growth in world population and industrialization urges the need for potable water in various regions, especially in hot and dry regions (Yusof et al., 2022). The main challenge in the productivity of potable water is the cost and availability of water sources. The entire world is expected to face 40% water deficit by 2030. Global concerns like COVID-19, global warming, and climate change have made the situation worse. Thus, it is crucial to develop effective methods to overcome the global problem of fresh water (Modi and Gamit, 2022). Solar energy is utilized in various applications such as cooking, desalination, drying, extraction and water heating, etc. (Kant and Kumar, 2022a). Desalination is the process that is used to convert brackish water into fresh water (Kumar et al., 2021). The solar still (SS) is a promising, practical, and cost-effective solar thermal energy-based device that can meet the needs of people for freshwater water who live in water-stressed areas. However, SS must deal with the problem of lower productivity, which prevents its widespread use on a larger scale (Shatar et al., 2023). Hence, the researchers opted to use internal and external additives to enhance efficiency and productivity of solar energy systems. There are several ways to utilize solar energy effectively; one of these methods is heat transfer. The basic idea behind this technology is to achieve high temperatures by concentrating sun radiation onto a thermal receiver using concentrators. Mirrors and lenses can be used to focus solar rays. Mirrors are preferred over lenses due to lower absorptivity and cost as rays are focused by reflection. Three types of solar thermal transfer systems have been used in manufacturing fields: flat mirror concentrators, cylindrical mirror concentrators, and parabolic concentrators (Hameed et al., 2023). Parabolic dish collector is a parabolic concentrator that comprises several pieces of mirrors. It concentrates sunlight on the focused receiver. The receiver converts the reflected solar radiation into thermal energy for heating the fluid to high temperatures and allowing it to be employed in steam generation.

Various studies have been performed to design various solar concentrator systems based on their shape, configurations, physics, use of PCM and nanomaterials, and many other aspects. Hameed et al. (2023) investigated the effectiveness of solar dish integrated water desalination system using different nanofluids. The thermal performance and productivity of system were evaluated using tap water, aluminum oxide (Al_2O_3), and zinc oxide (ZnO) nanofluids. The results showed that Productivity of system was increased by 43.53% and 21.89% using Al_2O_3 nanofluid as compared to tap water and ZnO nanofluid, respectively. The thermal efficiency of the system was also increased by 9.91% using Al_2O_3 nanofluid as compared to tap water. Chaouchi et al. (2007) designed and built a small solar desalination unit equipped with solar dish and developed a theoretical model to evaluate its performance. There was a relative error of 42% in theoretical and experimental values of distillate flow rate. The results showed that design of the system required to be modified with special emphasis on the absorber. Mawire and Taole (2014) performed energy and exergy analysis of cylindrical cavity receiver for solar dish concentrator. The efficiencies and rates of exergy for the receiver were discovered to be less than the efficiencies and rates of energy. Receiver's heat loss factor and parabolic dish system's optical efficiency were determined as 4.6 W/K, and 52%, respectively for high solar radiation conditions. Alhawsawi et al. (2023) evaluated the energetic performance of the hybrid solar dish/Stirling-powered single effect distillation system (SDSPSEDS) in terms of, solar-to-electricity conversion efficiency, net electric power, distilled freshwater productivity, Stirling engine rejected heat, and overall co-generation efficiency. The modelling results showed that electrical energy produced by system was 234.7, 194.6, 159.8, and 190.0 kWh/day. However, freshwater productivity was obtained as 787.4, 657.3, 545.0, and 650.0 kg/day in June, September, January, and March, respectively.

Stefanovic et al. (2018) optimized inlet temperature and flow rate for a solar dish collector with a spiral-coil absorber in order to maximize the collector's exergetic performance. The optimal fluid temperature and flow rate were 212.3 °C, and 314.6 L/h, respectively with thermal and exergetic efficiencies of 49.83% and 21.42%. Elashmawy and Alshammari (2020) proposed a new system for harvesting water from low humid regions using a tubular solar still powered by a parabolic concentrator system. Tubular solar still was equipped with solar parabolic concentrator to evaporate the water from calcium chloride in the proposed system. System can produce 0.51 L/kg of calcium chloride. The thermal efficiency and cost of water production were 24.61% and US$0.15/L, respectively. Productivity and efficiency of proposed system were increased by 292.4% and 82.3%, respectively, while water production cost was reduced by 25%. Abulkhair et al. (2023) enhanced yield and thermal performance of conventional pyramid distiller using low-cost materials such as v-corrugated aluminum basin, paraffin wax, black cotton fabric, carbon black nanofluid, and CB nanoparticles. Moreover, thermo-economic performance of modified pyramid distiller was evaluated and compared with conventional distiller. Yield, thermal efficiency and exergy efficiency of modified distiller (addition of CB nanoparticles with paraffin wax) was improved by 76.9%, 101.5% and 109.7%, respectively as compared to conventional distiller. Dafle and Shinde (2012) designed, developed and evaluated the performance of steam generation system equipped with Scheffler reflector (16 m^2 area). Overall performance of system and maximum temperature of steam at outlet of biller was obtained as 57.41% and 107 °C, respectively.

Malwad and Tungikar (2020) performed energy analysis to determine the thermal performance of receiver with three different configurations such as bare receiver, cavity receiver and tube coil on inner wall of cavity for solar Scheffler reflector. Under identical working conditions, the energy efficiency of the receiver with the first, second, and third configurations was 57.71%, 60.93%, and 68.03%, respectively. Farooq et al. (2022) analyzed thermal performance of parabolic trough collector (PTC) for two nanofluids (Al_2O_3 and CuO) at different flow rates using computational fluid dynamics (CFD). Maximum efficiency of system was improved by 6.99% and 12.90% for CuO nanofluid as compared to Al_2O_3 nanofluid. Furthermore, Temperature distribution behaviour of absorber tube materials such as steel, copper, and aluminum was also examined. The maximum output temperature for copper was 311 K, whereas exit temperatures for steel and aluminum were 307 K and 308 K, respectively. Mohanraj et al. (2021) evaluated the performance and asses the techno feasibility analysis of heat pump water heater assisted regenerative solar still with latent heat storage. Results showed that system is capable of producing 14 kg/day fresh water and 100 L hot water for domestic use. Cost of fresh water production was obtained as 0.0547 US$/L.

India has committed to reducing the emissions intensity of its GDP by 45 percent from 2005 levels by 2030, in accordance with the revised Nationally Determined Contribution (NDC) that it submitted to the United Nations Framework Convention on Climate Change (UNFCCC) in August 2022 under the COP26 agreement. Government of India has committed to addressing climate change through several initiative and programmes such as National Action Plan on Climate Change (NAPCC) which consist of missions in specific areas of energy efficiency, solar energy, water, Himalayan ecosystem, sustainable agriculture, sustainable habitat, health, green India, and strategic knowledge for climate

change. National Solar Mission under the NAPCC is one of the key initiatives to promote sustainable growth while addressing India's energy security ("Commitments made under COP 26," n.d.).

Existing literature primarily focuses on analyzing solar concentrator systems based on receiver shape, configuration, and physics. Limited research has been conducted on the performance of water desalination systems equipped with solar dish collectors using various nanofluids. To the best of our knowledge, there is a notable gap in the literature regarding the energy, exergy, economic, and environmental analysis of solar dish integrated water desalination systems. Therefore, the primary objective of this study is to comprehensively analyze the performance of a solar dish collector integrated water desalination system, considering productivity, energy, exergy, economic, and environmental parameters. Additionally, the study explores the interplay between energy and exergy with the economic (exergo-economic), environment aspects (enviro-economic), and energy matrices to enhance our understanding of the proposed system's performance.

2. Methodology

2.1. Experimentations

2.1.1. Nanofluid preparation

Al_2O_3 and ZnO nanoparticles are employed in this study to produce nanofluids as they have lower cost compared to other nanoparticles like gold nanopowder, carbon nanotubes, and diamond nanopowder. The thermophysical properties of the Al_2O_3 and ZnO nanoparticles are depicted in Table 1.

Nanofluids were produced in Green Energy Lab, Delhi Technological University, Delhi (India) using two-step process. Al_2O_3 and ZnO nanoparticles were directly dissolved in water within a Pyrex flask with a PVC cover to prepare the nanofluid. The flask was set up inside an ultrasonic water bath in a stainless-steel basket. Water was poured into basin of ultrasonic device to a depth of about 3 cm above the mixture level in flask. Then, degas mode was switched on to remove air from the mixture. After that, the flask was sealed with the cap as it continually oscillated for 8 h for Al_2O_3 and 10 h for ZnO at 60–70 °C with a frequency and power efficiency of 37 kHz and 100%, respectively. As a result, the nanoparticles can be evenly distributed. It is important to note that no surfactants were added to the nanofluid to prevent their impact on thermophysical properties of nanofluids. Ultrasonic bath's container was covered with a cap to stop the evaporation of water during sonication.

Thermal conductivity of Al_2O_3 was improved by 10.34% by using 0.10% concentration of Al_2O_3. Hence, Al_2O_3 and ZnO nanoparticle concentration of 0.10% (w/w) was used to prepare the nanofluid. Eq. (1) was used to calculate the amount of nanoparticles required to produce 0.10% w/w concentrated nanofluid (Modi and Gamit, 2022).

$$\textit{Weight of nano particle needed}\ (g) = \frac{\textit{Concentration}(\%)}{100} \times \textit{amount of water}\ (mL) \times \textit{bulk density of nanoparticles}\ (g/mL) \quad (1)$$

Amount of water was taken as 2500 mL in Eq. (1) to determine weight of nanoparticles required. Groundwater was mixed with 0.45 g and 4.25 g of nanoparticles to produce 0.10% w/w concentrated Al_2O_3 and ZnO nanofluid, respectively.

Table 1
Thermophysical properties of nanoparticles.

S. No.	Properties	Al_2O_3	ZnO
1	Thermal Conductivity (W/mK)	40	29
2	Density (kg/m^3)	3950	5500
3	Molar mass (g/mol)	101.96	81.38
4	Specific heat (J/kg °C)	773	490
5	Bulk density (g/mL)	0.18	1.7

2.1.2. Thermal property measurement

A particular mixture of water and nanoparticles was prepared in a beaker using ultrasonication to ensure the perfect blending of samples. Afterward, the mixture was carefully transferred into a mold inside the Hot Disk Device for thermal conductivity measurement. This device had an inbuilt sensor intended specifically for thermal analysis. Mold ensured the creation of uniform flat disk for accurate measurements. The sample was placed between two surfaces of hot disk sensor in the test setup device. The temperature variation rate was carefully observed upon activation of the instrument. The thermal conductivity of each sample was then calculated using these data. Initially, thermal conductivity of underground water was 0.598 W/mK and after preparing the water nanofluid the thermal conductivity of Al_2O_3 and ZnO nanofluid was calculated as 0.657 and 0.639 W/mK, respectively. These results showed that thermal conductivity of Al_2O_3 and ZnO nanofluid was improved by 10% and 7%, respectively compared to water as depicted in Table 2.

2.1.3. Experimental setup and observations

Solar water desalination system (SWDS) is installed on rooftop of Centre for Energy and Environment, Delhi Technological University, Delhi (28.7041° N, 77.1025° E). Experiments were carried out with and without nanofluids for nine different days in May–June 2023. First experiment was conducted using underground water and other two experiments were conducted using Al_2O_3 and ZnO nanofluid. Data was recorded from 8 a.m. to 6 p.m. for every trial day. Each experiment was repeated three times to verify that results were consistent. SWDS mainly consists of four parts: solar dish collector (2.5 m^2), boiler, water tank and condenser. An iron pole is installed in concrete construction to support the solar dish reflector. Solar dish is fabricated and prepared by covering an aluminum frame structure with mirrors of glass (0.05 × 0.05 m) as reflecting surface. Solar radiation falling on surface of reflector is reflected and concentrated at the base of receiver (boiler). The stainless-steel boiler was 0.12 m in diameter and 0.28 m in height. All faces of boiler except its base are insulated with glass wool (2.5 cm) to reduce heat loss to environment. Base of the boiler is painted with thermal black coating (absorptivity 0.90) to reduce reflection losses.

A ball valve fixed to the top of boiler is used to regulate water inlet flow and outlet steam flow through boiler. At the base boiler, a brass tube with a diameter and length of 25 mm and 5 cm, respectively was welded before being rotated upward. This pipe is connected to a 25 cm long and 25 mm diameter Pyrex pipe that serves as a sight glass to show the water level inside the boiler. It is also utilized to eliminate the concentrated salts produced by raw water desalination at the end of experiment. Radiation heats the water in the boiler, turning it into steam. Flexible nylon pipe of 1.25 cm in diameter is used to connect the boiler to condenser. An automatic double-axis tracking mechanism is implemented to enhance system's efficiency. Sun was tracked using four light difference resistors (LDR sensor) and two motors. The microcontroller receives a signal from the LDR sensor, which turns the motors in the appropriate direction. First motor is used to rotate solar dish in left-right direction and other one is used to move solar dish in up-down direction. Fig. 1 depicts the SWDS schematic view. Table 3 shows a detailed explanation of SWDS.

A salty cold-water tank of 50 L is located 2 m above the ground, which is higher than the receiver height of 1.65 m and condenser height of 1 m. A ball valve is connected to bottom of tank to control head of pressure and flow of water from water tank to receiver and condenser in

Table 2
Thermal conductivity of nanofluid.

Samples	Thermal conductivity (W/mK)	Increment (%)
Under groundwater	0.598	–
Al_2O_3 nanofluid	0.657	10
ZnO nanofluid	0.639	7

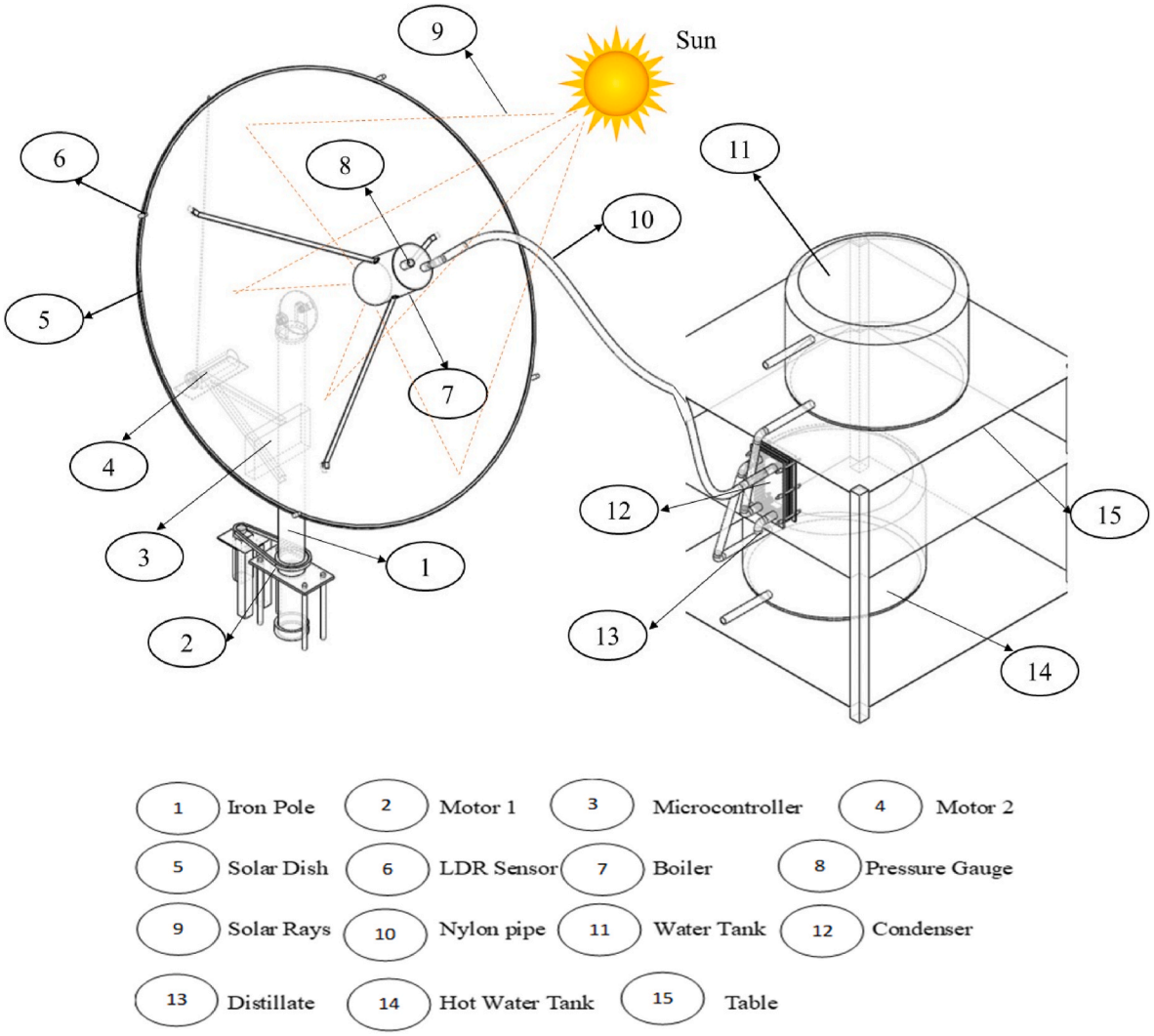


Fig. 1. Schematic of solar water desalination system.

order to maintain the desired water level and flow rate inside the receiver and condenser, respectively. PT 100 calibrated thermocouples with proper insulation are used to record the temperature at different positions of boiler and condenser. A Pressure gauge is attached to boiler to record the inside pressure. All thermocouples and pressure gauge is connected to 16-channel data logger (digital display unit). Solar radiation was recorded using solar power meter (Mico-936). Wind velocity and ambient humidity were measured by wane type anemometer (Testo-416) and thermal hygrometer (Testo-608-H1), respectively.

2.2. Uncertainty analysis

The standard uncertainty of the instruments used for observation is calculated from Eq. (2).

$$Standard\ Uncertainty = \frac{Accuracy\ of\ instrument}{\sqrt{3}} \tag{2}$$

Table 4 summarizes the instruments used in experiment along with their accuracy, range, and uncertainty. When any function Y depends on the number of input parameters or independent variables, then the uncertainty in the measured value of the function Y is given by Eq. (3) (Ahmad et al., 2023; Kumar and Sahoo, 2023):

$$U(Y) = \left[\left(\frac{\partial Y}{\partial x_1}\right)^2 u^2(x_1) + \left(\frac{\partial Y}{\partial x_2}\right)^2 u^2(x_2) + \ldots\ldots\ldots + \left(\frac{\partial Y}{\partial x_n}\right)^2 u^2(x_n)\right]^{1/2} \tag{3}$$

where, U(Y) is the uncertainty in the measured function and u(x1) … …u (x_n) is the uncertainties in the independent variables (x_1 … …x_n), affecting the function Y.

2.3. Optical design of solar dish concentrator

The purpose of this mathematical framework is to enable geometrical sizing to obtain the geometrical design parameters, such as focal length, receiver diameter, receiver position, and geometric design concentration, for various geometrical dish sizes. The optimal receiver height (H_r), rim angle ($\varnothing_r$), distance between focal point and receiver (d_f), and diameter of receiver (d) are geometrical parameters used in defining the geometry of receiver/concentrator system in proposed model.

Initially, the concentrator's focal length (f) can be calculated as a function of rim angle ($\varnothing_r$) and concentrator diameter (D) using Eq. (4) (Elashmawy, 2020; Elashmawy and Alshammari, 2020; Hameed et al., 2023).

Table 3
Detail description of SWDS.

S. No.	Components	Size (m)	Materials	Properties of Materials (Farooq et al., 2022; Modi and Gamit, 2022)
1.	Solar dish reflector Plane mirrors	Diameter = 1.8 Depth = 0.20 Aperture area = 2.5m^2 No of mirrors = 1600 (0.05 × 0.05)	Aluminum Glass	Density = 2700 kg/m^3, Thermal conductivity = 202.4 W/mK Reflectivity = 0.95
2.	Boiler	Capacity = 7 litters Height = 0.28 Diameter = 0.16 Thickness = 0.004	Stainless steel	Density = 8020 kg/m^3, Thermal conductivity = 16.27 W/mK
3.	Boiler base paint	Matt-type	Black color	Absorptivity = 0.95 Emissivity = 0.08
3.	Condenser (counter flow plate type heat exchanger)	Inlet diameter = 0.012 Outlet diameter = 0.012	Mild steel	Density = 7700 kg/m^3
4.	Insulation	Thickness = 0.025	Glass Wool	Thermal Conductivity = 0.055 W/mK

$$f = \frac{D}{4\tan\left(\frac{\varnothing_r}{2}\right)} \tag{4}$$

Rim radius is the distance between concentrator surface and receiver position and is determined (Zayed et al., 2020):

$$R_r = \frac{2f}{1 + cos\ \varnothing_r} \tag{5}$$

The spread of reflected beams from the concentrator should be taken into consideration in order to determine the aperture size of receiver for a specific concentrator diameter due to its significant impact on the optical behaviour of system. Other error sources, such as tracking drive system, sun tracking sensor, physical structure, positioning of receiver, and specular reflectance, affect the optical behaviour of the SWDS in addition to spread of reflected beam.

Eq. (6) provides the concentrator's overall optical error in terms of the SDSS's auxiliary component errors (Alhawsawi et al., 2023).

$$\sigma_t = \sqrt{(2\sigma_{str})^2 + (2\sigma_{sen})^2 + (2\sigma_{drive})^2 + (2\sigma_{ralign})^2 + (2\sigma_{mref})^2 + (2\sigma_{sun})^2} \tag{6}$$

Then, total spread of beam (Δ b,r) in a plane perpendicular to the reflected beam is as (Ranjbar et al., 2021):

$$\Delta_{b,s} = 2R_r\ \tan\left(n\frac{\sigma_t}{2}\right) \tag{7}$$

Receiver diameter (d) can be calculated by estimating beam spread projection on focal plane of receiver (Ouyang et al., 2021):

$$d = \frac{\Delta_{b,s}}{Cos(\varphi_r)} \tag{8}$$

The distance between focal point and receiver (d_f), and the focal point position (H) are calculated as (Ouyang et al., 2021):

$$d_f = \frac{d}{D}\left(f - \frac{D^2}{16f}\right) \tag{9}$$

$$H = f - d_f \tag{10}$$

The opto-geometrical solar concentration (C_{GD}) is a crucial geometric factor that has a considerable impact on the cavity receiver's energetic efficiency. It can be calculated using Eq. (11) by considering rim angle ($\varnothing_r$), solar angle (α_s), and dispersion angle (δ_{dis}) (Lashari et al., 2021).

$$C_{GD} = \left[\frac{Sin\ \varnothing_r\ Cos\left(\varnothing_r + \left(\frac{\alpha_s + \delta_{dis}}{2}\right)\right)}{Sin\left(\frac{\alpha_s + \delta_{dis}}{2}\right)}\right]^2 \tag{11}$$

The rim angle and focal length are responsible for completely defining the cross-sectional geometry of a dish reflector. In the present study, a rim angle of 45° for the parabolic dish concentrator design, as suggested by Lashari et al., 2021)(Lashari et al., 2021), has been chosen to achieve the maximum concentration ratio. The optimal receiver height (H) indicates the location of cavity receiver, where the DNI concentrates and produces highest temperatures. The receiver material, through which the heat transfer fluid passes, might be damaged by the high temperatures. Therefore, the receiver is considered to be at optimum distance (d_f) from the focal point in order to reduce receiver material overheating and provide consistent DNI. As a result, Table 5 shows optimized geometric parameters for the proposed collector/receiver system.

Table 5
Characteristics of solar dish concentrator and heat receiver.

Components	Characteristics	Values
Solar dish concentrator	Aperture diameter (D)	1.8 m
	Depth (h)	0.3 m
	Rim angle ($\varnothing_r$)	45
	Rim radius (R_r)	1.26 m
	Focal length (f)	1.08 m
	Aperture area of solar dish (A_d)	2.54 m^2
Heat receiver	Diameter (d)	0.12 m
	Optical error (σ_t)	11.34 mrad, (0.649)
	Spread of beam	1.69 m
	Distance of focal point from receiver (d_f)	0.19 m
	Optimal height of receiver (H)	0.88 m
	Receiving surface area (Ar)	0.01 m^2
	Optic geometrical concentration factor (C_{GD})	2551

Table 4
Technical specification of the instruments used in the experiments.

Instruments	Observed Parameter	Accuracy	Range	Standard Uncertainty	Uncertainty in Observed Parameter
Thermocouple (TC-KT-02)	Temperature	±0.1 °C	−20 to 200 °C	0.057 °C	±0.25 °C
Solarimeter (PMV-210)	Solar radiation	10 W/m^2	0–2000 W/m^2	5.77 W/m^2	±5.77 W/m^2
Digital Hygrometer (MEXTECH TM-1)	Ambient temperature and relative humidity	3%	5–85%	1.73%	±1.7%
Scaled Flask	Distilled Water	±10 ml	0–5000 ml	5.77 ml	±5 ml

2.4. Energy analysis

The concept of energy analysis has been developed to study and understand proposed systems in light of thermodynamics first law. This analysis calculates energy distribution across various components as well as performance of the proposed water desalination system. Solar energy incident on solar dish collector is computed as (Kant and Kumar, 2022b):

$$E_c = I_g A_c \tag{12}$$

Optical efficiency of collector is calculated as (Alhawsawi et al., 2023):

$$\eta_{optical} = \alpha\rho\tau S \tag{13}$$

Energy collected by reflector drops due to various losses such as optical losses, and scattering of rays reflected from dish surface. As a result, the energy focused on surface of receiver is computed by Eq. (14) as a function of solar irradiation (I_g), reflectivity of mirror (ρ), area of the concentrator (A_c) and shape factor (S) (Alhawsawi et al., 2023).

$$E_r = \eta_{optical} I_g A_c \tag{14}$$

Useful thermal energy output from the receiver is determined (Stefanovic et al., 2018):

$$E_u = \dot{m} C_p (T_o - T_i) \tag{15}$$

Thermal efficiency of receiver and solar collector is evaluated using Eq. (16) and Eq. (17), respectively (Stefanovic et al., 2018).

$$\eta_r = \frac{E_u}{E_r} \tag{16}$$

$$\eta_c = \frac{E_u}{E_c} \tag{17}$$

Overall thermal efficiency of SWDS is calculated as (Rani et al., 2021):

$$n_{th,o}(\%) = \frac{\sum \dot{m} L}{A_c \sum I_g(t)} \times 100 \tag{18}$$

2.5. Exergy analysis

Maximum amount of useful work obtained by system when it is in equilibrium with its environment is known as exergy. Exergy analysis is used in solar dish concentrator systems in order to reduce exergy losses and get the best design parameters. Useful exergy production is computed as (Stefanovic et al., 2018):

$$Ex_{,u} = (\dot{m} \times L)\left[1 - T_a\left(\frac{T_a}{T_w}\right)\right] \tag{19}$$

Exergy of sun is determined by Petela model (Kant and Kumar, 2022b; Stefanovic et al., 2018):

$$Ex_{,s} = I_g A_c \left[1 + \frac{1}{3}\left(\frac{T_a}{T_s}\right)^4 - \frac{4}{3}\left(\frac{T_a}{T_s}\right)\right] \tag{20}$$

Exergy efficiency of SWDS is evaluated as (Kant and Kumar, 2022b):

$$\eta_{ex}(\%) = \frac{Ex_{,u}}{Ex_{,s}} \times 100 \tag{21}$$

2.6. Energy matrices

The energy matrices are used as a decision-making tool to determine feasibility of a solar system from an energy perspective by comparing energy produced with energy consumed by system for its fabrication. Energy payback time and energy production factor are the two energy matrices that are evaluated in this study.

2.6.1. Energy payback time (EPBT)

The energy payback time (EPBT) is a period required to recuperate the energy spent (embodied energy) during the production of the SSDS components. EPBT is evaluated (Reddy et al., 2018; Sharon et al., 2020):

$$EPBT = \frac{E_{emb}}{E_{en,o}} \tag{22}$$

Thermal energy output obtained per year from distillation system can be calculated (Kumar, 2013):

$$E_{en,o} = \frac{\dot{m} L}{3600} \tag{23}$$

2.6.2. Energy production factor (EPF)

EPF predicts the performance of the system. It is computed as the amount of energy produced over life span divided by the system's embodied energy.

$$EPF = \frac{E_{en,o} \times n}{E_{emb}} \tag{24}$$

2.7. Economic analysis

The cost of freshwater produced per liter (CPL) is often determined using economic analysis. This analysis enables decision-makers to pick a system that is beneficial from both a technical and financial perspective. In order to get a clearer understanding of the feasibility of system from an energy and environmental perspective, this research additionally assesses the exergo-economic and enviro-economic factors.

The interest rate for commercial loans in India i.e 7% is used as the basis for interest rate (i) selection. Life of system (n) was taken as 30 years. The capital recovery factor (CRF) is calculated as (Shatar et al., 2023):

$$CRF = i(1+i)^n / [(1+i)^n - 1] \tag{25}$$

First annual cost of system is the product of total initial cost (TC) and CRF and is determined by Eq. (26) (Modi and Gamit, 2022).

$$FAC = TC \times (CRF) \tag{26}$$

The salvage value (S) was taken as 20% of TC and annual salvage value is determined using Eq. (28) (Modi and Gamit, 2022; Shatar et al., 2023).

$$S = 0.2 \times TC \tag{27}$$

$$ASV = SFF \times S \tag{28}$$

where, sinking fund factor is determined as:

$$SFF = i / [(1+i)^n - 1] \tag{29}$$

Annual maintenance cost was assumed as 15% of FAC and uniform annual cost (UAC) is calculated using Eq. (30) (Colakoglu and Durmayaz, 2022).

$$UAC = FAC + AMC - ASV \tag{30}$$

Cost per liter (CPL) of fresh water produced is the ratio of UAC and annual productivity (P) and is determined as (El-Sebaii and El-Naggar, 2017):

$$CPL = \frac{UAC}{P} \tag{31}$$

2.8. Exergoeconomic analysis

The relationship between energy and exergy produced by solar desalination system and overall annual cost of system is determined using the exergoeconomic analysis. Exergoeconomic parameter based

on energy approach is determined as (Shatar et al., 2023):

$$R_{en} = \frac{E_{n,o}}{UAC} \tag{32}$$

in order to maximize the exergy benefit, the exergoeconomic parameter was applied and calculated as (Abbasi et al., 2022; Fang et al., 2021a):

$$R_{ex} = \frac{E_{x,o}}{UAC} \tag{33}$$

2.9. Environmental analysis

The embodied energy method is a well-known designation for the environmental assessment framework since it is adaptable and able to compare systems easily. Embodied energy is the total amount of energy required for production of all of the materials and system parts (Vijayan et al., 2020). This kind of examination makes it possible to identify materials with high embodied energy values and subsequently swap them out for more beneficial ones. The main problem with this kind of research is a lack of knowledge of system boundaries. Materials like stainless steel, glass wool, brass, mild steel, and aluminum are used for construction of the solar distillation system. The solar distillation system is fabricated and developed using energy from coal-fired power plants. However, only solar thermal energy is used when the system is in operation, facilitating the recovery of energy used during construction of system.

The amount of carbon dioxide released and neutralized by the solar desalination system is determined by environmental analysis. This consideration is crucial to make sure that the technology being utilized does not result in increased CO_2 emissions, which would be opposed to the use of renewable technology. Additionally, this analysis displays the system's net carbon emissions, which is important for choosing carbon-neutral technology. As a result, the equations utilized to determine the CO_2 emission and mitigation are shown in the section.

2.9.1. CO_2 emissions

The coal power plant produced roughly 2 kg CO_2/kWh of CO_2 considering transmission and distribution losses (Shatar et al., 2023). Lifetime CO_2 emission, CO_2 of the solar desalination system is determined as (Kant and Kumar, 2023):

$$CO_2 \text{emissions(tons) over life span} = \frac{E_{emb} \times 2}{1000} \tag{34}$$

2.9.2. CO_2 mitigation

Environmental parameters are defined as CO_2 mitigation over the system's lifetime (tons). The computation of environmental parameters uses an energy and exergy approach. As a result, the amount of CO_2 that has been mitigated over the system's lifetime (tons) using an energy-based approach is computed as (Fang et al., 2021b; Saxena et al., 2022):

$$\varnothing_{en,CO_2} = \frac{E_{en,o} \times 2 \times n}{1000} \tag{35}$$

Exergoenvironmental parameter is the CO_2 mitigated over system life span (tons) using exergy approach and is estimated as:

$$\varnothing_{ex,CO_2} = \frac{E_{ex,o} \times 2 \times n}{1000} \tag{36}$$

2.10. Enviroeconomic analysis

The enviroeconomic approach determines annual revenue from CO_2 mitigations over life span of system. Energoenviroeconomic parameter evaluates CO_2 mitigation revenue based on energy approach and is represented as (Elbar et al., 2019; Fang et al., 2021b):

$$Z_{en,CO_2} = z_{CO_2} \times \varnothing_{en,CO_2} \tag{37}$$

Exergoenviroeconomic parameter determines and evaluates CO_2 mitigation revenue based on exergy approach and is calculated as (Singh et al., 2018):

$$Z_{ex,CO_2} = z_{CO_2} \times \varnothing_{ex,CO_2} \tag{38}$$

Carbon prices of US$14.5/tonnes of CO_2 have been used in previous studies. However, carbon credit's present value has been revised in the post-COVID era (the year 2022) as per the California Carbon Allowance, which averaged roughly US$28 per tonnes of CO_2 (Shatar et al., 2023).

3. Results and discussion

3.1. Performance of system

Experiments were performed with (Al_2O_3 and ZnO nanofluid) and without using nanofluids (groundwater) on nine different consecutive days (28/05/2023–05/06/2023). Every experiment was repeated three times and an average value of three trials was used for graphical representation. Although the experiments were conducted on different days from 8 a.m. to 6 p.m., it is important to note that there was very little variation in ambient temperature and solar radiation at every trial hour on every trial day. Thus, the data reported here only pertain to the solar radiation, and ambient temperature recorded on the first experimental day. Variation in solar radiation, ambient temperature and receiver fluid temperature with and without using nanofluids with respect time are illustrated in Fig. 2. Ambient and receiver temperatures showed a similar pattern as solar radiation. The average solar radiation and ambient temperature observed during first experiment were 808 W/m^2 and 31.5 °C respectively. The maximum receiver water temperature without nanofluid (T_w) and with Al_2O_3 (T_w- Al_2O_3) and ZnO (T_w-ZnO) nanofluids was about 114.3, 122.2 and 118.2 °C, respectively. The results showed that average receiver water temperature using Al_2O_3 (T_w-Al_2O_3) and ZnO (Tw-ZnO) nanofluids was higher than that of average receiver water temperature without using nanofluids (Tw) by 7.64% and 3.60%, respectively. This happens because water is lowest thermal conductive fluid as compared to Al_2O_3 and ZnO nanofluids. It is clear that the investigation of the receiver fluid temperature also shows the dependency on thermal conductivity of fluid.

Average hourly and cumulative yield of SWDS over the day for Al_2O_3, ZnO nanofluids and without nanofluid (groundwater) is depicted in Fig. 3. The hourly yield was initially increased to its maximum value at 15hr. and subsequently decreased. The SWDS with nanofluids produced more freshwater than SWDS without nanofluids. The maximum hourly yield of 1250 ± 6.2, 1200 ± 18.02 and 1100 ± 5 mL was obtained for Al_2O_3, ZnO nanofluids and without nanofluid, respectively at 15hr. The maximum daily cumulative yield for Al_2O_3, ZnO nanofluids and without nanofluid were obtained as 7460, 6995 and 6215 mL, respectively. Results showed that cumulative yield of SWDS with Al_2O_3,

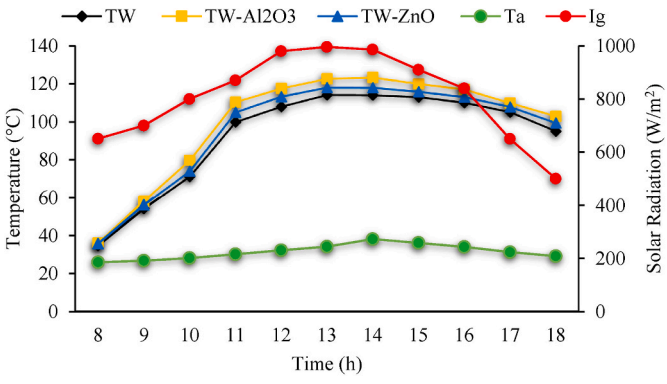


Fig. 2. Solar radiation and receiver fluid temperature with and without nanofluids versus time.

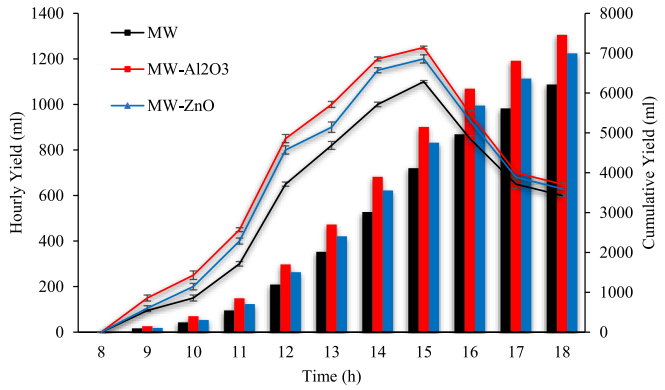


Fig. 3. Hourly and cumulative yield with and without nanofluid.

and ZnO nanofluids was 20.03% and 12.55% more than that of SWDS without nanofluids, respectively as nanofluids are more conductive than water.

Hourly efficiency for SWDS over the day for Al_2O_3, ZnO nanofluids and water is illustrated in Fig. 4. Efficiency is higher for Al_2O_3, ZnO nanofluids than that without nanofluid. Hourly efficiency was initially increased to its maximum value at 15hr and subsequently decreased. Maximum hourly efficiency of 33.95, 32.59 and 29.87% was obtained for Al_2O_3, ZnO nanofluids and water, respectively at 15hr as yield was maximum at this instant of time.

Energy and exergy efficiency of SWDS with Al_2O_3, ZnO nanofluids and water is depicted in Fig. 5. Energy efficiency of SWDS for Al_2O_3, ZnO nanofluids and water was obtained as 20.76, 19.46 and 17.29%, respectively whereas the corresponding exergy efficiency was obtained as 4.32, 4.08 and 3.52%, respectively. The energy efficiency of system for Al_2O_3, and ZnO nanofluids was 20.06 and 12.55% more than water whereas corresponding exergy efficiency was 22.72 and 15.90% more than that of water as corresponding yield was more.

3.2. Energy matrices

The embodied energy for SWDS is calculated in order to ensure that SWDS proposed in this research is both economically and energetically impressive. Table 6 shows the embodied energy values for the materials and parts of SWDS. Total embodied energy for SWDS is 1162 kWh.

Fig. 6 shows the energy payback time (EPBT) and energy production factor (EPF) of the SWDS with and without nanofluids based on energy and exergy approach. Energy payback time (EPBT) evaluation is required in order to confirm the sustainability of energy systems. Based on the findings from the several experiment days, the yield, energy, and energy output are determined for SWDS with and without nanofluids.

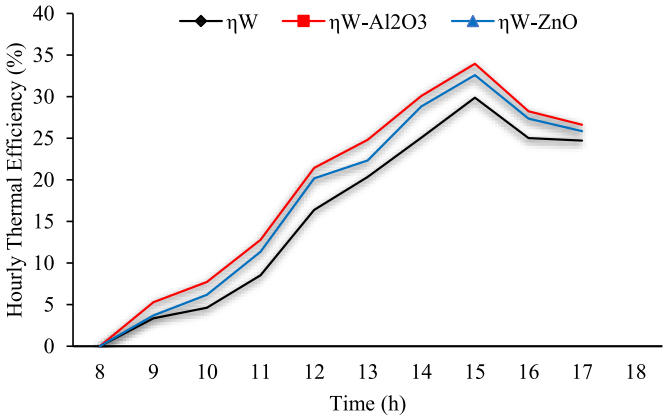


Fig. 4. Instantaneous thermal efficiency of system with and without nanofluid.

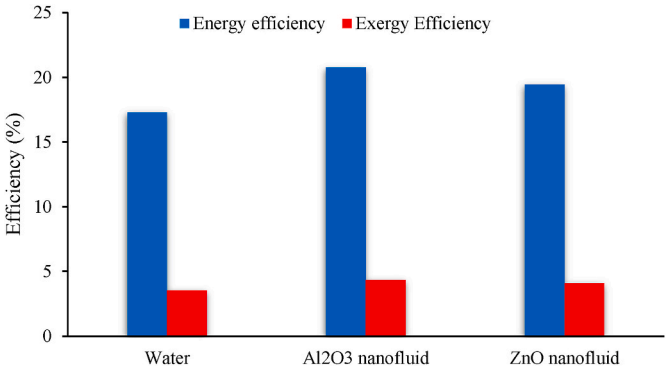


Fig. 5. Average energy and exergy efficiency of system with and without nanofluid.

The SWDS with Al_2O_3 nanofluids produced maximum energy and exergy output. The outcome demonstrates that the payback period for SWDS with Al_2O_3 nanofluids was shorter than SWDS with water and ZnO nanofluid from energetic and exergetic points of view. The EPBT values for SWDS with water, Al_2O_3 nanofluids and ZnO nanofluid based on energy approach were 1.10, 0.91 and 0.98 years, respectively. However, the corresponding values based on exergy approach were 5.69, 4.60 and 4.88 years, respectively. The energy production factor (EPF) based on energy and exergy point of view for SWDS with Al_2O_3 nanofluids was higher than SWDS with water and ZnO nanofluid. EPF values for SWDS water, Al_2O_3 nanofluids and ZnO nanofluid were 22.59, 27.18 and 25.45 based on energy approach. Whereas, based on exergy approach, the corresponding values were 4.38, 5.42 and 5.11, respectively. EPF for SWDS with Al_2O_3 nanofluids was enhanced by 20 and 6.79% than SWDS with water and ZnO nanofluid, respectively.

3.3. Economic analysis

Table 7 lists the total initial cost and various economic parameters for SWDS with and without nanofluids. The uniform annual cost was highest for SWDS with ZnO nanofluid and lowest for SWDS without using nanofluid. Cost per liter (CPL) values for SWDS without using nanofluid, with Al_2O_3 nanofluid and ZnO nanofluid were 0.051, 0.050 and 0.054 US$/L, respectively. CPL value was highest for SWDS with ZnO nanofluid and lowest for SWDS with Al_2O_3 nanofluid. CPL was enhanced by 7.25% using Zno nanofluid as compared to water. Whereas, CPL was reduced by 1.96% using Al_2O_3 nanofluid as compared to water.

3.4. Exergoeconomic analysis

Exergoeconomic analysis takes into account the financial aspects for evaluating SWDS with and without nanofluids. Table 8 illustrates the exergoeconomic evaluation's findings for the desalination system using water, Al_2O_3 nanofluid and ZnO nanofluid. It can be concluded that the SWDS with Al_2O_3 nanofluid had a higher exergoeconomic parameter value than the equivalent value of a SWDS using water. Ener-ergoeconomic parameters for SWDS without nanofluid, Al_2O_3 nanofluid and ZnO nanofluid were 12.20, 12.52 and 10.92 kWh/US$, respectively. Whereas, corresponding values of exergoeconomic parameter were 2.37, 2.50 and 2.19 kWh/US$, respectively. Exergoeconomic evaluation indicates that the incorporation of Al_2O_3 nanofluid and ZnO nanofluid is desirable. This conclusion is a result of high energy output of SWDS with nanofluids offsetting the additional cost of nanofluids.

3.5. Environmental analysis

CO_2 emissions and mitigations for SWDS with and without nano-fluids are depicted in Table 9. SWDS with Al_2O_3 nanofluid released highest CO_2 emissions over the course of its 30-year lifetime. CO_2

Table 6
Embodied energy of solar water desalination system (SWDS).

Components	Material	Energy Intensity (MJ/kg)	Weight of Components (kg)	Embodied Energy (MJ)	Embodied Energy (kWh)
Frame Structures	Mild Steel	38.8	15	582	161.66
Solar Dish	Aluminium	190	10	1900	527.76
Mirrors	Glass	25.8	5	129	35.83
Boiler	Stainless Steel	56.7	5	283.5	78.74
Side glass	Brass	62	1.5	93	25.83
Connection Hose	Silicon	252	0.6	151.2	41.99
Top Cap	Mild Steel	38.8	3	116.4	32.33
Gasket	Rubber	67.5	0.5	33.75	9.374
Heat exchanger	Mild Steel	38.8	15	582	161.66
Connecting Pipes	PVC	70	0.5	35	9.72
Water Tanks	PVC	70	4	280	77.77
			Total	**4185.85**	**1162.70**

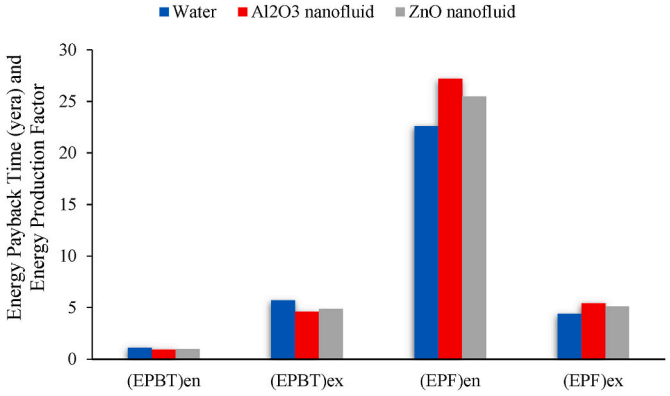


Fig. 6. Energy payback time and energy production factor for SWDS with and without nanofluid based on energy and exergy approach.

Table 7
Cost per liter values for SWDS with and without nanofluid.

Parameters (Unit)	SWDS without nanofluid	SWDS with Al_2O_3 nanofluid	SWDS with ZnO nanofluid
Total initial cost, TC (US$)	1199.85	1407.18	1510.86
Interest rate, i (%)	7	7	7
Life of system, n (Years)	30	30	30
Capital recovery factor, CRF	0.065	0.065	0.065
First annual cost, FAC (US$)	78.05	91.53	98.28
Shrinking fund factor, SFF	0.015	0.015	0.015
Annual Maintenance cost, AMC (US$)	11.7	13.73	14.74
Annual salvage value, ASV (US$)	3.611	4.23	4.54
Uniform annual cost, UAC (US$)	86.14	101.03	108.47
Annual Yield, P (L)	1676.7	2014.2	1887.3
Cost per liter, CPL (US $/L)	0.051	0.050	0.0547

Table 8
Exergoeconomic parameter for SWDS with and without nanofluid.

Parameters	Water	Al_2O_3 nanofluid	ZnO nanofluid
Annual $E_{n,o}$ (kWh)	1050.90	1264.47	1184.80
Annual $E_{x,o}$ (kWh)	204.12	252.45	237.87
Uniform annual cost, UAC (US$)	86.14	101.03	108.47
R_{en} (kWh/US$)	12.20	12.52	10.92
R_{ex} (kWh/US$)	2.37	2.50	2.19

Table 9
Environmental parameters for SWDS with and without nanofluid.

Parameters	Water	Al_2O_3 nanofluid	ZnO nanofluid
E_{emb} (kWh)	1162.70	1162.70	1162.70
Annual $E_{n,o}$ (kWh)	1050.90	1264.47	1184.80
Annual $E_{x,o}$ (kWh)	204.12	252.45	237.87
CO_2 Emissions (Tonnes)	2.33	2.33	2.33
CO_2 Mitigation (Tonnes), $\varnothing_{en},CO_2$	63.05	75.87	71.09
CO_2 Mitigation (Tonnes), $\varnothing_{ex},CO_2$	12.25	15.15	14.27
Net CO_2 Mitigation (Tonnes)	60.73	73.54	68.76
Total Carbon Credit Earned (US$)	1700.40	2059.20	1925.35

mitigations for SWDS without nanofluid, with Al_2O_3 nanofluid and ZnO nanofluid based on energetic approach were 63.05, 75.87 and 71.09, respectively. Whereas, corresponding values for CO_2 mitigations based on exergetic approach were 12.25, 15.15, and 14.27 tonnes, respectively. The amount of CO_2 mitigated by SWDS with Al_2O_3 nanofluid based on energetic approach was 20.33 and 6.27% more compared to SWDS using water and ZnO nanofluid, respectively. However, corresponding values of CO_2 mitigations based on exergetic approach were 23.67 and 6.16% more compared to SWDS using water and ZnO nanofluid, respectively. Total carbon credit earned for SWDS without nanofluid, Al_2O_3 nanofluid and ZnO nanofluid were 1700.40, 2059.20 and 1925.35 US$, respectively. It is observed that SWDS with Al_2O_3 nanofluid mitigated more CO_2 compared to water and ZnO nanofluid since yield and energy output obtained for SWDS with Al_2O_3 nanofluid was highest compared to SWDS without nanofluid and SWDS with ZnO nanofluid.

3.6. Enviroeconomic analysis

Enviroeconomic Parameters for SWDS with and without nanofluid based on energy and exergy approach are depicted in Table 10. Enviroeconomic parameters for SWDS with water, Al_2O_3 nanofluid and ZnO nanofluid based on energetic approach were 1765.4, 2124.8 and 1990.24 US$, respectively. Whereas, corresponding values of enviroeconomic parameters based on exergetic approach were 342.72, 423.92 and 399.56 US$, respectively. Highest carbon credits earned were 2124.08 and 423.92 for SWDS using Al_2O_3 nanofluid based on energetic and exergetic approaches, respectively.

Table 10
Enviroeconomic parameters for SWDS with and without nanofluid based on energy and exergy perspective.

Parameters	Water	Al_2O_3 nanofluid	ZnO nanofluid
CO_2 Mitigation (Tonnes), $\varnothing_{en},CO_2$	63.05	75.86	71.08
CO_2 Mitigation (Tonnes), $\varnothing_{ex},CO_2$	12.24	15.14	14.27
$ZCO_{2,en}$ (US$)	1765.4	2124.08	1990.24
$ZCO_{2,ex}$ (US$)	342.72	423.92	399.56

3.7. Scope of commercialization

Global demand for freshwater is increasing due to population growth and water scarcity. Conventional desalination systems are energy intensive. These techniques have a low efficiency and a high cost due to the high energy consumption. To address these problems, it is crucial to explore sustainable and eco-friendly desalination system. Hence solar dish integrated water desalination system was developed in present research work. Limited research has been conducted on the performance of water desalination systems equipped with solar dish collectors using various nanofluids. The novelty of present research work is integration of parabolic dish with automatic double-axis tracking mechanism and the use of nanofluids to enhance the productivity and efficiency of solar water desalination system. The present system is energy efficient, and environment friendly as solar energy is used as energy source. Present developed system will mitigate 2.41 tons of CO_2 per year.

Therefore, there is considerable potential to develop large-scale solar dish-integrated water desalination systems for commercial applications.

4. Conclusions

The present study investigated the performance of SWDS with and without nanofluids. A comparison was made between SWDS without using nanofluids and SWDS with different nanofluid (Al_2O_3 and ZnO) in order to determine the effects nanofluids on major parameters such as energy, exergy, freshwater productivity, energy-matrices, economic, exergoeconomic, environmental and enviroeconomic. The summarized findings of present study are listed as following:

- Annual freshwater productivity of SWDS with water, Al_2O_3 and ZnO nanofluid was 1676.7, 2014.2 and 1887.3 L, respectively.
- Annual freshwater productivity of SWDS was enhanced by 20.13% and 12.56% using Al_2O_3 and ZnO nanofluid, respectively compared to SWDS with water.
- Energy efficiency of SWDS for Al_2O_3, and ZnO nanofluids was 20.06 and 12.55% more than water whereas corresponding exergy efficiency was 22.72 and 15.90% more than that of water.
- EPBT values for SWDS with water, Al_2O_3 nanofluids and ZnO nanofluid based on energy approach were 1.10, 0.91 and 0.98 years, respectively. Whereas, based on exergy approach, the corresponding values were 5.69, 4.60 and 4.88 years, respectively.
- EPF for SWDS with Al_2O_3 nanofluids was enhanced by 20 and 6.79% than SWDS with water and ZnO nanofluids, respectively.
- Lowest CPL of US$0.050 was obtained for SWDS with Al_2O_3 nanofluids which was lower than the SWDS with water by 1.96%.
- Enerergoeconomic parameters for SWDS without nanofluid, Al_2O_3 nanofluid and ZnO nanofluid were 12.20, 12.52 and 10.92 kWh/US $, respectively. Whereas, corresponding values of exergoeconomic parameter were 2.37, 2.50 and 2.19 kWh/US$, respectively.
- Highest amount of net CO_2 mitigation was 72.54 tonnes of CO_2 for the SWDS with Al_2O_3 nanofluid.
- Enviro-economic results revealed the highest amount of carbon credit earned with SWDS using Al_2O_3 nanofluid based on energy and exergy approach was US$ 2124.08 and 423.92, respectively.

The solar water desalination system (SWDS) utilizing Al_2O_3 nanofluid demonstrated superior performance in energy and exergy output. Furthermore, the SWDS with Al_2O_3 nanofluid exhibited enhanced cost-effectiveness and environmental friendliness when compared to the SWDS employing water.

CRediT authorship contribution statement

Ravi Kant: Writing – original draft, Validation, Methodology, Formal analysis. **Anil Kumar:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Mukul Kumar Dubey:** Visualization, Resources, Investigation.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgment

Authors would like to thank Sardar Patel Renewable Energy Research Institute, Vallabh Vidyanagar, (India) and Centre for Energy and Environment, Delhi Technological University, Delhi (India) for providing the facilities and infrastructure to conduct the present study.

Nomenclature

A_c	Concentrator area (m^2)
E_{emb}	Embodied energy (kWh)
$E_{en,o}$	Thermal energy output (kWh)
E_r	Energy on receiver (kW)
E_u	Useful energy (kW)
Ex	Exergy (kWh)
i	Interest rate (%)
$\dot{m}$	Yield (kg/h)
n	Life of system (years)
R	Exergoeconomic parameter (kWh/US$)
S	Shape Factor
T	Temperature (°C)
φ	Environmental parameter (tonnes)
Z	Eenviroeconomic parameter (US$)

Subscript

a	Ambient
en	Energy

ex	Exergy
o	Useful
s	Sun
th	thermal

Greek letters

α	Absorbtivity of receiver
ρ	Reflectivity of concentrator
τ	Transmissivity
η	Efficiency

Abbreviations

ASV	Annual salvage value (US$)
CPL	Cost per liter (US$/L)
CRF	Capital recovery factor
EPBT	Energy payback time (years)
EPF	Energy production factor
FAC	First annual cost (US$)
SWDS	Solar dish integrated water desalination system
TC	Total initial cost (US$)
UAC	Uniform annual cost

References

Abbasi, N., Manesh, M.H.K., Yazdi, M., 2022. Dynamic analysis of a novel quadruple combined cycle based on integrated solar tower-gas turbine-supercritical CO2 and organic rankine cycles. J. Energy Resour. Technol. Trans. ASME 144, 1–10. https://doi.org/10.1115/1.4051743.

Abulkhair, H., Sharshir, S.W., Moustafa, E.B., Alsaiari, A., Moujdin, I.A., Elsheikh, A.H., 2023. Thermal performance enhancement of a modified pyramid distiller using different modifications with low-cost materials. Sustain. Energy Technol. Assessments 57, 103191. https://doi.org/10.1016/j.seta.2023.103191.

Ahmad, A., Prakash, O., Kumar, A., Hussain, M.S., 2023. Drying kinetics and performance analysis of thermal storage-based hybrid greenhouse dryer for uniform drying of tomato flakes. J. Therm. Sci. Eng. Appl. 15, 1–10. https://doi.org/10.1115/1.4056242.

Alhawsawi, A., Zayed, M.E., Moustafa, E., Banoqitah, E., Elsheikh, A.H., 2023. Hybridizing solar dish Stirling power system with single-effect desalination for sustainable electricity and freshwater co-generation: mathematical modeling and performance evaluation. Case Stud. Therm. Eng. 45, 102997 https://doi.org/10.1016/j.csite.2023.102997.

Chaouchi, B., Zrelli, A., Gabsi, S., 2007. Desalination of brackish water by means of a parabolic solar concentrator. Desalination 217, 118–126. https://doi.org/10.1016/j.desal.2007.02.009.

Colakoglu, M., Durmayaz, A., 2022. Multiobjective optimization of a novel solar tower-based gas turbine-driven multi-generation plant with energy, exergy, economic, and environmental impact analysis. J. Energy Resour. Technol. 144, 1–14. https://doi.org/10.1115/1.4051805.

Commitments made under CoP 26 [WWW Document], n.d. URL https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1882840#:~:text=(As per the updated NDC,the help of transfer of).

Dafle, V.R., Shinde, N.N., 2012. Design, development & performance evaluation of concentrating monoaxial scheffler technology for water heating and low temperature industrial steam application. Int. J. Eng. Res. Appl. 2, 848–852.

El-Sebaii, A.A., El-Naggar, M., 2017. Year round performance and cost analysis of a finned single basin solar still. Appl. Therm. Eng. 110, 787–794. https://doi.org/10.1016/j.applthermaleng.2016.08.215.

Elashmawy, M., 2020. Improving the performance of a parabolic concentrator solar tracking-tubular solar still (PCST-TSS) using gravel as a sensible heat storage material. Desalination 473, 114182. https://doi.org/10.1016/j.desal.2019.114182.

Elashmawy, M., Alshammari, F., 2020. Atmospheric water harvesting from low humid regions using tubular solar still powered by a parabolic concentrator system. J. Clean. Prod. 256, 120329 https://doi.org/10.1016/j.jclepro.2020.120329.

Elbar, A.R.A., Yousef, M.S., Hassan, H., 2019. Energy, exergy, exergoeconomic and enviroeconomic (4E) evaluation of a new integration of solar still with photovoltaic panel. J. Clean. Prod. 233, 665–680. https://doi.org/10.1016/j.jclepro.2019.06.111.

Fang, S., Mu, L., Tu, W., 2021a. Application design and assessment of a novel small-decentralized solar distillation device based on energy, exergy, exergoeconomic, and enviroeconomic parameters. Renew. Energy 164, 1350–1363. https://doi.org/10.1016/j.renene.2020.09.075.

Fang, S., Mu, L., Tu, W., 2021b. Application design and assessment of a novel small-decentralized solar distillation device based on energy, exergy, exergoeconomic, and enviroeconomic parameters. Renew. Energy 164, 1350–1363. https://doi.org/10.1016/j.renene.2020.09.075.

Farooq, M., Farhan, M., Ahmad, G., Tahir, Z. ul R., Usman, M., Sultan, M., Saad Hanif, M., Imran, M., Anwar, S., El-Sherbeeny, A.M., Ali Shakir, M., 2022. Thermal performance enhancement of nanofluids based parabolic trough solar collector (NPTSC) for sustainable environment. Alex. Eng. J. 61, 8943–8953. https://doi.org/10.1016/j.aej.2022.02.029.

Hameed, H.G., Neamah Diabil, H.A., Al-fahham, M.A., 2023. Performance study on a solar concentrator system for water distillation using different water nanofluids. Heliyon 9, e16535. https://doi.org/10.1016/j.heliyon.2023.e16535.

Kant, R., Kumar, A., 2023. Energy-economic and exergy-environment performance evaluation of solar energy integrated essential oil extraction system. Sol. Energy 265, 112101. https://doi.org/10.1016/j.solener.2023.112101.

Kant, R., Kumar, A., 2022a. Review on essential oil extraction from aromatic and medicinal plants : techniques , performance and economic analysis. Sustain. Chem. Pharm. 30, 100829 https://doi.org/10.1016/j.scp.2022.100829.

Kant, R., Kumar, A., 2022b. Thermodynamic analysis of solar assisted steam distillation system for peppermint oil extraction. J. Food Process. Eng. e14228, 1–12. https://doi.org/10.1111/jfpe.14228.

Kumar, A., Kant, R., Samsher, 2021. Review on spray-assisted solar desalination: concept, performance and modeling. Arabian J. Sci. Eng. 46, 11521–11541. https://doi.org/10.1007/s13369-021-05846-7.

Kumar, S., 2013. Thermal-economic analysis of a hybrid photovoltaic thermal (PVT) active solar distillation system: role of carbon credit. Urban Clim. 5, 112–124. https://doi.org/10.1016/j.uclim.2013.07.001.

Kumar, V., Sahoo, R.R., 2023. Energy-economic and exergy-environment performance evaluation of compact heat exchanger with turbulator passive inserts using THDNF. J. Therm. Sci. Eng. Appl. 15, 1–14. https://doi.org/10.1115/1.4056240.

Lashari, A.A., Shaikh, P.H., Leghari, Z.H., Soomro, M.I., Memon, Z.A., Uqaili, M.A., 2021. The performance prediction and techno-economic analyses of a stand-alone parabolic solar dish/stirling system, for Jamshoro, Pakistan. Clean. Eng. Technol. 2, 100064 https://doi.org/10.1016/j.clet.2021.100064.

Malwad, D., Tungikar, V., 2020. Experimental performance analysis of an improved receiver for Scheffler solar concentrator. SN Appl. Sci. 2, 1–14. https://doi.org/10.1007/s42452-020-03848-y.

Mawire, A., Taole, S.H., 2014. Experimental energy and exergy performance of a solar receiver for a domestic parabolic dish concentrator for teaching purposes. Energy Sustain. Dev. 19, 162–169. https://doi.org/10.1016/j.esd.2014.01.004.

Modi, K.V., Gamit, A.R., 2022. Investigation on performance of square pyramid solar still using nanofluid and thermal energy storage material: an experimental and theoretical study. J. Clean. Prod. 381, 135115 https://doi.org/10.1016/j.jclepro.2022.135115.

Mohanraj, M., Karthick, L., Dhivagar, R., 2021. Performance and economic analysis of a heat pump water heater assisted regenerative solar still using latent heat storage. Appl. Therm. Eng. 196, 117263 https://doi.org/10.1016/j.applthermaleng.2021.117263.

Ouyang, P., Xu, Y.-P., Qi, L.-Y., Xing, S.-M., Fooladi, H., 2021. Comprehensive evaluation of flat plate and parabolic dish solar collectors' performance using different operating fluids and MWCNT nanofluid in different climatic conditions. Energy Rep. 7, 2436–2451. https://doi.org/10.1016/j.egyr.2021.04.046.

Rani, A., Kant, R., Suresh, S., Kumar, A., 2021. Experimental investigation on thermal behavior of hybrid single slope solar still. J. Therm. Eng. 7, 677–689. https://doi.org/10.18186/THERMAL.889191.

Ranjbar, B., Mehrpooya, M., Marefati, M., 2021. Parametric design and performance evaluation of a novel solar assisted thermionic generator and thermoelectric device hybrid system. Renew. Energy 164, 194–210. https://doi.org/10.1016/j.renene.2020.09.068.

Reddy, K.S., Sharon, H., Krithika, D., Philip, L., 2018. Performance, water quality and enviro-economic investigations on solar distillation treatment of reverse osmosis

reject and sewage water. Sol. Energy 173, 160–172. https://doi.org/10.1016/j.solener.2018.07.033.

Saxena, A., Cuce, E., Kabeel, A.E., Abdelgaied, M., Goel, V., 2022. A thermodynamic review on solar stills. Sol. Energy 237, 377–413. https://doi.org/10.1016/j.solener.2022.04.001.

Sharon, H., Reddy, K.S., Krithika, D., Philip, L., 2020. Viability assessment of solar distillation for desalination in coastal locations of Indian sub-continent – thermodynamic, condensate quality and enviro-economic aspects. Sol. Energy 197, 84–98. https://doi.org/10.1016/j.solener.2019.12.080.

Shatar, N.M., Sabri, M.F.M., Salleh, M.F.M., Ani, M.H., 2023. Energy, exergy, economic, environmental analysis for solar still using partially coated condensing cover with thermoelectric cover cooling. J. Clean. Prod. 387, 135833 https://doi.org/10.1016/j.jclepro.2022.135833.

Singh, D.B., Kumar, N., Kumar, S., Dwivedi, V.K., Yadav, J.K., Tiwari, G.N., 2018. Enhancement in exergoeconomic and enviroeconomic parameters for single slope solar still by incorporating N identical partially covered photovoltaic collectors. J. Sol. Energy Eng. 140 https://doi.org/10.1115/1.4039632.

Stefanovic, V.P., Pavlovic, S.R., Bellos, E., Tzivanidis, C., 2018. A detailed parametric analysis of a solar dish collector. Sustain. Energy Technol. Assessments 25, 99–110. https://doi.org/10.1016/j.seta.2017.12.005.

Vijayan, S., Arjunan, T.V., Kumar, A., 2020. Exergo-environmental analysis of an indirect forced convection solar dryer for drying bitter gourd slices. Renew. Energy 146, 2210–2223. https://doi.org/10.1016/j.renene.2019.08.066.

Yusof, M.F., Zainol, M.R.R.M.A., Sandu, A.V., Riahi, A., Zakaria, N.A., Shaharuddin, S., Aziz, M.S.A., Noor, N.M., Vizureanu, P., Zawawi, M.H., Ikhsan, J., 2022. Clean water production enhancement through the integration of small-scale solar stills with solar dish concentrators (SDCs)—a review. Sustain. Times 14. https://doi.org/10.3390/su14095442.

Zayed, M.E., Zhao, J., Elsheikh, A.H., Li, W., Elaziz, M.A., 2020. Optimal design parameters and performance optimization of thermodynamically balanced dish/Stirling concentrated solar power system using multi-objective particle swarm optimization. Appl. Therm. Eng. 178, 115539 https://doi.org/10.1016/j.applthermaleng.2020.115539.

Signal, Image and Video Processing
https://doi.org/10.1007/s11760-024-03161-w

ORIGINAL PAPER

ERCU-Net: segmentation of road potholes using enhanced residual convolutional block based on U-Net for ADAS

Ruchi Tripathi[1] · S. Indu[1] · Rohit Kumar[1]

Received: 21 January 2024 / Revised: 2 March 2024 / Accepted: 17 March 2024
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

Abstract
Potholes are one of the most common forms of road damage, and they can have a negative impact on vehicle conditions, comfort while driving, and road safety. For the Advanced Driver-Assistance System (ADAS) to ensure safe driving, accurate segmentation of road potholes is a crucial feature. The main issue with the previous state-of-the-art (SOTA) techniques is that they are not able to distinguish the visual appearance and margins of potholes close to the road areas. In order to achieve superior segmentation results and proper pothole margins, we have proposed an ERCU-Net model based on U-Net for pothole segmentation in this paper. The network consists of a novel residual convolutional block structure and employs a dropout layer in the residual convolutional block to minimize over-fitting concerns. A region-based IoU loss is proposed to reduce the regions of overlap between the anticipated segmentation and the actual segmentation. The performance of our proposed ERCU-Net model is compared with five SOTA CNNs in this paper. The result is also validated using Grad-CAM. The simulation results indicate that our proposed ERCU-Net model obtains the highest mean IoU 88.45% and the best overall pothole segmentation accuracy 98.55% in comparison to the existing state-of-the-art models.

Keywords Deep learning · Semantic segmentation · Road potholes · UNet · Advanced driver assistance system

1 Introduction

A Pothole is a depression on a road surface, which emerges as a result of the presence of both water and traffic [1]. Road users may be put in danger by the presence of potholes. In addition to being inconvenient, potholes on the road pose a safety risk because they can negatively impact driving comfort, the condition of vehicles, and traffic safety [2]. Vehicle traffic can cause the upper surface material of the road to move out and it generates small weak spots on the road. As time passes these spots worsen due to the continuous movement of vehicles over that damaged surface and cause to emergence of new potholes [3]. Road inspection is frequently done by trained maintenance personnel who drive along the roads for visual inspection of the road surface to observe and report any discovered road issues. Automatic monitoring of these tasks can improve the safety of staff and also provide an effective solution for road maintenance hazards. Furthermore, the development of ADAS systems and significant investment in autonomous driving technologies have made it possible to integrate several sensors into cars and enable multi-modal perception of the environment [4].

Among other forms of road hazards (cracks, rutting, worn markings, etc.), potholes are the most dangerous hazards on roads all around the world, causing drivers discomfort, car damage, and expensive maintenance expenditures for both public and private roadways. Vehicles are equipped with dampening systems that can lower vibrations to diminish the detrimental effects of road potholes [5]. Many efficient approaches have been put out recently using a variety of sensors, including, visual cameras [6], thermal imaging [7], RGB-D imaging [8], and Lidar [9]. In the field of visual sensors, contemporary road-pothole segmentation approaches have taken RGB images, disparity images, and depth images into consideration. The benefits and drawbacks of each modality of data are unique [10].

In recent years, pothole segmentation tasks have shown encouraging results when using deep learning-based approaches.

✉ S. Indu
s.indu@dtu.ac.in

Ruchi Tripathi
ruchitripathi05@gmail.com

Rohit Kumar
rohit.kumar@dtu.ac.in

[1] Delhi Technological University, Delhi, India

Published online: 24 April 2024

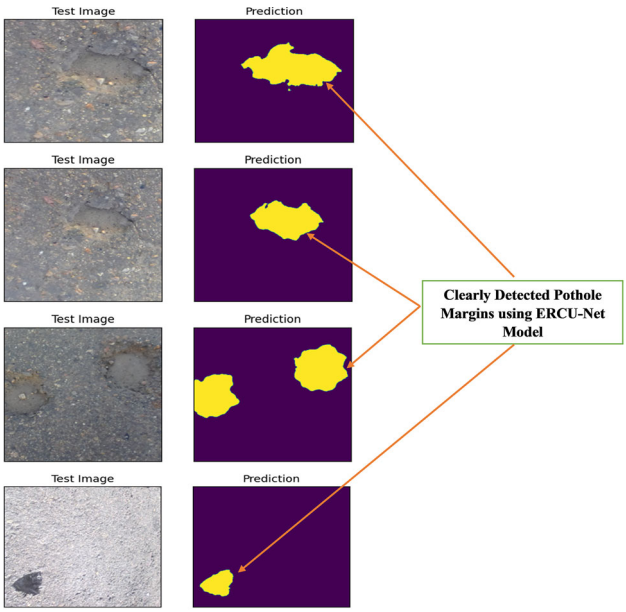


Fig. 1 Clearly visible pothole margins using proposed ERCU-Net

The fact that current techniques do not work well near pothole margins is a significant problem. We hypothesize that the causes involve just convolutional layers to take out the features. We could better protect edge information with our Enhanced Residual Convolutional block based on U-Net (ERCU-Net). We have proposed an ERCU-Net model for pothole segmentation by leveraging the advantage of residual learning with UNet. Additionally, adding residual modules in UNet boosts gradient propagation and enhances the network's capacity for generalization. Sample pothole test images with predicted output using the ERCU-Net model are shown in Fig. 1. where we can see visible pothole margins using the proposed model.

We believe that combining residual convolutional block with UNet will yield state-of-the-art results and improve the pothole segmentation impact. Our contributions can be summarized as follows:

1. We have proposed a residual convolutional block-based U-shaped encoder-decoder semantic segmentation network, replacing the convolutional block of UNet with the residual convolutional block, which led to superior segmentation effects than the original UNet.
2. In the proposed model, residual-based connections with a 1x1 convolutional block enhance the performance of the segmentation model, improve gradient flow during training, and reduce the problem of disappearing gradients, enabling the network to learn more effectively.
3. Before training the dataset, we introduced the contrast enhancement algorithm CLAHE to enhance its appearance for visual interpretation and boost the efficiency of the pothole segmentation task.
4. We have introduced a residual convolutional block for our network verified the effectiveness of different activation layers in the residual block through ablation studies and validated our result through Grad CAM.

The rest of this paper is organized as follows: A few cutting-edge existing methods for segmenting potholes are summarized in Sect. 2. The proposed model and its analysis are described in Sect. 3. In Sect. 4, the datasets, experiments, and results are thoroughly described. The conclusion and possible future directions are discussed in Sect. 5.

2 Related work

Review of pothole segmentation and detection work is presented in this section. Road pothole detection can be classified into three categories—Classical 2D image processing-based method, 3D Point cloud modeling and segmentation-based method, and machine/ deep learning-based methods [11]. 2-D image processing-based techniques are quite sensitive to environmental influences. The 3D road point cloud modeling and segmentation processes frequently interpolate a road surface point cloud into an explicit mathematical model, like a quadratic surface [12]. A 3D reconstruction-based method predicts the shape of potholes and determines their volume using stereo vision technology and it is the most accurate [13], but this method includes the high cost of pothole detection compared to alternative approaches and the difficulty of detection when a pothole is filled with water or dirt [14]. Two deep learning-based pothole detection techniques were proposed by Dhiman and Klette [15], who used stereo vision to estimate the 3-D point cloud information of the environment.

Machine/deep learning-based approaches for end-to-end road pothole detection train CNNs on annotated image data [6]. The first end-to-end single-modal CNN created for semantic segmentation was the fully convolutional network (FCN) [16]. Convolutions are created from all fully connected (FC) layers. A downsampling path and an upsampling path often make up an FCN. Convolutions, rectified linear units (ReLUs), and max pooling layers make up the converging path's usual CNN design [12]. Convolutions and concatenations are used in a series of convolutions and concatenations along the expanded path to integrate the desired visual attributes and spatial information [17]. In order to extract context information for improved segmentation efficiency, PSPNet [18], DeepLabv3+ [19], and DenseASPP [20] use a pyramid pooling module. Moreover, GSCNN [21] uses a two-branch structure made up of a shape branch and a

regular branch, which can significantly enhance the semantic predictions on the boundaries [22].

As an alternative to the single-modal networks stated above, numerous data-fusion networks have also been proposed to increase the accuracy of semantic segmentation by extracting and fusing the features from multiple visual information modalities [23, 24]. To fuse the feature maps obtained from the RGB and depth branches, for example, FuseNet [25] and depth-aware CNN [26] leverage the well-known encoder-decoder architecture but apply distinct procedures. Additionally, by combining the features acquired from RGB and thermal images, RTFNet [27] was created to enhance semantic segmentation performance. Moreover, an element-wise addition fusion approach and an encoder-decoder architecture are used. Pan et al. [28] combined multispectral images taken by an unmanned aerial vehicle to detect potholes and cracks in asphalt roads. Zhen et al. [29] have proposed a multimodal Attention Fusion network MAFNet for the segmentation of potholes. However, compared to the ground truth, in the case of MAFNet, several edges still have some incorrectly detected pixels. Gao et al. [30] proposed a new UNet based on multi-residual attention (MRA-UNet), which is used in road crack segmentation. Guo et al. [31] proposed a novel Crack Transformer (CT) model that models the long-range dependency and noisy pavement crack pixels at linear computational complexity and with high accuracy and efficiency. The model makes use of the Swin-Transformer and SegFormer with the hierarchical architecture as the Encoder and a simplified Decoder structure in SegFormer. For reliable and accurate pixel-level pavement crack identification, Guo et al. [32] suggested a Transformer-based semantic segmentation network that combines the UperNet with the attention module as the Decoder and the Swin Transformer as the Encoder. Dong et al. [33] has presented an enhanced sliding window algorithm that combines the YOLOv5 processing method for processing the two-dimensional images generated from point cloud data for automatic pavement crack detection with the random sample consensus (RANSAC) technique for directly extracting the crack information from point clouds. Zhang et al. [34] came up with a novel early warning system that combines event detection with data on the severity of foreign objects. To be more precise, the created approach can extract the track area using the upgraded railway identification method and recognize objects using the cutting-edge YOLOv5 model.

3 Proposed methodology

In the proposed ERCU-Net model residual connections are used with a shortcut path containing the 1×1 convolution with batch normalization and the activation function. The preceding feature map's depth is increased by applying the 1×1 convolution to the shortcut connectivity in the residual unit. The symmetric architecture of ERCU-Net includes expanding and contracting paths. The expanding path reconstructs spatial information, and the contracting path acquires context and learns high-level features. This symmetry makes it easier to encode and decode visual features, such as the pothole's edge, creating clearer boundaries.

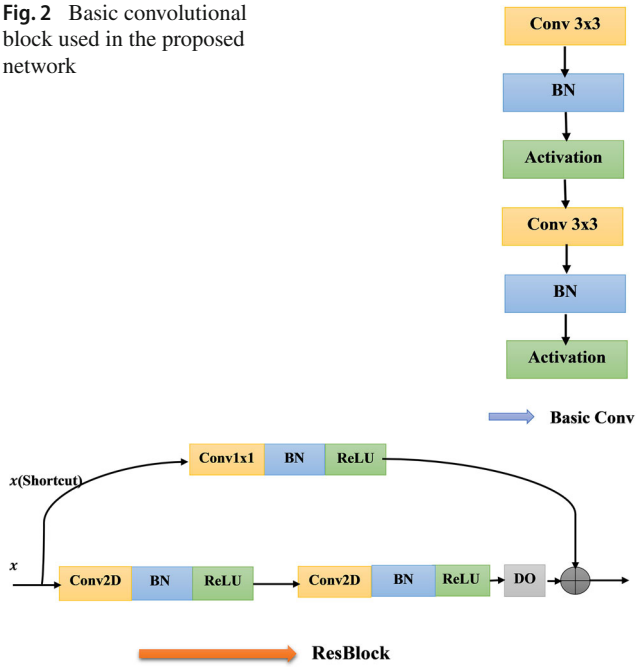


Fig. 2 Basic convolutional block used in the proposed network

Fig. 3 Proposed residual convolutional block with modified shortcut path

3.1 Proposed network architecture

The U-shaped topology of U-Net can well solve image segmentation tasks. However, due to the limitations of traditional convolution operations, U-Net cannot realize global semantic information [35]. Inspired by [36], a semantic segmentation neural network, which incorporates the benefits of residual learning and U-Net, is proposed to address the problem of correct detection of pixels at the edges of pothole areas. The proposed model employs skip connections to link the networks for the encoder and decoder. The residual convolution block uses a residual connection containing 1×1 convolution in its shortcut path, which allows the network to learn residual features and helps to improve the pothole segmentation performance by preserving more information from the input image.

Mathematically, a residual connection can be defined as follows:

$$H(x) = F(x) + x \tag{1}$$

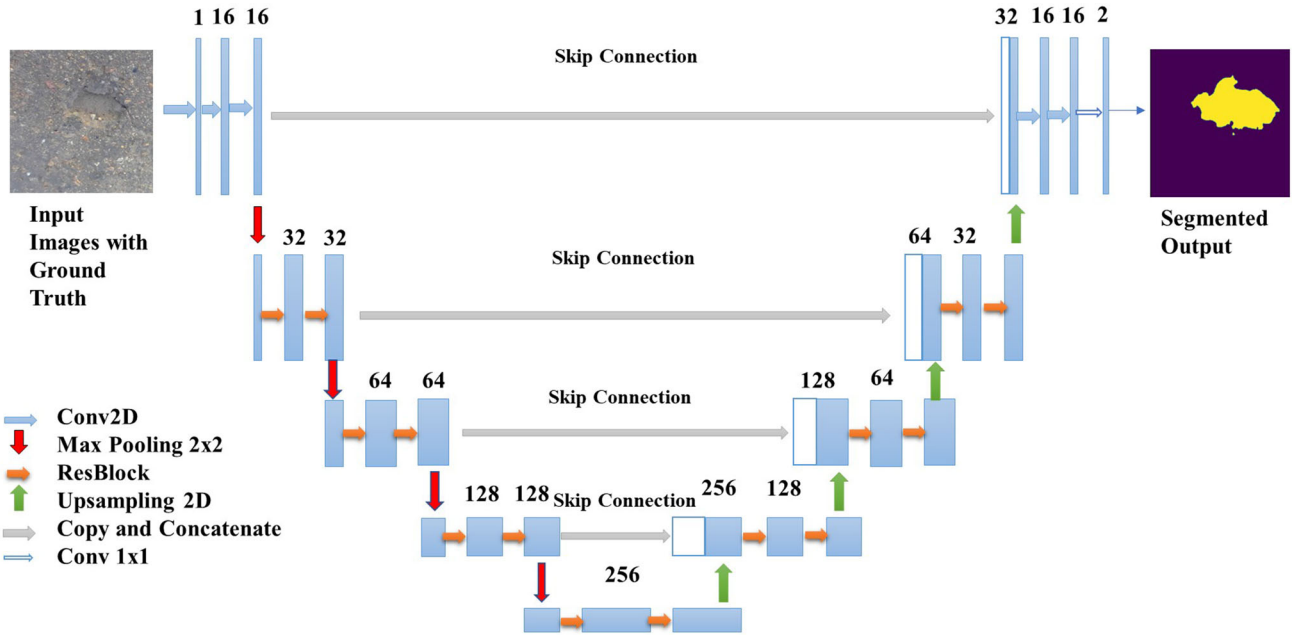


Fig. 4 Schematic of the proposed ERCU-Net architecture

where x is the input to the layer, $F(x)$ is the function of the layer, and $H(x)$ is the output of the layer. The structure of our basic convolutional block and residual convolutional block used in our proposed network is shown in Figs. 2 and 3 respectively.

The complete proposed network architecture is shown in Fig. 4. The Input RGB images of the pothole with ground truth are fed to the preprocessing block and preprocessed using the CLAHE algorithm. Certain issues, like contrast variability, feature extraction, and computing efficiency, are meant to be addressed using this algorithm. We have included pre-processed images using the CLAHE algorithm for more clarity in Fig. 5, which shows the prediction of CLAHE pre-processed images for our proposed model. After this preprocessing, patch extraction takes place, which splits the images into patches. After training, the same steps are followed during the testing stage. The overall proposed methodology is shown in Fig. 6. Skip connections and residual convolutional blocks in an ERCU-Net enable the transmission of both high-level and low-level information, as a result of testing the proposed model, we obtain a clearer pothole margin. In the proposed work we have trained three versions of the U-Net model—Simple U-Net, Attention U-Net, and Residual Attention U-Net along with the proposed ERCU-Net.

Algorithm 1 depicts the training procedure for the proposed ERCU-Net model. The algorithm iterates the training epochs, modifies the model's parameters, and evaluates the model's performance on the validation set.

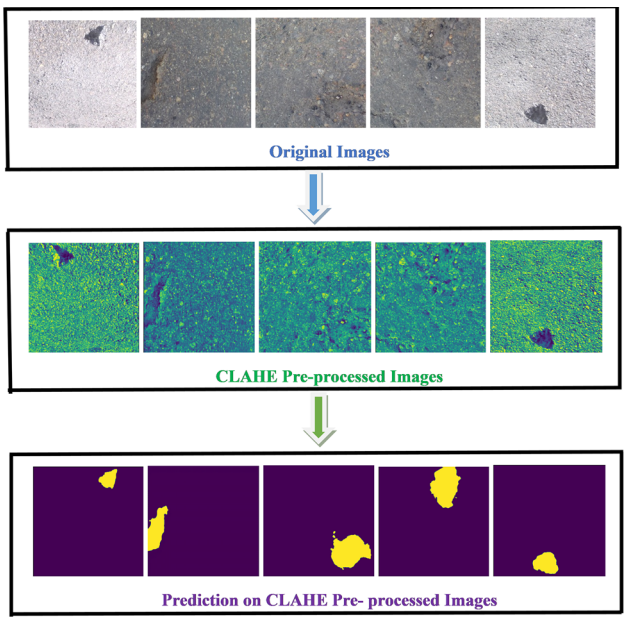


Fig. 5 Prediction on CLAHE pre-processed images

3.2 Encoder

We have utilized a modified U-Net encoder block that consists of two 3×3 convolutions applied repeatedly, each followed by a rectified linear unit (ReLU), to create a single 3×3 convolution, following this convolutional block, there is a 2×2 max pooling operation with stride 2 for input stage downsampling. The proposed residual convolutional block is used in the rest stages of the downsampling process.The $1 \times$

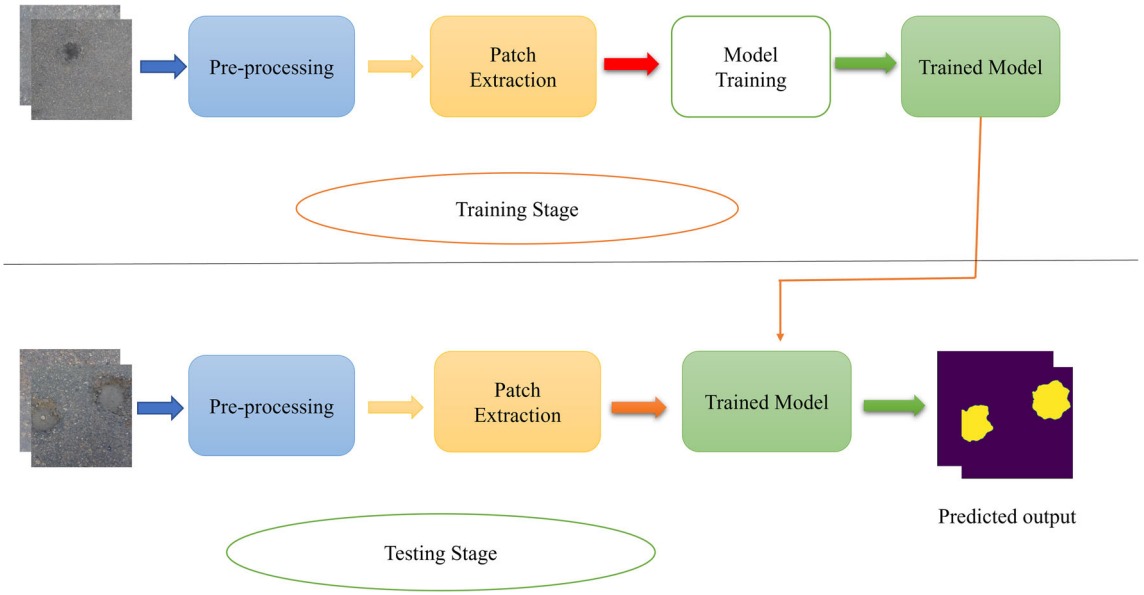


Fig. 6 Proposed methodology

Algorithm 1: Training Algorithm of ERCU-Net Model

Input: Pre-processed training dataset: $(x_i, y_i)_{i=1}^{N}$, where x_i - input image and y_i - corresponding ground truth mask for potholes
Output: Trained ERCU-Net model
Initialize ERCU-Net model
Define loss function
Choose adam optimizer with learning rate $\eta = 1e-3$
Split dataset $\leftarrow$ Train-val(70:30)
for *each training epoch* **do**
  for *each mini-batch in the training set* **do**
    Sample a mini-batch of input images (x_j) and corresponding masks (y_j)
    Perform forward pass through the ERCU-Net model to obtain predicted masks $\hat{y}_j$
    Compute the loss $\mathcal{L}_j$
    Compute the gradients $\frac{\partial \mathcal{L}_j}{\partial \text{model parameters}}$
    Update model parameters: model parameters $\leftarrow$ model parameters $- \eta \cdot \frac{\partial \mathcal{L}_j}{\partial \text{model parameters}}$
  end
  Evaluate model performance
end

1 convolution layer in the shortcut path architecture provides dimensionality reduction by lowering the number of filters while keeping crucial, feature-related information.

3.3 Decoder

The decoder has five steps. The feature map resolution doubles for every step. In terms of resolution, the output matches the input exactly. The final step adds a sigmoid layer to determine the probability of each pixel falling into the background or pothole class. The detailed configuration of the encoder and decoder blocks used in the proposed model is shown in Fig. 7.

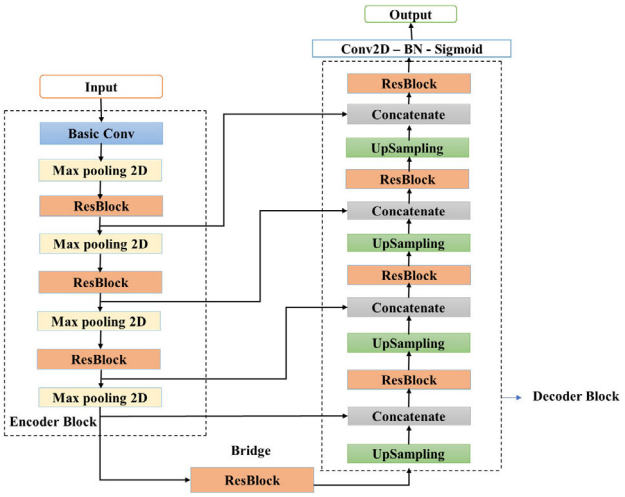


Fig. 7 Detailed configuration of encoder and decoder block

4 Experimental results and discussions

4.1 Dataset-geometric transformation augmentation

For the experiments, we employ the freely available Pothole-600 dataset [22]. In total, there are 600 RGB images of 400 x 400 resolution in the dataset.

We employ the method of [22] for dividing images, 240 RGB images are included in the training set and there are 180 images in each of the testing and validation sets. Flipping, rotating, and cropping the RGB images in the initial training set allows us to create an augmented training set. The enlarged training set has a total of 960 images, which is 4×240. The total number of RGB images for training is $240 + 960 = 1200$ [29].

4.2 Training detail

The proposed architecture was implemented with the use of the Keras framework. Our network is trained and tested using Google Colab, which has a Tesla K80 accelerator, 13 GB of RAM, and 12 GB of GDDR5 VRAM.

We used residual convolutional blocks in the ERCU-Net model and trained it from scratch. To avoid overfitting during a neural network model's training, we choose a dropout rate of 20 percent in each stage of model architecture in downsampling as well as upsampling path. With an initial learning rate of 1e−3, we train the network for 150 epochs using the adaptive moment estimation (Adam) optimization function.

4.3 Evaluation metrics

To assess the efficacy of semantic segmentation, we use two quantitative metrics: intersection over union (IoU) and accuracy. The following is how they are determined:

$$\text{Accuracy} = \frac{\text{True Positives}}{\text{True Positives} + \text{True Negatives}} \tag{2}$$

$$\text{IoU} = \frac{\text{True Positive}}{\text{True Positives} + \text{False Negatives} + \text{False Positives}} \tag{3}$$

In the ground-truth images of the dataset, there are two classes: pothole and unlabeled background (Fig. 8). The average values, which are designated as Average Accuracy and mIoU, are the mean values across the two classes [29].

4.4 Ablation analysis

4.4.1 Ablation on the effect of activation layer in residual block

In the ablation study, we tested the network with different activation layers-Softplus, Softsign, ReLU, and GeLU activation layer, in the residual convolutional block shown in Fig. 3, and discovered that using the ReLU activation layer in the residual convolutional block produces a better mean iou and accuracy than the other three. The ablation using different activation layers in the residual convolutional block is shown in Table 1.

4.4.2 Ablation study for selection of baseline

Our decision to adopt the residual-based UNet model as our baseline for the pothole segmentation problem was inspired by the model's underutilized potential in the field of pothole segmentation. The performance of various models, i.e., Simple UNet, Attention UNet, Residual Attention UNet, and ERCU-Net is examined in terms of accuracy and trainable parameters as demonstrated in Fig. 9.

The performance of various models is also examined in terms of inference speed as shown in Fig. 10.

4.5 Comparative evaluation with other cutting-edge models

We compared our proposed ERCU-Net model with SegNet [37], DeepLabv3+ [19], ESPNet [38], GSCNN [21] and MAFNet [29]. We have also made a comparison of different Unet variants on the pothole-600 dataset. The effectiveness of our proposed model is shown by contrasting the network's performance with that of other cutting-edge models. Table 2 compares the results of our model with other models that are considered to be state-of-the-art. ERCU-Net outperforms other cutting-edge models on the same dataset.

Figure 11 presents validation loss for different models on the Pothole-600 augmented dataset.

4.5.1 Qualitative demonstrations

Figure 8 presents the qualitative demonstrations. Compared to the others, our proposed ERCU-Net performs pothole segmentation with more accuracy and the best IoU score. Our network performs substantially better on pothole edges than the other networks. It improves the effectiveness of road maintenance procedures, promotes data-driven decision-making, and allows for efficient communication regarding the state of the roads and the need for repairs.

4.6 Grad-CAM visualization

Utilizing the Gradient-weighted Class Activation Mapping (Grad-CAM) technique, the heatmap is generated. The pothole regions are displayed as a heatmap by Grad-CAM, which uses the target's gradients to create the coarse localization map. The heatmap is created by utilizing the last convolution layer. Figure 12 demonstrates Grad-CAM visualization of the proposed ERCU-Net model.

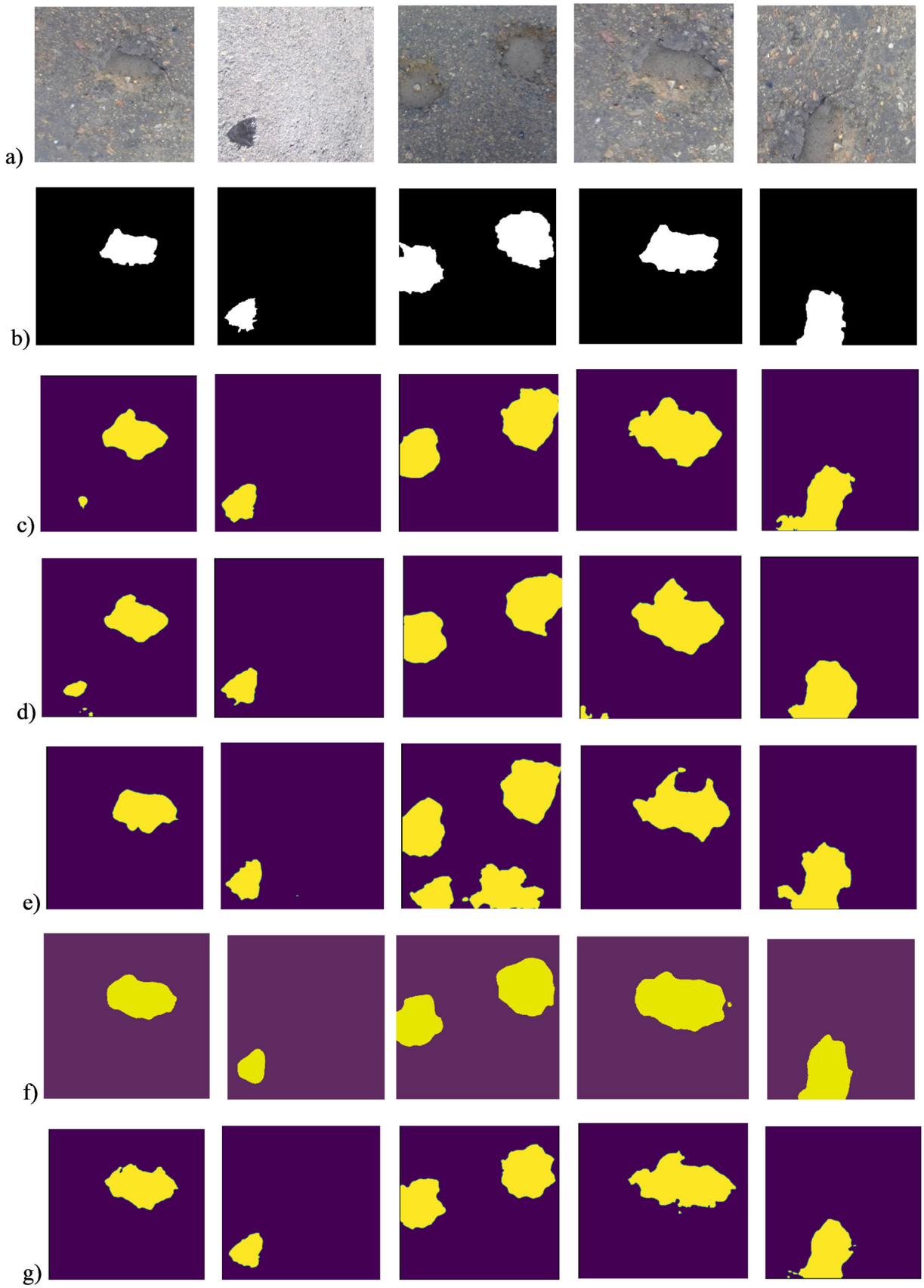


Fig. 8 Sample qualitative demonstrations for the networks on the Pothole-600 test dataset. **a** RGB images, **b** ground truth, **c** simple U-Net, **d**, attention U-Net, **e** residual attention U-Net, **f** MAFNet, **g** ERCU-Net

Table 1 Effect of activation layer in the residual convolutional block on mean IoU and average accuracy

Activation layer	mIoU	Average accuracy
Softplus	0.71	0.87
Softsign	0.86	0.97
GeLU	0.85	0.96
ReLU	0.88	0.99

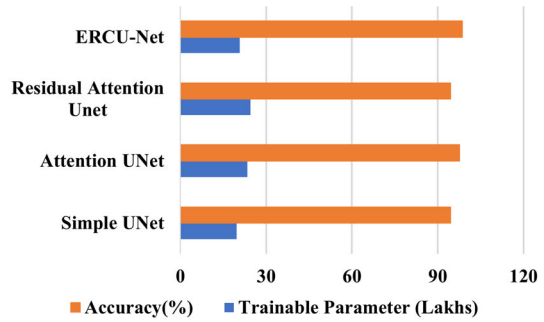


Fig. 9 Accuracy versus trainable parameter

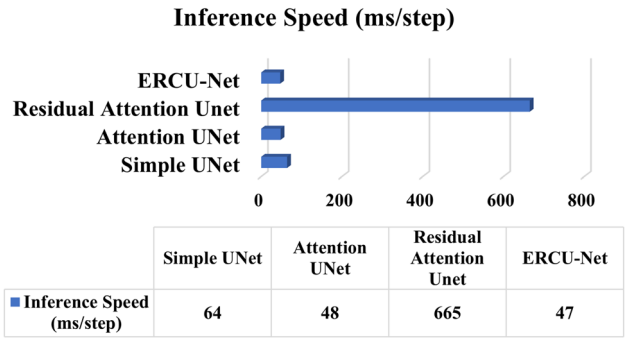


Fig. 10 Inference speed of trained models on pothole-600 dataset

5 Conclusion and future scope

In conclusion, we have proposed the use of the ERCU-Net architecture for pothole segmentation and demonstrated its effectiveness on the Pothole-600 dataset of road images. Our findings demonstrate that the ERCU-Net beats other cutting-edge techniques in terms of average accuracy and mIoU, making it a promising method for tasks requiring pothole segmentation. But there are still certain limitations in place. In situations where the training data distribution differs greatly from real-world circumstances, the model might not be able to generalize successfully. Additionally, a variety of unfavorable factors in real-world scenarios, such as different textures, patterns, marks, debris, etc., can produce background noise and clutter that hamper the segmentation of potholes and reduce the accuracy of detection. There is still scope for further research in this area. One potential direction is to investigate the real-time inference speed of the model on different datasets. Another direction is to explore multimodal fusion approaches for improving segmentation accuracy and robustness to environmental variability in pothole segmentation tasks.

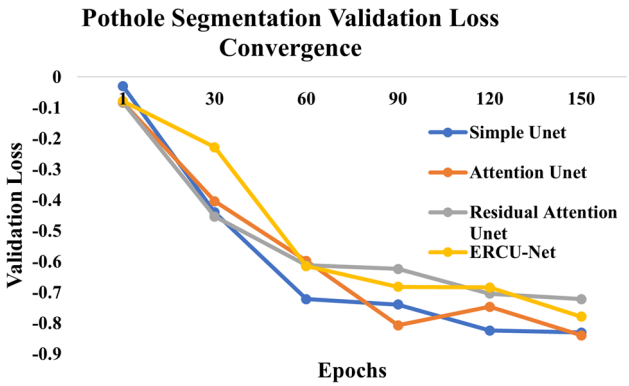


Fig. 11 Pothole segmentation validation loss convergence

Table 2 Comparative result analysis % on the pothole-600 testing dataset

Model name	Mean IoU	Average accuracy
SegNet [37]	30.71	89.35
DeepLabv3+ [19]	55.31	97.39
ESPNet [38]	49.33	96.13
GSCNN [21]	52.71	97.71
MAFNet [29]	87.29	97.62
Simple Unet	86.72	94.49
Attention Unet	86.76	97.62
Residual Attention Unet	82.63	94.54
ERCU-Net	**88.45**	**98.55**

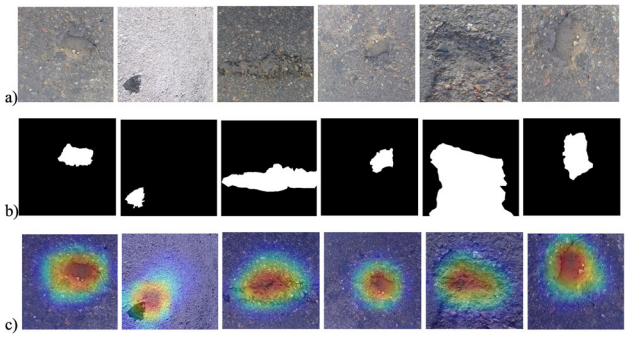


Fig. 12 Grad-CAM visualization of proposed ERCU-Net Model, **a** RGB images, **b** ground truth images, **c** corresponding heatmap images

Author Contributions The authors confirm their contribution to the paper as follows: Study conception, Interpretation of results, Review, Supervision: Prof.S. Indu, Dr. Rohit Kumar; Data collection, data processing, analysis, and interpretation of results, draft manuscript preparation: Ruchi. All authors reviewed the results and approved the final version of the manuscript.

Funding There was no specific grant for this research from any funding organization in the public, private, or nonprofit sectors.

Data availability The data that support the findings of this research are openly available at https://sites.google.com/view/pothole-600/dataset.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethical approval Not applicable.

Consent for publication Not applicable.

References

1. Fan, R., Ozgunalp, U., Wang, Y., Liu, M., Pitas, I.: Rethinking road surface 3-d reconstruction and pothole detection: from perspective transformation to disparity map segmentation. IEEE Trans. Cybern. **52**(7), 5799–5808 (2021)
2. Fan, R., Ozgunalp, U., Hosking, B., Liu, M., Pitas, I.: Pothole detection based on disparity transformation and road surface modeling. IEEE Trans. Image Process. **29**, 897–908 (2019)
3. Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., Fieguth, P.: A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. Adv. Eng. Inform. **29**(2), 196–210 (2015)
4. Janai, J, Güney, F, Behl, A, Geiger, A, et al.: Computer vision for autonomous vehicles: problems, datasets and state of the art. Found. Trends® Comput. Graph. Vis. **12**(1–3), 1–308 (2020)
5. Luo, L., Feng, M.Q., Jianping, W., Leung, R.Y.: Autonomous pothole detection using deep region-based convolutional neural network with cloud computing. Smart Struct. Syst. **24**(6), 745–757 (2019)
6. Fan, R., Liu, M.: Road damage detection based on unsupervised disparity map segmentation. IEEE Trans. Intell. Transp. Syst. **21**(11), 4906–4911 (2019)
7. Gupta, S., Sharma, P., Sharma, D., Gupta, V., Sambyal, N.: Detection and localization of potholes in thermal images using deep neural networks. Multimedia tools and applications **79**, 26265–26284 (2020)
8. Chen, Y.L., Jahanshahi, M.R., Manjunatha, P., Gan, W.P., Abdelbarr, M., Masri, S.F., Becerik-Gerber, B., Caffrey, J.P.: Inexpensive multimodal sensor fusion system for autonomous data acquisition of road surface conditions. IEEE Sens. J. **16**(21), 7731–7743 (2016)
9. Ravi, R., Habib, A., Bullock, D.: Pothole mapping and patching quantity estimates using lidar-based mobile mapping systems. Transp. Res. Record **2674**(9), 124–134 (2020)
10. Yang, Y., Cao, S., Huang, S., Wan, W.: Multimodal medical image fusion based on weighted local energy matching measurement and improved spatial frequency. IEEE Trans. Instrum. Meas. **70**, 1–16 (2020)
11. Ma, N., Fan, J., Wang, W., Wu, J., Jiang, Y., Xie, L.: Fan, R Computer vision for road imaging and pothole detection: a state-of-the-art review of systems and algorithms. Transp. Saf. Environ. **4**(4), tdac026 (2022)
12. Fan, R., Wang, H., Wang, Y., Liu, M., Pitas, I.: Graph attention layer evolves semantic segmentation for road pothole detection: a benchmark and algorithms. IEEE Trans. Image process. **30**, 8144–8154 (2021)
13. Fan, R., Ai, X., Dahnoun, N.: Road surface 3d reconstruction based on dense subpixel disparity map estimation. IEEE Trans. Image Process. **27**(6), 3025–3035 (2018)
14. Sattar, S., Li, S., Chapman, M.: Developing a near real-time road surface anomaly detection approach for road surface monitoring. Measurement **185**, 109990 (2021)
15. Dhiman, A., Klette, R.: Pothole detection using computer vision and learning. IEEE Trans. Intell. Transp. Syst. **21**(8), 3536–3550 (2019)
16. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
17. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241. Springer (2015)
18. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890 (2017)
19. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818 (2018)
20. Yang, M., Yu, K., Zhang, C., Li, Z., Yang, K.: Denseaspp for semantic segmentation in street scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3684–3692 (2018)
21. Takikawa, T., Acuna, D., Jampani, V., Fidler, S.: Gated-SCNN: Gated shape CNNs for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5229–5238 (2019)
22. Fan, R., Wang, H., Bocus, M.J., Liu, M.: We learn better road pothole detection: from attention aggregation to adversarial domain adaptation. In: Computer Vision—ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16, pp. 285–300. Springer (2020)
23. Wang, H., Fan, R., Sun, Y., Liu, M.: Applying surface normal information in drivable area and road anomaly detection for ground mobile robots. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2706–2711. IEEE (2020)
24. Fan, R., Wang, H., Cai, P., Liu, M.: Sne-roadseg: Incorporating surface normal information into semantic segmentation for accurate freespace detection. In: European Conference on Computer Vision, pp. 340–356. Springer (2020)
25. Hazirbas, C. Ma, L., Domokos, C., Cremers, D.: Fusenet: incorporating depth into semantic segmentation via fusion-based CNN architecture. In: Computer Vision—ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part I 13, pp. 213–228. Springer (2017)
26. Wang, W., Neumann, U.: Depth-aware CNN for RGB-D segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 135–150 (2018)
27. Sun, Y., Zuo, W., Liu, M.: Rtfnet: Rgb-thermal fusion network for semantic segmentation of urban scenes. IEEE Robot. Autom. Lett. **4**(3), 2576–2583 (2019)
28. Pan, Y., Zhang, X., Cervone, G., Yang, L.: Detection of asphalt pavement potholes and cracks based on the unmanned aerial vehicle multispectral imagery. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. **11**(10), 3701–3712 (2018)
29. Feng, F.Z., Guo, Y., Liang, Q., Bhutta, M.U.M., Wang, H., Liu, M., Sun, Y.: Mafnet: Segmentation of road potholes with multimodal attention fusion network for autonomous vehicles. IEEE Trans. Instrum. Meas. **71**, 1–12 (2022)

30. Gao, X., Tong, B.: MRA-UNet: balancing speed and accuracy in road crack segmentation network. Signal Image Video Process. **17**(5), 2093–2100 (2023)
31. Guo, F., Qian, Y., Liu, J., Yu, H.: Pavement crack detection based on transformer network. Autom. Constr. **145**, 104646 (2023)
32. Guo, F., Liu, J., Lv, C., Yu, H.: A novel transformer-based network with attention mechanism for automatic pavement crack detection. Constr. Build. Mater. **391**, 131852 (2023)
33. Dong, Q., Wang, S., Chen, X., Jiang, W., Li, R., Gu, X.: Pavement crack detection based on point cloud data and data fusion. Philos. Trans. R. Soc. A **381**(2254), 20220165 (2023)
34. Zhang, Z., Chen, P., Huang, Y., Dai, L., Feng, X., Hao, H.: Railway obstacle intrusion warning mechanism integrating yolo-based detection and risk assessment. J. Ind. Inf. Integr. **38**, 100571 (2024)
35. Li, B., Liu, S., Wu, F., Li, G.H., Zhong, M., Guan, X.: RT-Unet: an advanced network based on residual network and transformer for medical image segmentation. Int. J. Intell. Syst. **37**(11), 8565–8582 (2022)
36. Zhang, Z., Liu, Q., Wang, Y.: Road extraction by deep residual u-net. IEEE Geosci. Remote Sens. Lett. **15**(5), 749–753 (2018)
37. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
38. Mehta, S., Rastegari, M., Caspi, A., Shapiro, L., Hajishirzi, H.: Espnet: efficient spatial pyramid of dilated convolutions for semantic segmentation. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 552–568 (2018)

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

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

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.

Experimental Investigation on Microstructure, Mechanical Behavior and Tribological analysis of Al2017/SiC/TiO Composed by double stir casting route

Shlok Agarwal
Delhi Technological University
Kaustubh Kambiri
kaustubhkambiri2001@gmail.com

Delhi Technological University
Roop Lal
Delhi Technological University

Article

Keywords: Stir casting, Aluminium metal matrix, composites, Reinforcement, Mechanical properties, Future potential

Posted Date: April 24th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3966938/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

This study is on synthesising aluminium-based metal matrix hybrid composites with silicon carbide (SiC) and Titanium dioxide (TiO₂) reinforcements by double stir casting technique. SiC and TiO₂ were chosen for their excellent mechanical, tribological and physical properties. As limited information exists on the mechanical properties of Al2017 composites with both SiC and TiO₂. The Al/SiC/TiO₂ hybrid nanocomposites were prepared with 5 wt.% of SiC nanoparticles, 5 wt.% of TiO₂ nanoparticles, and a mixture of both at 5,5 wt.%. Microstructure analysis confirmed a uniform distribution of SiC and TiO₂ particles within the aluminium matrix. The Al2017/SiC composite exhibited a hardness of 46.33 HRB, while the Al2017/TiO₂ composite displayed a comparable hardness of 44.33 HRB. However, the combined Al2017/SiC/TiO₂ composite showed a decreased hardness of 40.33 HRB suggesting the importance of considering the composition and combination of reinforcement materials when designing metal matrix composites. The observed decrease in hardness in the Al2017/SiC/TiO₂ composite can be attributed to various factors, including the interaction and compatibility between SiC and TiO₂ particles within the aluminium matrix. Further research is required to understand the underlying mechanisms and enhance the mechanical behaviour of the Al2017 composites.

1. Introduction

It is challenging to choose the right material for a given application; lighter materials typically have lower strengths, whereas brittle materials have lesser toughness and fatigue resistance. Thus the hunt for new and better materials is ongoing due to the newest technological demands for greater energy efficiency, durability, lighter, and less expensive materials. For technological applications, it is nearly hard to find a single monolithic material with the necessary property profile.[1]

Since the early 1960s, new and improved construction materials have been sought with the advances in modern technology that have an interest in the rapid development of metal matrix composites in the aerospace and automotive sectors. Extensive research and development efforts have been made in the realm of composites as a result of the high demands placed on materials to increase overall performance. Metal matrix composites based on aluminium are often employed in the field of composites. Innovations in material processing have made it feasible to enhance wear resistance, attain high strength-to-weight ratios, increase stiffness, and maintain strength at high temperatures in order to satisfy new demands.[2]

2. Aluminium Metal Matrix Composites (AlMMCs)

A composite material is defined as a mixture of at least two materials that produces preferred qualities when compared to the properties of the component segments utilised separately. Every material has its unique chemical, physical, and mechanical qualities, as compared to metallic alloys. When compared to mass materials, composite materials' primary benefit is their excellent quality and solidity, along with their low density. This enables the finished part's weight to be reduced.[3]

Aluminium matrix composites are regarded as a novel and emerging class of materials having unique characteristics for specific applications. Advanced engineered materials, or AlMMCs, have better qualities than other traditional aluminium. Because of these characteristics, they have recently attracted considerable interest for a range of prospective applications in the automotive, aerospace, and other structural fields. In order to obtain the material with the needed qualities, extensive research and development has been done on Al-based MMCs using every conceivable alloy and various reinforcements.[4]

3. Stir Casting

In stir casting, the molten metal matrix is stirred with a stirrer. The material used to make the stirrer often has a greater melting point than the matrix temperature. Graphite stirrers are typically used in stir casting. The stirrer primarily consists of two parts: a cylindrical rod and an impeller. One end of the rod is attached to the motor shaft, and the other end is linked to the impeller. The stirrer is typically kept upright and rotates at different speeds thanks to a motor. Afterward, the molten metal is put into a die for casting. Stir casting is suited for producing composites with reinforcement volume fractions up to 30%. [5]

Such setup can be seen in Fig. 1

A major concern associated with the stir casting is segregation of reinforcement particles due to various process parameters and material properties resulting in the non-homogeneous metal distribution. The numerous process variables include things like the degree to which metal particles are moist, relative density, settling velocity, etc. The speed of the stirrer, its angle, the presence of vortices, and other factors all have an impact on how the particles are distributed inside the molten metal matrix. Where G.G. Sozhamannan et al.(2012) [6] fabricated aluminium metal composites at different processing temperatures with different holding time, which had different mechanical properties.

3.1 Double Stir-Casting Technique

There have been several reports on the use of two-step (double) stir casting to enhance cast metal matrix composites. K.K. Alaneme et al.(2012)[7] proposed that producing AMCs without using two-step stirring causes less homogeneity of the particles and greater porosity levels, which may be over the permitted limits. To promote wettability between the melt of the aluminium alloy and the SiC particles, the procedure calls for pre-treating the SiC particulates in a dry oven at 1,100°C..Additionally, to reduce the moisture content, P. Shanmughasundaram et al.(2011)[8] employed fly ash particles that had been warmed to 600℃ for two hours in a different muffle furnace. The graphite crucible was charged with aluminium, and the furnace temperature was increased to a liquidus temperature of 670℃ in order to completely melt the aluminium scraps. The melt temperature was then reduced to 620℃ in order to achieve a semi-solid condition. The crucible was filled with warmed fly ash particles and 1.5 wt% magnesium. To encourage the wetting activity between the Al matrix and the fly ash reinforcement

particles, Mg was added to the melt. After 10 minutes, the molten Al composite slurry reached a semi-solid condition and was agitated with a stirrer at a speed of 300 rpm. According to K. K. ALANEME et al. (2013)[9] research, the AA 6063/Al2O3p composites made using the two-step stir casting method used in this study had low porosity values of less than 3.6%. Although the strain to fracture and fracture toughness values of the composites declined with increasing volume percent alumina, the tensile strength, yield strength, and hardness values rose. Comparing well with other aluminium-based alumina reinforced composites described in literature, the mechanical characteristics derived from the AA 6063/Al2O3p composites created.

3.2 Material

For this study aluminium 2017 was used with compositions given below in Table 1.

Table.1. Composition of Al used as matrix material (wt. %)

Al%	Si%	Cu%	Fe%	Zn%	Mg%
92.784	0.722	4.466	0.311	0.056	0.623

Sn%	Ti%	V%	Ni%	Mn%	Cr%
0.077	0.031	0.01	0.049	0.0807	0.035

According to J. R. Davis (2001)[10], In alloys from the 2xxx class, copper serves as the main alloying element, with magnesium serving as a minor component. To achieve their ideal qualities, these alloys must undergo solution heat treatment; in this state, their mechanical properties are comparable to and occasionally even surpass those of low-carbon steel. Precipitation heat treatment, also known as ageing, is used to further improve mechanical characteristics. This process reduces elongation while increasing yield strength; it has a less significant impact on tensile strength. Screw machine products, fittings, fasteners, and machine components are the principal uses for Al2017. According to K. Mroczka et al. (2012)[11] research on the 2017A aluminium alloy under various heat treatment circumstances, the alloy should be treated at a temperature of 500°C (the allowable temperature range is between 490°C and 520°C). After 72 hours, the material had reached its maximum hardness and ultimate tensile strength due to natural ageing. After six hours, artificial ageing at 180°C results in a significant increase in hardness. The ultimate tensile strength of the material improves by 20% as a result of natural ageing (compared to the supersaturation state). Mrówka-Nowotnik Grayna et al.(2021)[12] came to the conclusion that the temperature of the solution heat treatment also influences the alloy's hardness during artificial ageing. It was discovered that as compared to natural ageing, artificial ageing enables the achievement of greater hardness levels. The hardest alloy is the alloy 2017 solution, which was heat treated at 510°C and aged at 175°C. The hardness is quite similar to the value attained during natural ageing after ageing at 120°C. It was discovered that during artificial ageing, the temperature of the solution heat treatment has no bearing on how quickly hardness increases to its maximum value.

4. Reinforcement

The reinforced metal matrix has the potential to create improved MMCs and exhibits superior characteristics when compared to base or reinforced material used alone. For micron-sized particles in general, 5 to 10% might be adequate, and for nanomaterials, even less than 5%, resulting in a significant improvement in mechanical behaviour. V.K. Sharma et al.(2012)[13] discovered that the wear resistance of manufactured composites rises as the flyash concentration increases. Composites with high flyash contents wore out 13.6% less than composites with low flyash concentrations.The sample with a medium flyash concentration (4%) had the lowest average coefficient of friction (0.12), while the sample with a high flyash level (6%) had the highest average coefficient of friction (0.161). As a result, the amount of flyash in the aluminium matrix is restricted to 4%. The addition of more reinforcement increases the coefficient of friction between the tribopairs.

4.1 SiC (Silicon Carbide)

The most widely employed reinforcement material in the development of AMMCs is silicon carbide (SiC). It also increases ultimate tensile strength while decreasing porosity. The qualities of the reinforcement vary depending on the grade of aluminium alloy used.[14]

Md. Habibur Rahmana(2014)[15] came to the conclusion that adding SiC to an Al matrix boosted the Vickers hardness and tensile strength of composites when compared to unreinforced Al. AlMMC with a SiC concentration of 20% had the highest hardness and tensile strength. The wear resistance of SiC reinforced AMCs increased as the SiC concentration in the Al matrix increased. Based on wear resistance, AMC reinforced with 20% SiC performed best. Cui Yana et al.(2008)[16] research also reveals that SiC/Al composites have a low density (2.94 g/cm^3), a high elastic modulus (220 GPa), a prominent thermal management function due to a low coefficient of thermal expansion ($8\times10^{-}$ K^{-1}) and high thermal conductivity (235 W/(mK)), and a unique preventability of resonance vibration. The multi-functional SiC/Al composites were manufactured into near-net-shape pieces using a number of established processes. Precision components for space-based optomechanical structures and airborne optoelectronic platforms have been manufactured in large quantities. Several common goods are now being tested in the field.

4.2 TiO (Titanium Dioxide)

According to Anna A. Murashkina et al.(2015)[17] research, TiO samples exhibit a mixed phase rutile-anatase crystalline structure, with the anatase component increasing linearly from 0.0 wt% for Al-TiO to 18 wt% for 1.1-Al-TiO. With increasing Al concentration, particle size decreases from 800 nm to 50 nm, and specific surface area increases from 1.7 m^2/g to 28 m^2/g. Adeolu A. Adediran et al. (2021)[18] demonstrated Simultaneous optimization of the characteristics 779.3°C, 574.2 rpm, and 22.5 min as the best stir casting parameters for temperature, speed, and duration, respectively, using Minitab 19 software. The interaction profile of the parameters according to the response surface was examined.

Contour plots for each interaction revealed distinct ranges of stirring settings for optimising each attribute.

4.3 Hybrid of SiC (Silicon Carbide)&TiO₂ (Titanium Dioxide)

Mechanical parameters of hybrid composites such as tensile strength (TS), hardness (BHN), density, and impact strength have been found to be identical to or higher than those of conventional composites. It has been said that hybrid composites provide higher flexibility and reliability in the structure of the future parts determined by the fortification mix and organization.[19]

G. Elangoa et al. (2013)[20] studied the wear behaviour of aluminium alloy LM25 reinforced with SiC particulate and further addition of TiO_2 particulate. The wear resistance and frictional properties of a hybrid metal matrix composite were investigated using a pin on disc wear tester concluding that reinforcing the metal matrix with SiC and TiO_2 minimises wear rate at room temperature.

Aluminium hybrid metal matrix nanocomposites (Al/SiC/TiO_2) were synthesised by M.R. Mattli et al. (2021)[21] using a microwave-assisted powder metallurgy procedure, and their developed microstructure and mechanical properties were examined. The Al/SiC/TiO_2 hybrid nanocomposites were prepared by reinforcing aluminium (Al) matrix with a fixed amount of silicon carbide (SiC) nanoparticles (5 wt.%) and varying concentrations of titanium dioxide (TiO_2) nanoparticles (3, 6, and 9 wt.%). The addition of TiO_2 nanoparticles increased the hardness and compressive strength of the Al/SiC/TiO_2 hybrid nanocomposites. The Al/SiC/TiO_2 hybrid nanocomposites with the highest concentration of TiO_2 nanoparticles (9 wt.%) showed the best mechanical properties.

A.D. Assi et al. (2020)[22] produced composites were reinforced with silicon carbide (SiC) and titanium dioxide (TiO_2) nanoparticles added in precise weight percentages (3, 6, 9 wt%) using a stir casting technique. All specimens were subjected to cold mechanical treatment (pressure forming) and heat-treated T6 (solution heat treatment artificial ageing). Mechanical tests were then performed on the samples. According to the SEM and X-RD data, the reinforced AA (6061) had improved in hardness, yield strength, ultimate tensile strength, and impact toughness.

S. Hariharan et al. (2020)[23] developed an aluminium-based hybrid composite material by stir casting with Al7075 as the matrix and SiC and TiO_2 as the reinforcements. Mechanical analysis of Al7075 + SiC + TiO_2 revealed greater hardness and tensile strength.

5. Experimental Procedure

Material Preparation

The Al2017 matrix alloy was used as the composites' base material. As reinforcement materials, silicon carbide (SiC) and titanium dioxide (TiO_2) particles were employed. To guarantee appropriate dispersion, the SiC and TiO_2 particles were pre-treated.

Double Stir Casting Technique

The Al2017/SiC/TiO composites were created using the double stir casting technique. The Al2017 alloy was melted in a crucible using a controlled temperature furnace. The pre-treated SiC and TiO particles were gradually added to the molten Al2017 alloy while the mixture was constantly stirred. To achieve consistent dispersion of the reinforcement particles, the stirring operation was carried out using mechanical stirrers at a predetermined speed. The stirring was done for a certain amount of time to ensure proper mixing and bonding between the matrix alloy and the reinforcing particles.

Surface Preparation

To eliminate any dirt, grease, or impurities, the surfaces of the aluminium samples were thoroughly washed with a suitable cleansing agent. To eliminate surface flaws, scratches, and oxidation, coarse grit sanding was performed with sandpaper with a grain size of 120–180.The entire surface was uniformly sanded, with special attention paid to any unusually rough or damaged areas. To further refine the surface and remove any leftover scratches or markings, fine grit sanding was performed using increasingly finer sandpapers. During the sanding operation, water or a suitable polishing fluid was employed as a lubricant to reduce heat buildup and improve surface quality. The stirring was done for a certain amount of time to ensure proper mixing and bonding between the matrix alloy and the reinforcing particles.

Mirror Polishing

Mirror finishing was performed on the smoothed surface using a high-quality aluminium polishing compound along with a clean cloth and rotatory polishing station, aluminium surface was polished in circular motions while applying light pressure and keeping a consistent speed. The polishing compound was applied uniformly throughout the whole surface, with special attention paid to any residual defects. Throughout the polishing process, the surface was inspected at regular intervals to check progress and ensure the intended outcomes. The polishing process was repeated until a mirror-like quality was obtained.

Microstructure Analysis

On polished cylindrical samples, an Olympus microscope was used for microstructure examination. This microscope provided superior optical resolution and magnification, enabling a thorough analysis of the dispersion and properties of SiC and TiO particles within the Al2017 matrix. The microscope was used to record high-resolution microstructures at various magnifications in order to gain a full understanding of the microstructural characteristics and bonding properties of the composite materials. The microstructure analysis yielded information on the distribution, shape, and bonding properties of the reinforcing particles.

Rockwell Hardness Testing

The Rockwell hardness values of the individual composite samples were determined. A Rockwell hardness testing machine with the Rockwell B scale was used. The cylindrical samples were subjected to a small load of 10 kg, followed by a main load of 100 kg. The indenter was a steel ball with a diameter of 1/16 of an inch. The cylindrical samples' polished surfaces were indented, and the related hardness values were recorded. To assure statistical reliability, each sample was indented many times.

Such setup can be seen in Fig. 2

Data Collection:

- The obtained microstructure and hardness values were recorded and documented for further analysis and interpretation.

6. Results and Discussions

The results of the experimental analysis on aluminium-based composites are presented in this section. This assessment is divided into three parts: surface analysis, hardness testing, and microstructure characterization.

These three factors were chosen because they are critical in defining the mechanical properties and performance of composites. The results of this analysis helps to understand the material's behaviour, identify any limitations, and provide a foundation for future optimization and advancement in research and applications.

6.1. Surface analysis

Surface analysis was used to evaluate the quality and characteristics of the produced samples." The results revealed that the porosity levels of the various composites varied. The TiO₂ particle-reinforced sample had the maximum porosity, followed by the hybrid composite and the SiC-reinforced sample. This finding indicates that the addition of TiO₂ particles has made achieving a fully dense structure more difficult. Porosity in composites can have a negative impact on mechanical qualities such as hardness. As a result, addressing the porosity issue is critical in order to improve the performance of composites.

On the other hand, the sample reinforced with SiC particles, demonstrated a remarkable observation after dry polishing, in which the composite was seen to shed abrasive particles. This is due to a less strong bond between the SiC particles and the aluminium matrix. However, the TiO₂ casting technique did not produce adequate results for the Al2017 alloy, indicating difficulties in obtaining uniform distribution and effective bonding of TiO₂ particles with the aluminium matrix. To improve the performance of the Al2017/TiO₂ composite, more research and optimisation of the casting parameters are required. These surface analysis findings highlight the necessity of regulating porosity and establishing strong interfacial interaction between the reinforcement particles and the matrix in hybrid composites.

6.2 Hardness

Hardness refers to a material's resistance to surface indentation. The microhardness of composites measures the strength of the interface bonding between reinforcing particles and matrix. The Rockwell hardness test was used to determine the hardness values of the produced samples. The hardness of the hybrid composite was 40.3333HRB, the hardness of the SiC-reinforced sample was 46.333 HRB, and the hardness of the TiO₂-reinforced sample was 43.333 HRB. The difference in hardness across samples reflects the effect of different reinforcements on the mechanical properties of the composites. The hybrid composite's decreased hardness seems to be related to elements such as the interaction and compatibility of SiC and TiO₂ particles within the aluminium matrix, as well as the presence of porosity. Such variation in hardness can be seen in Fig. 3

6.3 Microstructures

The microstructure study of the samples revealed distinct characteristics. The SiC sample revealed a network-like architecture with interconnected particles. It also had numerous fine black dots, which indicated the dispersion of SiC particles within the aluminium matrix. The TiO₂ sample, on the other hand, had a more homogeneous microstructure with finer grain boundaries and less apparent black spots. The microstructure of the hybrid sample was identical to that of the TiO₂ sample, indicating a homogenous distribution of TiO₂ particles inside the aluminium matrix.

The changes in microstructure across the samples can have a substantial impact on their mechanical qualities and performance. The SiC sample's linked structure may lead to improved mechanical strength and load-bearing capabilities.The highly homogeneous microstructure of the TiO₂ and hybrid samples indicates good particle dispersion and possible improvements in characteristics like wear resistance and dimensional stability. Microstructure for the following can be seen in Fig. 4.

7. Conclusion

Possible reasons for this outcome could be related to the particle distribution, particle interactions, or the effect of different reinforcement mechanisms. It is important to note that the interaction between SiC and TiO₂ particles might have influenced the overall microstructure and mechanical behaviour of the composite. The combined presence of these two types of particles may have affected their distribution and bonding within the matrix, resulting in a reduction in hardness compared to individual reinforcement additions.

Potential reasons for the observed decrease in hardness when SiC and TiO₂ are added simultaneously:

Particle Interactions

The interaction between SiC and TiO₂ particles may have led to agglomeration or clustering, resulting in uneven distribution within the matrix. This clustering could have hindered effective load transfer and

reduced the overall hardness.

Reinforcement Mechanisms

SiC and TiO₂ have different reinforcement mechanisms. SiC is known for its excellent hardness and high wear resistance, while TiO₂ is primarily used for its beneficial frictional properties. The combined addition of these particles might have influenced the dominant reinforcement mechanism, potentially leading to a compromise in hardness.

Matrix Compatibility

The Al2017 matrix may have different levels of compatibility with SiC and TiO₂. The interactions between the matrix and the two types of particles could have varied, leading to a difference in the effectiveness of reinforcement and subsequent hardness.

The addition of SiC and TiO₂ to the Al2017 matrix reduced hardness as compared to individual additions. Potential particle interactions, changes in reinforcement mechanisms, and differences in matrix compatibility can all be attributed to this drop. These findings highlight the significance of carefully evaluating the combination and interactions of different reinforcing particles when attempting to obtain the desired mechanical characteristics in composite materials.

To further enhance the understanding of the observed trend, future studies could focus on optimising the particle distribution, exploring alternative processing techniques, or investigating the effect of varying weight fractions of SiC and TiO₂ in the composite. Such research could help customise the mechanical properties of Aluminum composites for various applications.

Declarations

Author Contribution

S.A. and K.K. wrote the main manuscript text and prepared figures 1-3. All authors reviewed the manuscript. R.L. supervised the experimental procedure and the analysis of the microstructures.

Data Availability

The data that support the findings of this study are available from the corresponding author upon special request.

References

1. A.K. Sharma, R. Bhandari, A. Aherwar, R. Rimašauskiene, C.P. Bretotean, A study of advancement in application opportunities of aluminium metal matrix composites, Materials Today: Proceedings 26

(2020) 2419–2424

2. P.D. Srivyas, M.S. Charoo, aluminium metal matrix composites a review of reinforcement; mechanical and tribological behaviour, International Journal of Engineering & Technology, 7 (2.4) (2018) 117–122
3. J. Hashim, L. Looney and M.S.J. Hashmi, Metal matrix composites: production by the stir casting method, Journal of Materials Processing Technology 92–93 (1999) 1–7
4. Amit Pal, Abhishek Verma, S.S Saxena and B.C. Kandpal,Stir Casting of Metal Matrix Composites – A Review, International Journal of Computer & Mathematical Sciences, ISSN 2347–8527, Volume 4, February 2015
5. A. Kumar, R.C. singh, R. Chaudhary, Recent progress in production of metal matrix composites by stir casting process: An overview, Materials Today: Proceedings, Volume 21, Part 3, 2020, Pages 1453–1457
6. G.G. Sozhamannan, S. Balasivanandha Prabu, V.S.K. Venkatagalapathy, Effect of Processing Parameters on Metal Matrix Composites: Stir Casting Process, Journal of Surface Engineered Materials and Advanced Technology, 2012, 2, 11–15
7. K.K. Alaneme and A.O. Aluko, Production and Age-Hardening Behaviour of Borax Premixed SiC reinforced Al-Mg-Si alloy Composites developed by Double Stir-Casting Technique, The West Indian Journal of Engineering, Volume 34, No 1/2, (2012), Pages 80–85, ISSN 0511–5728
8. P. Shanmughasundaram, R. Subramanian, G. Prabhu, Some Studies on Aluminium–Fly Ash Composites Fabricated by Two Step Stir Casting Method, European Journal of Scientific Research, ISSN 1450-216X, Vol.63 No.2 (2011), pp.204–218
9. K. K. ALANEME, M. O. BODUNRIN, MECHANICAL BEHAVIOUR OF ALUMINA REINFORCED AA6063 METAL MATRIX COMPOSITES DEVELOPED BY TWO STEP – STIR CASTING PROCESS, ACTA TECHNICA CORVINIENSIS- Bulletin of Engineering Tome V (Year 2013), ISSN 2067–3809
10. J.R. Davis, Alloying: Understanding the Basics, ASM International, p351-416,(2001)
11. K. Mroczka, A. Wójcicka, P. Kurtyka, 2017A aluminium ALLOY IN DIFFERENT HEAT TREATMENT CONDITIONS, Acta Metallurgica Slovaca, Vol. 18, (2012), No. 2–3, p. 82–91
12. Mrówka-Nowotnik Grażyna, Kamil Gancarczyk, Andrzej Nowotnik, Kamil Dychtoń and Grzegorz Boczkal, Microstructure and Properties of As-Cast and Heat-Treated 2017A Aluminium Alloy Obtained from Scrap Recycling, Acta Metallurgica Slovaca (2021), 14, 89.
13. Vipin K. Sharma, R.C. Singh, Rajiv Chaudhary, Effect of flyash particles with aluminium melt on the wear of aluminium metal matrix composites, Engineering Science and Technology, an International Journal 20 (2017) 1318–1323
14. Dunia Abdul Saheb, aluminium SILICON CARBIDE AND aluminium GRAPHITE PARTICULATE COMPOSITES, ARPN Journal of Engineering and Applied Sciences, (2011), VOL. 6, NO. 10
15. Md. Habibur Rahmana, H. M. Mamun Al Rashed, Characterization of silicon carbide reinforced aluminium matrix composites, Procedia Engineering 90 (2014) 103–109

16. Cui Yana, Wang Lifeng, Ren Jianyue, Multi-functional SiC/Al Composites for Aerospace Applications,Chinese Journal of Aeronautics 21(2008) 578–584
17. Anna A. Murashkina, Petr D. Murzin, Aida V. Rudakova, Vladimir K. Ryabchuk, Alexei V. Emeline and Detlef W. Bahnemann, Influence of the Dopant Concentration on the Photocatalytic Activity: Al-Doped TiO2, J. Phys. Chem. C 2015, 119, 24695–24703
18. Adeolu A. Adediran, Abayomi A. Akinwande, Oluwatosin A. Balogun, Bayode J. Olorunfemi & M. Saravana Kumar, Optimisation studies of stir casting parameters and mechanical properties of TiO2 reinforced Al7075 composite using response surface methodology, *Scientific Reports* volume11, Article number: 19860 (2021)
19. S. Venugopal, L. Karikalan, Microstructure and physical properties of hybrid metal matrix composites AA6061-TiO2-SiC via stir casting techniques, Materials Today: Proceedings 37 (2021) 1289–1294
20. G. Elangoa, B.K.Raghunath, Tribological Behavior of Hybrid (LM25Al + SiC + TiO2) Metal Matrix Composites, Procedia Engineering 64 (2013) 671–680
21. Manohar Reddy Mattli, Penchal Reddy Matli, Adnan Khan, Rokaya Hamdy Abdelatty, Moinuddin Yusuf, Abdulla Al Ashraf, Rama Gopal Kotalo and Rana Abdul Shakoor, Study of Microstructural and Mechanical Properties of Al/SiC/TiO2 Hybrid Nanocomposites Developed by Microwave Sintering, Crystals 2021, 11, 1078
22. Abdullah Dhayea Assi, Hassan A.Abdulhadi, Salman Hussien Omran, Effect Of Adding SIC And TIO2 Nanoparticles To AA6061 By Stir Casting Technique On The Mechanical Properties Of Composites, Journal of Mechanical Engineering Research and Developments, ISSN: 1024–1752, (2020), Vol. 43, No. 6, pp. 167–183
23. S. Hariharan, M. Kartheeswaran, M. Essakiraj, R. Harikrishnan, M. Ayyanarraja, R. Samuel hansen, A Review of aluminium (Al7075) SiC & TiO2 with Stir Casting, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 9, Issue 2, (2020)

Figures

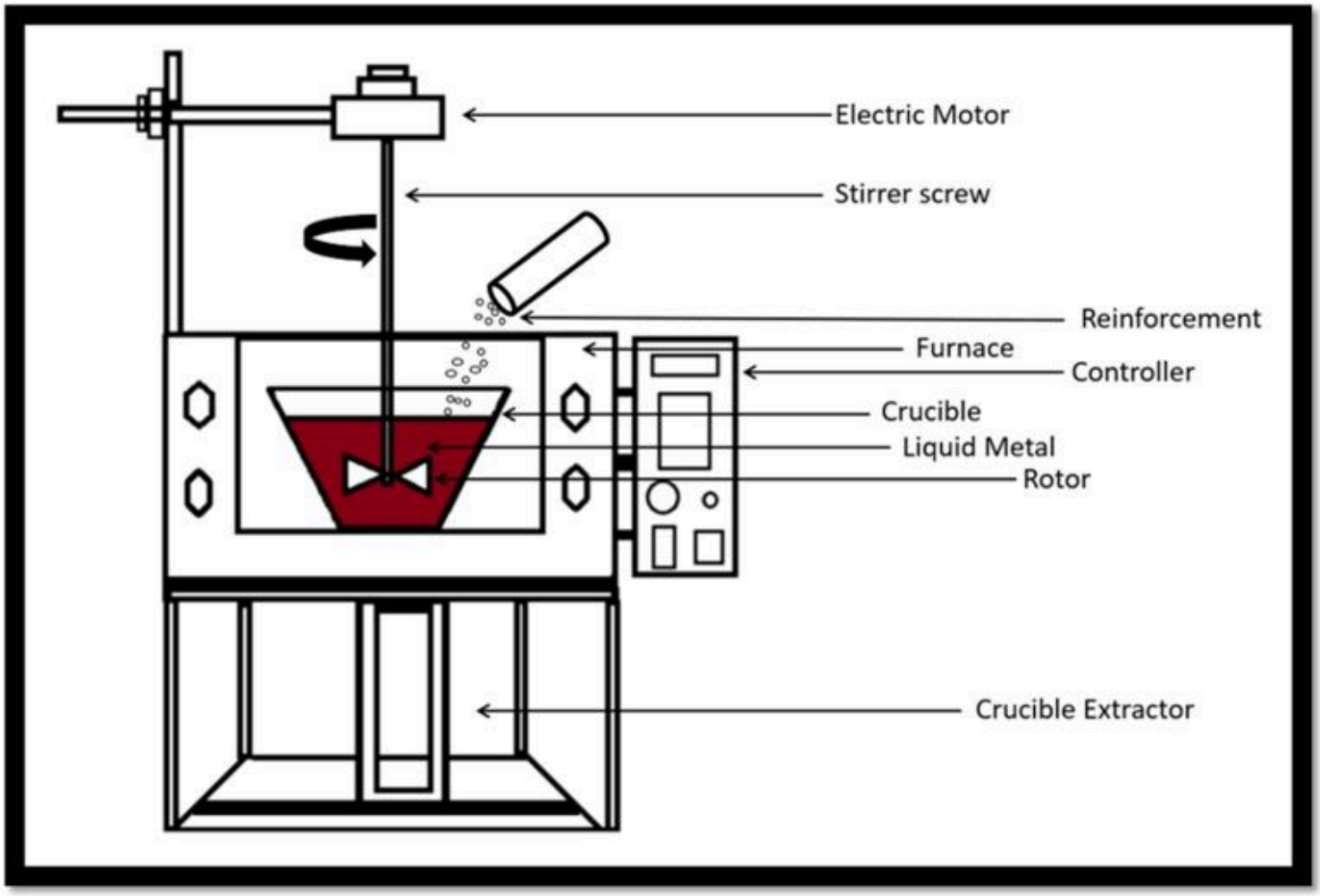


Figure 1

Schematic set up diagram of stir casting process

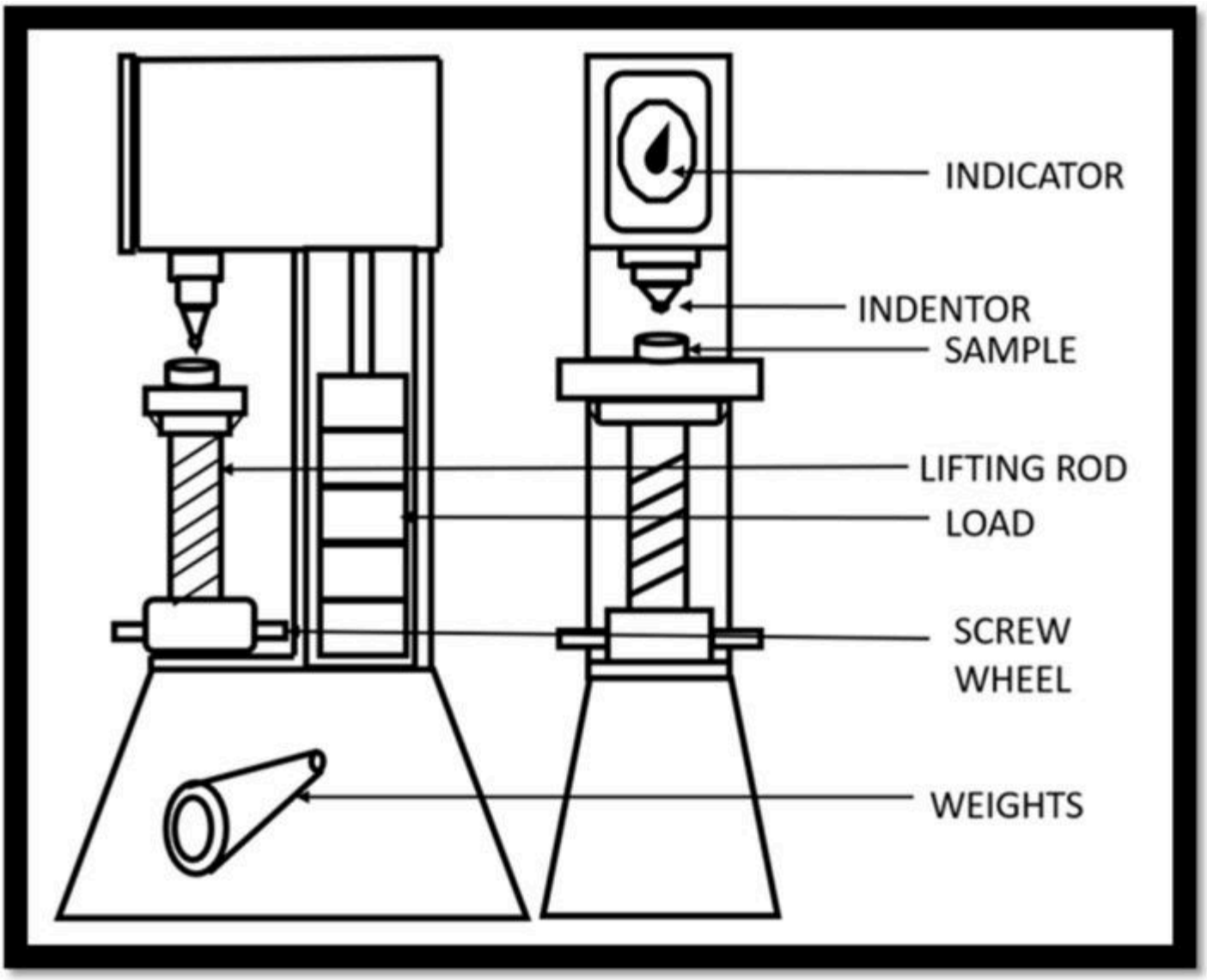


Figure 2

Schematic set up diagram of Rockwell Hardness Tester

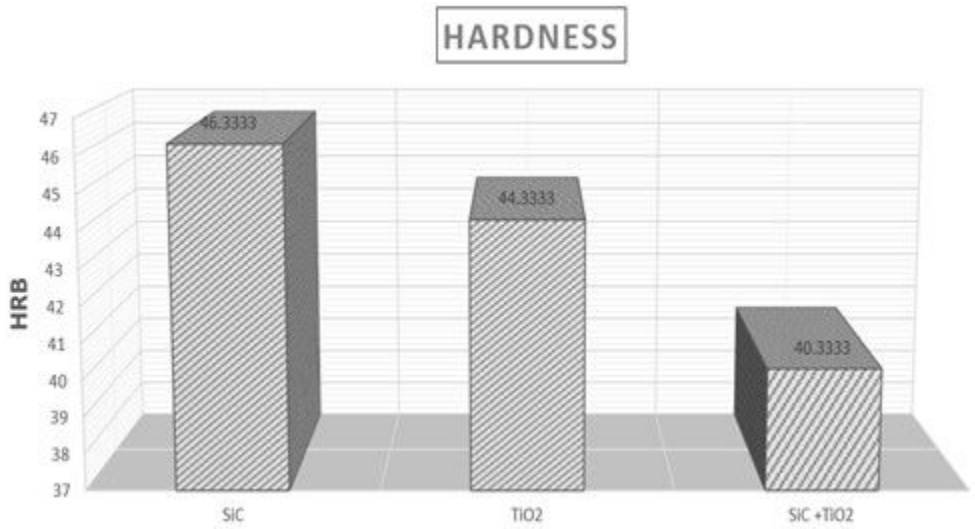


Figure 3

Variation in Rockwell hardness

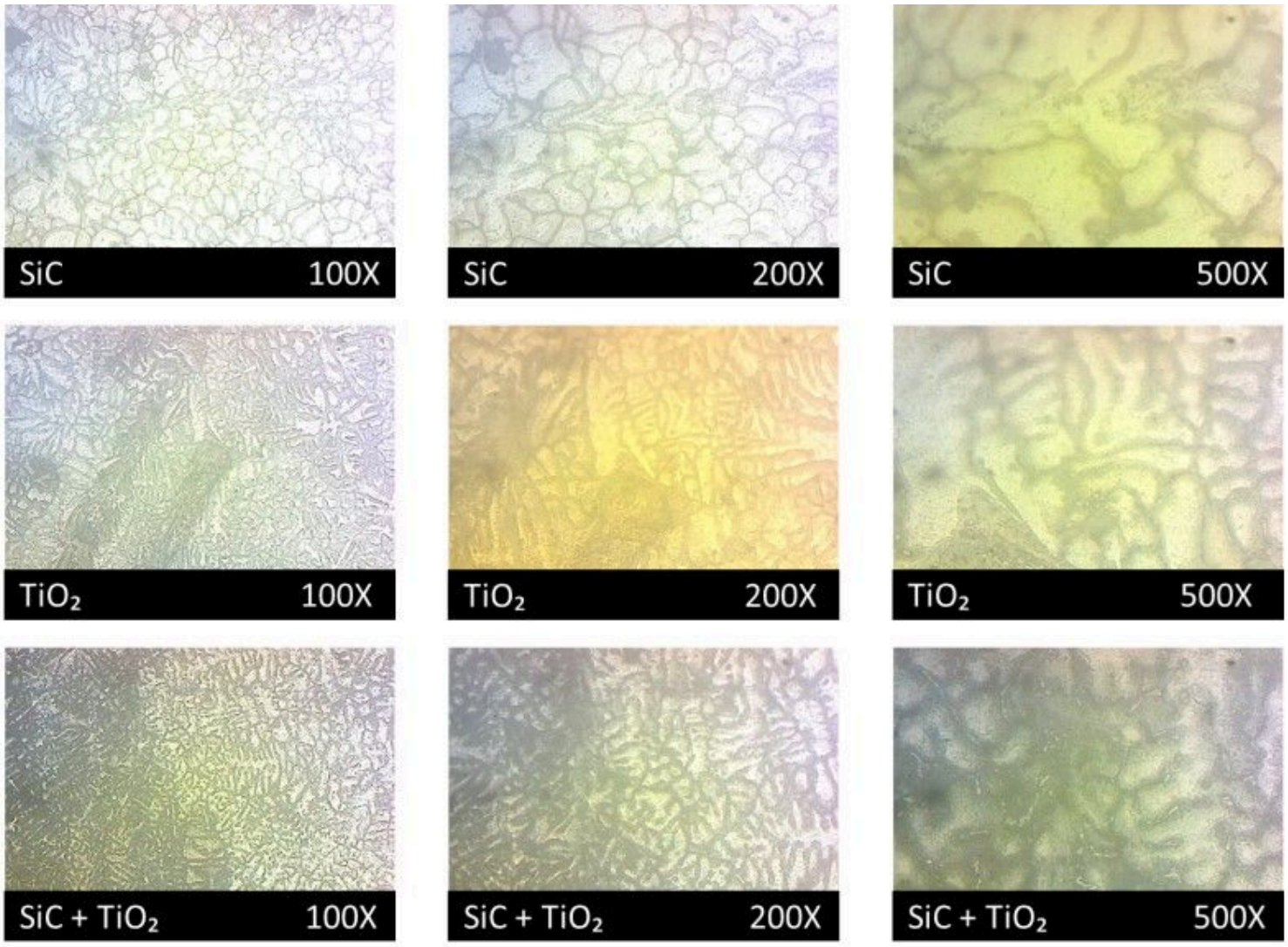


Figure 4

Microstructures

arXiv:2405.15296v1 [astro-ph.CO] 24 May 2024

Exploring interacting bulk viscous model with decaying vacuum density

Vinita Khatri* and C. P. Singh†
Department of Applied Mathematics,
Delhi Technological University, Delhi-110042, India

Milan Srivastava‡
Department of Mathematics,
Sharda School of Basic Sciences and Research,
Sharda University, Greater Noida, India

In the present work, we study a cosmological model composed of a viscous dark matter interacting with decaying vacuum energy in a spatially flat Universe. In the first part, we find the analytical solution of different cosmological parameters by assuming the physically viable forms of bulk viscosity and decaying vacuum density with the interaction term. The second part is dedicated to constrain the free parameters of the interacting viscous model with decaying vacuum energy by employing latest observational data of *Pantheon+*, Cosmic Chronometer and $f(z)\sigma_8(z)$. We find that the interacting model just deviate very slightly from well-known concordance ΛCDM model and can alleviate effectively the current H_0 tension between local measurement by R21 and global measurement by Planck 2018, and the excess in the mass fluctuation amplitude σ_8 essentially vanish in this context. We report the Hubble constants as $H_0 = 72.150^{+0.989}_{-0.779}$, and $72.202^{+0.796}_{-0.937}$ $kms^{-1}Mpc^{-1}$, deceleration parameters as $q_0 = -0.533 \pm 0.024$, and -0.531 ± 0.024, and equation of state parameters as $w_0 = -0.689 \pm 0.016$, and -0.687 ± 0.016 for ΛCDM and interacting models, respectively. It is found that the interacting model is in good agreement with ΛCDM. Further, we discuss the amplitude of matter power spectrum σ_8 and its associated parameter S_8 using $f(z)\sigma_8(z)$ data. Finally, the information selection criterion and Bayesian inference are discussed to distinguish the interacting model with ΛCDM model.

I. INTRODUCTION

In modern cosmology, understanding of two dark components: dark energy (DE) and dark matter (DM) of the Universe is so far the most challenging research area in cosmology. The discovery of accelerated expansion in 1998 [1, 2] has motivated to comprehend the composition of these dark entities. Dark matter, which interacts only gravitationally with ordinary matter, explains the measurement of rotation curves of spiral galaxies [3]. The most important theories for DM are scalar fields [4, 5], supersymmetry models [6]. On the other hand, dark energy, which has negative pressure, is considered for the accelerated expansion of the Universe. The most interesting candidates are the cosmological constant (CC), phantom energy, quintessence, Chaplygin gas, modified theories of gravity, etc. (refer to Refs.[7–9]). The cosmological constant as DE so-called Lambda-cold-dark matter, abbreviated as ΛCDM is still the best candidate to explain the cosmological observations. However, ΛCDM model suffers with some theoretical problems [10, 11] when its origin is considered as quantum vacuum fluctuations.

Many alternative theories have been proposed beyond the ΛCDM to solve these conundrums. Some works in literature have shown that some dynamical DE models are able to resolve the problems of ΛCDM. In this respect, the cosmological models with time-varying vacuum energy are the promising models to resolve the issues. Although there is no fundamental theory to describe a time-varying vacuum, a phenomenological technique has been proposed to parametrized CC. Shapiro and Solá [12] and Solá [13] proposed a running vacuum models (RVMs) on the basis of renormalization group (RG) formalism of Quantum field theory (QFT) in curved spacetime. In the context of RVMs, it is considered that the vacuum energy density evolves slowly with the cosmic expansion. It has been illustrated that both the background and linear perturbation levels of the cosmic evolutions can be described by the RVMs. The vacuum energy is typically determined in curved space-times using renormalization group procedures, which depend on the Hubble parameter H and its time derivative and is given by the form $\rho_\Lambda(H, \dot{H}) = M_{Pl}^2\Lambda(H, \dot{H})$ [14]. In past literature, many authors [15–20] have studied the time-varying vacuum energy models. In recent years, the RVMs have been carefully confronted against many cosmological data, which receive a significant success [21–40].

On the other hand, viscous fluid cosmology is an interesting research area to understand the accelerated expansion of the Universe. It has been observed that irreversible processes in the evolution of the Universe may also be responsible for explaining the recent accelerated expansion. There are two types of viscosity: bulk and shear, however, bulk viscosity is the one that plays an important role in the evolution of the Universe as it satisfies the cosmological principle. Bulk viscous models have

* vinitakhatri_2k20phdam501@dtu.ac.in
† cpsingh@dce.ac.in (Corresponding author)
‡ milan.srivastava@sharda.ac.in

been studied using two approaches: the Eckart [41] and Israel-Stewart [42] theories. Eckart [41] proposed a non-causal theory of viscous of first order which was later modified by Landau and Lifshitz [43]. In this theory, all equilibrium states are unstable [44] and the signals can propagate through the fluids faster than the speed of light [45]. The Israel-Stewart formalism is a full causal theory of second order which avoids the casuality problem. When the relaxation time goes to zero, the causal theory reduces to the Eckart's first order theory. Despite of the causality problem, Eckart theory is widely used due to its simplicity. In both theories, the bulk viscous term is included in Einstein field equations through an effective pressure, written in the form $\bar{p} = p + \Pi$, where p is pressure of matter contents such as dust-matter, DE or relativistic species (photons and neutrinos) and Π refers the bulk viscous pressure. In Eckart theory, the bulk viscous pressure is assumed as $\Pi = -3\zeta H$, where ζ is the bulk viscous coefficient and H the Hubble parameter. Some works related to Eckart theory have studied the evolution of the Universe at late time by assuming bulk viscous coefficient with a constant [46–51] and polynomial [52–55] as functions of redshift or in terms of energy density. Some more authors [56–76] have explored the viability of a bulk viscous Universe to explain the present accelerated expansion of the Universe.

Decaying vacuum energy density (VED) models and viscous fluid models are two appealing theoretical models that have been separately studied by many authors to solve some of problems facing by standard ΛCDM model. Despite the success of decaying vacuum energy and viscous fluid models, it should be noted that they have, separately, limitations in describing the entire cosmological evolution. Recently, in Refs.[77–80] the authors have studied the cosmological model with combination of these two notions in a single cosmological setting and have investigated their cosmological implications using observational data.

The ΛCDM's problems also motivate research into new physics beyond this standard model. In standard cosmology, it is usually assumed that DM and DE do not interact with each other. However, there is no physical basis for this assumption. In this regard, a popular approach is to investigate the cosmological models where the interactions between the DM and DE take place. In this interaction theory, DM and DE are not separately conserved but they exchange energy (and /or momentum). However, the total energy (and/or momentum) is conserved. It is worth constraining the interacting models against the available wealth of precision cosmological data. This has motivated a large number of studies based on models where DM and DE share interactions, usually referred to as interacting dark energy (IDE) models. Several studies in the literature have been devoted to explore whether DM-DE interactions may help to resolve the enduring H_0 tension [81–88]. In Refs.[89–96], authors have studied the interacting viscous dark fluids for explaining the early and late time evolution of the Universe. The above cited works on bulk viscosity show that interacting viscous fluid models can be one of the possible mechanism to explain the present acceleration of the Universe.

In light of the aforementioned discussions, the main purpose of the present work is to investigate a cosmological model for a spatially flat Friedmann-Lemaître-Robertson-Walker Universe including two components: a non-perfect and interacting viscous dark matter, and decaying vacuum energy that interact with viscous dark matter in Eckart's approach. We analyze the dynamics of interacting model by constraining the free parameters by performing a Markov chain Monte Carlo (MCMC) using the latest observational data. Further, we investigate how our interacting viscous model with decaying VED affects the perturbation level. To investigate the role of this model in structure formation, we employ the perturbation equation to determine the growth of matter fluctuations. Finally, we study the evolutions of various cosmological parameters and compare the perfect fluid case that corresponds to the ΛCDM model through the model selection criteria and Bayesian evidence analysis.

The paper is organised as follows. In Sec. II, we present the general features of the proposed cosmological model for a spatially flat Friedmann-Lemaître-Robertson-Walker Universe where dissipative effects are present with interacting decaying VED. Section III deals with the analysis on Structure formation and perturbation equations. We present in Sec. IV the cosmological probes that are used to constrain the model. Section V gives the results and discussions on various cosmological parameters with the trajectories and compares the proposed model with the concordance ΛCDM using information criterion and the Bayesian inference. The main results of the work are summarized and discussed in Sec. VI. Two more solutions are also presented in Appendix.

II. INTERACTING BULK VISCOUS MODEL WITH DECAYING VED

We start with a homogeneous and isotropic Universe described by the Friedmann-Lemaître-Robertson-Walker (FLRW) metric

$$ds^2 = -dt^2 + a^2(t)\left[dr^2 + r^2(d\theta^2 + sin^2\theta d\phi^2)\right], \quad (1)$$

where $a(t)$ represents the cosmic scale factor of the Universe. The Einstein field equations with a time-varying cosmological constant term, Λ take the form

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi G(T_{\mu\nu} + g_{\mu\nu}\rho_\Lambda) \quad (2)$$

where $R = g_{\mu\nu}R_{\mu\nu}$ is the Ricci scalar, $\rho_\Lambda = \Lambda/8\pi G$ is the energy density associated to Λ-term, so-called the vacuum energy density (VED), G is the Newton gravitational constant and $T_{\mu\nu}$ is the energy-momentum tensor of matter. It should be emphasised that we employ the geometric units $8\pi G = c = 1$.

In this paper, we propose to study the cosmological dynamics of the Universe which include the interaction between the dark matter component including dissipation through a bulk viscous coefficient and a vacuum energy density described by running coupling depending on the Hubble parameter (hereafter we refer interacting viscous $\Lambda(t)$ model). Then, the energy-momentum tensor is given by

$$T_{\mu\nu} = (\rho + \tilde{p})u_\mu u_\nu + g_{\mu\nu}\tilde{p} \tag{3}$$

where $\rho = \rho_m + \rho_\Lambda$ is the sum of total energy density of the fluid contributed from ρ_m, the energy density of DM and ρ_Λ, the energy density of vacuum, u_μ is the associated four-velocity and $\tilde{p} = p_m + \Pi + p_\Lambda$ is the sum of total pressure of fluid contributed from p_m, the equilibrium pressure, Π, the non-equilibrium pressure due to bulk viscosity and p_Λ, the pressure due to vacuum. It is assumed that deviation of any system from the local thermodynamical equilibrium is the source of bulk viscosity. In accordance with the second law of thermodynamics, the restoration of thermal equilibrium is a dissipative process that produces entropy. As a result of this entropy generation, the bulk viscous term causes an expansion in the system.

The viscous fluid in homogeneous and isotropic cosmological models is determined by its bulk viscosity. The Eckart's formalism serves as the basis for this theory[41]. It is basically obtained from the second order theory of non-equilibrium thermodynamics in the limit of vanishing relaxation time which was proposed by Israel and Stewart [42]. Inspired by the viscosity behavior in fluid mechanics, being proportional to the speed, we assume $\Pi = -3\zeta H$, where H is the Hubble parameter and ζ is the bulk viscous coefficient, which is assumed to be positive on thermodynamic grounds. Furthermore, we consider the non-relativistic matter with $p_m = 0$ to be the bulk viscous fluid. Therefore, the sum of the vacuum energy pressure, $p_\Lambda = -\rho_\Lambda$ and viscous pressure $\Pi = -3\zeta H$ are the components contributing to the total pressure. These two extra ingredients have been introduced to get a more realistic fluid description of DM and DE and also a suitable comparison with ΛCDM model.

In the presence of a non-gravitational interaction between viscous dark matter and decaying vacuum energy, which is characterized by a coupling function $Q(t)$, also known as the interacting rate, the friedmann equation and conservation equations can be written as

$$3H^2 = \rho_m + \rho_\Lambda, \tag{4}$$

$$\dot{\rho}_m + 3H\rho_m = 9\zeta H^2 + Q(t), \tag{5}$$

$$\dot{\rho}_\Lambda + 3H(\rho_\Lambda + p_\Lambda) = -Q(t), \tag{6}$$

where dot denotes derivative with respect to the cosmic time t. In Eqs.(5) and (6), $Q(t)$ denotes the interaction function providing the rate of energy transfer between viscous dark matter and decaying VED. It is to be noted that $Q(t) < 0$ gives energy transfer from viscous DM to decaying vacuum energy where as $Q(t) > 0$ gives energy transfer from decaying vacuum energy to viscous DM. Once the interaction function $Q(t)$ is specified, the background dynamics of the model can be found using (4)-(6). We can define $Q(t)$ by two ways: either by deriving the interaction function from some fundamental physics at the Lagrangian level or by assuming at phenomenological level and testing using observational data. Due to the absence of a fundamental physical theory, we will consider the second approach by assuming the interaction function as [97, 98]

$$Q = 3\alpha\rho_m H, \tag{7}$$

where the term α denotes the dimensionless coupling parameter and included in the fitting vector of free parameters which is to be constrained by the observational dataset.

For decaying VED, we assume a phenomenological application of renormalization group analysis, which can be written as [33, 40]

$$\rho_\Lambda = c_0 + 3\nu H^2, \tag{8}$$

where $c_0 = 3H_0^2(\Omega_\Lambda - \nu)$ is the additive constant and fixed by the boundary condition $\rho_\Lambda(H_0) = \rho_{\Lambda 0}$. In this scenario, ν is the dimensionless vacuum parameter and is naturally anticipated to have extremely small magnitude i.e. $|\nu| \ll 1$. Thus, the positive magnitude of ν enables the vacuum's cosmic evolution. In this instance, we will fit ν to the cosmological data set by taking it as a free parameter.

Substituting (7) into (5), we get

$$\dot{\rho}_m = 9\zeta H^2 - 3(1-\alpha)\rho_m H \tag{9}$$

From (4)-(9), we get the following Hubble evolution equation,

$$\dot{H} + \frac{3}{2}(1-\alpha)H^2 = \frac{3}{2}\frac{\zeta}{(1-\nu)}H + \frac{1}{2}\left(\frac{1-\alpha}{1-\nu}\right)c_0. \tag{10}$$

The above evolution equation (10) has H and ζ as unknown quantities. We can get the solution of H only if the functional form of ζ is specified. We consider the bulk viscous coefficient ζ as proportional to the expansion rate of the Universe, i.e., to the Hubble parameter H, which can be expressed as [48, 52, 59, 67, 99, 100]

$$\zeta = \zeta_1 H \tag{11}$$

where ζ_1 is a dimensionless constant to be estimated from the observations. Substituting this relation into (10), we get

$$\dot{H} + \frac{3}{2}\left(\frac{(1-\nu)(1-\alpha) - \zeta_1}{1-\nu}\right)H^2 - \frac{1}{2}\left(\frac{1-\alpha}{1-\nu}\right)c_0 = 0. \tag{12}$$

The preceding equation with a variable change from t to $x = \log a$ together with $c_0 = 3H_0^2(\Omega_\Lambda - \nu)$ can be written as

$$\frac{dh^2}{dx} + 3\left(\frac{(1-\nu)(1-\alpha)-\zeta_1}{1-\nu}\right)h^2 = 3\left(\frac{(\Omega_\Lambda-\nu)(1-\alpha)}{1-\nu}\right), \tag{13}$$

where $h = H/H_0$ represents the Hubble parameter which is dimensionless and $\Omega_\Lambda = \rho_\Lambda/3H_0^2$. Assuming $(1-\nu)(1-\alpha)-\zeta_1 > 0$ and employing the redshift relation to the normalised scale factor, $a = (1+z)^{-1}$, Eq.(13) gives

$$E^2(z) = \tilde{\Omega}_{m,0}(1+z)^{3\left(\frac{(1-\nu)(1-\alpha)-\zeta_1}{(1-\nu)}\right)} + \tilde{\Omega}_{\Lambda,0}, \tag{14}$$

where

$$\tilde{\Omega}_{\Lambda,0} = \frac{(\Omega_\Lambda-\nu)(1-\alpha)}{(1-\nu)(1-\alpha)-\zeta_1}, \tag{15}$$

and

$$\tilde{\Omega}_{m,0} = 1 - \frac{(\Omega_\Lambda-\nu)(1-\alpha)}{(1-\nu)(1-\alpha)-\zeta_1} = 1 - \tilde{\Omega}_{\Lambda,0}. \tag{16}$$

In Eq.(14), $E(z) = H/H_0$ is dimensionless Hubble parameter and $\Omega_{i,0}$ (i = DM,DE) represents the current value of density parameter of viscous DM and decaying vacuum energy. For $E(0) = 1$, we have $\tilde{\Omega}_{m,0} + \tilde{\Omega}_{\Lambda,0} = 1$. The scale factor of expansion can be calculated by using $H = \dot{a}/a$ and integrating the equation (14), the solution for the scale factor $a(t)$ is given by

$$a(t) = \left(\frac{\tilde{\Omega}_{m,0}}{\tilde{\Omega}_{\Lambda,0}}\right)^{\frac{(1-\nu)}{3[(1-\alpha)(1-\nu)-\zeta_1]}} \left(\sinh^{2/3}\left(\frac{t}{\tilde{t}}\right)\right)^{\frac{(1-\nu)}{(1-\alpha)(1-\nu)-\zeta_1}}, \tag{17}$$

where $\tilde{t} \equiv 2/(3H_0\sqrt{\frac{[(1-\nu)(1-\alpha)-\zeta_1](1-\alpha)(\Omega_\Lambda-\nu)}{(1-\nu)^2}})$. It can be observed from (17) that the scale factor reduces to $a(t) = (\Omega_{m,0}/\Omega_{\Lambda,0})^{1/3}\ \sinh^{2/3}(t/\tilde{t})$ for $\alpha = 0$, $\zeta_1 = 0$ and $\nu = 0$, which is the analytical solution of the scale factor for ΛCDM model. It can be observed from above equation that the scale factor varies as $a \propto t^{\frac{2(1-\nu)}{3[(1-\nu)(1-\alpha)-\zeta_1]}}$ during early times which the power-law expansion of the Universe. For late-time evolution the scale factor varies as $a \propto \exp\sqrt{\frac{(\Omega_\Lambda-\nu)}{3(1-\zeta_1-\nu)}}H_0 t$, which implies the de Sitter Universe.

To investigate the decelerated and accelerated phases of the expansion of the Universe as well as its transition during the evolution, we explore a crucial cosmological parameter, called 'deceleration parameter'. The definition of deceleration parameter q is

$$q = -\frac{\ddot{a}}{a}\frac{1}{H^2} = -\left(1 + \frac{\dot{H}}{H^2}\right). \tag{18}$$

Using (14) into (18), the deceleration parameter q in terms of redshift z is given by

$$q(z) = -1 + \frac{3}{2(1-\nu)}\left(\frac{((1-\alpha)(1-\Omega_\Lambda)-\zeta_1)(1+z)^{3\left(\frac{(1-\nu)(1-\alpha)-\zeta_1}{(1-\nu)}\right)}}{\frac{(\Omega_\Lambda-\nu)(1-\alpha)}{(1-\nu)(1-\alpha)-\zeta_1} + \left(1 - \frac{(\Omega_\Lambda-\nu)(1-\alpha)}{(1-\nu)(1-\alpha)-\zeta_1}\right)(1+z)^{3\left(\frac{(1-\nu)(1-\alpha)-\zeta_1}{(1-\nu)}\right)}}\right) \tag{19}$$

The aforementioned equation demonstrates how the redshift affects the dynamics of q . We note that the value of $q(z)$ approaches -1 in the future (negative redshift). Furthermore, we determine the present value of q for $z = 0$, denoted as (q_0), which is given by

$$q_0 = -1 + \frac{3}{2(1-\nu)}\left[(1-\alpha)(1-\Omega_\Lambda)-\zeta_1\right]. \tag{20}$$

The accelerating and decelerating phase of the Universe can be inferred from the negative and positive signs of the deceleration parameter, respectively. For $q_0 > 0$, the Universe exhibits expanding behaviour and goes through a deceleration phase. For $-1 < q_0 < 0$, represents the current state of the Universe which is the expanding and accelerating Universe.

For the sake of completion, we further calculate effective equation of state parameter w_{eff} as a function of redshift z,

$$w_{eff}(z) = -1 + \frac{1}{(1-\nu)}\left(\frac{((1-\alpha)(1-\Omega_\Lambda)-\zeta_1)(1+z)^{3\left(\frac{(1-\nu)(1-\alpha)-\zeta_1}{(1-\nu)}\right)}}{\frac{(\Omega_\Lambda-\nu)(1-\alpha)}{(1-\nu)(1-\alpha)-\zeta_1} + \left(1 - \frac{(\Omega_\Lambda-\nu)(1-\alpha)}{(1-\nu)(1-\alpha)-\zeta_1}\right)(1+z)^{3\left(\frac{(1-\nu)(1-\alpha)-\zeta_1}{(1-\nu)}\right)}}\right) \tag{21}$$

At $z = 0$, the present value of w_{eff} is determined by

$$w_{eff}(z=0) = -1 + \frac{1}{(1-\nu)}\left[(1-\alpha)(1-\Omega_\Lambda)-\zeta_1\right]. \tag{22}$$

III. GROWTH OF PERTURBATIONS IN INTERACTING MODEL

The presence of cosmological fluctuations influences the background cosmology in which the perturbations

evolve. Hence for the complete analysis of our interacting viscous $\Lambda(t)$ model we must take into account the effects on the large scale structure (LSS) formation data, which we incorporate into our observational analysis, in order to fully confront the model. As a result, we must take the matter density perturbations into consideration. Details on this portion of the analysis has been provided in many references [101, 102]. Here, we simply cite the resulting differential equation, which is entirely consistent with the analysis of [80] and references therein. For the interacting viscous $\Lambda(t)$ model, we use the standard perturbations equation for the linear matter density contrast $\delta_m \equiv \delta\rho_m/\rho_m$ as given below:

$$\delta_m'' + \left(\frac{3}{a} + \frac{H'(a)}{H(a)}\right)\delta_m' - \frac{4\pi G\rho_m}{H^2(a)}\frac{\delta_m}{a^2} = 0 \tag{23}$$

Here $()' \equiv d/da$ represents the derivative with respect to the scale factor. The aforementioned approximation to structure formation is sufficient for considering the primary consequences originate from the distinct expression of the Hubble function as compared to the ΛCDM. Thus, the preceding second-order differential equation (23) seems to be accurate. We examine the Hubble function (14) as it was determined in Section II. The smoothness of the matter perturbation in the interacting viscous $\Lambda(t)$ model is described by the equation (23).

The weighted linear growth, $f(z)\sigma_8(z)$, is typically used to compare the theoretical calculations with the structure formation data in the linear regime. Here, $\sigma_8(z)$ represents the r.m.s. mass fluctuations on $R_8 = 8h^{-1}$ Mpc at redshift z and $f(z) = d\ln\delta_m(a)/d\ln(a)$ is the growth factor. The details of the calculation are provided in coming Section-IV(C).

IV. DATA AND METHODOLOGY

In this section, we use a variety of observational data and methodology in order to constrain the model parameters of ΛCDM and interacting viscous Λ(t) models which comprise observations from: (i) Pantheon+ dataset; (ii) Hubble dataset(Cosmic Chronometers); and (iii) $f(z)\sigma_8(z)$ dataset. The following subsections provide an overview of each data set.

A. Pantheon+ dataset

We consider an updated Pantheon+ sample [103] which is the successor of the Pantheon sample. This compilation comprises of 1701 light curves gathered by 1550 type Ia supernovae (SNe), providing detailed information for cosmological analysis. In addition to the most recent SNe sightings, the Pantheon+ collection expands upon earlier SNe compilations and offers a broad range of redshifts ranging from $z = 0.00122$ to 2.2613. The Chi-squared related to the Pantheon+ dataset is given by

$$\chi^2_{Pan+} = \sum_{i=1}^{1701} \vec{D}^T C^{-1}_{Pan+}\vec{D} \tag{24}$$

where $\vec{D} = m_{Bi} - M - \mu_{model}$. Here, m_{Bi} and M are the observed peak magnitude in the rest frame of the B band and absolute B-band magnitude respectively. The theoretical distance modulus μ_{model} is defined by

$$\mu_{model}(z_i) = 5\log_{10}\left(\frac{D_L(z_i)}{1Mpc}\right) + 25, \tag{25}$$

where $D_L(z)$ represents the luminosity distance, that is described by $D_L(z) = (1+z)H_0\int_0^z \frac{dz'}{H(z')}$. In Eq. (24), C denotes the total covariance matrix which takes the form $C = D_{stat} + C_{sys}$, where the diagonal matrix D_{stat} and covariant matrix C_{sys} denote the statistical uncertainties and the systematic uncertainties.

In contrast to the Pantheon dataset, the degeneracy between the absolute magnitude M and H_0 is broken in $Pantheon+$. In order to accomplish this, the vector $\vec{D}$ in Eq. (24) is written in terms of the distance moduli of SNe in Cepheid hosts. As a result, M can have an independent constraint, which leads to the following expression:

$$\vec{D'} = \begin{cases} m_{Bi} - M - \mu_i^{Cepheid} & i \in Cepheid \\ m_{Bi} - M - \mu_{model}(z_i) & Otherwise \end{cases} \tag{26}$$

where $\mu_i^{Cepheid}$ denotes the distance modulus, determined independently using Cepheid calibrators, and corresponds to the Cepheid host of the ith SNe. Hence, we can rewrite Eq.(24) as follows:

$$\chi^2_{Pan+} = \sum_{i=1}^{1701} \vec{D'}^T C^{-1}_{Pan+}\vec{D'} \tag{27}$$

B. Cosmic chronometers

The cosmic chronometers (CC) is another observational data obtained through the differential-age method. We can find the Hubble parameter values from CC method at distinct redshifts based on the relative age of passively evolving galaxies. For our analysis, we employ a compilation of 32 data points of the Hubble parameter derived through the differential age technique [104] in the redshift range $0.07 \le z \le 1.965$ with errors. The Chi-squared function for CC is given by

$$\chi^2_{CC} = \sum_{i=1}^{32} \frac{[H(z,\theta) - H^{obs}(z_i)]^2}{\sigma^2_{H(z)}} \tag{28}$$

where $H(z)$ represents the theoretical prediction of the Hubble parameter calculated in Eq. (14) and $H^{obs}(z)$ are the observational data with errors $\sigma_{H(z)}$.

TABLE I. Flat prior used on various parameters during statistical analysis.

Parameters	Priors
H_0	$[50, 100]$
Ω_Λ	$(0, 1]$
ζ_1	$(0, 1]$
ν	$(0, 1]$
α	$(0, 1]$
σ_8	$[0.6, 1.2]$

C. $f(z)\sigma_8(z)$ data

Finally, in this work we use the recent "Gold -17" compilation consisting of 18 independent measurements of $f(z)\sigma_8(z)$. These data points are based on Redshift Space Distortion (RSD) measurements from various observations of the Large Scale Structure (LSS) and complied in Table III of Ref. [105].

The linear growth rate of matter perturbations is defined as follows:

$$f(z) = \frac{d\ln\delta_m}{d\ln a} \equiv -(1+z)\frac{\delta'_m(z)}{\delta_m(z)} \tag{29}$$

Here, (') denotes the derivative with respect to redshift. Further, in the linear regime $\sigma_8(z)$ is the redshift-dependent parameter which measures the growth of root-mean-square mass fluctuations in spheres with radius $8h^{-1}$ Mpc scales which is given as:

$$\sigma_8(z) = \sigma_8(z=0)\frac{\delta_m(z)}{\delta_m(z=0)} \tag{30}$$

Combining (29) and (30), the weighted linear growth is determined by:

$$f\sigma_8(z) = -(1+z)\frac{\sigma_8(z=0)}{\delta_m(z=0)}\frac{d\delta_m}{dz} \tag{31}$$

The Chi-squared function of $f(z)\sigma_8(z)$ can be calculated as [106]:

$$\chi^2_{f(z)\sigma_8(z)} = \sum_{i=1}^{18}\frac{[f\sigma_8(z_i) - (f\sigma_{8,i})_{ob}]^2}{\sigma_i^2} \tag{32}$$

Based on the EMCEE [107] python module, we employ the Markov Chain Monte Carlo(MCMC) statistical technique to examine the parameter space of our cosmological models and minimise the χ^2 function for both ΛCDM and interacting viscous $\Lambda(t)$ models. We consider the joint analysis by assuming the sum of all χ^2 functions:

$$\chi^2_{tot} = \chi^2_{Pan+} + \chi^2_{CC} + \chi^2_{f\sigma_8(z)}. \tag{33}$$

To perform this analysis, we choose uniform priors for the parameters of models which is listed in Table I.

TABLE II. The values of parameters for ΛCDM and interacting viscous $\Lambda(t)$ models for combination of Pantheon+, Hubble(CC) and $f(z)\sigma_8(z)$ observational datasets.

Parameter	ΛCDM	Interacting viscous $\Lambda(t)$
H_0	$72.150^{+0.989}_{-0.779}$	$72.202^{+0.796}_{-0.937}$
Ω_Λ	$0.689^{+0.016}_{-0.017}$	$0.681^{+0.014}_{-0.017}$
M	$-19.292^{+0.026}_{-0.023}$	$-19.290^{+0.024}_{-0.026}$
ζ_1	$-$	$0.010^{+0.006}_{-0.007}$
ν	$-$	$0.005^{+0.003}_{-0.003}$
α	$-$	$-0.005^{+0.003}_{-0.004}$
σ_8	$0.750^{+0.034}_{-0.031}$	$0.739^{+0.034}_{-0.030}$
S_8	$0.765^{+0.030}_{-0.030}$	$0.762^{+0.030}_{-0.030}$
q_0	$-0.533^{+0.024}_{-0.024}$	$-0.531^{+0.024}_{-0.024}$
z_{tr}	$0.643^{+0.041}_{-0.041}$	$0.647^{+0.038}_{-0.044}$
w_0	$-0.689^{+0.016}_{-0.016}$	$-0.687^{+0.016}_{-0.016}$
$t_0(Gyr)$	$13.74^{+0.019}_{-0.021}$	$13.75^{+0.023}_{-0.021}$

V. RESULT AND DISCUSSION

In this section, we present the results of our observational analysis on ΛCDM and interacting viscous $\Lambda(t)$ models, including both the constraints and cosmological parameters from the current data. In the first subsection, we will present the parameters constraints achieved by observational analysis using the data discussed in Sect.IV. In second subsection, we will implement the Bayesian inference, and in last subsection, we will examine the model comparison through information criterion.

A. Parameter constraints

The best-fit values of parameters of ΛCDM and interacting viscous $\Lambda(t)$ models using a combined $CC + f(z)\sigma_8(z) + Pan+$ data are summarized in Table II. Figures 1-3 show the 68.3% and 95.4% confidence regions and marginalized likelihood distributions for ΛCDM and interacting viscous $\Lambda(t)$ models, respectively. The GetDist code [108] is utilised to retrieve their mean values and the aforementioned Figs. 1-3. In what follows we discuss the constraints on different cosmological as well as model parameters.

It is noted that the H_0 determined by Riess et al.[109] is $73.04 \pm 1.04\ kms^{-1}Mpc^{-1}$, so-called R21 where as Planck Collaboration [110] predicts $H_0 = 67.4 \pm 0.5\ kms^{-1}Mpc^{-1}$ at 5σ confidence level. This discrepancy is recognized as "Hubble tension". For ΛCDM model we obtain $H_0 = 72.150^{+0.989}_{-0.779}\ kms^{-1}Mpc^{-1}$. This is in tension with both Planck [110] and R21 [109] results at 4.67σ and 0.65σ respectively. For interacting viscous $\Lambda(t)$ model, we obtain $H_0 = 72.202^{+0.796}_{-0.937}\ kms^{-1}Mpc^{-1}$ which is tension with both Planck and R21 results at 4.80σ and 0.86σ respectively. In other words, the H_0 measurement of interacting viscous $\Lambda(t)$ model is consistent with R21. The evolutions of $H(z)$ with redshift z are shown in

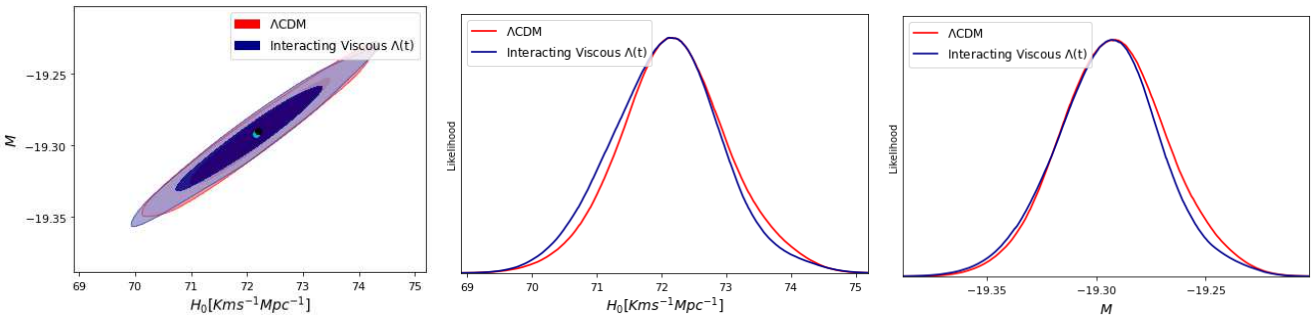


FIG. 1. The 68.3% (dark shaded) and 95.4% (light shaded) confidence contours and likelihoods for H_0 and M for the ΛCDM and interacting viscous $\Lambda(t)$ models using combined dataset. The best-fit values for these models are represented by the cyan and black dot respectively.

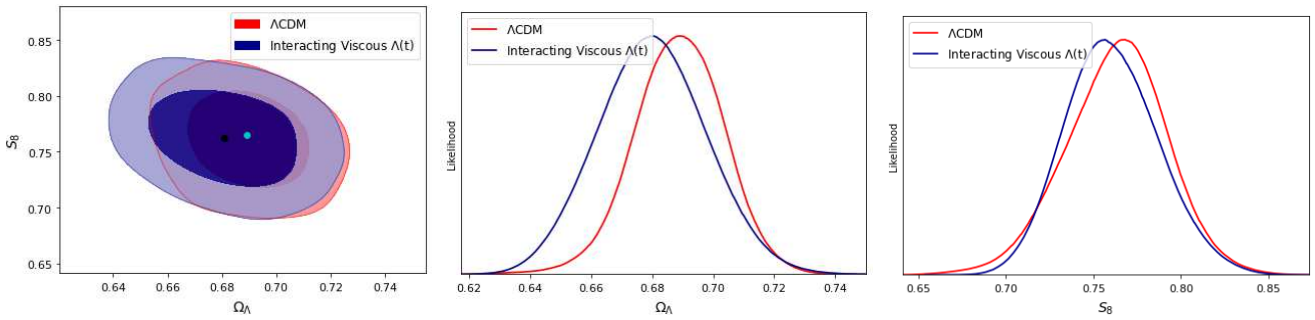


FIG. 2. The 68.3% (dark shaded) and 95.4% (light shaded) confidence contours and likelihoods for Ω_Λ and S_8 for the ΛCDM and interacting viscous $\Lambda(t)$ models using combined dataset. The best-fit values for these models are represented by the cyan and black dot respectively.

Fig.4 which predict that the trajectories cover majority of dataset with error bars of $H(z)$. This means that the interacting viscous $\Lambda(t)$ model agrees with the combination of dataset as well as ΛCDM model.

In the interacting viscous $\Lambda(t)$ model, we obtain the viscosity coefficient, ζ_1 as $0.010^{+0.006}_{-0.007}$ and ν as $0.005^{+0.003}_{-0.003}$. The coupling parameter of the interaction term, α is calculated as $-0.005^{+0.003}_{-0.004}$. The negative value of α indicates that there is a possibility of the dark matter to decay into the dark energy.

Using the best-fit values, the evolutions of deceleration parameter with redshift for ΛCDM and interacting viscous $\Lambda(t)$ models with errors are plotted in Fig.5. The trajectories show that the models have transition from decelerating phase to accelerating phase at transition redshift $z_{tr} = 0.643^{+0.041}_{-0.041}$ and $z_{tr} = 0.647^{+0.038}_{-0.044}$, respectively. The present value of deceleration parameter are found to be $q_0 = -0.533^{+0.024}_{-0.024}$ and $q_0 = -0.531^{+0.024}_{-0.024}$. These values of q_0 are within the range of the observational results, i.e., $q_0 = -0.64 \pm 0.12$ [110]. In late time of evolution, $q(z)$ approaches to -1 in both models which is the future de Sitter phase.

In Fig.6, we plot the effective EoS parameter as a function of redshift for best-fit values of ΛCDM and interacting viscous $\Lambda(t)$ models. It is observed that $w_{\rm eff} \to -1$ in the late-time in both the models, which implies that the models correspond to de Sitter phase in late-time. Using the best-fit, we get the present EoS parameters $w_{\rm eff}(z = 0) = -0.689 \pm 0.016$ and $w_{\rm eff}(z = 0) = -0.687 \pm 0.016$, respectively.

Next, we explore another tension between the theoretical prediction of the growth rate of matter perturbations with the observational growth rate data points for ΛCDM and interacting viscous $\Lambda(t)$ models. For both models the constraints on σ_8 and S_8 are given in Table II. The amplitude of the matter power spectrum (σ_8) and its associated parameter $S_8 = \sigma_8\sqrt{(1-\Omega_\Lambda)/0.3}$ may be determined using the $f(z)\sigma_8(z)$ dataset and it is further possible to calculate the σ_8/S_8 tension. The combined dataset gives $\sigma_8 = 0.750^{+0.034}_{-0.031}$ and $S_8 = 0.765^{+0.030}_{-0.030}$ in the ΛCDM model and $\sigma_8 = 0.739^{+0.034}_{-0.030}$ and $S_8 = 0.762^{+0.030}_{-0.030}$ in interacting viscous $\Lambda(t)$ model. Our results are perfectly consistent with the combined use of the SDSS and KiDS/Viking data which gives $\sigma_8 = 0.760^{+0.025}_{-0.020}$ and $S_8 = 0.766^{+0.020}_{-0.014}$ [111]. The interacting viscous $\Lambda(t)$ model is in 2.21σ tension in σ_8 and 2.11σ tension in S_8 with the corresponding values of Plank [110]. These tensions are not as large compared to H_0 tension as discussed above. The confidence contours in (S_8, Ω_Λ) and (S_8, σ_8) parameter space and corresponding likelihoods are shown in Figs. 2 and 3, which show no deviation from ΛCDM contours. This confirms that the measurement of

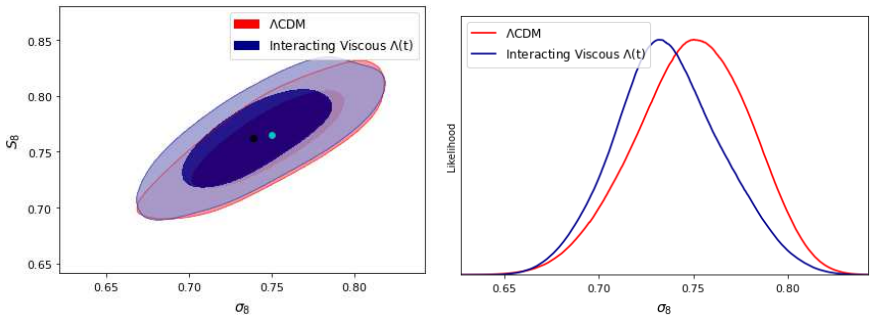


FIG. 3. The 68.3% (dark shaded) and 95.4% (light shaded) confidence contours and likelihoods for S_8 and σ_8 for the ΛCDM and interacting viscous $\Lambda(t)$ models using combined dataset. The best-fit values for these models are represented by the cyan and black dot respectively.

σ_8 and S_8 using $CC+Pan+$ with $f(z)\sigma_8(z)$ data is fully consistent with ΛCDM. The trajectories of $f(z)\sigma_8(z)$ for ΛCDM and interacting viscous $\Lambda(t)$ models are plotted in Fig.7. It can be observed that both the models are consistent with the observational data points.

Let us discuss the interacting viscous $\Lambda(t)$ model by defining a dimensionless parameter, known as jerk parameter, j. This parameter is purely kinematical and is associated with the third order derivative of the scale factor $a(t)$. It is defined as $j = \frac{\dddot{a}(t)}{aH^3}$ [112]. In terms of the deceleration parameter, it can be expressed as follows [112]:

$$j(z) = q(2q+1) + (1+z)\frac{dq}{dz}. \tag{34}$$

It can provide us the deviation of any model from the ΛCDM model. It is noted that jerk parameter always has constant value $j = 1$ for ΛCDM model. We report $j_0 = 0.992$ in interacting viscous $\Lambda(t)$ model which is quite comparable to the ΛCDM model. We plot the evolution of jerk parameter with respect to the redshift as shown in Fig.8. We observe that the trajectory of $j(z)$ deviates from ΛCDM in early phase where as it approaches to $j = 1$ as $z \to -1$. Thus, jerk parameter points us the effects of interacting viscous $\Lambda(t)$ model over the ΛCDM model.

B. Model selection

In this section, we will discuss mainly reduced Chi-square and model selection criterion to observe the compatibility of the proposed model.

Let us first calculate the reduced Chi-square, χ^2_{red}, which is defined as $\chi^2_{red} = \chi^2_{min}/dof$. Here, dof denotes the degree of freedom which is equal to the number of observational data points (N) minus the number of parameters (K). It is to be noted that we have used $N = 1751$ and $K = 3$ in ΛCDM and $K = 6$ in interacting viscous $\Lambda(t)$ model with same number of data points. Table II presents the χ^2 and χ^2_{red} of ΛCDM and interacting viscous $\Lambda(t)$ models, respectively. It is noted that

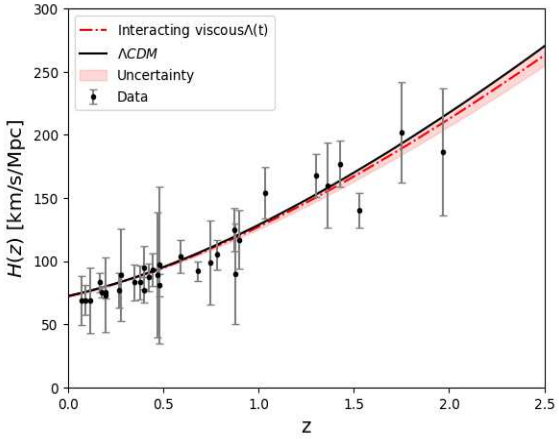


FIG. 4. The evolution of hubble function $H(z)$ with redshift z. The solid black line corresponds to the ΛCDM model and the dashed red line corresponds to the interacting viscousΛ(t) model($\zeta = \zeta_1 H$). The $H_{obs}(z)$ data are also plotted with their error bars.

a value $\chi^2_{red} < 1$ gives the best-fit with the data. We observed that both the models have the reduced χ^2 less than unity (for ΛCDM, $\chi^2_{red} = 0.447$ and for interacting viscous $\Lambda(t)$ model, it is $\chi^2_{red} = 0.448$). This shows that both the models are in very good fit with the observational data.

We discuss the another method which provides the statistical comparison of the proposed model with ΛCDM model. In this regard, there are two model selection criterion, namely, Akaike information criteria (AIC) and Bayesian information criteria (BIC). These selection criterion allow to compare models with different degree of freedom. They are defined as [113, 114]

$$AIC = \chi^2_{min} + \frac{2KN}{N-K-1}, \qquad BIC = \chi^2_{min} + K\ln N, \tag{35}$$

where N is the size of the data sample and K is the number of free parameters. In these approaches, the model with low values of AIC(BIC) is preferred by data.

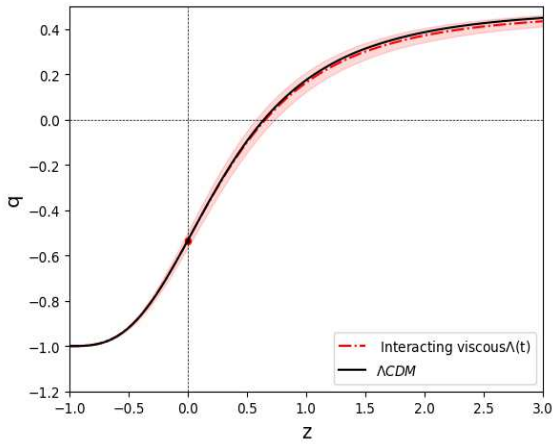


FIG. 5. The evolution of $q(z)$ with redshift z for interacting viscous$\Lambda(t)$ model and ΛCDM model using the best fit values. Present value q_0 is represented by dot.

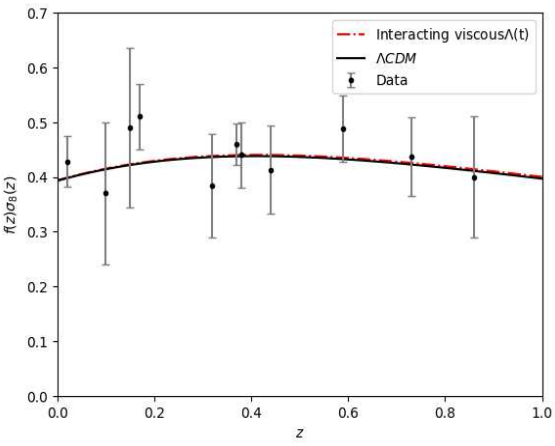


FIG. 7. The evolution of $f(z)\sigma_8(z)$ with redshift z for interacting viscous$\Lambda(t)$ model and ΛCDM model using the best fit values.

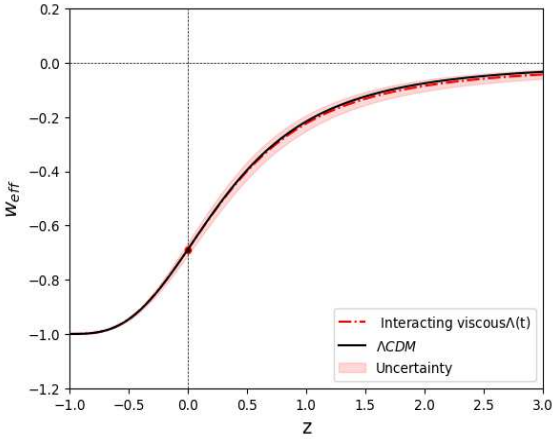


FIG. 6. The evolution of w_{eff} with redshift z for interacting viscous$\Lambda(t)$ model and ΛCDM model using the best fit values. Present value of $w_{eff}(z=0)$ is represented by dot.

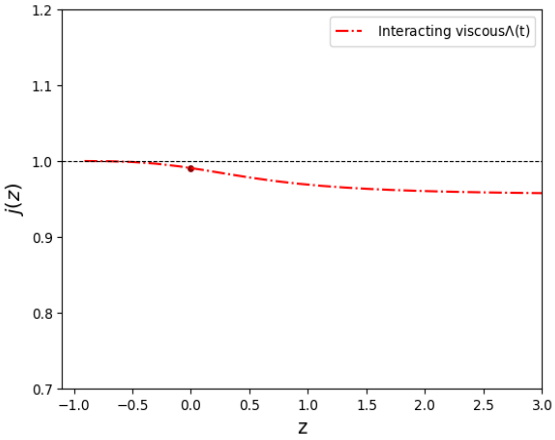


FIG. 8. The evolution of $j(z)$ with redshift z for interacting viscous$\Lambda(t)$ model the best fit values. Present value of $j_0(z=0)$ is represented by dot and the horizontal line $j=1$ represents ΛCDM model.

Considering the AIC (BIC) of ΛCDM model as reference model, denoted as $AIC_\Lambda(BIC_\Lambda)$, we compute $\Delta AIC = AIC_{model} - AIC_{\Lambda CDM}(\Delta BIC = BIC_{model} - BIC_{\Lambda CDM})$.

Following the rules described in Refs.[80, 96], let us discuss our results. Table IV displays the values of AIC (BIC) and corresponding ΔAIC (ΔBIC) for interacting viscous $\Lambda(t)$ model. According to our results, we have $\Delta AIC = 6.670$ and $\Delta BIC = 23.039$. This shows that there is "less support in favor" of the interacting viscous $\Lambda(t)$ model as far as AIC is concerned where as BIC gives "strong evidence against" the interacting viscous $\Lambda(t)$ model.

C. Bayesian Inference

The Bayesian evidence serves as the foundation for assessing a model's performance in light of the data. In Bayesian data analysis, the correlation between the data, the model or hypotheses, and the prior knowledge is characterized by the joint probability distributions. Given the observed data, the conditional probability distribution of the unknowns can be used to uniquely infer the posterior distribution using Bayes theorem. The key statistic of Bayesian model selection is the Bayesian evidence $\mathcal{E}$ [115, 116], which is employed in a model comparison problem by integrating the product of likelihood and posterior over the entire parametric space of the model.

TABLE III. "Jeffreys' scale" for evaluating the strength of evidence between two comparing models, M_i versus M_j. The right column provides the convention for denoting the different levels of evidence above these thresholds.

$\ln \mathcal{B}_{ij}$	Strength of Evidence
< 1.0	Inconclusive
1.0	Weak evidence
2.5	Moderate evidence
5.0	Strong evidence

It is defined as [117–120]

$$\mathcal{E}(D|M) = \int_M \mathcal{L}(D|\Theta, M)\mathcal{P}(D|\Theta, M)d\Theta. \tag{36}$$

On the right-hand side, Θ is a set of free parameters for the given data D, M stands for the model, the vertical bar reads as 'given', and $\mathcal{L}$ and $\mathcal{P}$ stand for likelihood and prior probability distribution function of those parameters before the data, respectively.

In cosmology, this concept has been utilized extensively [121, 122]. When comparing two models, M_i versus M_j, one is interested in the ratio of the models' evidences known as *Bayes factor*, which is given by:

$$\mathcal{B}_{ij} \equiv \frac{\mathcal{E}_i}{\mathcal{E}_j} \tag{37}$$

Here, B_{ij} indicates the support for model i over model j.

Bayes factors are typically interpreted using the Jeffreys' scale [115] which measures the strength of the evidence, given in Table III. We accomplish this by estimating the values of the logarithm of Bayes factor ($\ln \mathcal{B}$) and the Bayesian evidence ($\ln \mathcal{E}$). Using the dataset and the priors mentioned in Table I, the values of $\mathcal{E}$ and $\mathcal{B}$ are determined. We assume the ΛCDM as the reference model.

The results for the Bayesian evidence ($\ln \mathcal{E}$) and Bayes factor ($\ln \mathcal{B}$) for both ΛCDM and interacting viscous Λ(t) models examined in this work are summarized in Table IV. We find that the $\ln \mathcal{B}$ for interacting viscous Λ(t) model with respect to the ΛCDM model is obtained as 2.577 and thus the Bayesian evidence analysis shows that our model is "moderately supported" by the considered priors and dataset on Jeffreys'scale.

TABLE IV. Summary of $\ln \mathcal{E}$, $\ln \mathcal{B}$, χ^2, χ^2_{red}, AIC and BIC for the ΛCDM and interacting viscous$\Lambda(t)$ models for the following:

Values	ΛCDM	Interacting viscous$\Lambda(t)$
$\ln \mathcal{E}$	-793.120 ± 0.473	-795.697 ± 0.532
$\ln \mathcal{B}$	–	2.577
χ^2	781.708	782.344
K	3	6
N	1751	1751
χ^2_{red}	0.447	0.448
AIC	787.722	794.392
ΔAIC	–	6.670
BIC	804.112	827.152
ΔBIC	–	23.039

VI. CONCLUSION

Inspired by dissipative phenomena and decaying vacuum energy, we have discussed some cosmological consequences of an alternative mechanism of accelerating Universe based on a class of interacting viscous model with decaying vacuum energy, referred as interacting viscous $\Lambda(t)$ model. The coupling between viscous fluid and vacuum energy density has been made through a coupling parameter, Q. Within the framework of Eckart thermodynamic theory the non-equilibrium pressure, Π is proportional to the Hubble parameter H with proportionality constant ζ, i.e., $\Pi = -3\zeta(t)H$. Thus, the effective pressure is assumed as sum of barotropic pressure, bulk viscous pressure and pressure due to vacuum energy, i.e., $\bar{p} = p_m + \Pi + p_\Lambda$. In the first part of work, we have obtained some main results for the scale factor, the Hubble parameter, the deceleration parameter, jerk parameter and EoS parameter by assuming the interaction term $Q = 3\alpha\rho_m H$. Although the nature of bulk viscosity and time-varying vacuum energy density are unknown, we have assumed $\zeta = \zeta_1 H$ for bulk viscous coefficient and $\rho_\Lambda = c_0 + 3\nu H^2$ for vacuum energy density. One can expect that the viscosity is affected by the expansion rate of the Universe and varying VED from the general covariance of the effective action in QFT. We have investigated the growth perturbation of interacting viscous $\Lambda(t)$ model. In the second part of this work, we have performed the Bayesian analysis using the latest background probes such as SNe Pantheon+, cosmic chronometer and $f(z)\sigma_8(z)$. We have compared the interacting viscous $\Lambda(t)$ model with ΛCDM using the Bayesian inference and model selection criterion such as AIC and BIC. In what follows, we summarize the main points of our analysis.

The constraints on model free parameters have been reported in Table II. The best-fit value of H_0 according to a combination of Pantheon+, CC and $f(z)\sigma_8(z)$ is $H_0 = 72.202^{+0.796}_{-0.937}\ kms^{-1}Mpc^{-1}$, which is tension with both Planck and R21 results at 4.80σ and 0.86σ ,respectively. In other words, the tension in H_0 measurement is almost resolved in interacting viscous $\Lambda(t)$ model with respect to R21. The best fit values of present q and w_{eff} are $q_0 = -0.531^{+0.024}_{-0.024}$ and $w_{\rm eff}(z=0) = -0.687 \pm 0.016$, respectively. From Fig.5, it has been observed that the interacting viscous $\Lambda(t)$ model exhibits transition from an early decelerated phase to late-time accelerated phase and the transition takes place at $z_{tr} = 0.647^{+0.038}_{-0.044}$ which is close proximity $z_{tr} = 0.643^{+0.041}_{-0.041}$ to that of ΛCDM model. It has been found that the value of $\chi^2_{red} < 1$ which shows that the interacting viscous $\Lambda(t)$ model is

in very good fit with the used data points. The jerk parameter remains positive and less than one in past, and tends to unity in late-time.

We have explored the σ_8 and S_8 parameters for ΛCDM and interacting viscous $\Lambda(t)$ models using the combined data set of Pantheon+, CC and $f(z)\sigma_8(z)$. The constraints on σ_8 and S_8 in interacting viscous $\Lambda(t)$ model are $\sigma_8 = 0.739^{+0.034}_{-0.030}$ and $S_8 = 0.762^{+0.030}_{-0.030}$ which are very close to ΛCDM model as reported in Table II. The tensions of our fitting results in σ_8 and S_8 with respect to Plank results [110] are 2.21σ and 2.11σ, respectively. The evolution of $f(z)\sigma_8(z)$ has been plotted in Fig. 7 which shows that it is consistent with the observational data points.

It is noted that cosmological models are testable from the abundance of observational data as discussed above. However, an important distinction must be made between parameter fitting and model selection. As we know that the parameter fitting simply tells us how well a model fit the data. Model selection such as Bayesian inference and AIC and BIC are necessary to discriminate the proposed model with the existing model. The Bayesian inference analysis demonstrated the interacting viscous Λ(t) model is moderately supported by the considered dataset and priors. Further, there has been increasing interest in applying information criterion such as AIC and BIC for model selection. We have examined the models using AIC and BIC for a fairer comparison. We have observed that the interacting viscous $\Lambda(t)$ model has "less support" according to the selection criteria ΔAIC. On contrary, with respect to ΔBIC, interacting viscous $\Lambda(t)$ model has "strong evidence against" the model with the considered datasets.

As a concluding remark we must point out that despite its intrinsic nature of bulk viscosity and decaying vacuum energy and its interaction are not well understood yet, the work presented in this paper suggests a possible description for resolving the H_0 and σ_8 tensions in cosmology. In principle, to give a robust approach for investigating the dark energy model beyond ΛCDM, background dynamics should be considered. Taking into account such interaction between the dark components may provide an opportunity to explain the present accelerating Universe. Indeed considering the interaction between the viscous fluid and decaying vacuum energy potentially enables us to resolve tensions in cosmological parameters.

ACKNOWLEDGMENTS

One of the author, VK would like to express gratitude to Delhi Technological University, India for providing Research Fellowship to carry out this work.

VII. APPENDIX

The main aim in this appendix is to present some more analytical solutions of the cosmological parameters of the interacting viscous model with decaying vacuum energy density based on the different choices of ζ.

Appendix A: Solution with $\zeta = \zeta_0$

It is the most simple parametric form of the bulk viscous coefficient. Many authors [50, 57, 64, 66, 71, 74, 75, 123–125] have studied the viscous cosmological models with constant bulk viscous coefficient. Using $\zeta = \zeta_0$ in Eq.(10), the Hubble evolution equation deduces the form

$$\dot{H} + \frac{3}{2}(1-\alpha)H^2 - \frac{3}{2}\frac{\zeta_0 H}{(1-\nu)}H = \frac{1}{2}\left(\frac{1-\alpha}{1-\nu}\right)c_0. \quad \text{(A1)}$$

Solving (A1) with the condition $H(t_0) = H_0$, we get

$$H = \frac{\zeta_0}{2(1-\nu)(1-\alpha)} + k\coth\left(\frac{3}{2}(1-\alpha)kt\right). \quad \text{(A2)}$$

where $k = \frac{{\zeta_0}^2 + 4(\Omega_\Lambda - \nu)(1-\nu)(1-\alpha)^2 {H_0}^2}{4(1-\nu)^2(1-\alpha)^2}$.
Integrating again with the condition $a(t_0) = 1$, we obtain the scale factor

$$a(t) = e^{\frac{\zeta_0}{2(1-\nu)(1-\alpha)}t}\left[\sinh\left(\frac{3}{2}(1-\alpha)kt\right)\right]^{\frac{2}{3(1-\alpha)}}, \quad \text{(A3)}$$

Using (A2), the deceleration parameter q and Effective EoS parameter w_{eff} are respectively given by

$$q = -1 + \frac{3}{2}\frac{(1-\alpha)k^2\csc^2 h(\frac{3}{2}(1-\alpha)kt)}{\left(\frac{\zeta_0}{2(1-\nu)(1-\alpha)} + k\coth(\frac{3}{2}(1-\alpha)kt)\right)^2}. \quad \text{(A4)}$$

$$w_{eff} = -1 + \frac{(1-\alpha)k^2\csc^2 h(\frac{3}{2}(1-\alpha)kt)}{\left(\frac{\zeta_0}{2(1-\nu)(1-\alpha)} + k\coth(\frac{3}{2}(1-\alpha)kt)\right)^2}. \quad \text{(A5)}$$

Appendix B: Solution with $\zeta = \zeta_0 + \zeta_1 H$

This form of bulk viscous coefficient has been discussed by several authors and the references therein [54, 59, 65]. For $\zeta = \zeta_0 + \zeta_1 H$, where ζ_0 and ζ_1 are constants, the Eq. (10) simplifies into

$$\dot{H} + \frac{3}{2}(1-\alpha)H^2 - \frac{3}{2}\frac{\zeta_1 H^2}{(1-\nu)} - \frac{3}{2}\frac{\zeta_0 H}{(1-\nu)} = \frac{1}{2}\left(\frac{1-\alpha}{1-\nu}\right)c_0, \quad \text{(B1)}$$

which can be integrated to calculate the Hubble function and the scale factor

$$H = \frac{\zeta_0}{2((1-\alpha)(1-\nu)-\zeta_1)} + k_1 \coth\left(\frac{3}{2}\frac{((1-\alpha)(1-\nu)-\zeta_1)}{(1-\nu)}k_1 t\right), \tag{B2}$$

$$a = e^{\frac{\zeta_0\, t}{2((1-\alpha)(1-\nu)-\zeta_1)}} \times \left[\sinh\left(\frac{3}{2}\frac{((1-\alpha)(1-\nu)-\zeta_1)}{1-\nu}k_1 t\right)\right]^{\frac{2(1-\nu)}{3((1-\alpha)(1-\nu)-\zeta_1)}}, \tag{B3}$$

where $k_1 = \frac{{\zeta_0}^2 + 4((1-\alpha)(1-\nu)-\zeta_1)((\Omega_\Lambda - \nu)(1-\alpha))H_0{}^2}{4((1-\nu)(1-\alpha)-\zeta_1)^2}$.

With this, we can calculate the deceleration parameter q and effective equation of state parameter w_{eff}, which are respectively given by

$$q = -1 + \frac{3(1-\zeta_1-\nu)\sigma_1^2 \csc^2 h(\frac{3}{2}(1-\zeta_1-\nu)\sigma_1 t)}{2\left(\frac{\zeta_0}{2(1-\zeta_1-\nu)} + \sigma_1 \coth(\frac{3}{2}(1-\zeta_1-\nu)\sigma_1 t)\right)^2} \tag{B4}$$

and

$$w_{eff} = -1 + \frac{(1-\zeta_1-\nu)\sigma_1^2 \csc^2 h(\frac{3}{2}(1-\zeta_1-\nu)\sigma_1 t)}{\left(\frac{\zeta_0}{2(1-\zeta_1-\nu)} + \sigma_1 \coth(\frac{3}{2}(1-\zeta_1-\nu)\sigma_1 t)\right)^2} \tag{B5}$$

[1] A. G. Riess et al., Astron. J. **116**, 1009 (1998)
[2] S. Perlmutter et al., Astrophys. J. **517**, 565 (1999)
[3] M. Persic, P. Salucci and F. Stel, Mon. Not. Roy. Astron. Soc. **281**, 27 (1996)
[4] J. Magaña and T. Matos, J. Phys. Conf. Ser. **378**, 012012 (2012)
[5] A. Hernández-Almada and M.A. García-Aspeitia, Int. J. Mod. Phys. D **27**, 1850031 (2018)
[6] S.P. Martin, Adv. Ser. Dir. High Energy Phys. **18**, 1 (1998)
[7] V.Sahni and A.Starobinsky, Int. J. Mod. Phys. A **9**, 373 (2000)
[8] P.J.E. Peebles and B. Ratra, Rev. Mod. Phys. **75**, 559 (2003)
[9] E.J. Copeland, M. Sami and S. Tsujikawa, Int. J. Mod. Phys. D **15**, 1753 (2006)
[10] S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989)
[11] T. Padmanabhan, Phys. Rept. **380**, 235 (2003)
[12] I.L. Shapiro and J. Solá, J. High Energy Phys. **202**, 006 (2002)
[13] J. Solá, J. Phys. A: Math. Theor.**41**, 164066 (2008)
[14] J Solá, J. Phys. Conf. Ser. **453**, 012015 (2013)
[15] J.C. Carvalho, J.A.S. Lima and I. Waga, Phys. Rev. D **46**, 2404 (1992)
[16] J.A.S. Lima, J.C. Carvalho, Gen. Relat. Gravit. **26**, 909 (1994)
[17] O. Bertolami, Nuovo Cim. B **93**, 36 (1986)
[18] M. Özer and M. O. Taha, Nucl. Phys. B **287**, 776 (1987)
[19] P. J. E. Peebles and B. Ratra, Astrophys. J. **325**, L17 (1988)
[20] J. M. Overduin and F. I. Cooperstock, Phys. Rev. D **58**, 043506 (1998)
[21] P. Wang and X. Meng, Class. Quantum Grav. **22**, 283 (2005)
[22] E. Elizalde, S. Nojiri, S.D. Odintsov and P. Wang, Phys. Rev. D **71**, 103504 (2005)
[23] H.A. Borges and S. Carneiro, Gen. Relativ. Gravit. **37**, 1385 (2005)
[24] S. Carneiro, C. Pigozzo, H.A. Borges and J.S. Alcaniz, Phys. Rev. D **74**, 023532 (2006)
[25] H.A. Borges, S. Carneiro, J.C. Fabris and C. Pigozzo, Phys. Rev. D **77**, 043513 (2008)
[26] S. Carneiro, M.A. Dantas, C. Pigozzo and J.S. Alcaniz, Phys. Rev. D **77**, 083504 (2008)
[27] S. Basilakos, Mon. Not. R. Astron. Soc. **395**, 2347 (2009)
[28] S. Basilakos, M. Plionis and J. Solà, Phys. Rev. D **80**, 083511 (2009)
[29] F.E.M. Costa and J.S. Alcaniz, Phys. Rev. D **81**, 043506 (2010)
[30] C. Pigozzo, M.A. Dantas, S. Carneiro and J.S. Alcaniz, JCAP **08**, 022 (2011)
[31] D. Bessada and O.D. Miranda, Phys. Rev. D **88**, 083530 (2013)
[32] J. Solà and A. Gomez-Valent, Int. J. Mod. Phys. D **24**, 1541003 (2015)
[33] J. Solá, A. Gómez-Valent and J. de Cruz Pérez, Phys. Lett.B **774**, 317 (2017)
[34] J. Solá, A. Gómez-Valent and J. de Cruz Pérez, Astrophys. J. **836**, 43 (2017)
[35] J. Solá Peracaula, J. de Cruz Pérez and A. Gómez-Valent, Mon. Not. Roy. Astron. Soc. **478**, 4357 (2018)
[36] J. Solá Peracaula, A. Gómez-Valent and J. de Cruz Pérez, Phys. Dark Univ. **25**, 100311 (2019)
[37] A. P. Jayadevan, M. Mukesh, A. Shaima, and T. K. Mathew, Astrophys. Space Sci. **364**, 67 (2019).
[38] J. Solá Peracaula, A. Gómez-Valent, J. de Cruz Pérez and C. Moreno-Pulido, Eur. Phys. Lett. **134**, 19001 (2021)
[39] C. P. Singh and J. Solá, Eur. Phys. J. C **81**, 960 (2021)
[40] V. Khatri and C. P. Singh, Phys. Dark Univ. **42**, 101300 (2023)
[41] C. Eckart, Phys. Rev. **58**, 267 (1940)
[42] W. Israel and J.M. Stewart, Phys. Lett. A **58**, 213 (1976)
[43] L.D. Landau and E.M. Lifshitz, Fluid Mechanics, Vol. **6**, Butterworth Heinermann Ltd. Oxford (1987)
[44] W.A. Hiscock and L. Lindblom, Phys. Rev.D **31**, 725 (1985)
[45] I. Müller, Z. Physik **198**, 329 (1967)
[46] G.L. Murphy, Phys. Rev. D **8**, 4231 (1973)
[47] T. Padmanabhan and S.M. Chitre, Phys. Lett. A **120**, 433 (1987)
[48] Ø. Grøn, Astrophys. Space Sci. **173**, 191 (1990)
[49] R. Maartens, Class. Quantum Grav. **12**, 1455 (1995)
[50] I. Brevik and O. Gorbunovae, Gen. Relativ. Gravit. **37**, 2039 (2005)
[51] B.D. Normann and I. Brevik, Mod. Phys. Lett. A **32**,

1750026 (2017)
[52] C.P. Singh, S. Kumar and A. Pradhan, Class. Quantum Grav. **24**, 455 (2007)
[53] M. Xin-He and D. Xu, Commun. Theor. Phys. **52**, 377 (2009)
[54] A. Avelino and U. Nucamendi, J. Cosmol. Astropart. Phys. **08**, 009 (2010)
[55] A. Hernández-Almada, Eur. Phys. J. C **79**, 751 (2019)
[56] J.C. Fabris, S.V.B. Goncalves and R. de Sá Ribeiro, Gen. Relativ. Grav. **38**, 495 (2006)
[57] M.-G. Hu and X.-H. Meng, Phys. Lett. B **635**, 186 (2006)
[58] J. Ren and X.-H. Meng, Phys. Lett. B **633**, 1 (2006)
[59] X.-H. Meng, J. Ren and M.-G. Hu, Commun. Theor. Phys. **47**, 379 (2007)
[60] J.R. Wilson, G.J. Mathews and G.M. Fuller, Phys. Rev.D **75**, 043521 (2007)
[61] G.J. Mathews, N.Q. Lan and C. Kolda, Phys. Rev.D **78**, 043525 (2008)
[62] A. Avelino and U. Nucamendi, AIP Conf. Proc. **1083**, 1 (2008)
[63] A. Avelino, U. Nucamendi and F.S. Guzman, AIP Conf. Proc. **1026**, 300 (2008)
[64] A. Avelino and U. Nucamendi, J. Cosmol. Astropart. Phys. **04**, 006 (2009)
[65] X.H. Meng and X. Dou, Commun. Theor. Phys. **52**, 377 (2009)
[66] N. Mostafapoor and Ø. Grøn, Astrophys. Space Sci. **333**, 357 (2011)
[67] C.P. Singh and P. Kumar, Eur. Phys. J. C **74**, 3070 (2014)
[68] A. Sasidharan and T.K. Mathew, Eur. Phys. J. C **75**, 348 (2015)
[69] K. Bamba and S.D. Odintsov, Eur. Phys. J. C **76**, 18 (2016)
[70] D. Wang, Y.-J. Yan and X.-H. Meng, Eur. Phys. J. C **77**, 660 (2017)
[71] C.P. Singh and A. Kumar, Eur. Phys. J. Plus **133**, 312 (2018)
[72] C.P. Singh and A. Kumar, Mod. Phys. Lett. A **33**, 1850225 (2018)
[73] C.P. Singh and M. Srivastava, Eur. Phys. J. C **78**, 190 (2018)
[74] C.P. Singh and A. Kumar, Astrophys. Space Sci. **364**, 94 (2019)
[75] C.P. Singh and S. Kaur, Astrophys. Space Sci **365**, 2 (2020)
[76] J. Hu and H. Hu, Eur. Phys. J. Plus **135**, 718 (2020)
[77] L. Herrera-Zamorano, A. Hernández-Almada and Miguel A. García-Aspeitia, Eur. Phys. J. C **80**, 637 (2020)
[78] A. Ashoorioon and Z. Davari, Astrophys. J. **959**, 120 (2023)
[79] N. Cruz, G. Gómez, E. González, G. Palma and A. Rincón, Phys. Dark Univ. **42**, 101351 (2023)
[80] C. P. Singh and V. Khatri, Phys. Rev.D **109**, 023508 (2024)
[81] J. S. Wang and F. Y. Wang, Astron. Astrophys. **564**, A137 (2014)
[82] R.C. Nunes, S. Pan and E. N. Saridakis, Phys. Rev. D **94**, 023508 (2016)
[83] E. Di Valentino, A. Melchiorri and O. Mena, Phys. Rev.D **96**, 043503 (2017)
[84] E. Di. Valentino, A. Melchiorri, O. Mena and S. Vagnozzi, Phys. Dark Univ. **30**, 100666 (2020)
[85] W. Yang, S. Pan, E. Di Valentino, et al., J. Cosmol. Astropart. Phys. **09**, 019 (2018)
[86] L. Wang, J. Zhang, D. He, J. Zhang and X. Zhang, Mon. Not. Roy. Astron. Soc. **514**, 1433 (2022)
[87] S. Gariazzo, E. Di Valentino, O. Mena, and R. C. Nunes, Phys. Rev. D **106**, 023530 (2022)
[88] Marcel A. van der Westhuizen and Amare Abebe, J. Cosmol. Astropart. Phys. **01**, 048 (2024)
[89] Chen Ju-Hua, Zhou Sheng and Wang Yong-Jiu, Chinese Phys. Lett. **28**, 029801 (2011)
[90] G.M. Kremer and O.A.S. Sobreiro, Braz. J. Phys. **42**, 77 (2012)
[91] A. Avelino, AIP Conf. Proc. **1473**, 98 (2012)
[92] A. Avelino, Y. Leyva and L. A. Ureña-López, Phys. Rev. D **88**, 123004 (2013)
[93] T. Harko and F. S. N. Lobo, Phys. Rev.D **87**, 044018 (2013)
[94] Y. Leyva and M. Sepúlveda, Eur. Phys. J. C **77**, 426 (2017)
[95] A. Atreya, J.R. Bhatt and A. Mishra, J. Cosmol. Astropart. Phys. **02**, 024 (2018)
[96] A. Hernández-Almada, Miguel A. García-Aspeitia, J. Magaña and V. Motta, Phys. Rev. D **101**, 063516 (2020)
[97] S. Nojiri and S.D. Odintsov, Phys. Rep. **505**, 59 (2011), arXiv:1011.0544
[98] I. Brevik, V.V. Obukhov and A.V. Timoshkin, Astrophys. Space Sci. **355**, 399 (2015)
[99] I. Brevik, Phys. Rev.D **65**, 127302 (2002)
[100] J. Hu and H. Hu, Eur. Phys. J. Plus **135**, 718 (2020)
[101] J. Solá, et al., Mon. Not. Roy. Astron. Soc. **478**, 4357 (2018)
[102] A. Gomez-Valent and J. Sola Peracaula, Mon. Not. Roy. Astron. Soc.**478**, 126 (2018), arXiv:1801.08501
[103] D. Scolnic, D. Brout, et al. , Astrophys. J. 938, **2**, 113(2022), arXiv:2112.03863
[104] M. Moresco et al., Living Rev. Relativ. **25**, 6 (2022), arXiv:2201.07241[astro-ph.cO]
[105] S. Nesseris, G. Pantazis and L. Perivolaropoulos, Phys. Rev. D **96**, 023542 (2017), arXiv:1703.10538
[106] A. Quelle and A.L. Maroto, Eur. Phys. J. C **80**, 369 (2020)
[107] Daniel Foreman- Mackey et al., PASP **125**, 306 (2013)
[108] A. Lewis, "GetDist: a Python package for analysing Monte Carlo samples", (2019), arXiv:1910.13970
[109] A. G. Riess et al., Astrophys. J. Lett. **934**, L7 (2022), arXiv:2112.04510[astro-ph.CO]
[110] N. Aghanim et al. (Planck Collaboration), Astrophys. Astron. **641**, A6 (2020), arXiv:1807.06209 [astro-ph.CO]
[111] C. Heymans et al., Astron. Astrophys. **646**, A140 (2021),arXiv:2007.15632 [astro-ph.CO]
[112] A. Al Mamon and K. Bamba, Eur. Phys. J. C **78**, 862 (2018)
[113] H. Akaike, IEEE Trans. Autom. Control **19**, 716 (1974)
[114] G. Schwarz, Ann. Stat. **6**, 461 (1978)
[115] H. Jeffreys, Theory of Probability, Oxford University Press, Oxford, (1961)
[116] D.J.C. MacKay, Information theory, Inference, and Learning Algorithm (Cambridge: Cambridge University Press)
[117] A. Liddle, P. Mukherjee, D. Parkinson and Y. Wang, Phys. Rev. D **74**, 123506 (2006)
[118] P. Mukherjee, D. Parkinson and A. Liddle, Astrophys.

J. **638**, L51 (2006)
[119] A. Liddle, Mon. Not. Roy. Astron. Soc. **377**, L74 (2007)
[120] R. Trotta, Mon. Not. Roy. Astron. Soc. **378**, 72 (2007)
[121] B. Santos, N. Chandrachani Devi and J.S. Alcaniz, Phys. Rev. D **95**, 123514 (2017)
[122] W. J. C. da Silva and R. Silva, Eur. Phys. J. C **81**, 403 (2021)
[123] C.P. Singh and Ajay Kumar, Grav. Cosmol. **25**, 58 (2019)
[124] T. Padmanabhan and S. M. Chitre, Phys. Lett. A **120**, 433 (1987)
[125] A. Montiel and N. Breton, J. Cosmol. Astropart. Phys. **08**, 023 (2011)

Colloid and Polymer Science
https://doi.org/10.1007/s00396-024-05257-8

RESEARCH

Fabrication and assessment of carboxymethyl guar gum-based sustainable films for packaging application

Subodh K. Juikar[1] · **Sudhir G. Warkar**[1]

Received: 25 December 2023 / Revised: 10 April 2024 / Accepted: 12 April 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

This study aimed towards the fabrication of biodegradable packaging films using carboxymethyl guar gum (CMG) as the base material. Various ratios of glycerol (Gly) as a plasticizer and glutaraldehyde (Glu) as a cross-linker were employed in the synthesis of these biopolymer-based films through the solution casting technique. Comprehensive characterization, such as attenuated total reflectance-Fourier transform infrared spectroscopy (ATR-FTIR), differential scanning calorimetry (DSC), thermogravimetric analysis (TGA), and scanning electron microscopy (SEM), was conducted. The films were further examined for tensile strength, percent elongation, thickness, solubility in water, gloss, haze, and transparency. The maximum tensile strength of 37.82 MPa, maximum percent elongation of 124.20%, maximum gloss of 25.52% at a 20° angle, maximum haze of 41.93%, maximum transparency of 88.53%, and a minimum water solubility of 60.42% were observed for the developed films. This reveals that CMG-based films hold promising potential as a novel biomaterial for future applications in packaging, such as edible coatings and water-soluble, biodegradable packaging films.

Keywords Biopolymer · Carboxymethyl guar gum · Glycerol · Solvent casting · Packaging film

Introduction

Packaging serves a pivotal role in safeguarding and extending the shelf life of products until they reach the end consumer. Conventional packaging relies heavily on synthetic polymers derived from petroleum, such as polyethylene (PE), polypropylene (PP), polyvinylchloride (PVC), and polystyrene (PS). These polymers are preferred for their cost-effectiveness and beneficial properties such as robust mechanical strength and effective barrier characteristics against aroma, gases, and water. Despite these benefits, the disposal and recycling of petroleum-based polymers pose significant challenges. The global annual generation of approximately 1.3 billion tons of plastic waste worsens environmental issues, underscoring the need for sustainable alternatives in packaging solutions [1, 2].

In recent years, due to the comeback of the environmental challenges posed by conventional packaging, research has shifted towards exploring the application of biopolymers for packaging purposes.

Various biopolymers due to desired properties provided by them are seen as a sustainable option for replacement of synthetic polymers for various packaging applications. The global trend of biopolymer production is expected to increase from 2.11 MT from 2020 to 2.87 MT in 2025 [3]. Biopolymers are sourced from natural materials and possess excellent biodegradability, making them more environmentally friendly [1, 2]. Biopolymers, also referred to as generation-IV packaging materials, are utilized in various forms, serving as coatings directly applied to products, as well as in the fabrication of films for items like bags, envelopes, containers, and lids. They present a promising alternative to petroleum-based synthetic polymers [2, 4–6]. Researchers are exploring a diverse range of biopolymers derived from natural resources, including polysaccharides, proteins, and lipids, as well as those synthesized from bio-derived monomers like polylactic acid, and also those synthesized by microorganisms, such as polyhydroxyalkanoates [5].

While these biopolymers exhibit excellent film-forming characteristics, their inherent brittleness and high water absorption capacity necessitate modification through the addition of plasticizers and cross-linkers for easier handling

✉ Sudhir G. Warkar
sudhirwarkar@gmail.com

[1] Department of Applied Chemistry, Delhi Technological University, Delhi, India

Published online: 24 April 2024

and commercial processing. Incorporating plasticizers like glycerol enhances film flexibility by reducing intermolecular forces, and the introduction of cross-linkers, like citric acid, improves the tensile strength of the films. This research marks a significant step towards sustainable packaging alternatives, addressing the ecological impact of traditional petroleum-based polymers [5–11].

The biopolymer-based films and coatings represent a significant stride towards sustainable packaging solutions, employing materials such as alginate, chitosan, cellulose, guar gum, and starches. Notably, in our exploration of biopolymer-based materials for packaging applications, it has been observed that carboxymethyl guar gum (CMG) has yet to be used for the synthesis of biopolymer-based packaging materials. This underscores the potential novelty and unexplored possibilities in incorporating CMG into the realm of eco-friendly packaging solutions [1, 2, 7–30].

Carboxymethyl guar gum is a water-soluble biopolymer, derived through the reaction of guar gum with monochloroacetic acid in the presence of sodium hydroxide (NaOH) [31–33]. This versatile derivative finds applications across various industries, like oil recovery, paints, minerals, personal care, textiles, and food [34]. It is revealed that, due to presence of carboxyl and hydroxyl groups in the CMG backbone, it can be easily cross-linked, grafted, or blended with other polymers, leading to effective film formation [35]. This excellent film-forming property of CMG opens up various applications, particularly in areas like packaging.

The current study proposes the development of various carboxymethyl guar gum-based films via solvent casting technique, for packaging applications. The varying amounts of glycerol (plasticizer) and glutaraldehyde (cross-linker) were employed in the fabrication of films, and their impact on the film properties was further studied. The films were characterized thoroughly using TGA, SEM ATR-FTIR, and DSC techniques. Various parameters such as haze, transparency, solubility in water, gloss, tensile strength, elongation, and thickness for all the synthesized film were also examined.

Materials and methods

Carboxymethyl guar gum (CMG, M/s. Maruti Udyog, Ahmedabad, India), glycerol (Merck Life Sciences Private Ltd., India), and gluteraldehyde solution (CDH, India) were used as obtained. Double distilled water was utilized for solution preparations and washing purposes.

Fabrication of CMG-based film

The series of a total of 64 variants of CMG-based films as shown in Table 1 were fabricated by employing the different concentrations of glycerol (0 ml, 0.1 ml, 0.25 ml, 0.5 ml, 1 ml, 1.25 ml, 1.5 ml, and 2 ml) as a plasticizer and glutaraldehyde (0 ml, 0.1 ml, 0.25 ml, 0.5 ml, 1 ml, 1.25 ml, 1.5 ml, and 2 ml) as a cross-linker, via solvent casting technique. The fixed amount (1 g) of CMG was dispersed in 100 ml of distilled water at room temperature. Then, the desired volume of glycerol was added to the dispersion under stirring and heating at 70 °C on magnetic stirrer. It was then followed by the addition of glutaraldehyde to the above mixture. After 45 min, the resultant mixtures were immediately cast in an acrylic tray with dimensions (200 mm × 150 mm × 5 mm) and allowed to air dried at 25 °C for 24 h and then transferred to a hot air circulating oven maintained at 60 ± 5 °C for another 24 h. The dried films were then peeled off and stored at 25 °C. Figure 1 shows the schematic representation of the fabrication of the CMG-based films.

The same procedure was repeated for casting the films at different temperature conditions (75 °C).

Characterization and testing

Differential scanning calorimetry (DSC)

The thermal behavior of the cross-linked CMG films fabricated at two different temperatures of 70 °C and 75 °C and CMG powder were analyzed using a Perkin Elmer (model

Table 1 Various proportions of reagents used for CMG-based film synthesis

Code no.	CMG (g)	Glycerol (plasticizer) (ml)	Glutaraldehyde (cross-linker) (ml)							
			1	2	3	4	5	6	7	8
A	1.00	0.00	0.00	0.10	0.25	0.50	1.00	1.25	1.50	2.00
B	1.00	0.10	0.00	0.10	0.25	0.50	1.00	1.25	1.50	2.00
C	1.00	0.25	0.00	0.10	0.25	0.50	1.00	1.25	1.50	2.00
D	1.00	0.50	0.00	0.10	0.25	0.50	1.00	1.25	1.50	2.00
E	1.00	1.00	0.00	0.10	0.25	0.50	1.00	1.25	1.50	2.00
F	1.00	1.25	0.00	0.10	0.25	0.50	1.00	1.25	1.50	2.00
G	1.00	1.50	0.00	0.10	0.25	0.50	1.00	1.25	1.50	2.00
H	1.00	2.00	0.00	0.10	0.25	0.50	1.00	1.25	1.50	2.00

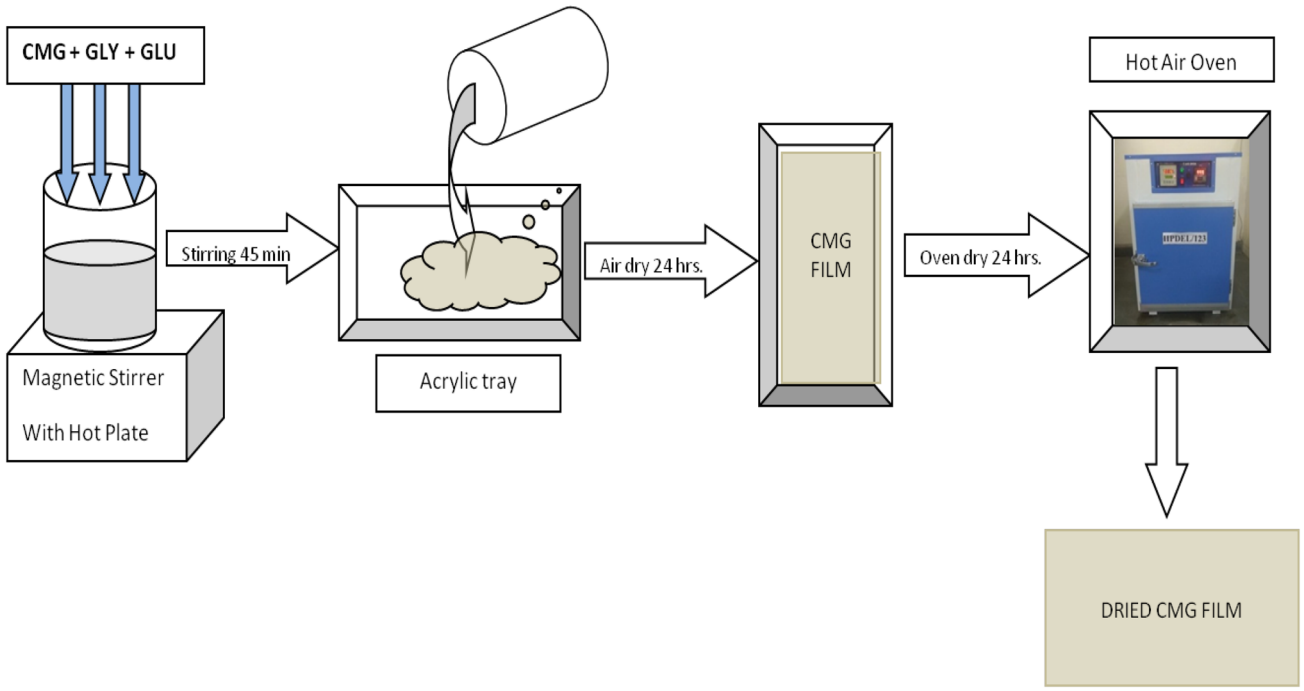


Fig. 1 Schematic representation of fabrication of the CMG-based film

no.: DSC 8000). The samples (2–3 mg) were heated at a rate of 10 °C/min, at a temperature range from 0 to 200 °C.

Thermogravimetric analysis (TGA)

TGA was done utilizing the Perkin Elmer instrument (Model No: TGA 4000). The sample (2–3 mg) was heated from a temperature range of 25 to 600 °C under a nitrogen atmosphere in a sample holder. The flow rate was maintained at 50 ml/min, and the heating rate was set at 10 °C/min. The weight change of the sample was recorded as a function of temperature.

Scanning electron microscopy (SEM)

SEM analysis was carried out using a Hitachi (model: S-3400N) at an accelerating voltage of 5 kV to examine the surface morphology of the fabricated CMG films.

Attenuated total reflectance-Fourier transform infrared spectroscopy (ATR-FTIR)

The FTIR spectrum of the films was recorded using a Shimadzu Corp. ATR-FTIR instrument (model: IR Affinity-1S). The spectra were recorded in the frequency range from 4000 to 400 cm^{-1}, at the spectral resolution of 5 cm^{-1}.

Mechanical properties

The samples were examined for tensile strength and elongation on a Universal Testing Machine (Shimadzu, Autograph AG–X 110 N) equipped with a 1-kN load cell capacity, as per the ASTM D 882–97 guidelines with minor modifications. The film dimensions were 25 mm × 150 mm, with a gauge length of 50 mm and a crosshead speed of 50 mm/min. The tensile strength was reported in MPa. The strain at the point of breakage was recorded and reported as percent elongation at break. Tests were conducted in triplicate.

Film thickness

The film thickness measurement was conducted on the Mitutoyo Japan Micrometer (Model 547-400S) as per the ASTM 6988 guideline, with a sample dimension of 50 mm × 50 mm. A total of 10 readings were randomly taken on each type of sample, and the average value is reported in microns (μ).

Solubility in water

The solubility of the film in water is referred to the percentage of soluble matter present in the film [20]. The films having dimensions of about 50 mm × 50 mm were immersed in 50 ml of double distilled water at room

temperature for 24 h. Later, the water was discarded, and the films were dried in a hot air oven set at 60 °C for 24 h. The solubility of the film was then calculated using the following Eq. (1).

$$Solubility(\%) = \frac{\text{Initial dry weight} - \text{Final dry weight}}{\text{Initial dry weight}} \times 100 \quad (1)$$

Gloss

Samples were analyzed on the Gloss meter (Model BYK Gardner) as per ASTM D 523/ASTM D2457 guidelines, at a 20° angle. A total of 10 readings were acquired for each type of sample, and the average value was reported.

Haze and transparency

The spherical haze meter, diffusion systems (M57D) was utilized for the measurement of the haze and transmittance of the film. Each sample was tested in triplicate, and the reported values signify the average percentage. The measurement was done according to the guidelines outlined in ASTM D 1003.

Results and discussions

Probable mechanism

The various 64 formulations of CMG-based films were synthesized using the different quantities of glycerol and glutaraldehyde, as outlined in Table 1. The glutaraldehyde interacts with the hydroxyl groups present in both CMG and glycerol, leading to the formation of acetyl bonds. This interaction leads to the successful development of a cross-linked CMC/Glu/Gly polymeric network. The probable mechanism underlying this process is depicted in Fig. 2 [36].

Characterization and testing

Differential scanning calorimetry (DSC)

The DSC analysis was carried out for the CMG film (Sample A-1), and CMG/Glu/Gly film (Sample E-5) fabricated at 70 °C, and 75 °C, as shown in Fig. 3. A broad peak at 91.63 °C and 97.21 °C is observed in CMG film and CMG/Glu/Gly film fabricated at 70 °C, respectively, while a sharper peak at 102.25 °C is noticed in CMG/Glu/Gly film fabricated at 75 °C. This sharp peak at 102.25 °C indicates an endothermic reaction associated with the melting of the film at a higher temperature. The difference in the peaks

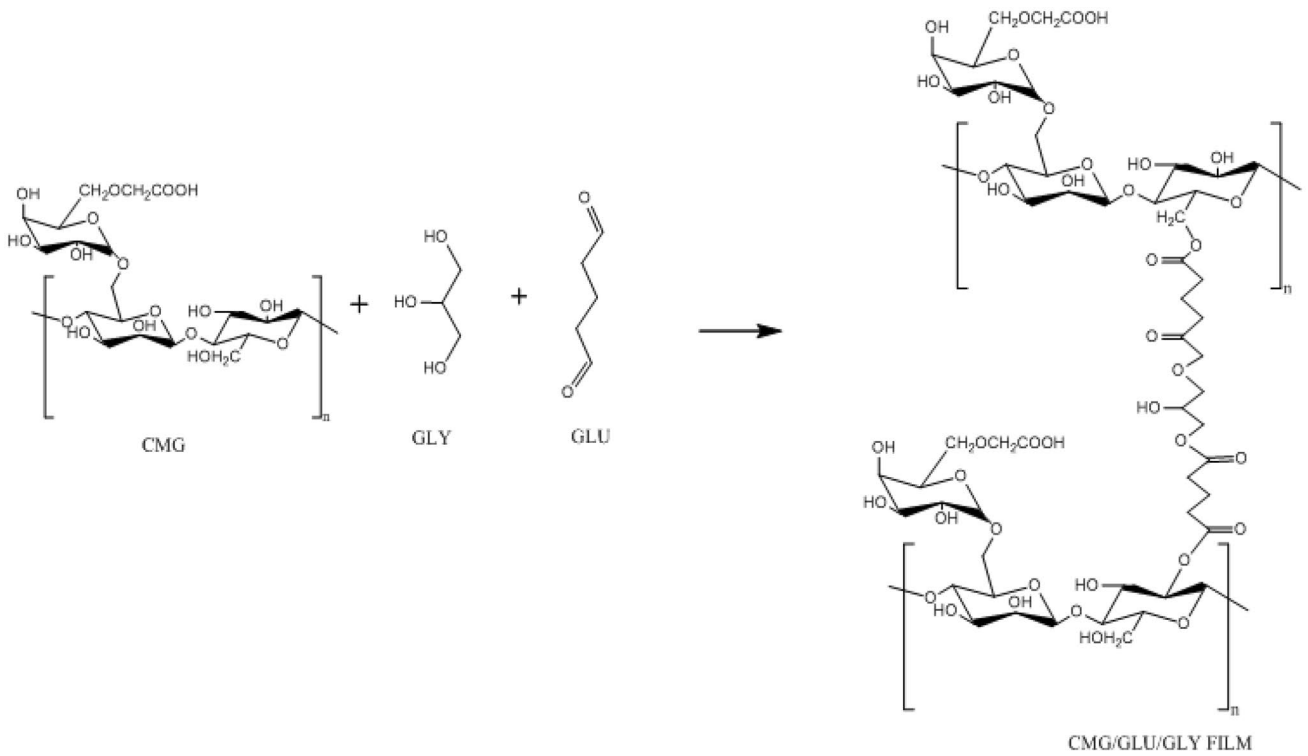


Fig. 2 Probable mechanism of CMG/Glu/Gly film formation

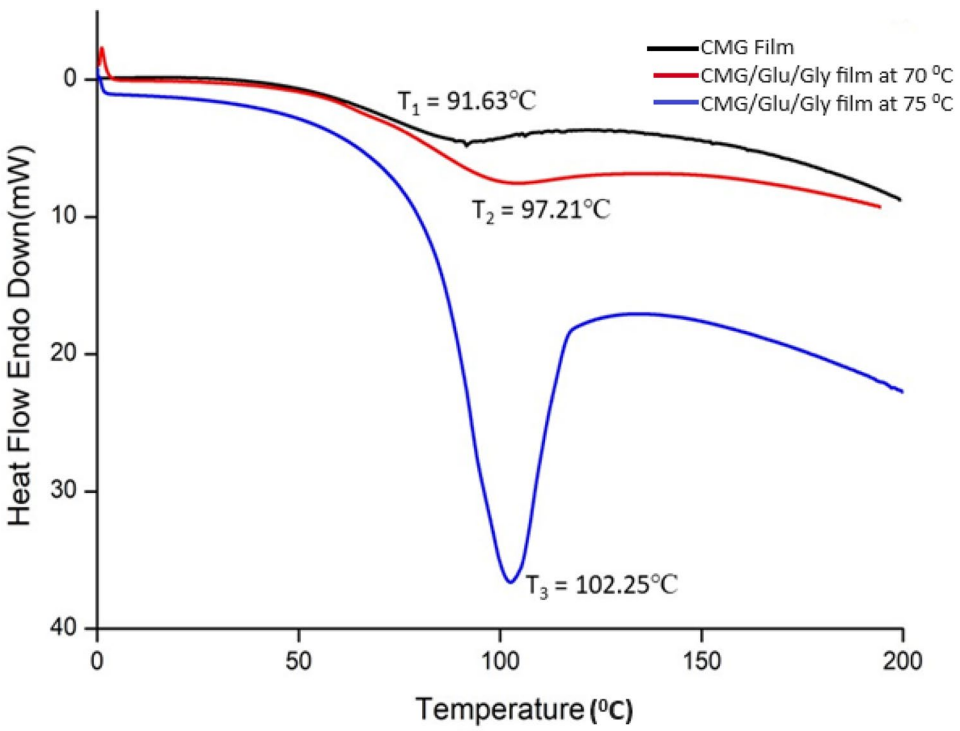


Fig. 3 DSC for CMG film and CMG/Glu/Gly films fabricated at 70 °C and 75 °C

revealed the improved cross-linking at the higher reaction temperature of 75 °C compared to the reaction temperature of 70 °C [11].

Thermogravimetric analysis (TGA)

TGA thermographs of CMG/Glu/Gly films (Sample E-5) fabricated at two different temperatures (70 °C and 75 °C) are shown in Fig. 4. The thermograms of both CMG-based films show three-stage weight loss. Initially, the same weight loss of 26% due to water evaporation from 30 to 245 °C was observed in CMG films prepared at 70 °C and 75 °C. Secondly, again, the same weight loss of 40% from 245 to 325 °C, due to the degradation of polymer backbone (-COOH and -OH group of CMG), was found in the films synthesized at reaction temperatures 70 °C and 75 °C.

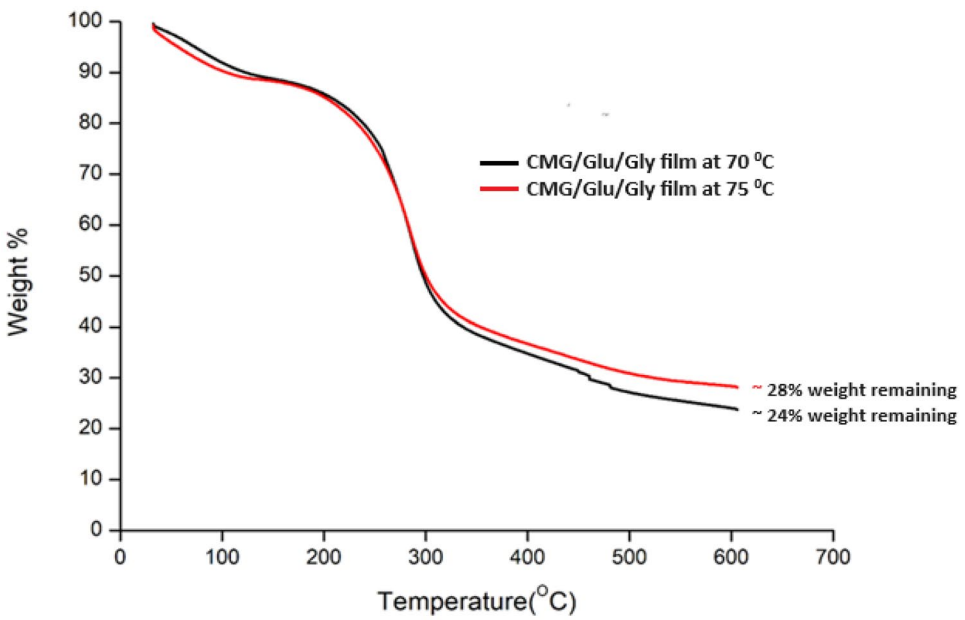


Fig. 4 TGA plot for optimized CMG/Glu/Gly films fabricated at reaction temperatures (70 °C and 75 °C)

Lastly, the weight loss of 20% and 16% from 325 to 600 °C signifies the breakage of the cross-linking present in the films fabricated at 70 °C and 75 °C, respectively [37–40].

Moreover, it was observed that up to the second stage, both films exhibited similar weight loss. However, beyond this temperature, the film produced at 70 °C shows a higher weight loss, compared to the film fabricated at 75 °C. This suggests that the film synthesized at the higher reaction temperature of 75 °C demonstrates greater stability, likely attributed to a higher degree of cross-linking. Consequently, the films synthesized at the 75 °C reaction temperature were selected for further analysis [11].

Scanning electron microscopy (SEM)

The surface morphology of the CMG (Sample A-1), CMG/Gly (Sample E-1), CMG/Glu (Sample A-5), and CMG/Glu/Gly (Sample E-5) films was analyzed by SEM. In Fig. 5A, the single-phase morphology was observed without any surface complexity, attributed to the solely presence of CMG. Figure 5B reveals the uneven surface morphology associated with the presence of cavities and air pockets resulting from the dispersion of the plasticizer within the polymer matrix. Figure 5C indicates a more uniform and denser surface, which signifies the successful cross-linking between CMG and the cross-linker. Figure 5D shows a comparatively more homogeneous surface than Fig. 5C, implying the combined effect of plasticizer and cross-linker on the overall surface morphology of the films [6].

Attenuated total reflectance-Fourier transform infrared spectroscopy (ATR-FTIR)

The ATR-FTIR spectra of the CMG (Sample A-1), CMG/Gly (Sample E-1), CMG/Glu (Sample A-5), and CMG/Glu/Gly (Sample E-5) film are shown in Fig. 6. In the FTIR spectra of the CMG film, characteristic peaks associated with O–H stretch were observed at 3300–3400 cm^{-1}, while the peak associated with C–H stretch appears at 2800–2900 cm^{-1}.

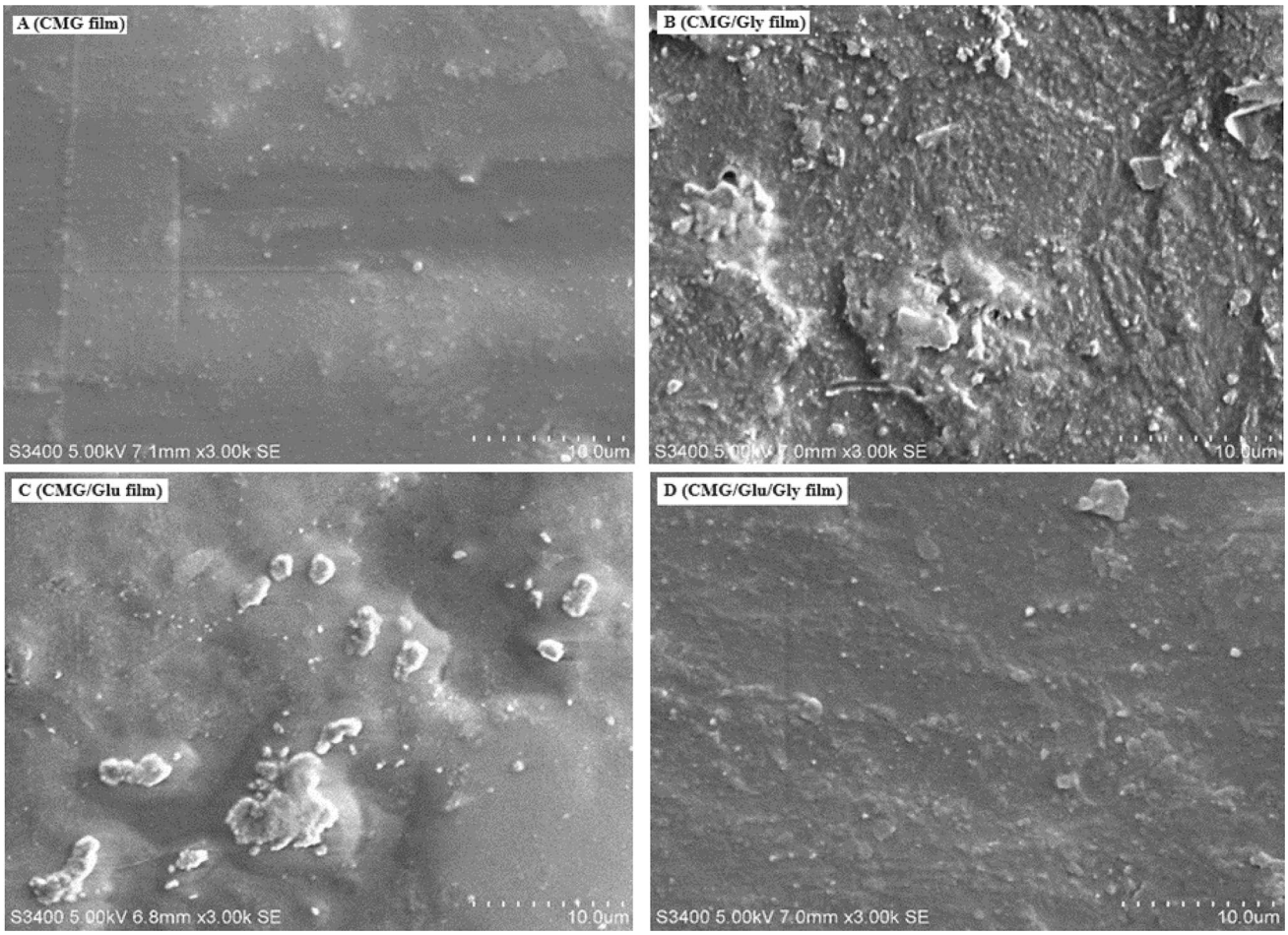


Fig. 5 SEM micrograph of **A** CMG film, **B** CMG/Gly film, **C** CMG/Glu film, and **D** CMG/Glu/Gly film, at 3000 magnifications

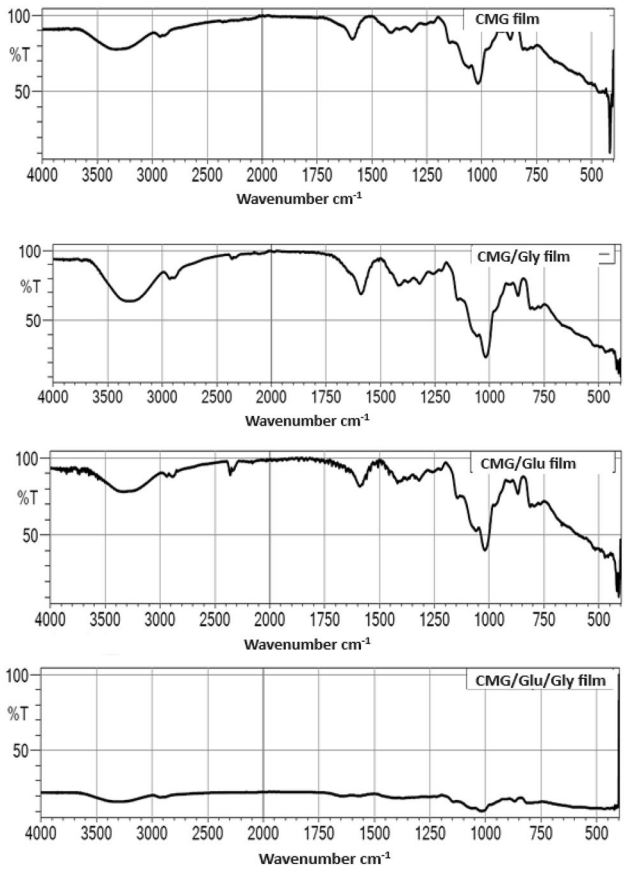


Fig. 6 FTIR spectra of CMG film, CMG/Gly film, CMG/Glu film, and CMG/Glu/Gly film

Moreover, the C = O stretch appears at 1600–1700 cm^{-1}, and the C–H bend appears at 1300–1400 cm^{-1} [41].

In FTIR spectra of CMG/Gly film, it was observed that the addition of glycerol resulted in a notable increase in the number of alcoholic hydroxyl groups. Consequently, the intensity of the peaks within the 3300–3400 cm^{-1}, corresponding to hydroxyl groups, exhibited a significant increase in amplitude [42].

In FTIR spectra of CMG/Glu film, an increase in the intensity of the peak around 2800–2900 cm^{-1} is observed, attributed to the larger number of CH_2 groups resulting upon the introduction of cross-linker (glutaraldehyde) [43].

Moreover, in the FTIR spectra of CMG/Glu/Gly film, the disappearance of peak around 2300–2400 cm^{-1}, present in CMG/Gly film and CMG/Glu film, confirms the successful formation of a cross-linked polymeric film.

Mechanical properties

All the 64 variants of CMG-based films as shown in Table 1 were tested for evaluating their mechanical properties.

The tensile strength test provides information about the mechanical strength of the material, crucial for packaging applications. The plasticizer's impact on tensile strength varies with changes in cross-linker concentration as observed in Fig. 7a. Without or with low cross-linker concentration (up to 0.25 ml), plasticizer increases film flexibility, leading to a decrease in tensile strength. For cross-linker concentrations (0.25–1 ml), tensile strength slightly increases due to glycerol's participation in cross-linking. Beyond 1-ml cross-linker, tensile strength sharply decreases, rendering the film sticky due to excess plasticizer.

Similarly, Fig. 7b illustrates the impact of cross-linker content on tensile strength. At low concentrations of cross-linker (0.1–0.25 ml), tensile strength increases slowly due to insignificant cross-linking. Beyond this range, a considerable increase in the tensile strength occurs with increased cross-linking reaching a maximum at 1.25 ml of cross-linker. Moreover, with an increase in the cross-linker content above

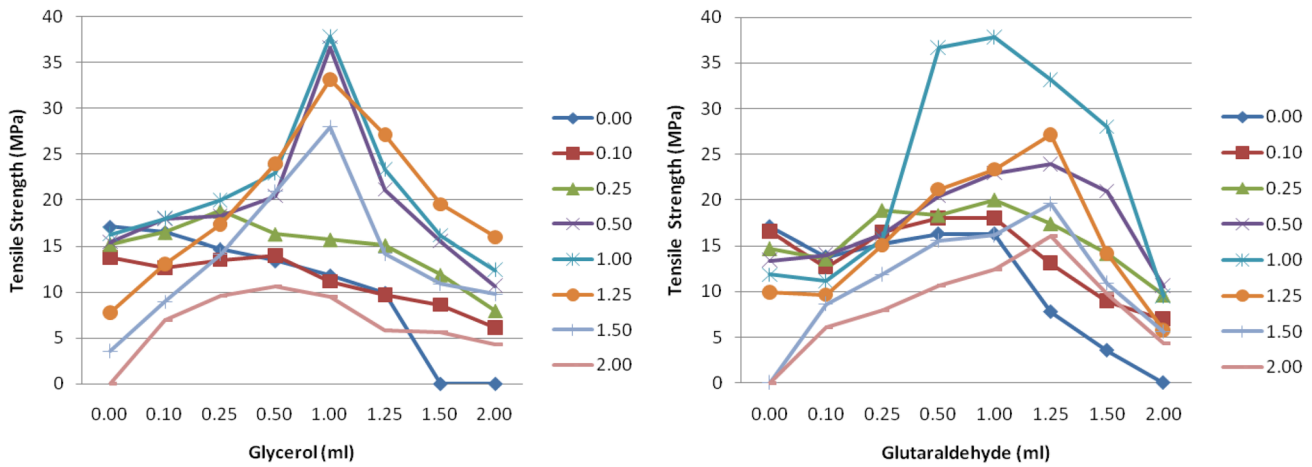


Fig. 7 Influence of **a** plasticizer and **b** cross-linker on the tensile strength of the CMG-based film

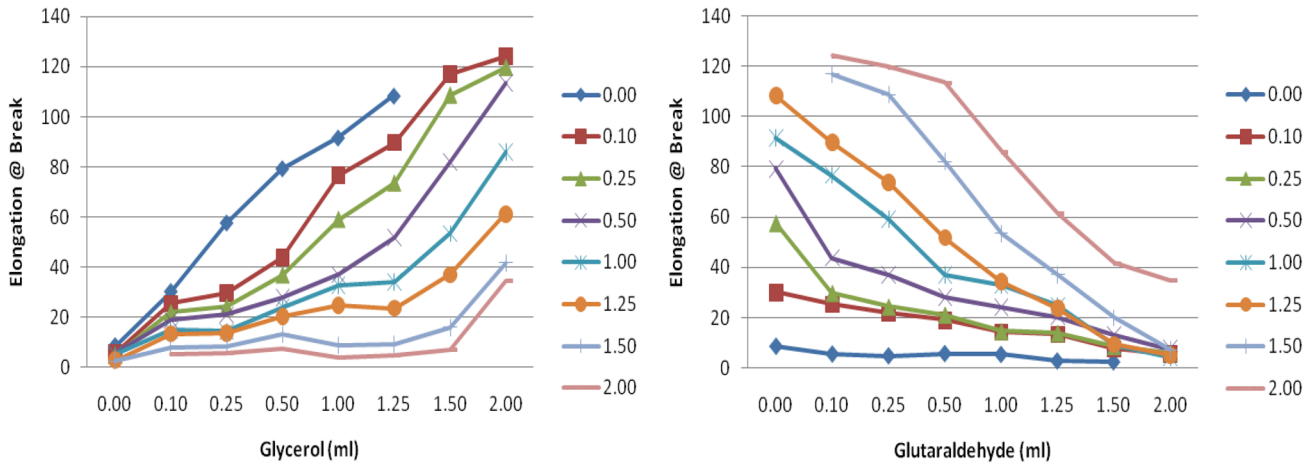


Fig. 8 Impact of **a** plasticizer and **b** cross-linker on elongation @ break of CMG-based films

1.25 ml, tensile strength decreases, possibly due to excessive cross-linker hindering available cross-linking sites. The maximum tensile strength of 37.82 MPa is observed in CMG-based film. Hence, CMG-based films can be employed for potential applications in the packaging sector [11, 42].

The elongation test provides information about the extent to which films can elongate or withstand strain before reaching the breaking point. It is an essential consideration for processes such as lamination and printing. The elongation percentage of the CMG-based films is depicted in Fig. 8. Notably, the film exhibited a maximum elongation of 124.2%. It is observed that with an increase in glycerol quantity, molecular mobility rises, leading to a decrease in intermolecular forces, resulting in an increase in film flexibility, consequently elongation increases. However, at lower glycerol quantities, where molecular mobility is limited, the plasticization effect is less significant. As seen in Fig. 8a, the addition of glycerol increases the percent elongation of the film by decreasing its brittleness due to plasticizing effect. At the lower quantity of glycerol, as the mobility of the molecules is limited, the plasticization effect is not significant but as the quantity of glycerol increases, the mobility of the molecules increases, and elongation increases due to an increase in the flexibility of the film by reducing the intermolecular forces. Similarly, with a quantity in the number of cross-linkers as shown in Fig. 8b, a decrease in the elongation at break was observed. This is due to the cross-linking effect, the films become stiffer and the flexibility of the cross-linked films decreases with an increased number of cross-links produced [11, 42, 44].

Film thickness

The average thickness of all the variants of CMG-based films as shown in Table 1 is shown in Fig. 9. The maximum

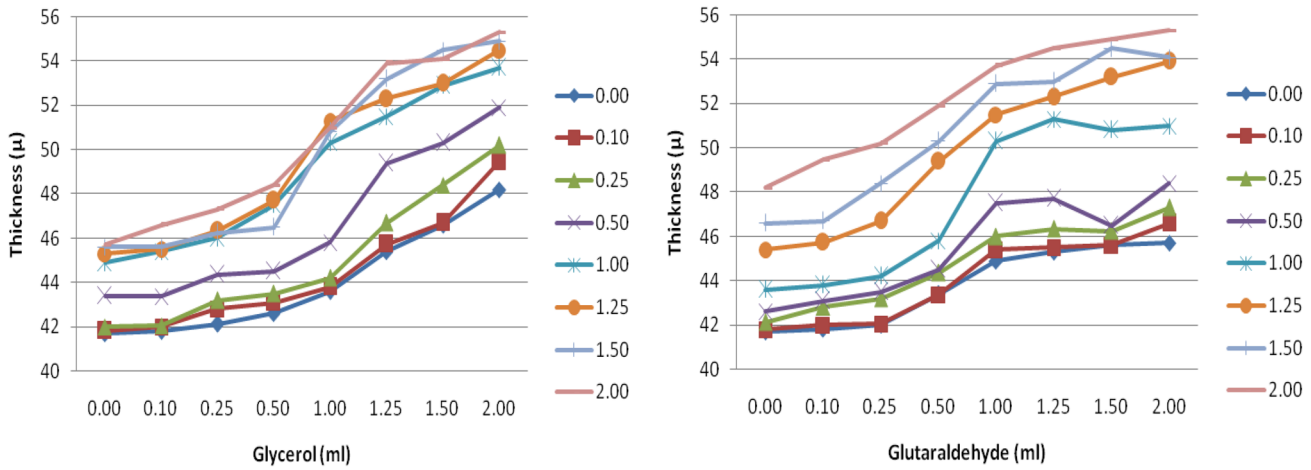


Fig. 9 Thickness of the CMG-based films as a function of **a** plasticizer and **b** cross-linker

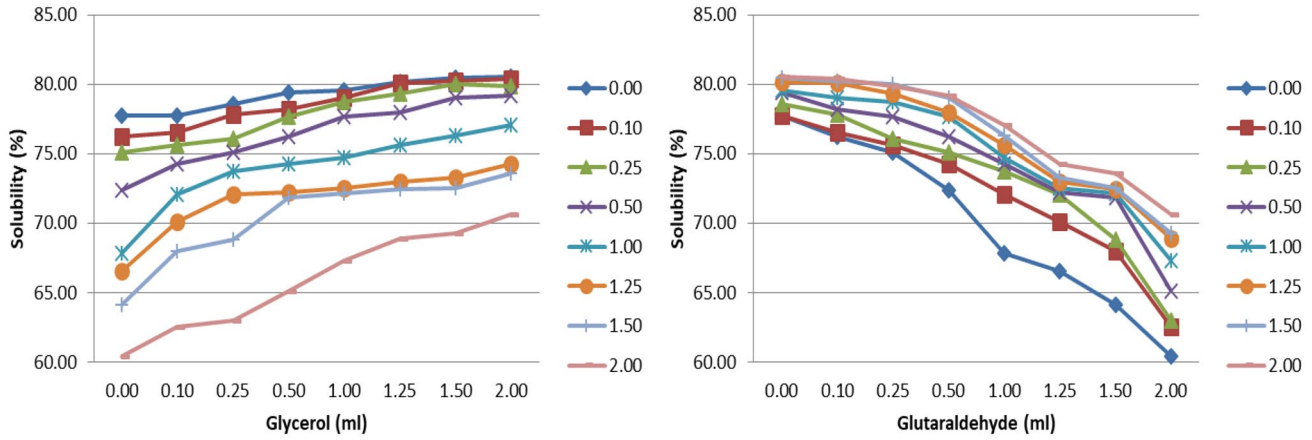


Fig. 10 Effect of **a** plasticizer and **b** cross-linker on the solubility of the CMG-based films in water

thickness of the film observed was 55.30 µ. The thickness of the film increases with the increase in the quantity of both plasticizer and cross-linker due to the distribution of intermolecular bonds leading to rearrangement providing more expanded structure due to the effect of plasticizer and in case of cross-linker due to an increase in the cross-linking which gives rise to denser film structure on cross-linking [12, 18, 45, 46].

Solubility in water

The solubility percentage of the synthesized CMG-based films as given in the Table 1 is calculated using Eq. (1), as illustrated in Fig. 10. The minimum observed solubility of the film was 60.42%. It is evident that the solubility decreases with an increase in the quantity of the cross-linker. This is attributed to higher cross-linking, which reduces hydrogen bonding in the film matrix, resulting in slower dissolution [42].

Moreover, the film's solubility increases with an increase in the plasticizer quantities due to the enhanced flexibility of the film and due to glycerol separating from the hydrogen bonding site of CMG and dissolving in water [14]. The water solubility of CMG-based films suggests potential applications in edible packaging and water-soluble packaging materials.

Gloss

The glossiness of polymeric films is an essential parameter for packaging applications. It provides information about the amount of light reflected when light is incident on the film surface from various angles. Figure 11 illustrates the gloss values of the CMG-based films as listed in Table 1, at a 20° angle with the CMG-based film displaying a maximum gloss of 25.52%. Notably, at lower quantities of concentration of plasticizer and cross-linker, the glossiness tends to increase, owing to the increased smoother surface of the

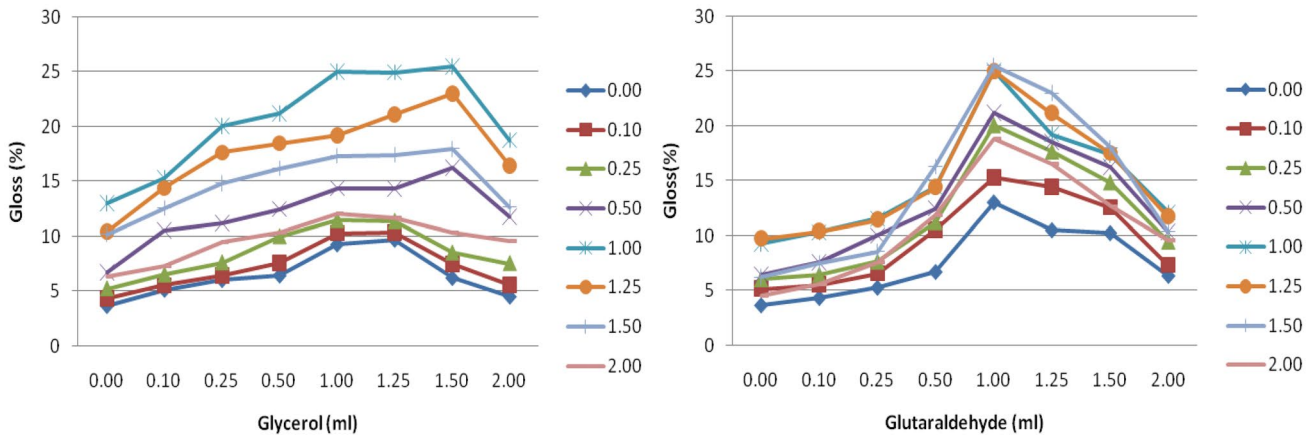


Fig. 11 Impact of **a** plasticizer and **b** cross-linker on the gloss of the CMG-based films at 20° angle

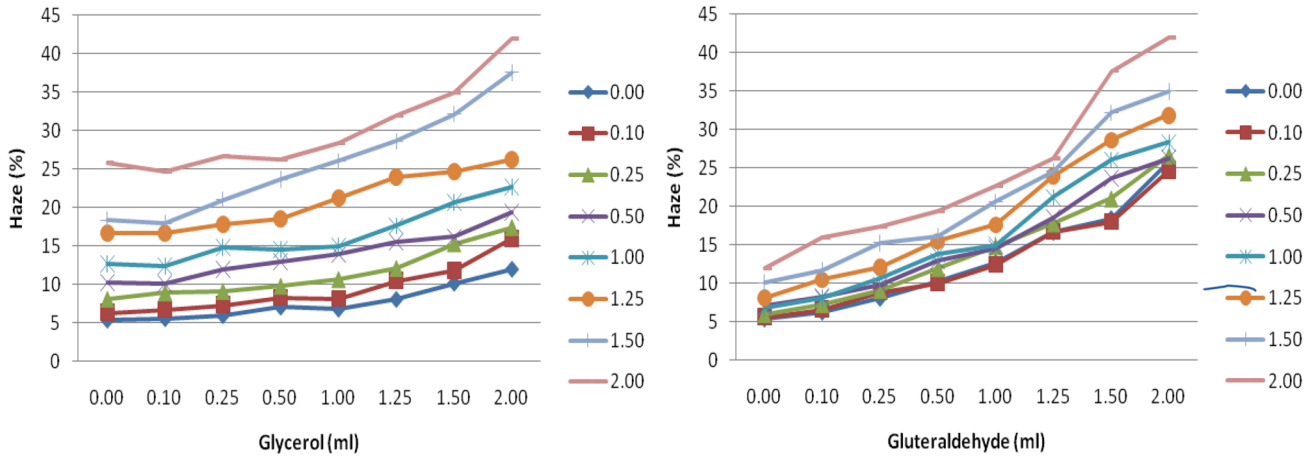


Fig. 12 Influence of **a** plasticizer and **b** cross-linker on the haze of the CMG-based films

film due to plasticization and the cross-linking. However, with the addition of high quantities of plasticizer and cross-linker, a reduction in glossiness occurs. This decrease may be attributed to the formation of rougher surfaces due to an increase in cross-linking [47].

Haze and transparency

The haze percentage of the synthesized CMG-based films given in Table 1 is presented in Fig. 12. The film exhibits a maximum haze of 41.93%. It is noteworthy that the haziness of the film increases with a rise in both the plasticizer and cross-linker content. This is due to the increase of diffusion of incident light caused by increased bulkiness at higher plasticizers and cross-linking of the film due to polymeric chain compaction to a higher cross-linking [16].

The transparency of the CMG- based films, prepared as given in Table 1, is given in Fig. 13. The transparency of any polymer plays a crucial role in its visible light barrier properties and can be correlated to its crystalline behavior. Optical transparency is inversely proportional to its crystallinity [41]. The transparency of the synthesized CMG-based films is presented in Fig. 13. The findings indicate that CMG films exhibit excellent optical transparency, with the maximum transparency observed at 88.53%.

Moreover, it is observed that as glutaraldehyde content in the film increases, the cross-linking increases, leading to a noticeable decrease in the transparency of the film. Additionally, the transparency of the film is also decreased by an increase in the quantity of plasticizer. This can be attributed to the decrease in transmittance caused change in the refractive index by the increased polymeric chain compaction [16].

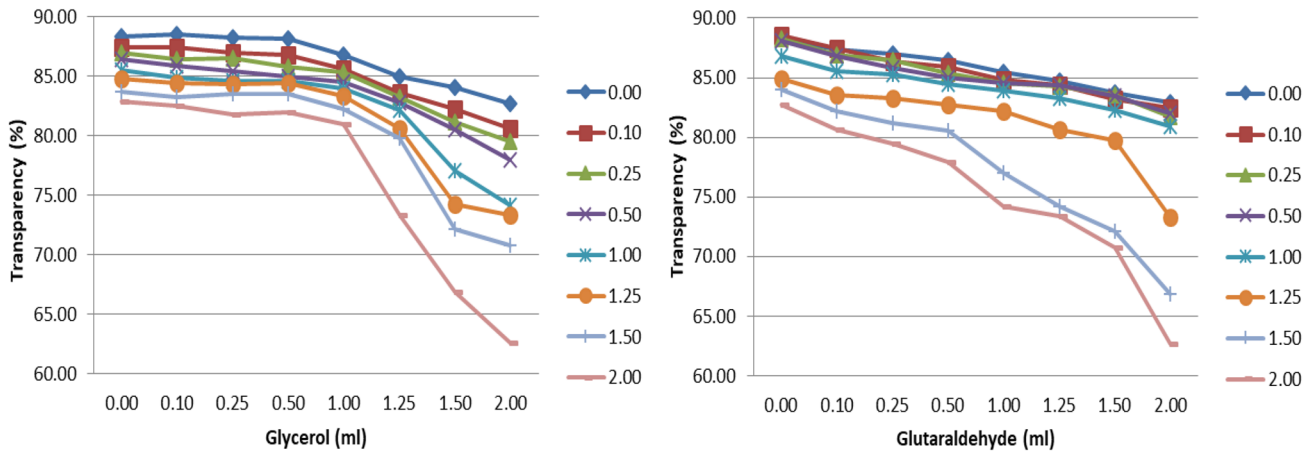


Fig. 13 Effect of **a** plasticizer and **b** cross-linker on the transparency of the films

Conclusion

In this study, films based on carboxymethyl guar gum (CMG) were successfully fabricated by utilizing glycerol as a plasticizer and glutaraldehyde as a cross linking agent using solution casting technique. The films were characterized for FTIR, DSC, and SEM. The result of the FTIR showed increase in the number of hydroxyl and CH_2 groups which confirms the cross linking of CMG with glycerol and glutaraldehyde. The DSC analysis shows that reaction temperature has significant effect on the formation of the CMG films as sharper peak at 102.25 °C when prepared at 75 °C due to better cross-linking in comparison with the blunt peak observed at 70 °C. The SEM micrographs showed improvement in improved surface in the films having both plasticizer and cross linker. Moreover, mechanical, optical, and solubility tests were also conducted on our CMG-based films. The maximum values observed for tensile strength, percent elongation, gloss at a 20° angle, haze, and transparency were 37.82 MPa, 124.20%, 25.52%, 41.93%, and 88.53%, respectively. All these properties revealed the possibility of our CMG films for packaging applications. The solubility test results indicated a minimum water solubility of 60.42% of these CMG films which highlights the potential applications of CMG-based films in areas such as edible coatings and water-soluble packaging. The maximum tensile strength value of 37.82 MPa suggests that, with further refinements, CMG-based films can compete with certain biopolymer-based films and even some commercially used synthetic polyethylene films. This study also points the CMG-based biopolymer materials as promising candidate for coatings in the packaging industry.

Acknowledgements The authors Subodh K. Juikar and Sudhir G. Warkar are thankful to the Delhi Technological University, New Delhi, India, and Subodh K. Juikar also thanks the Indian Institute of Packaging, Mumbai, India, for its infrastructure facilities.

Author contributions Conceptualization: [Subodh K. Juikar], [Sudhir G. Warkar]; Methodology: [Subodh K. Juikar]; [Sudhir G. Warkar], Formal analysis and investigation: [Subodh K. Juikar]; Writing-original draft preparation: [Subodh K. Juikar]; Writing-review and editing: [Sudhir G. Warkar]; Funding acquisition: NA; Resources: [Subodh K. Juikar]; Supervision: [Sudhir G. Warkar].

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

Declarations

Ethical approval Not applicable.

Competing interests The authors declare no competing interests.

References

1. Davis G, Song JH (2006) Biodegradable packaging based on raw materials from crops and their impact on waste management. Ind Crops Prod 23:147–161. https://doi.org/10.1016/j.indcrop.2005.05.004
2. Lazic VL, Hromiš NM (2018) Biopolymer packaging materials for food shelf-life prolongation. In: Biopolymers for food design. pp 223–277
3. Juikar SK, Warkar SG (2023) Biopolymers for packaging applications: an overview. Packag Technol Sci 36:229–251. https://doi.org/10.1002/pts.2707
4. Harding KG, Gounden T, Pretorius S (2017) "Biodegradable" plastics: a myth of marketing? Procedia Manuf 7:106–110. https://doi.org/10.1016/j.promfg.2016.12.027
5. Othman SH (2014) Bio-nanocomposite materials for food packaging applications: types of biopolymer and nano-sized filler. Agric Agric Sci Procedia 2:296–303. https://doi.org/10.1016/j.aaspro.2014.11.042
6. Ramakrishnan N, Sharma S, Gupta A, Alashwal BY (2018) Keratin based bioplastic film from chicken feathers and its characterization. Int J Biol Macromol 111:352–358. https://doi.org/10.1016/j.ijbiomac.2018.01.037
7. Janik H, Sienkiewicz M, Przybytek A et al (2018) Novel biodegradable potato starch-based compositions as candidates in packaging industry, safe for marine environment. Fibers Polym 19:1166–1174. https://doi.org/10.1007/s12221-018-7872-1
8. Chadijah S, Rustiah WO, Munir MID (2018) Determination of the optimum concentration cellulose baggase in making film bioplastic. In: Journal of Physics: Conference Series
9. Irissin-Mangata J, Bauduin G, Boutevin B, Gontard N (2001) New plasticizers for wheat gluten films. Eur Polym J 37:1533–1541. https://doi.org/10.1016/S0014-3057(01)00039-8
10. Rouilly A, Geneau-Sbartaï C, Rigal L (2009) Thermo-mechanical processing of sugar beet pulp. III. Study of extruded films improvement with various plasticizers and cross-linkers. Bioresour Technol 100:3076–3081. https://doi.org/10.1016/j.biortech.2008.12.047
11. Reddy N, Yang Y (2010) Citric acid cross-linking of starch films. Food Chem 118:702–711. https://doi.org/10.1016/j.foodchem.2009.05.050
12. da Silva MA, Bierhalz ACK, Kieckbusch TG (2009) Alginate and pectin composite films crosslinked with Ca2+ ions: effect of the plasticizer concentration. Carbohydr Polym 77:736–742. https://doi.org/10.1016/j.carbpol.2009.02.014
13. Bergo P, Sobral PJA (2007) Effects of plasticizer on physical properties of pigskin gelatin films. Food Hydrocoll 21:1285–1289. https://doi.org/10.1016/j.foodhyd.2006.09.014
14. Tong Q, Xiao Q, Lim LT (2008) Preparation and properties of pullulan-alginate-carboxymethylcellulose blend films. Food Res Int 41:1007–1014. https://doi.org/10.1016/j.foodres.2008.08.005
15. Mu C, Guo J, Li X et al (2012) Preparation and properties of dialdehyde carboxymethyl cellulose crosslinked gelatin edible films. Food Hydrocoll 27:22–29. https://doi.org/10.1016/j.foodhyd.2011.09.005
16. Saberi B, Thakur R, Vuong QV et al (2016) Optimization of physical and optical properties of biodegradable edible films based on pea starch and guar gum. Ind Crops Prod 86:342–352. https://doi.org/10.1016/j.indcrop.2016.04.015
17. Medina Jaramillo C, Gutiérrez TJ, Goyanes S et al (2016) Biodegradability and plasticizing effect of yerba mate extract on cassava starch edible films. Carbohydr Polym 151:150–159. https://doi.org/10.1016/j.carbpol.2016.05.025
18. Saberi B, Chockchaisawasdee S, Golding JB et al (2017) Physical and mechanical properties of a new edible film made of pea starch and guar gum as affected by glycols, sugars and polyols. Int J Biol Macromol 104:345–359. https://doi.org/10.1016/j.ijbiomac.2017.06.051
19. Chabbi J, Jennah O, Katir N et al (2018) Aldehyde-functionalized chitosan-montmorillonite films as dynamically-assembled,

switchable-chemical release bioplastics. Carbohydr Polym 183:287–293. https://doi.org/10.1016/j.carbpol.2017.12.036
20. Ginting MHS, Hasibuan R, Lubis M et al (2018) Supply of avocado starch (Persea americana mill) as bioplastic material. IOP Conf Ser Mater Sci Eng 309. https://doi.org/10.1088/1757-899X/309/1/012098
21. Maulida, Kartika T, Harahap MB, Ginting MHS (2018) Utilization of mango seed starch in manufacture of bioplastic reinforced with microparticle clay using glycerol as plasticizer. In: IOP Conference Series: Materials Science and Engineering
22. Santana RF, Bonomo RCF, Gandolfi ORR et al (2018) Characterization of starch-based bioplastics from jackfruit seed plasticized with glycerol. J Food Sci Technol 55:278–286. https://doi.org/10.1007/s13197-017-2936-6
23. Shellikeri A, Kaulgud V, Yaradoddi J et al (2018) Development of neem based bioplastic for food packaging application. IOP Conf Ser Mater Sci Eng 376. https://doi.org/10.1088/1757-899X/376/1/012052
24. Mostafa NA, Farag AA, Abo-dief HM, Tayeb AM (2018) Production of biodegradable plastic from agricultural wastes. Arab J Chem 11:546–553. https://doi.org/10.1016/j.arabjc.2015.04.008
25. Hasan M, Rahmayani RFI, Munandar (2018) Bioplastic from chitosan and yellow pumpkin starch with castor oil as plasticizer. In: IOP Conference Series: Materials Science and Engineering
26. Isroi, Rahman A, Syamsu K (2018) Biodegradability of oil palm cellulose-based bioplastics. In: IOP Conference Series: Earth and Environmental Science
27. Nandi S, Guha P (2018) Modelling the effect of guar gum on physical, optical, barrier and mechanical properties of potato starch based composite film. Carbohydr Polym 200:498–507. https://doi.org/10.1016/j.carbpol.2018.08.028
28. Liang T, Wang L (2018) Preparation and characterization of a novel edible film based on Artemisia sphaerocephala Krasch. gum: effects of type and concentration of plasticizers. Food Hydrocoll 77:502–508. https://doi.org/10.1016/j.foodhyd.2017.10.028
29. Zuo G, Song X, Chen F, Shen Z (2019) Physical and structural characterization of edible bilayer films made with zein and corn-wheat starch. J Saudi Soc Agric Sci 18:324–331. https://doi.org/10.1016/j.jssas.2017.09.005
30. Fathanah U, Lubis MR, Nasution F, Masyawi MS (2018) Characterization of bioplastic based from cassava crisp home industrial waste incorporated with chitosan and liquid smoke. In: IOP Conference Series: Materials Science and Engineering
31. Hasan AMA, Abdel-Raouf ME (2018) Applications of guar gum and its derivatives in petroleum industry: a review. Egypt J Pet 27:1043–1050. https://doi.org/10.1016/j.ejpe.2018.03.005
32. Gong H, Liu M, Chen J et al (2012) Synthesis and characterization of carboxymethyl guar gum and rheological properties of its solutions. Carbohydr Polym 88:1015–1022. https://doi.org/10.1016/j.carbpol.2012.01.057
33. Manna PJ, Mitra T, Pramanik N et al (2015) Potential use of curcumin loaded carboxymethylated guar gum grafted gelatin film for biomedical applications. Int J Biol Macromol 75:437–446. https://doi.org/10.1016/j.ijbiomac.2015.01.047
34. Gupta AP, Verma DK (2014) Preparation and characterization of carboxymethyl guar gum nanoparticles. Int J Biol Macromol 68:247–250. https://doi.org/10.1016/j.ijbiomac.2014.05.012
35. Dalei G, Das S (2022) Carboxymethyl guar gum: a review of synthesis, properties and versatile applications. Elsevier 176:111433. https://doi.org/10.1016/j.eurpolymj.2022.111433
36. Qamruzzaman M, Ahmed F, Mondal MIH (2022) An overview on starch-based sustainable hydrogels: potential applications and aspects. J Polym Environ 30:19–50. https://doi.org/10.1007/s10924-021-02180-9
37. Meena P, Singh P, Warkar SG (2023) Development and assessment of carboxymethyl tamarind kernel gum-based pH-responsive hydrogel for release of diclofenac sodium. Eur Polym J 197:112340. https://doi.org/10.1016/j.eurpolymj.2023.112340
38. Meena P, Singh P, Warkar SG (2023) Fabrication and evaluation of stimuli - sensitive xanthan gum - based hydrogel as a potential carrier for a hydrophobic drug ibuprofen. Colloid Polym Sci. https://doi.org/10.1007/s00396-023-05198-8
39. Tushar YS, Meena P, Warkar SG (2023) Synthesis and characterization of novel xanthan gum - based pH - sensitive hydrogel for metformin hydrochloride release. Colloid Polym Sci. https://doi.org/10.1007/s00396-023-05135-9
40. Kajal KR, Meena P, Warkar SG (2023) Development and characterization of pH-responsive CMTKG/PAM/PEG hydrogel for oral administration of etophylline. Colloid Polym Sci. https://doi.org/10.1007/s00396-023-05152-8
41. Chauhan D, Kumar A, Warkar SG (2022) Synthesis, characterization and metal ions sensing applications of meta-benziporphodimethene-embedded polyacrylamide/carboxymethyl guargum polymeric hydrogels in water. Environ Technol 43:991–1002. https://doi.org/10.1080/09593330.2020.1812730
42. Gao Y, Zheng H, Wang J et al (2022) Physicochemical properties of zein films cross-linked with glutaraldehyde. Polym Bull 79:4647–4665. https://doi.org/10.1007/s00289-021-03723-9
43. Alix S, Marais S, Morvan C, Lebrun L (2008) Biocomposite materials from flax plants: preparation and properties. Compos Part A Appl Sci Manuf 39:1793–1801. https://doi.org/10.1016/j.compositesa.2008.08.008
44. Turasan H, Barber EA, Malm M, Kokini JL (2018) Mechanical and spectroscopic characterization of crosslinked zein films cast from solutions of acetic acid leading to a new mechanism for the crosslinking of oleic acid plasticized zein films. Food Res Int 108:357–367. https://doi.org/10.1016/j.foodres.2018.03.063
45. Mehboob S, Ali TM, Sheikh M, Hasnain A (2020) Effects of cross linking and/or acetylation on sorghum starch and film characteristics. Int J Biol Macromol 155:786–794. https://doi.org/10.1016/j.ijbiomac.2020.03.144
46. Russo R, Malinconico M, Santagata G (2007) Effect of cross-linking with calcium ions on the physical properties of alginate films. Biomacromol 8:3193–3197. https://doi.org/10.1021/bm700565h
47. Bas S, Soucek MD (2013) Comparison of film properties for crosslinked core-shell latexes. React Funct Polym 73:291–302. https://doi.org/10.1016/j.reactfunctpolym.2012.05.005

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.

Machine Vision and Applications (2024) 35:68
https://doi.org/10.1007/s00138-024-01544-0

RESEARCH

$FDT - Dr^2T$: a unified Dense Radiology Report Generation Transformer framework for X-ray images

Dhruv Sharma[1] · Chhavi Dhiman[1] · Dinesh Kumar[1]

Received: 16 November 2023 / Revised: 24 March 2024 / Accepted: 17 April 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
Medical Image Captioning (MIC), is a developing area of artificial intelligence that combines two main research areas, computer vision and natural language processing. In order to support clinical workflows and decision-making, MIC is used in a variety of applications pertaining to diagnosis, therapy, report production, and computer-aided diagnosis. The generation of long and coherent reports highlighting correct abnormalities is a challenging task. Therefore, in this direction, this paper presents an efficient $FDT - Dr^2T$ framework for the generation of coherent radiology reports with efficient exploitation of medical content. The proposed framework leverages the fusion of texture features and deep features in the first stage by incorporating ISCM-LBP + PCA-HOG feature extraction algorithm and Convolutional Triple Attention-based Efficient XceptionNet ($C - TaXNet$). Further, fused features from the FDT module are utilized by the Dense Radiology Report Generation Transformer (Dr^2T) model with modified multi-head attention generating dense radiology reports by highlighting specific crucial abnormalities. To evaluate the performance of the proposed $FDT - Dr^2T$ extensive experiments are conducted on publicly available IU Chest X-ray dataset and the best performance of the work is observed as 0.531 BLEU@1, 0.398 BLEU@2, 0.322 BLEU@3, 0.251 BLEU@4, 0.384 CIDEr, 0.506 ROUGE-L, 0.277 METEOR. An ablation study is carried out to support the experiments. Overall, the results obtained demonstrate the efficiency and efficacy of the proposed framework.

Keywords Deep-features · Texture features · Transformer · Medical image captioning · XceptionNet · Computer vision · Natural language processing · Tripple attention

1 Introduction

Image captioning [1] provides insights into a given image semantically and syntactically in the form of natural language sentences. This research area also finds application in the medical domain for the generation of medical reports [2]. Medical Image Captioning (MIC) [3] can assist doctors by accelerating the reporting process and reducing their workload. Figure 1 presents an example of the captioning of medical images. Medical image captioning highlights the relationships between image objects and clinical findings, which makes it a very challenging task. The generation of medical reports not only reduces the doctor's workload and accelerates clinical workflow but also aids in the efficient exploitation of medical content. Therefore, this produces faster and more accurate interpretations of findings, and offers important assistance to doctors.

Radiology and pathology medical images are widely used for the detection and treatment of different diseases [4, 5]. Generation of medical reports in multiple sentences from these images is a tedious task as this includes comprehensive examination e.g., X-ray images necessitate a detailed interpretation of visible information, including the airway, lung, cardiovascular system, and disability. Therefore, such interpretation requires the use of fundamental physiological knowledge in addition to a comprehensive comprehension of results from laboratories, electrocardiograms, and pulmonary function tests. Therefore, there has been a significant increase in interest in both clinical medicine and artificial intelligence for automatic report-generating technology.

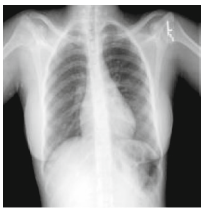
✉ Dinesh Kumar
dineshkumar@dtu.ac.in

Dhruv Sharma
dhruv.0906@yahoo.in

Chhavi Dhiman
chhavi.dhiman@dtu.ac.in

[1] Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India

Published online: 21 May 2024



Normal Chest with no pulmonary enema. The cardiac silhouette and mediastinum size are within normal limits. There is no focal consolidation. There are no traces of pleural effusion. There is no evidence of pneumothorax.

Fig. 1 An example of radiology report generated

Also, for a large number of radiological images manual writing of reports becomes a significant load for radiologists. Additionally, for radiologists with less experience, some anomalies in radiological images may be disregarded and not recorded in the reports [6]. Therefore, automatically generated radiography reports become a necessity to alleviate radiologists from their excessive burden and to alert novice radiologists to anomalies.

With the increase in the advancements in technology, different methods [2, 7, 8] are developed to automatically generate medical reports based on the data available. These methods generated incorrect reports for some cases with erroneous and repeated sentences. Some models [2, 9, 10] generated incoherent sentences with incorrect order of words, which were not clinically acceptable. Also, these methods failed to report some rare but important abnormalities due to ambiguities and incorrect detection of objects from radiological images. To overcome these challenges, this paper presents an efficient framework for the generation of paragraph-based reports sometimes also known as dense MIC. The proposed framework is split into two stages (as depicted in Fig. 2): (1) FDT-Image feature extraction module and (2) Dense Radiology Report Generation Transformer (Dr^2T) model for coherent medical report generation. The proposed FDT module efficiently fuses or concatenates deep features and texture features to extract effective features from the data available. The fused features obtained from the first stage are further leveraged by the transformer-based module for efficient and coherent report generation. The key contributions of the proposed work are summarized as:

- The paper presents an FDT feature extraction module that extracts salient features from medical images concatenation of deep features (Convolutional Triple Attention-based Efficient XceptionNet, $C-TaXNet$) and texture or hand-crafted features Image Segmentation-based Central Multiscale-LBP (ISCM-LBP + HOG-PCA).
- The paper also presents a Dense Radiology Report Generation Transformer (Dr^2T) with Modified-MHA which selects the best head from the structure and concatenates best-head with the multi-head representations to explore more correlations for higher-order feature representations.
- Extensive experiments are conducted on the publically available IU Chest X-ray dataset. The proposed framework provides state-of-the-art results in terms of BLEU, METEOR, CIDEr, and ROUGE-L evaluation metrics.
- An ablation study is also carried out to exhibit the effect of different feature representation methods and generate medical reports using the baseline transformer and the proposed Dr^2T.

The organization of the rest of the paper is as follows: Sect. 2 discusses previous works on medical image captioning. Section 3 provides a detailed explanation of the proposed methodology. Furthermore, Sect. 4 presents the dataset and implementation details with the results obtained. Finally, we conclude the work in Sect. 5.

2 Related works

The task of image captioning [1, 11] has gained immense popularity in the recent past. The early methods [12, 13] are based on an encoder–decoder architecture that generates single-sentence natural language descriptions of images. With the advancements made in technology, paragraph-based descriptions of images came into existence [14, 15].

MIC highlights the relationships between image objects and clinical findings. Radiologists with or without experience find writing medical reports a time-consuming and tedious task [2]. Also, the generation of error-free reports is also an intimidating challenge. In this direction, similar to image paragraph captioning few works [2, 16, 17] adopted hierarchical-RNN structures for the generation of fluent medical reports. These methods though generated fluent reports still suffer from data deviation which leads to the generation of general reports without prominent abnormalities [2, 17–19]. Also, some methods [8, 16, 20] have been proposed to alleviate data deviation.

Yuan et al. [16] fine-tuned the encoder to extract the most common medical concepts from x-ray images using radiology reports as training data. Another work [17] incorporated CNN's and LSTM in a recurrent manner to generate high-level conclusive impressions. Jing et al. [18] introduced two RNNs for writing accurate reports with normal and abnormal sentences. Further, [8] presented a hybrid model that enhanced the ability of the model by generating normal and abnormal sentences. Srinivasan et al. [21] utilized abnormality detection and a transformer with dual encoders and decoders for the generation of medical reports. Liu et al. [22] demonstrated a domain-aware automated system for creating chest X-ray radiology reports that conditionally generates sentences based on the themes that will be covered in the report. Lovelace et al. [23] developed a radiology report generation model using the transformer architecture

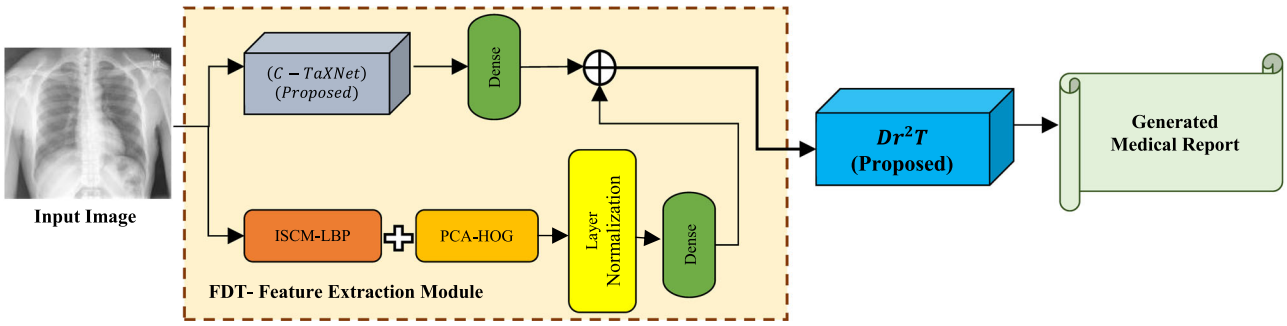


Fig. 2 Block diagram of the proposed FDT- Dr^2T framework

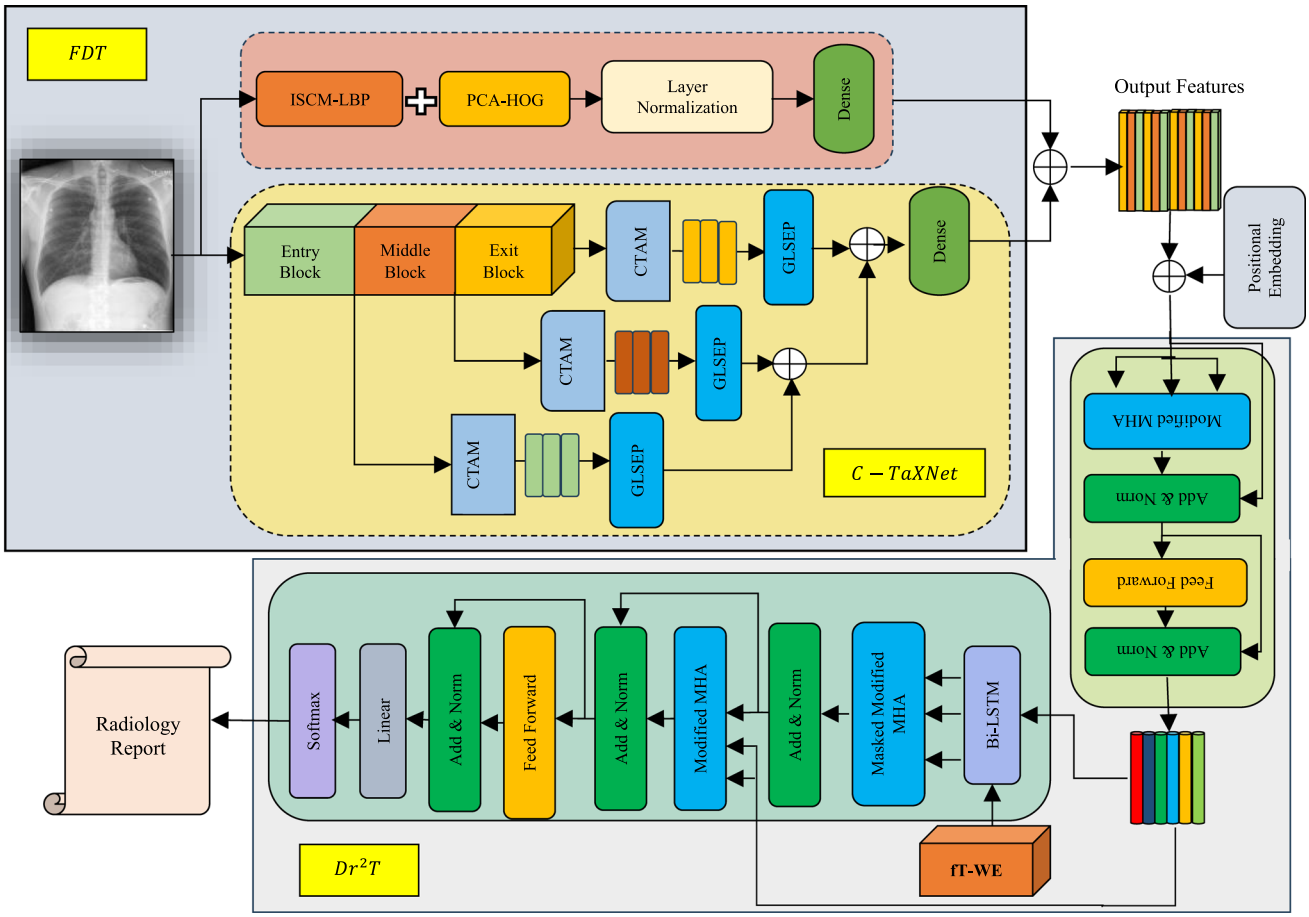


Fig. 3 Proposed FDT- Dr^2T framework

that outperforms competitor baselines in terms of both clinical coherence and standard language generation criteria. The above-mentioned methods generated incoherent sentences with incorrect order of words, which are not clinically acceptable. Also, some methods fail to describe some rare but important abnormalities due to ambiguities and incorrect detection of objects from medical images. To generate coherent and semantically and syntactically correct reports the proposed model utilized a fusion of deep and texture features for extraction of enhanced features. Further, these features are utilized by a Modified-MHA based transformer network for the generation of medical reports by considering higher-order multi-modal interactions.

3 Proposed methodology

This section provides the details for the proposed FDT- Dr^2T framework as depicted in Fig. 3.

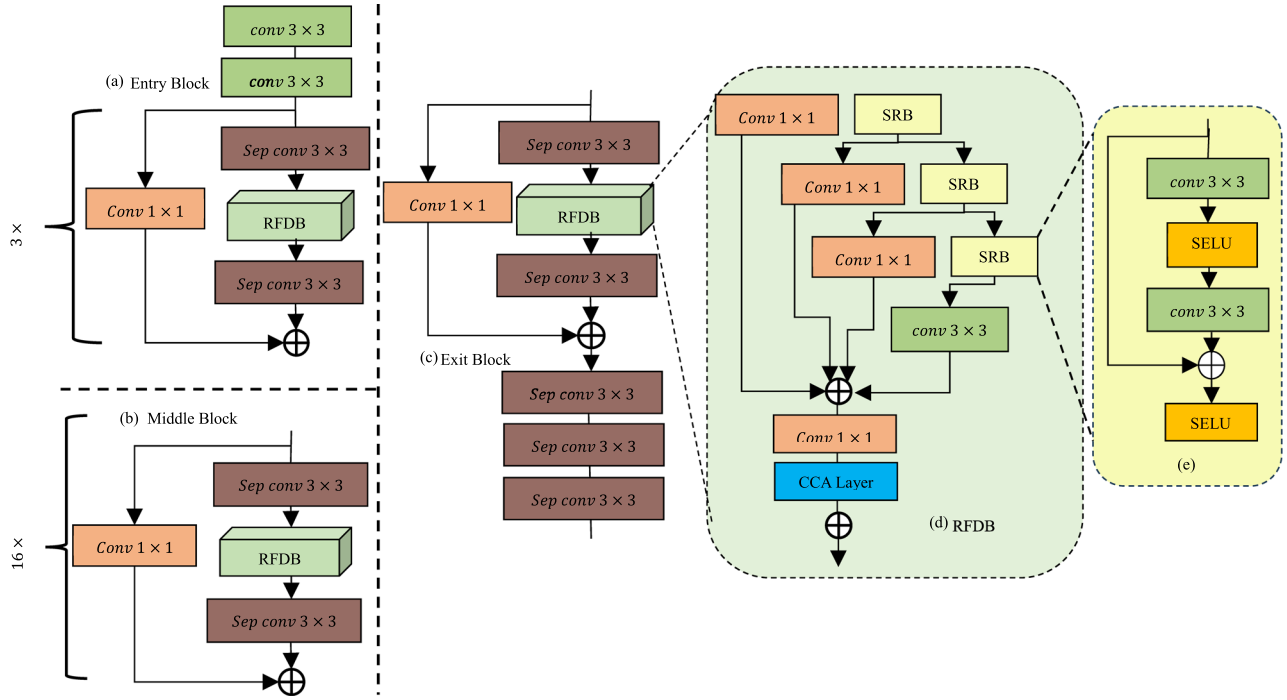


Fig. 4 Modified XceptionNet **a** entry block, **b** middle block, and **c** exit block, **d** RFDB block, **e** SRB block

3.1 FDT-image feature extraction

Medical image feature extraction aims to find the most compact and informative set of features. Deep-feature extraction incorporates a modified XceptionNet [24] module with a Residual Feature Distillation Block (RFDB) [25] and a Shallow Residual Block (SRB). Further, the proposed modified XceptionNet module also leverages a triplet attention module for the extraction of multi-level rich deep -features. Further, for the extraction of texture features, the ISCM-LBP [26] algorithm is incorporated with reduced-dimensional HOG features [27]. This helps to preserve a trade-off between model performance and parameters and ensures the appropriate level selection for the extraction of pertinent texture features.

3.1.1 Convolutional triple attention-based efficient XceptionNet ($C - TaXNet$)

The proposed $C - TaXNet$ model (Fig. 3) leverages a modified version XceptionNet module with RFDB and SRB as depicted in Fig. 4. RFDB utilizes multiple features distillation connections to learn more discriminative feature representation. For more fine-grained residual learning SRB is utilized without introducing additional parameters [28]. Further, the SRB is equipped with SELU activation function [29] described mathematically as:

$$g(x) = \begin{cases} \lambda x, & x > 0 \\ \alpha\lambda(e^x - 1), & x \leq 0 \end{cases} \tag{1}$$

where α and λ are the activation constants with approximate values of $\alpha = 1.6732$ and $\lambda = 1.0507$. Also, the proposed $C - TaXNet$ utilizes the concept of Convolutional Triple Attention Module (CTAM) [30] (almost parameter-free) that model channel attention and spatial attention [31]. The internal structure of the CTAM module is presented in Fig. 5. Consider $f_{input} \in \mathbb{R}^{C \times H \times W}$ be the feature vector of the corresponding input image which serves as an input to the Entry block of the $C - TaXNet$. The features extracted from each Xception block are made to pass through CTAM which is a three-stage architecture. For the first stage, height and channel dimension interactions are calculated by rotating the input 90° along the height dimension. The next stage calculates the interaction by rotating the input tensor to 90° anti-clockwise along W axis. For the final stage, the channels of the input tensor are reduced by using Z-pool. The Z-pool concatenates the AveragePool and MaxPool across that dimension. Mathematically, Z-pool can be represented as:

$$z - pool(x) = [MaxPool_{0d}(x) AvgPool_{0d}(x)] \tag{2}$$

where $0d$ is the zeroth dimension across which the AveragePool and MaxPool operation takes place. Further, the refined triple attended output is represented by:

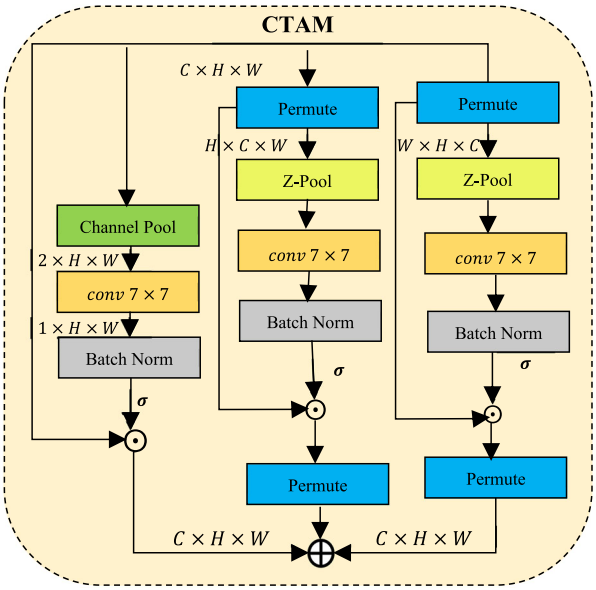


Fig. 5 Conventional triple attention module

$$y = \frac{1}{3}\left(\overline{\hat{x}_1\sigma(\varphi_1(\hat{x}_1^*))} + \overline{\hat{x}_2\sigma(\varphi_2(\hat{x}_2^*))} + x\sigma(\varphi_3(\hat{x}_3^*))\right) \tag{3}$$

In the above equation, σ represents the sigmoid activation function; φ_1, φ_2, and φ_3 represent the standard 2-D convolutional layers in three branches of triplet attention. Also, after the application of CTAM at each block of the Efficient XceptionNet we apply Global Log-Sum-Exp Pooling (GLSEP) [32] individually and concatenate the reduced features into a single vector.

3.1.2 ISCM-LBP + PCA-HOG feature extraction algorithm

This section provides the details for the hand-crafted feature extraction algorithm. Traditional LBP features [33] are sensitive to noise and incorporate center window pixels as the threshold, without any enhancement in the center pixel and considering the relationship between the center pixel and the neighborhood pixel [34]. When the central pixel intensity value is too large or too small then there might be a possibility that the detailed features are missed. Therefore, ISCM-LBP algorithm is incorporated that utilizes standard deviation to improve the resilience of the center pixel and the function of the neighborhood pixel. This further enhances the strength of the center pixel and the role of the neighborhood pixel. The specific operations are as follows:

- The 3×3 window is subdivided into sub-windows of size 3×3 to calculate the average gray value of each window

Finally, standard deviation σ is calculated for 9 windows.

$$\sigma = \sqrt{\frac{\sum_{j=1}^{9}(\tilde{\mathcal{X}}_j - \widetilde{\mathcal{X}})^2}{9}} \tag{4}$$

where $\tilde{\mathcal{X}}_j$ average gray value of each window and $\tilde{\mathcal{X}}$ is the average value of the entire window.

- For threshold value, for $\ell \leq \sigma$, the threshold q_t is the median of the nine pixels in the 3×3 window and LBP is evaluated by the formula

$$LBP(X_c, Y_c) = \sum_{P=0}^{7} 2^P \mathbb{S}(q_p - q_c) \tag{5}$$

$$\mathbb{S}(N) = \begin{cases} 1, & N \geq 0 \\ 0, & N < 0 \end{cases} \tag{6}$$

Further, if $\ell > \sigma$, let m, and n be the maximum and minimum value from the 9-pixel values. The threshold is evaluated using:

$$q_t = \frac{m + n + \tilde{q}_c}{3} \tag{7}$$

- Finally, for median M of nine pixels, ISCM-LBP features are evaluated using the formula:

$$ISCM - LBP = \sum_{P=0}^{7} 2^P \mathbb{S}(\tilde{q}_p - q_t) \tag{8}$$

$$q_t = \begin{cases} M, & \ell \leq \sigma \\ \frac{m+n+\tilde{q}_c}{3}, & \ell > \sigma \end{cases} \tag{9}$$

$$\mathbb{S}(N) = \begin{cases} 1, & \tilde{q}_p - q_t \geq 0 \\ 0, & \tilde{q}_p - q_t < 0 \end{cases} \tag{10}$$

The ISCM-LBP only extracts the texture information features and somehow neglects some edge density features from images. Therefore, we incorporate dimensionality reduction of histogram of gradient orientation i.e., PCA-HOG. This helps in the extraction of large dimensional HOG features which are reduced with the utilization of PCA mapping. Finally, the ISCM-LBP and HOG-PCA features are serially concatenated that helps to retain all the texture feature details.

Flowchart depicting the extraction process for the ISCM-LBP + PCA-HOG algorithm is presented in Fig. 6. Furthermore, the features obtained from $C - TaXNet$ and ISCM-LBP + PCA-HOG are fed to dense blocks which are the learnable feature encodings that converts the deep-features and hand-crafted features into learnable vectors $\mathbb{f}_{deep}$ and $\mathbb{f}_{hcf}$.

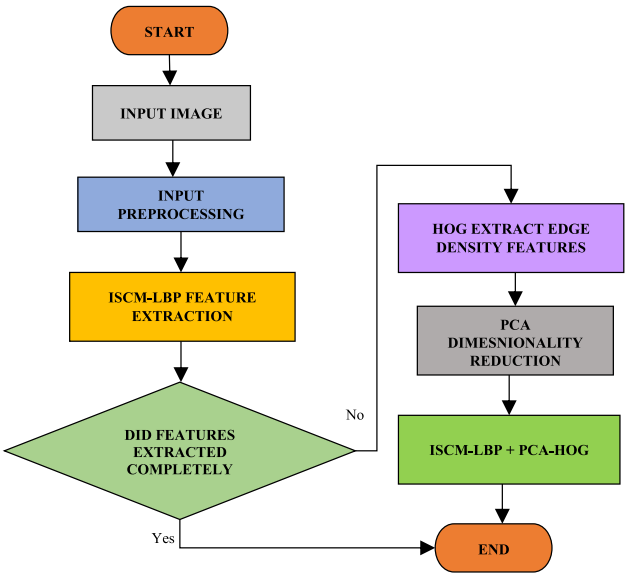


Fig. 6 Flowchart for ISCM-LBP + PCA-HOG algorithm

3.2 Dense Radiology Report Generation Transformer (Dr^2T)

The image features obtained from the FDT-Image feature extraction module are fed to the transformer-based architecture for the generation of dense medical reports. Figure 3 presents the detailed structure of the proposed Dr^2T transformer model. The traditional transformer model incorporates Multi-Head Attention (MHA) and calculates a weighted average of values based on the similarity between queries and keys. The attention scores assigned to each value determine its importance. To extract more minute details in the form of deep local features we present a modified MHA as shown in Fig. 7, which selects the best-head from the structure and concatenates the best-head and the multi-head representations to explore more correlations between the higher-order feature representations.

3.2.1 Dr^2T encoder

The encoder structure of the Dr^2T converts a set of visual attributes from the input image into a number of encodings. The encodings or embeddings obtained are represented in the form of a linear transformation of local features as keys $\hbar$, query q, and values v respectively. With this, we first calculate the multi-head self-attention scores $m\hbar_i$ using scaled-dot product attention (SPA). Mathematically,

$$mathcalligramh_i = [\hbar_1, \hbar_2 \ldots, \hbar_i] \tag{11}$$

$$SPA = softmax\left(\frac{q_i \hbar_i^T}{\sqrt{d_i}}\right) v_i \tag{12}$$

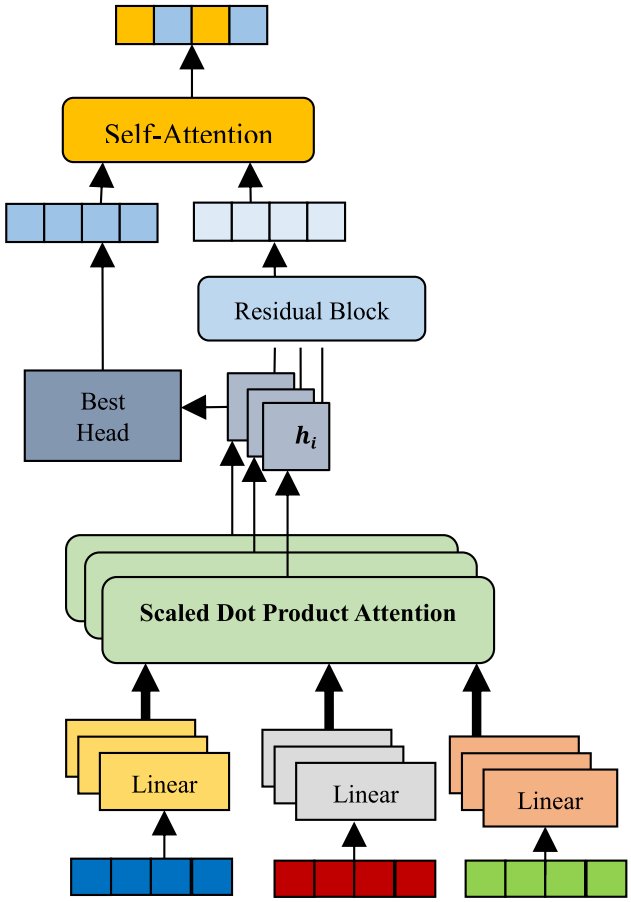


Fig. 7 Proposed modified MHA

Further, we select the best head ($m\hbar_b$) out of the calculated multi-head self-attention scores. Also, these heads $m\hbar_i$ are made to pass through to a residual block (RB) in order to avoid gradient explosion.

$$S_t = \sum RB(m\hbar_i) \tag{13}$$

The output generated from the RB and the best-selected head is fed to the self-attention network to provide a better correlation among the features obtained. The features obtained after the self-attention network is represented by:

$$\mathcal{S} = S - A(m\hbar_b, S_t) \tag{14}$$

The Modified-MHA features are added, normalized and fed to an FFN to get higher order visual features f_v as output which acts as input to Dr^2T Decoder module.

3.2.2 Dr^2T decoder

The Dr^2T decoder aims for the generation of dense medical reports either in the form of single or multiple sentences with the use of the high-level features obtained from the Dr^2T

encoder. A combination of visual features f_v and textual embeddings f_t are used to provide higher-order intra and inter-modal interactions between the image and text pairs. Also, the textual embeddings are extracted through fastText word embeddings (fT-WE) [35]. Therefore, textual and visual embeddings are concatenated via Bi-LSTM to perform multi-modal reasoning and high-level feature representation. This whole fT-WE with visual embedding (from Dr^2T encoder) and Bi-LSTM layer supports detailed encryption of visual data and provides inter-and intra-model interactions between the text and visual features.

At the decoding stage, we further integrate the Masked Modified-MHA and Modified-MHA with add and norm layers and feed-forward network layer. Finally, the medical report is generated by the generation of words through FC and softmax layers. The proposed Dr^2T model is trained by minimizing the cross-entropy loss:

$$\mathcal{L}_{XE}(\theta) = -\sum_{t=1}^{\mathcal{T}} \log(p_\theta(\omega_t^* | \mathbb{S}_{1:t-1}^*)) \tag{15}$$

where $\mathcal{T}$is the number of words in sentence; θ denotes all the parameters in the model; $\mathbb{S}_{1:t-1}^*$ is the ground truth.

4 Results and discussions

To evaluate the performance of the proposed framework, publically available dataset IU X-Ray [36] is used for the generation of paragraph-based reports. IU Chest X-Ray contains about 7470 pairs of images and reports. Each report consists of the following sections: impressions, findings, tags, comparison and indication. The report for the image includes the impression and findings only. Following standard procedure [37], images are normalized and resized to 224 × 224, making them appropriate for extracting visual features. The text data was pre-processed to produce 572 unique tags and 1915 unique words by changing all tokens to lowercase and deleting any non-alpha tokens. On average, there are 2.2 tags, 5.7 sentences, and 6.5 words per sentence attached to each image. Additionally, we discovered that the top 1000 terms account for 99.0% of all word occurrences in the dataset; as a result, we only used the top 1000 terms listed in dictionaries. Finally, we chose 500 images at random for validation and 500 images as a test. Further, to validate the effectiveness of the proposed Dr^2T we evaluated BLUE-n [38], METEOR [39], CIDEr [40], and ROUGE-L [41] scores.

4.1 Implementation details

For the first phase of the experiment, the input images are resized into 480 × 480 resolution and Adam [42] optimizer is used for training of $C-TaXNet$ and texture feature extraction module. The feature extraction FDT module is trained for 100 epochs with a learning rate of $10e-4$ and batch size of 32. For second phase, the Dr^2T model is trained for further 70 epochs with a learning rate of $1e-4$ for the encoder and decoder. We apply a linear decay scheduler with a warmup ratio of 0.01 to control the learning rate. For regularization, we set dropout to 0.1 and use a weight decay of 0.01. For the generation of paragraph, we set the dimensions of all hidden states and word embeddings as 512 and the maximum output length is set as 45. For both FDT and Dr^2T modules, we adopt a softmax cross-entropy loss.

Table 1 Comparison of the performance of the proposed $C-TaXNet$ and other Deep-feature Extraction Framework

Model	Accuracy (%)	Loss
ResNet [28]	89.64	3.547
VGGNet [43]	90.32	3.10
AlexNet [44]	89.10	3.09
Inception-V3 [45]	94.35	2.341
DenseNet [46]	94.13	2.61
XceptionNet [24]	96.80	1.928
Proposed ($C-TaXNet$)	98.74	1.29

4.2 Quantitate comparison with state-of-the-art

For the deep-feature extraction module, the proposed $C-TaXNet$ is compared with different generic networks (ResNet, VGGNet, Inception-V3, and XceptionNet) in terms of accuracy and loss. Table 1 presents the results obtained for the proposed $C-TaXNet$, it is evident that the proposed network provides significant improvements in terms of accuracy with the addition of triple attention module. Also, Fig. 8 presents the attention maps generated from the proposed $C-TaXNet$. The proposed model attends to the input image efficiently by highlighting the minute image regions that are of importance.

Table 2 presents the dimension of extracted features and feature extraction speed for the proposed ISCM-LBP + HOG-PCA algorithm. From Table 2 it is evident that with the improvement in speed of feature extraction a reduction in data redundancy is observed. The results prove that more accurate image features with strong anti-interference capabilities and minimal computing complexity are extracted by the ISCM-LBP approach. Additionally, PCA-HOG extracts the image edge features and ensures that, even at accelerated extraction speeds, the feature dimension is successfully lowered without compromising the edge features that have already been retrieved. With serial fusion of ISCM-LBP and HOG-PCA relatively complete image features with edge and

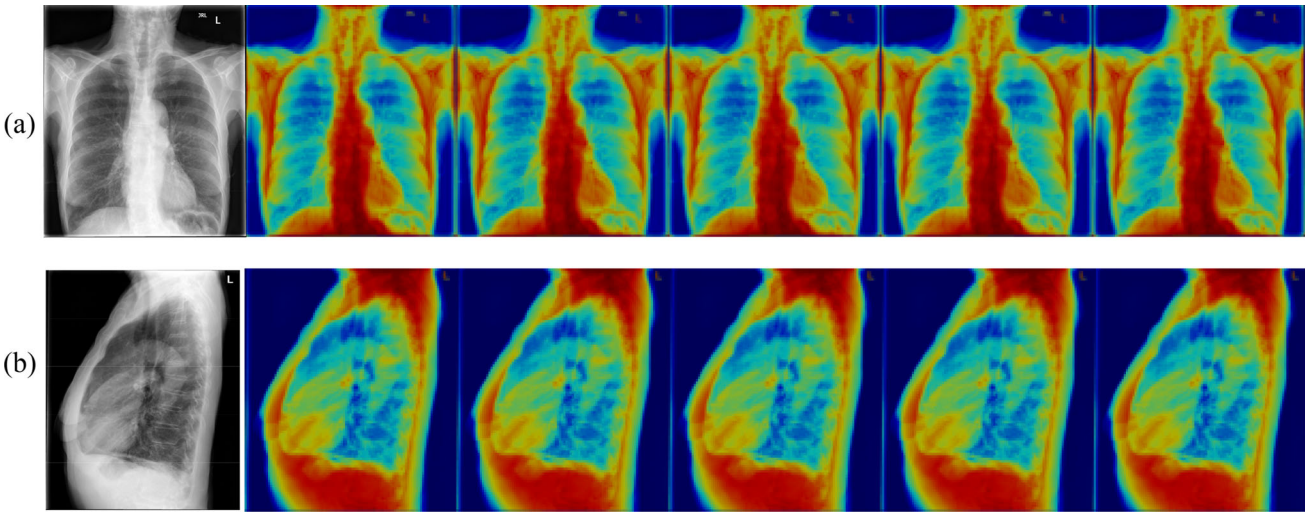


Fig.8 Attention visualization generated by the proposed $C-TaXNet$

Table 2 Comparison of the performance of different texture feature extraction algorithms

Method	Feature dimension	Feature extraction time (s)
LBP	1982	0.0121
ISCM-LBP	2770	0.0068
HOG	3221	0.0039
HOG-PCA	1026	0.0070
ISCM-LBP + HOG-PCA	2951	0.0105

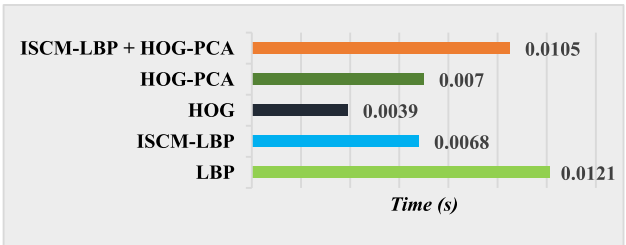


Fig. 9 Feature extraction time taken by different texture feature extraction algorithms

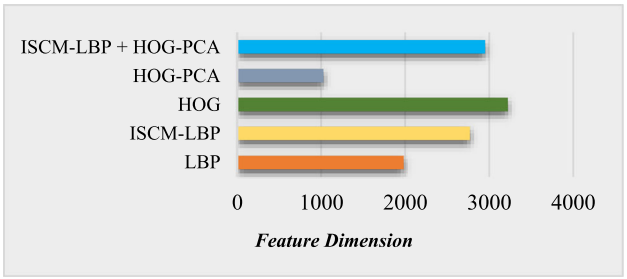


Fig. 10 Comparison of performance of different texture feature extraction algorithm

texture features are obtained. Figures 9 and 10 compares different texture feature extraction algorithms on the basis of the feature dimension and time taken (s) for the extraction of features.

Table 3 reports the quantitative results obtained on the proposed FDT-DMICT. We compare our proposed approach with a wide range of state-of-the-art medical report generation methods, i.e., Variational topic inference (VTI) [37], a graph-based method [47], cross-modal memory (CMR) [48], CMAS [18], Eddie-Transformer [49], DeltaNet [50], IIHT [51], PPKED [52], Co-Att [2], RTMIC [53]. From the results reported it is evident that the $FDT-Dr^2T$ framework outperforms the recent LXMERT [54], ITHN [55], CMCA [56], ORGAN [57] state-of-the-art as well by a large margin across all evaluation metrics. Furthermore, the proposed work outperforms different hierarchical-based approaches [7, 8, 18, 20]. The proposed FDT – Dr^{2}T reports abnormalities by detecting opacity by observing the contrast, detecting hyperexpanded lungs by determining the change in size, and identifying the location of airspace disease.

4.3 Qualitative results for the proposed $FDT-Dr^2T$

Figure 11 presents the reports generated by the proposed $FDT-Dr^2T$. The figure describes the report generated corresponding to the input image and the ground truth. From the qualitative results portrayed in Fig. 11, we can infer that the proposed $FDT-Dr^2T$ framework is better at capturing abnormalities. For the third image, our framework easily detects and reports "T-spine osteophytes in the thoracic

Table 3 Comparison of quantitative results obtained for the proposed $FDT - Dr^2T$ framework with state-of-the-art

Method	B-1	B-2	B-3	B-4	M	CIDEr	R-L
VTI [37]	0.493	0.360	0.291	0.154	0.218	–	0.375
Wang et al. [47]	0.450	0.301	0.213	0.158	–	–	0.384
CMR [48]	0.475	0.309	0.222	0.170	0.191	–	0.375
TieNet [7]	0.330	0.194	0.124	0.081	–	–	–
Srinivasan et al. [21]	0.464	0.301	0.212	0.158	–	–	–
RTMIC [53]	0.350	0.234	0.143	0.096	–	–	–
Li et al. [8]	0.438	0.298	0.208	0.151	–	–	0.322
R2Gen [19]	0.470	0.304	0.219	0.165	0.187	–	0.371
Eddie-Transformer [49]	0.466	0.307	0.218	0.158	–	–	0.358
CMAS [18]	0.464	0.301	0.210	0.154	–	–	0.362
Delta-Net [50]	0.485	0.324	0.238	0.184	–	–	0.379
IIHT [51]	0.513	0.375	0.297	0.245	0.264	–	0.492
PPKED [52]	0.483	0.315	0.224	0.168	–	0.351	0.376
Co-Att [2]	0.517	0.386	0.306	0.247	0.217	0.327	0.447
LXMERT [54]	0.498	0.32	0.229	0.169	0.205	–	0.379
ITHN [55]	0.491	0.328	0.231	0.183	0.210	–	0.387
CMCA [56]	0.496	0.349	0.268	0.215	0.209		0.392
ORGAN [57]	0.510	0.346	0.255	0.195	0.205		0.399
Proposed ($FDT - Dr^2T$)	0.531	0.398	0.322	0.251	0.277	0.384	0.506

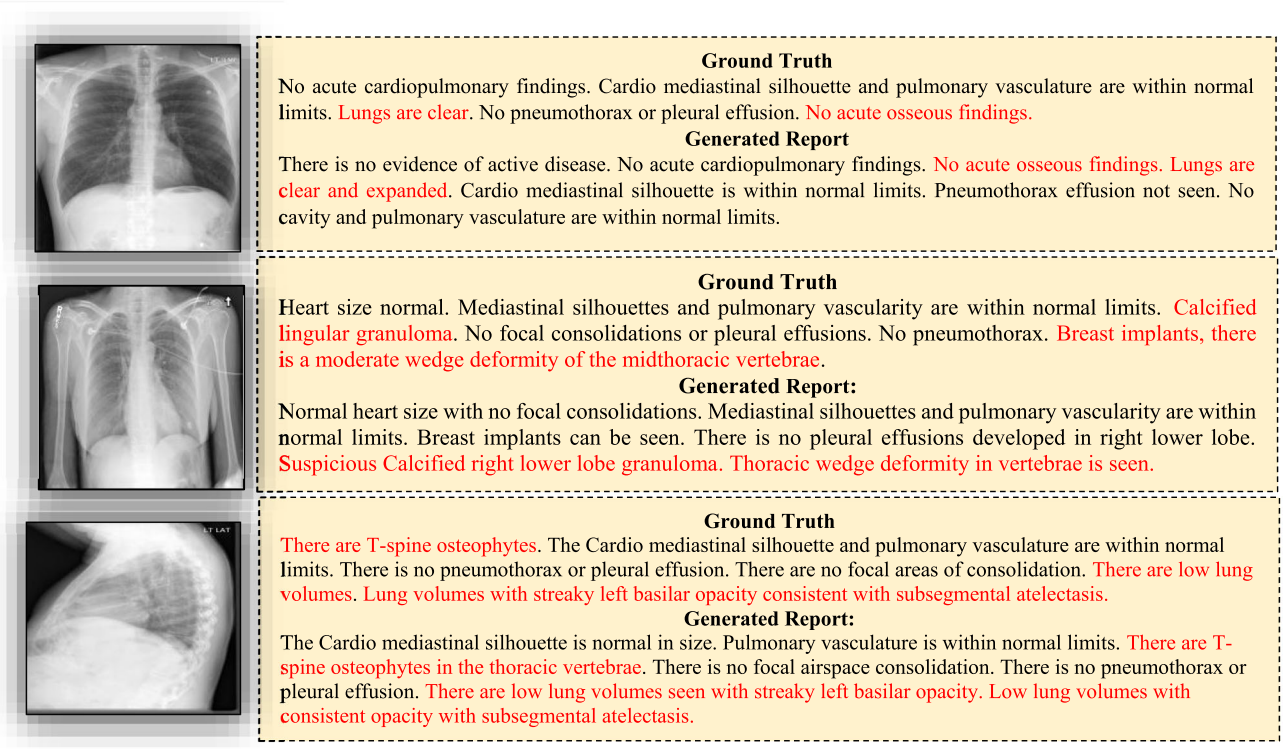


Fig. 11 Qualitative results for the proposed $FDT - Dr^2T$, (red marks highlight different abnormalities)

vertebrae". Therefore, the proposed framework has a remarkable ability to accurately generate comprehensive reports for abnormal cases. Further, with this, the proposed framework provides accurate locations of the abnormalities with a well balance of normal sentences and abnormal sentences. Also, we can infer from these results that the $FDT - Dr^2T$ can efficiently alleviate the visual data deviation problem by efficiently detecting changes in the contrast, size, and disease location, which helps in the coherent generation of medical reports. Further, there were very few cases observed where the proposed framework does not generate descriptions as per the grammar rules like in Fig. 11, the generated sentence is "Pneumothorax effusion not seen".

4.4 Ablation studies

An ablation study is conducted which exhibits the effect different modules for the efficient generation of paragraph-based medical reports. This study effectively highlights the influence of baseline Transformer and Dr^2T. This study also helps to gain a better understanding of the overall behavior of the ablated models on the performance of the system. Table 4 reports the results obtained for different models i.e., (M-1 to M-6), and Fig. 12 provides the qualitative results obtained for all the ablated models. We start with the baseline transformer model [56] with Inception-V3 and LBP + HOG features for the FDT module (M-1). The results obtained with M-1 do not provide significant results. In second model i.e., M-2, we replace Inception-V3 with XceptionNet, a slight improvement is observed in B-1, B-2, METEOR, and ROUGE-L scores. Further, in M-3, the baseline transformer model is replaced with the proposed Dr^2T, and significant qualitative and quantitative results are obtained in comparison with M-1 and M-2. With baseline transformer and XceptionNet with ISCM-LBP + HOG-PCA deep and texture features module (M-4), there is no significant improvement observed with respect to M-3 in the evaluation scores. Therefore, to improve the qualitative and quantitative results we exploited the proposed $C - TaXNet$ for deep-image feature extraction with ISCM-LBP + HOG-PCA and baseline transformer i.e., M-5. To further obtain coherent medical reports with correct abnormalities we employed the proposed FDT module ($C - TaXNet$ and ISCM-LBP + HOG-PCA) and the proposed Dr^2T transformer. M-6 provides better correlations for higher-order feature representations from medical images and generates coherent reports with rare and important abnormalities.

The qualitative results obtained for different models makes it evident that a coherent report is generated with the use of proposed $FDT - Dr^2T$ highlighting the abnormalities in the report generated. Furthermore, Fig. 13 presents the influence of the proposed Dr^2T and the baseline transformer on the with respect to evaluation parameters.

5 Conclusion

This paper an efficient two-stage framework, $FDT - Dr^2T$ for generation of dense and coherent radiology or medical reports. This model efficiently learns image features from the first stage to select the highest scoring head from the structure and concatenates best-head and the multi-head representations.

to explore more correlations and provide better higher-order feature representations. Also, the proposed framework is able to detect changes in the contrast, size, and disease location of medical images which help in the coherent generation of medical reports. Furthermore, the proposed $FDT - Dr^2T$ framework imitates the working patterns of radiologists by highlighting the specific crucial abnormalities in reports.

The effectiveness of the proposed framework is analyzed on the IU-Chest X-ray dataset and provided state-of-the-art results in terms of $BLEU@N$, $METEOR$, $CIDEr$, $ROUGE-L$ evaluation metrics. To support the experimental analysis an ablation study is also carried out that exhibits the effect of different feature representation methods and generates medical reports using the baseline transformer and the proposed Dr^2T. Further, there were very few cases (around 2–3%) observed where the proposed framework does not generate descriptions to the extent expected. These cases generated sentences with missing words and some grammatical errors. This could be alleviated by utilizing fused character or word embedding models or by incorporating attention-based embedding models.

In future, the proposed framework can be utilized for other fatal diseased parts like breast and brain, etc. subject to the availability of their captioning datasets. Also, the proposed model can be utilized for the generation of radiology reports for 3D imaging data i.e. brain MRI dataset. We can incorporate different modalities of medical images for the generation of more detailed radiology reports. Furthermore, explainable AI solutions paired with examination by qualified physicians appear to be quite useful in understanding the outcomes of complex deep learning models and facilitating the evaluation process.

Table 4 Ablation study results to study the influence of different FDT networks/algorithms with the baseline and the proposed Dr^2T transformer

Model	FDT		Transformer	B-1	B-2	B-3	B-4	M	CIDEr	R-L
	Deep-Features	Texture Features								
M-1	Inception-V3	LBP+HOG	Baseline [58]	0.311	0.184	0.126	0.092	0.164	0.302	0.384
M-2	XceptionNet	LBP+HOG	Baseline [58]	0.342	0.207	0.154	0.112	0.169	0.304	0.384
M-3	XceptionNet	LBP+HOG	$\boldsymbol{Dr^2T}$	**0.427**	**0.303**	**0.216**	**0.172**	**0.181**	**0.313**	**0.391**
M-4	XceptionNet	ISCM-LBP+HOG-PCA	Baseline [58]	0.429	0.296	0.216	0.172	0.183	0.315	0.390
M-5	$C-TaXNet$	ISCM-LBP+HOG-PCA	Baseline [58]	0.461	0.327	0.249	0.214	0.218	0.321	0.414
Proposed (M-6)	$\boldsymbol{C-TaXNet}$	**ISCM-LBP+HOG-PCA**	$\boldsymbol{Dr^2T}$	**0.531**	**0.398**	**0.322**	**0.251**	**0.277**	**0.384**	**0.506**



M-1: No acute cardiopulmonary abnormality. Lungs are clear. Pulmonary vasculature is within normal limits. No pneumothorax or pleural effusion.
M-2: No acute cardiopulmonary findings. Lungs are clear. No evidence of pleural effusion. Pulmonary vasculature is within normal limits. Lungs are clear. No pneumothorax or pleural effusion.
M-3: No acute cardiopulmonary findings. Pulmonary vasculature is within normal limits. No evidence of focal air space. There is no pneumothorax or pleural effusion. The heart is not enlarged.
M-4: No acute cardiopulmonary findings. Cardio mediastinal silhouette is unremarkable. Pulmonary vasculature is within normal limits. Lungs are clear. No pneumothorax effusion. No acute bony abnormality.
M-5: No acute cardiopulmonary abnormality. No acute bony abnormality. Pulmonary vasculature is within normal limits. Lungs are clear. No pneumothorax effusion. The heart is not enlarged.
M-6 ($\boldsymbol{FDT-Dr^2T}$): There is no evidence of active disease. No acute cardiopulmonary findings. No acute osseous findings. Lungs are clear and expanded. Cardio mediastinal silhouette is within normal limits. Pneumothorax effusion not seen. No cavity and pulmonary vasculature are within normal limits.

Fig. 12 Qualitative results obtained for different ablated models

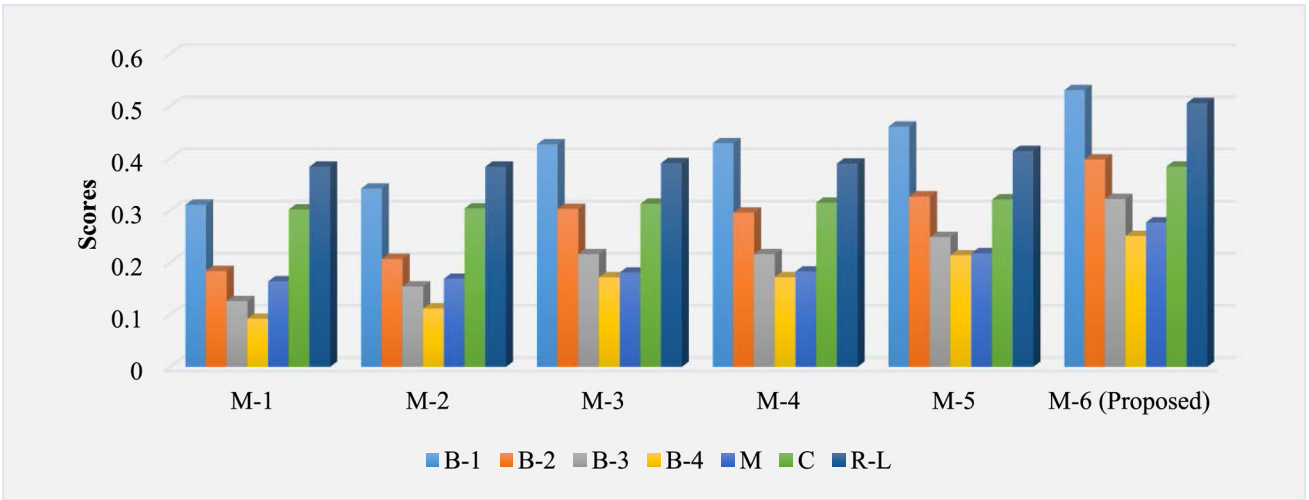


Fig. 13 Comparison highlighting the influence of different ablated models with respect to different evaluation parameters

Author contributions All authors contributed equally for the work submitted

Declarations

Conflict of interest The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

1. Sharma, D., Dhiman, C., Kumar, D.: Evolution of visual data captioning methods, datasets, and evaluation metrics: a comprehensive survey. Expert Syst. Appl. **221**, 119773 (2023)
2. Jing, B., Xie, P., Xing, E.: On the automatic generation of medical imaging reports. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia (2018)

3. Beddiar, D.-R., Oussalah, M., Seppänen, T.: Automatic captioning for medical imaging (MIC): a rapid review of literature. Artif. Intell. Rev. **56**, 4019–4076 (2023)
4. Shin, H.-C., Roberts, K., Lu, L., Demner-Fushman, D., Yao, J., Summers, R.M.: Learning to read chest X-rays: recurrent neural cascade model for automated image annotation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA (2016)
5. Coche, E., Ghaye, B., Mey, J., Duyck, P.: Difficulties in the Interpretation of Chest Radiography, Comparative Interpretation of CT and Standard Radiography of the Chest. Springer, Berlin (2011)
6. Brady, A., Laoide, R.O., McCarthy, P., McDermott, R.: Discrepancy and error in radiology: concepts, causes and consequences. Ulst. Med. J. **81**, 3–9 (2012)
7. Wang, X., Peng, Y., Lu, L., Lu, Z., Summers, R.M.: TieNet: text-image embedding network for common thorax disease classification and reporting in chest X-rays. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA (2018)
8. Li, Y., Liang, X., Hu, Z., Xing, E.P.: Hybrid retrieval-generation reinforced agent for medical image. In: Conference on Neural Information Processing Systems, Montreal Convention Centre (2018)
9. Li, C.Y., Liang, X., Hu, Z., Xing, E.P.: Knowledge-Driven Encode, Retrieve, Paraphrase for Medical Image Report Generation. arXiv:1903.10122 (2019)
10. Li, M., Wang, F., Chang, X., Liang, X.: Auxiliary Signal-Guided Knowledge Encoder–Decoder for Medical Report Generation. arXiv:2006.03744 (2020)
11. Sharma, D., Dhiman, C., Kumar, D.: XGL-T transformer model for intelligent image captioning. Multimed. Tools Appl. **83**, 4219–4240 (2024)
12. Vinyals, O., Toshev, A., Bengio, S., Erhan, D.: Show and tell: a neural image caption generator. In: IEEE Conference on Computer Vision and Pattern Recognition, Boston (2015)
13. Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., Zhang, L.: 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)
14. Krause, J., Johnson, J., Krishna, R., Fei-Fei, L.: A hierarchical approach for generating descriptive image paragraphs. arXiv:1611.06607 (2016)
15. Yang, L.-C., Yang, C.-Y., Hsu, J.Y.-J.: Object relation attention for image paragraph captioning. In: The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21) (2021)
16. Yuan, J., Liao, H., Luo, R., Luo, J.: Automatic radiology report generation based on multi-view image fusion and medical concept enrichment. arXiv:1907.09085 (2019)
17. Xue, Y., Xu, T., Long, L.R., Xue, Z., Antani, S., Thoma, G.R., Huang, X.: Multimodal recurrent model with attention for automated radiology report generation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018 (2018).
18. Jing, B., Wang, Z., Xing, E.: Show, describe and conclude: on exploiting the structure information of chest X-ray reports. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy (2019)
19. Chen, Z., Song, Y., Chang, T.-H., Wan, X.: Generating radiology reports via memory-driven transformer. arXiv:2010.16056 (2020)
20. Zhang, Y., Wang, X., Xu, Z., Yu, Q., Yuille, A., Xu, D.: When radiology report generation meets knowledge graph. arXiv:2002.08277v1 (2020)
21. Srinivasan, P., Thapar, D., Bhavsar, A., Nigam, A.: Hierarchical X-ray report generation via pathology tags and multi head attention. In: Asian Conference on Computer Vision, Kyoto, Japan (2021)
22. Liu, G., Hsu, T.-M.H., McDermott, M., Boag, W., Weng, W.-H., Szolovits, P., Ghassemi, M.: Clinically accurate chest X-ray report generation. In: Proceedings of the 4th Machine Learning for Healthcare Conference, PMLR (2019)
23. Lovelace, J., Mortazavi, B.: Learning to generate clinically coherent chest X-ray reports. In: Findings of the Association for Computational Linguistics: EMNLP 2020 (2020)
24. Chollet, F.: Xception: Deep Learning with Depthwise Separable Convolutions. arXiv:1610.02357 (2016)
25. Liu, J., Tang, J., Wu, G.: Residual Feature Distillation Network for Lightweight Image Super-Resolution. arXiv:2009.11551 (2020)
26. Liu, Q., Song, Y., Tang, Q., Bu, X., Hanajima, N.: Wire rope defect identification based on ISCM-LBP and GLCM features. Vis. Comput. **40**(2), 545–557 (2023)
27. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA (2005)
28. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. arXiv:1512.03385 (2015)
29. Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S.: Self-Normalizing Neural Networks. arXiv:1706.02515 (2017)
30. Misra, D., Nalamada, T., Arasanipalai, A.U., Hou, Q.: Rotate to Attend: Convolutional Triplet Attention Module. arXiv:2010.03045 (2020)
31. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: Convolutional Block Attention Module. arXiv:1807.06521 (2018)
32. Korschens, M., Bodesheim, P., Denzler, J.: Beyond global average pooling: alternative feature aggregations for weakly supervised localization. In: Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Online (2022)
33. Hadid, A.: The local binary pattern approach and its applications to face analysis. In: First Workshops on Image Processing Theory, Tools and Applications, Sousse, Tunisia (2008)
34. Pan, Z., Hu, S., Wu, X., Wang, P.: Adaptive center pixel selection strategy in local binary pattern for texture classification. Expert Syst. Appl. **180**, 115123 (2021)
35. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching Word Vectors with Subword Information. arXiv:1607.04606 (2017)
36. Fushman, D.D., Kohli, M.D., Rosenman, M.B., Shooshan, S.E., Rodriguez, L., Antani, S., Thoma, G.R., McDonals, C.J.: Preparing a collection of radiology examinations for distribution and retrieval. J. Am. Med. Inform. Assoc. **23**(2), 304–310 (2016)
37. Najdenkoska, I., Zhen, X., Worring, M., Shao, L.: Variational Topic Inference for Chest X-Ray Report Generation. arXiv:2107.07314 (2021)
38. Papineni, K., Roukos, S., Ward, T., Zhu, W.-J.: Bleu: a method for automatic evaluation of machine. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (2002)
39. Banerjee, S., Lavie, A.: 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)
40. Vedantam, R., Zitnick, C.L., Parikh, D.: Cider: consensus-based image description evaluation. In: IEEE Conference on Computer Vision and Pattern Recognition, Boston, (2015)
41. Lin, C.-Y.: ROUGE: a package for automatic evaluation of summaries. In: Association for Computational Linguistics, Barcelona, Spain (2004)
42. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic. arXiv:1412.6980 (2014)
43. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 (2014)
44. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems 25 (NIPS 2012) (2012)

45. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA (2016)
46. Huang, G., Liu, Z., Maaten, L.V.D., Weinberger, K.Q.: Densely Connected Convolutional Networks. arXiv:1608.06993 (2018)
47. Wang, S., Tang, L., Lin, M., Shih, G., Ding, Y., Peng, Y.: Prior knowledge enhances Radiology Report Generation. arXiv:2201.03761 (2022)
48. Chen, Z., Shen, Y., Song, Y., Wan, X.: Cross-Modal Memory Networks for Radiology Report Generation. arXiv:2204.13258 (2022)
49. Nguyen, H.T., Nie, D., Badamdorj, T., Liu, Y., Hong, L., Truong, J., Cheng, L.: EDDIE-transformer: enriched disease embedding transformer for X-ray report generation, Kolkata, India. In: IEEE 19th International Symposium on Biomedical Imaging (ISBI) (2022)
50. Wu, X., Yang, S., Qiu, Z., Ge, S., Yan, Y., Wu, X., Zheng, Y., Zhou, S.K., Xiao, L.: DeltaNet: conditional medical report generation for COVID-19 diagnosis, Gyeongju, Republic of Korea. In: Proceedings of the 29th International Conference on Computational Linguistics (2022)
51. Fan, K., Cai, X., Niranjan, M.: IIHT: Medical Report Generation with Image-to-Indicator Hierarchical Transformer. arXiv:2308.05633 (2023)
52. Liu, F., Wu, X., Ge, S., Fan, W., Zou, Y.: Exploring and distilling posterior and prior knowledge for radiology report generation, online. In: EEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021)
53. Xiong, Y., Du, B., Yan, P.: Reinforced Transformer for Medical Image Captioning, Machine Learning in Medical Imaging. Springer, Berlin (2019)
54. Wijerathna, V., Raveen, H., Abeygunawardhana, S., Ambegoda, T.D.: Chest X-ray caption generation with CheXNet. In: Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka (2022)
55. Voutharoja, B.P., Wang, L., Zhou, L.: Automatic Radiology Report Generation by Learning with Increasingly Hard Negatives. arXiv:2305.07176 (2023)
56. Song, X., Zhang, X., Ji, J., Liu, Y., Wei, P.: Cross-modal contrastive attention model for medical report generation. In Proceedings of the 29th International Conference on Computational Linguistics, Gyeongju, Republic of Korea (2022)
57. Hou, W., Xu, K., Cheng, Y., Li, W., Liu, J.: ORGAN: observation-guided radiology report generation via tree reasoning. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, Toronto, Canada (2023)
58. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: NIPS (2017)

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.



Dhruv Sharma received his B.Tech. degree from GGSIPU University in 2015 and M.Tech. degree from Ambedkar Institute of Advanced Communication Technology and Research, New Delhi, India, in 2017. He is currently pursuing the Ph.D. degree with the Department of Electronics and Communication Engineering from Delhi Technological University, New Delhi, India. His current research interest includes machine learning, deep learning, computer vision, natural language processing and image captioning.



Chhavi Dhiman (M'16) received her B.Tech. degree from Indira Gandhi Delhi Technical University for Women (IGDTUW), Delhi, India, in 2011, M.Tech. and Ph.D. degree from Delhi Technological University (DTU), Delhi, India, in 2014 and 2019 respectively. She is currently working as Assistant Professor, in the Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India. Her research interest includes Machine Learning, Deep Learning, Pattern Recognition, Computer Vision, Human Action Identification and Classification. She is a reviewer of various Journals/Transactions of IEEE, Elsevier, and Springer. She has been awarded "Premium Research Award" by Delhi Technological University, Delhi, India in 2021.



Dinesh Kumar received his B.Tech. (ECE) and M.Tech. (ECE) (Gold Medalist) from National Institute of Technology (NIT) Kurukshetra (erstwhile REC Kurukshetra) and received Ph.D. degree in the field of Biometrics (Face Recognition) from GGS Indraprastha University, Delhi. He has more than 30 years of teaching and research experience and currently working as Professor in the Electronics and Communication Engineering Department of Delhi Technological University, Delhi. He has research publications in National/International Journals and Conferences including IEEE, Elsevier, Springer, Wiley. His current research interests include image processing, biometrics, soft computing, clustering, biomedical imaging, machine learning, image captioning. He is on the panel of reviewers for IEEE, Elsevier and Springer journals/Transactions.

Feature Aggregation with Latent Generative Replay for Federated Continual Learning of Socially Appropriate Robot Behaviours

Nikhil Churamani[1,*], Saksham Checker[2], Hao-Tien Lewis Chiang[3], and Hatice Gunes[1]

***Abstract*— For widespread real-world applications, it is beneficial for robots to explore Federated Learning (FL) settings where several robots, deployed in parallel, can learn independently while also sharing their learning with each other. This work explores a simulated living room environment where robots need to learn the social appropriateness of their actions. We propose Federated Root (FedRoot), a novel weight aggregation strategy which disentangles feature learning across clients from individual task-based learning. Adapting popular FL strategies to use FedRoot instead, we present a novel FL benchmark for learning the social appropriateness of different robot actions in diverse social configurations. FedRoot-based methods offer competitive performance compared to others while offering sizeable (up to 86% for CPU usage and up to 72% for GPU usage) reduction in resource consumption. Furthermore, real-world interactions require social robots to dynamically adapt to changing environmental and task settings. To facilitate this, we propose Federated Latent Generative Replay (FedLGR), a novel Federated Continual Learning (FCL) strategy that uses FedRoot-based weight aggregation and embeds each client with a *generator* model for pseudo-rehearsal of learnt feature embeddings to mitigate forgetting in a resource-efficient manner. Our benchmark results demonstrate that FedRoot-based FCL methods outperform other methods while also offering sizeable (up to 84% for CPU usage and up to 92% for GPU usage) reduction in resource consumption, with FedLGR providing the best results across evaluations.**

I. Introduction

As advances in Artificial Intelligence (AI) gear social robots towards a ubiquitous application, they are expected to be deployed across environmental and contextual settings, interacting with several users at a time and learning different tasks [1]. Each robot, operating in its unique application settings, should be able to adapt to the dynamics of its ever-changing environment while also being sensitive to the changing preferences of the users it interacts with [2]. Such an understanding of social dynamics and norms [3] can help robots effectively navigate complex social settings while also offering enriching interaction experiences to their users. Moreover, given the vast array of potential applications [4], robots can benefit from sharing knowledge and individual experiences with one another. Under such complex and diverse application settings, there is a need to move beyond centralised platforms towards more *distributed learning paradigms*, enabling robots to keep learning *continually* while also sharing their learning with others.

Moving beyond *centralised* learning paradigms, where individual robots only gather data and send it to a central server to be aggregated and used to train a single global model for application, Federated Learning (FL) [5] (see Fig. 1; left) offers an efficient *distributed learning* paradigm where individual robots can learn *independently* from their own *unique experiences*, updating their learning models using only the data collected by them *locally*. Over time, these local updates can be *aggregated* across the centralised server, in the form of model updates that can inform training the unified global model. FL allows for a *privacy-preserving learning paradigm* where local data is never shared with a centralised server. FL solutions have been used popularly in embedded or EdgeAI devices [6] that benefit from *distributed* learning settings [7] gathering and processing their own data but sharing their learning towards training a global aggregated model [8]. More recently, FL has been explored for robotic and autonomous systems [1], [9] allowing for collaborative learning across robots, learning from each others' experience while maintaining end-user privacy.

Another challenge faced by social robots operating in dynamic human-centred environments is to effectively discern novel information from past experiences and adapt their learning models to accommodate this new knowledge [10]. As the real-world is not static and changes dynamically [11], each individual robot may learn with incremental and/or sequential streams of data where data is not *independent and identically distributed* (i.i.d). This may result in the robot forgetting past knowledge, overfitting to the current task or data. Even though FL allows for model aggregation across client robots, such forgetting locally will cause the globally aggregated model to also forget past knowledge. Continual Learning (CL) [11] can help address this problem by enabling individual robots to adapt their learning with incrementally acquired data from non-stationary or changing environments [2]. This allows robots to accumulate new information locally while preserving previously seen knowledge. Combining the principles of FL and CL, Federated Continual Learning (FCL) [1], [12] (see Fig. 1; right) allows for individual robots, learning incrementally in their unique settings, to also benefit from other robots' learning. Each robot periodically sends only their model parameters to the centralised server where the knowledge from all agents is aggregated into a unified model which is sent back.

[1]N. Churamani and H. Gunes are with the Department of Computer Science and Technology, University of Cambridge, UK

[2]S. Checker is with Delhi Technological University, India. S. Checker contributed to this work while undertaking a remote visiting studentship at Department of Computer Science and Technology, University of Cambridge.

[3]H.T.L. Chiang is with Google DeepMind.

This work is supported by Google under the GIG funding scheme. The authors also thank Jie Tan and Carolina Parada for their valuable feedback.

*Corresponding Author: Nikhil.Churamani@cl.cam.ac.uk

Code: https://github.com/nchuramani/FedLGR_SocRob

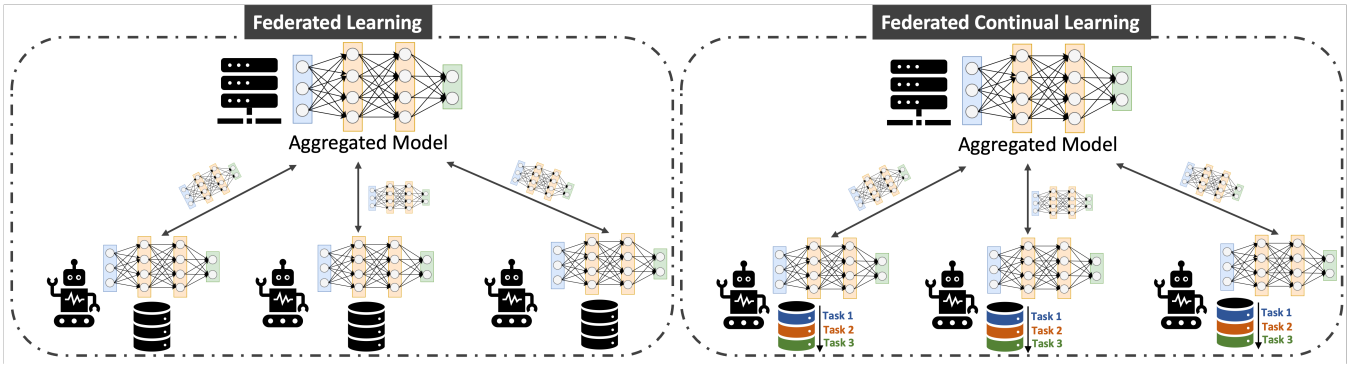


Fig. 1: **Federated Learning (FL)** (left): Local models are aggregated on the server without sharing data. **Federated Continual Learning (FCL)** (right): Individual robots incrementally learn tasks, periodically sharing model updates with each other.

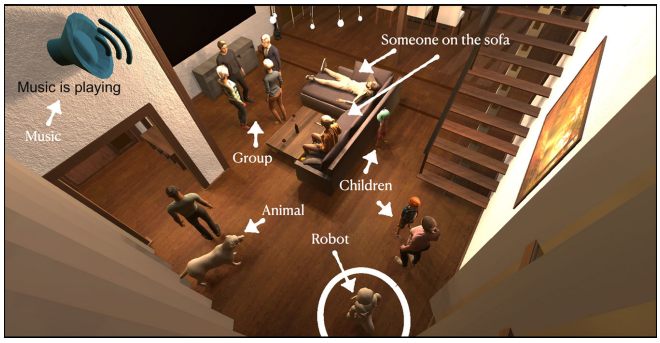


Fig. 2: MANNERS-DB: A Living Room scenario with a robot [13].

In this paper, we explore a simulated home environment with humans and animals where the robot is required to learn the social appropriateness of different tasks. We explore the MANNERS-DB dataset [13] as it has been used previously for CL-based [13] and zero-shot learning [14]. MANNERS-DB provides social-appropriateness ratings for different high-level robot actions in simulated living room settings (see Fig. 2), where the robot learns these actions in different social configurations. To simulate several robots deployed in parallel under similar living room settings, the dataset (training data) is split evenly across these robots (clients), with a shared test-set. We extend our preliminary work [15] benchmarking FL and FCL on the MANNERS-DB using the predefined $29-$d scene descriptor vectors to learning end-to-end, directly from scene renders. We propose Federated Root (FedRoot) averaging, a novel FL strategy which disentangles feature-based learning across clients from individual task-based learning. Under FedRoot, only the feature extraction layers are aggregated between *clients* however task-based learning is kept *strictly local* enhancing the resource-efficiency and privacy for FL. Adapting popular FL strategies to use FedRoot instead, we present a novel FL benchmark for learning the social appropriateness of different robot actions in diverse social configurations. Extending FedRoot to FCL settings, we propose Federated Latent Generative Replay (FedLGR), a novel FCL strategy that combines FedRoot with LGR [16] by adding a local generator model to the client for efficient pseudo-rehearsal of learnt features in order to mitigate forgetting. For successive tasks, pseudo-samples of previously seen tasks are generated and mixed with the current task's data to simulate *i.i.d* data settings. We benchmark FedLGR against popular CL adaptations of FL strategies, with and without FedRoot-based aggregation, on their ability to learn socially appropriate robot behaviours in a task-incremental manner. Both FedRoot and FedLGR reduce the memory and resource footprint of their FL and FCL counterparts, without compromising model performance.

II. Related Work

A. Socially Appropriate Robot Behaviours

Human interactions are governed by different social norms laying out expected behaviours from individuals which may be considered socially appropriate by others [17]. For social robots operating in human-centred environments, where interaction contexts and individual user preferences can vary widely, these robots must be able to comprehend and respond to the unique dynamics of each contextual setting [18]. This not only enhances their ability to navigate complex social situations but also improves the users' impressions of the robots and their acceptance in social Human-Robot Interaction (HRI) [19]. Furthermore, social robots need to *continually* learn from their interactions and adjust their behaviours accordingly [20]. Whether it is effectively navigating complex social environments [21], learning approach and positioning behaviours [22], [23] or learning task-specific behaviours [13], it is essential for robots to consider the social-appropriateness of their behaviours in order to comply with social norms [3]. With distributed deployment increasingly becoming a reality, there is a need to investigate *federated* application frameworks that allow robots to learn in their own unique environments while also informing a global learning of *generalisable* social norms and preferences [1]. This works aims to move in this direction in a more resource-efficient and privacy preserving manner.

B. Federated Learning

Distributed learning settings can be particularly desirable for social robots, to understand and learn socially appropriate behaviours, depending upon the context of the interaction, environmental factors as well as individual user preferences [1]. Federated Learning (FL) [5] enables a network of distributed client devices (such as robots) to

learn individually with locally gathered data while updating their model parameters towards aggregating a global learning model (server) that combines these updates from each individual device, over successive update rounds, and distributes this model back to individual clients. This way, each client shares their knowledge with the other clients, without sharing any local data, and gets updated with global model which combines the knowledge from other clients. Strategies such as Federated Averaging (FedAvg) [8] offer a straightforward approach for weight aggregation by collating individual client model weights and computing a weighted average in the form of the global model. However, FedAvg is sensitive to data imbalances and concepts drift challenges, especially in non-i.i.d data settings. Several improvements have been proposed on FedAvg, for instance, FedBN [24] that adapts FedAvg by keeping the parameters for all the *BatchNorm* layers 'strictly local', that is, all other model weights are aggregated across clients apart from the *BatchNorm* parameters. Similar to FedBN, FedProx [25] also proposes improvements over FedAvg by allowing for only partial aggregation of weights by adding and tracking a *proximal term* to FedAvg. FedOpt [26] offers a 'general optimisation framework' where each client uses a *client optimiser* to optimise on local data while the server updates apply a gradient-based *server optimiser* to the aggregated model weights. FedDistill [27], on the other hand, aims to improve the ability of the clients to deal with *heterogeneous* data conditions by using *knowledge distillation* [28]. Each client maintains two models: (i) a local copy of the global model and (ii) a personalised model that acts as a teacher to the student global model. The updated student model is then aggregated across clients. In this work, we adapt these methods for a regression-based task, learning to predict the social appropriateness of different robot actions.

C. Federated Continual Learning

As robots in real-world settings continually encounter novel information, under non-i.i.d conditions, their ability to remember previously learnt tasks may progressively deteriorate, resulting in forgetting [29]. Continual Learning (CL) [11], [30] strategies may enable robots to learn and adapt throughout their 'lifetime', balancing *incremental* learning of novel information with the retention of past knowledge. Federated Continual Learning (FCL) [12] combines FL and CL principles, enabling individual robots to *incrementally* learn *without forgetting* on streams of gathered data. Individual robots learn a series of local tasks while periodically updating the parameters of a global *aggregated* model that combines updates from individual clients, over successive update rounds. Several recent works [12] explore FCL for vision [31] and natural language processing [32] applications, however very little work has been done to explore its application for social robots [1]. Adapting FL approaches by adding CL-based objectives can offer straightforward solutions for FCL. For instance, regularisation-based methods such as Elastic Weight Consolidation (EWC) [33], EWCOnline [34], Synaptic Intelligence (SI) [35] or Memory Aware Synapses (MAS) [36] can be used to apply penalties on weight updates between old and new tasks to help mitigate forgetting. Rehearsal strategies such as Naive Rehearsal (NR) [37] can be used to maintain local memory buffers for each robot to store and rehearse previously seen data to preserve knowledge. Alternatively, an efficient pseudo-rehearsal of data [38] or features [16] can help mitigate forgetting by maintaining generators that model local data distributions for previously seen tasks. In this work, we augment FL with regularisation-based CL objectives for multi-label regression, predicting social appropriateness of the different robot actions under different contextual settings.

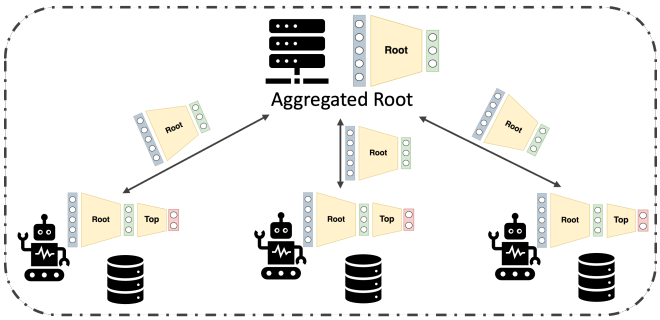


Fig. 3: **Federated Root (FedRoot)**: Local model split into (i) *Root* for feature extraction and (ii) *Top* for task-based learning. Only model *Root* is aggregated across clients while *Top* remains local.

III. Methodology

A. FedRoot: Federated Root for Feature Aggregation

With several robotic clients, deployed in parallel, FL ensures end-user privacy as data remains *strictly local* for each client and is never shared with others. Yet, as the learning models, both in terms of learnt features as well as task-based learning layers are shared with the server, there might still be 'information/data leak' [39] that may reveal how end-users interact with individual robots. Furthermore, as the robots may be equipped with Deep Neural Network (DNN)-based learning models that process high-dimensional input data, aggregating individual model weights into a global model can be computationally expensive [40]. To address the above-mentioned challenges, we propose Federated Root (FedRoot) (see Fig. 3) as a novel model aggregation FL strategy that splits client learning models into two modules:

1) Root (R): The *Root* constitutes the feature extraction layers of the learning model. For example, for Convolutional Neural Network (CNN)-based learning models, as implemented in this work, the *Conv* layers would constitute the *root* of the model that learn meaningful and relevant features from input images. For robots operating in home scenarios, as discussed here, the *root* learns efficient and descriptive scene embeddings that can help summarise each scene.

2) Top (T): The *Top* of the model, on the other hand, constitutes the task-based learning (classification or regression) Fully Connected (FC) layers. Using the *Root*-extracted features, the *top* of the model is used for task-based predictions, for example, learning to predict the social appropriateness of different robot actions.

Splitting model learning into two parts allows for robots to effectively share their learning with other clients while also protecting end-user privacy. The model is trained in an end-to-end manner, however, only the *root*, that is, the feature extraction layers, from individual client models gets aggregated across clients, over repeated aggregation rounds. This allows for the clients to improve their feature-embeddings, making them more robust, benefiting from the diversity of data settings experienced by each individual client. Yet, client-specific data as well as the task-based learning from the data, for instance end-user preferences on social appropriateness of individual robot behaviours, is kept 'strictly local'. Since FedRoot is proposed as a weight aggregation strategy, we adapt popular FL approaches (see Section II-B) to use FedRoot instead.

B. FedLGR: Federated Latent Generative Replay

Real-world application adds another complexity for robotic clients with respect to *incremental learning* where they are exposed to information in an incremental manner and have to constantly learn new tasks while preserving past knowledge [2], [29]. While replay-based CL may provide a straightforward approach for remembering past information by storing and periodically *replaying* past data [37], this may not always be possible due to memory constraints or privacy and compliance reasons. *Pseudo-rehearsal* approaches such as Deep Generative Replay (DGR) [41], on the other hand, use probabilistic or generative models that learn the inherent data statistics of previously seen data and draw *pseudo-samples* as and when needed, to be replayed to the model. These generative models, however, become hard to train as the number of tasks increases as the generator needs to learn to reconstruct high-quality discriminative pseudo-samples, adding a significant computational expense to the model. Generative feature rehearsal approaches such as Latent Generative Replay (LGR) [16] offer resource-efficient pseudo-rehearsal strategies that combine the benefits of DGR [41], with using low-dimensional latent features [42].

Under FCL learning settings clients not only need to incrementally learn on the acquired 'local data' ensuring they do not forget past knowledge, they also need to aggregate this learning in a centralised global model. To achieve this, we propose Federated Latent Generative Replay (FedLGR) (see Fig. 4) as a novel FCL strategy that applies LGR-based pseudo-rehearsal [16] under *federated learning* settings making use of FedRoot-based weight aggregation.

FedLGR uses a *scholar*-based architecture [16] that consists of three modules: (i) a *Root* to extract latent feature representations, (ii) a *Top* to learn task-discriminative information, and (iii) a *Generator* to reconstruct and rehearse *latent features*. Under FedLGR, each client locally uses LGR to *continually* and *incrementally* learn the social appropriateness of robot actions under different contextual settings. This learning is then shared with other clients by aggregating *Root* weights using FedRoot. At each time-step, the learning for each client *locally* consists of the following steps:

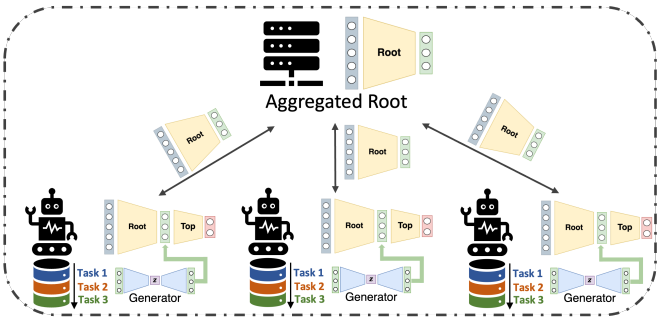


Fig. 4: **Federated Latent Generative Replay (FedLGR)**: Local model split into (i) *Root* for feature extraction, (ii) *Top* for task-based learning and (iii) *Generator* for *pseudo-rehearsal* of features. Only *Root* is aggregated across clients while *Top* and *Generator* remain local.

1) The *root* and *top* for the *scholar* are updated with the current task's data, however at different learning rates. While the *top* is trained rapidly to learn task-discriminative information, the *root* is updated *slowly* ensuring the new data does not completely *overwrite* previously seen information. This is an important step as LGR [16] is based on the assumption that if the latent-space distribution of *root*-extracted features remains relatively static between model updates, the extracted features can be effectively used to rehearse past knowledge [42].
2) For training the *generator*, the *Root*-extracted feature representations ($R(x)$) for the current task's data (x) are interleaved with *generated* features (R'_x) for all previously seen tasks. Training the *generator* on both $R(x)$ and R'_x (only $R(x)$ is used for Task 1) ensures that the updated *generator* encodes both new and old tasks. R'_x is passed to the *top* to obtain labels $T(R'_x)$.
3) Once the *generator* is updated, the *top* of the *solver* is updated using the current task data $\langle R(x), y \rangle$, interleaved with labelled *latent pseudo-samples* $\langle R'_x, T(R'_x) \rangle$ generated for previously seen tasks. The *root* is then updated *slowly* using only current task's data.

Training on each task consists of several rounds of weight aggregations, where FedRoot is used to aggregate *root* module weights across clients. The *top* and the *generator* for each client are kept 'strictly local' ensuring task-discriminative information is never shared.

IV. Experiments

A. MANNERS-DB Dataset

For our proof-of-concept evaluations on FL and FCL for learning socially appropriate robot behaviours, we explore the MANNERS-DB dataset [13] that consists of Unity-generated scenes of the Pepper robot co-inhabiting a living room (see Figure 2) with other humans and animals under different social settings. MANNERS-DB has been explored for CL-based [13] and zero-shot [14] learning, however, not for FCL. For each scene, the robot can perform 8 different actions, that is, *vacuuming, mopping, carrying warm food, carrying cold food, carrying big objects, carrying small objects, carrying drinks and cleaning/starting conversations*

either *within a circle of influence* or in the *direction of operation* (see Figure 2). Crowd-sourced annotations are provided for the social appropriateness of each of these actions for every scene ($\approx$ 1000 scenes), labelled on a 5-point Likert scale, ranging from *very inappropriate* to *very appropriate*. For both FL and FCL evaluations, the data is split into training and test splits in the ratio of $3:1$. The training data is further split amongst the different clients (2 or 10) with a shared test-set used for evaluation. For FCL evaluations, the training set of individual clients is further split into two tasks, that is, samples depicting the robot operating with an circle (**Task 1**) and in the direction of the arrow (**Task 2**). For training the models, *normalised* RGB images are used, resized to $(128\times128\times3)$ due computational restrictions of training multiple clients in parallel on GPU. Since MANNERS-DB is a relatively small dataset consisting of $\approx$ 1000 images, we evaluate the different approaches both *without* and *with* data-augmentation. We use random ($p = 0.5$) horizontal flipping and random ($p = 0.5$) rotation (up to $10°$) in either direction to augment the dataset. For a fair comparison, results are presented individually for *without* and *with* augmentation comparisons.

B. Implementation Details

For each approach, a Convolutional Neural Network (CNN)-based model is used consisting of two parts: (i) a *Conv* module that uses the popularly used MobileNet-V2 [43] backbone to extract scene features, followed by an *AdaptiveAvgPool* layer for dimensionality reduction, and (ii) an *FC* module where the $1280-$d *flattened* scene features are passed through a FC layer consisting of 32 units followed by the $8-$unit output layer. Each layer in the *FC* module is followed by *BatchNorm* layer and uses a *linear* activation with the output layer predicting the *social appropriateness* of all 8 robot actions. All models are trained using the *Adam* optimiser (β_1=0.9, β_2=0.999) with *learning rate* of lr=$1e^{-}3$. All methods are implemented using the *Flower*[1] and PyTorch Python libraries. The FL and FCL experiments are run for $2-10$ clients, attuned to a potential real-world evaluation to be conducted in the future using physical robots. For brevity, results for only 2 and 10 clients are presented. Model hyper-parameters such as the regularisation coefficients for FCL approaches are optimised using a grid-search. For without augmentation experiments, the *BatchSize* is empirically set to 16 for *two clients* and 8 for *ten clients* due to the overall low number of samples per client. When using with augmentation, *BatchSize* is set to 16. All models are trained using an Mean Squared Error (MSE) loss, computed as an average across the 8 actions. Each client agent trains on its own split of the training data but a shared test-set is used for model evaluation. We report the average MSE loss averaged across the 8 actions and all clients.

C. Evaluation Metrics

Predicting the social-appropriateness for each robot action requires transforming FL and FCL methods to use regression-based objectives, using the following metrics:

- *Root Mean Squared Error (RMSE):* We report the RMSE values which are calculated for the test-set, averaged across the 8 actions across all clients. RMSE scores are used to compare our results with the baseline provided with the MANNERS-DB dataset [13].
- *Pearson's Correlation Coefficient (PCC):* In addition to an absolute metric such as the RMSE, we also evaluate the predicted social appropriateness values relative to the ground-truth using the PCC [44] scores. PCC scores are calculated on the test-set individually for each of the 8 actions, for each client. Average scores across all actions and all clients are reported.
- *CPU Usage (s):* To evaluate the resource efficiency of FL and FCL approaches, CPU usage is reported as the CPU time (in seconds) allocated to training each client, on average. This is calculated as the CPU time utilised by each client, across all rounds of weight aggregation.
- *GPU Usage (%):* Similar to CPU usage, GPU usage is also important to evaluate the resource efficiency of the proposed methods. Since we have a single GPU simulation set-up, we report the percentage (%) of the GPU allocated per client, on average. GPU usage is also calculated using Nvidia's `nvidia-smi` tool, logged at different intervals during the training process, and *averaged* over the entire training time for each client.

TABLE I: Federated Learning Results for the MANNERS-DB Dataset for Two (left) and Ten (Right) Clients. **Bold** values denote best while [bracketed] denote second-best values.

	Two Clients			*Ten Clients*		
Method	Loss ▼	RMSE ▼	PCC ▲	Loss ▼	RMSE ▼	PCC ▲
	W/O Augmentation					
FedAvg	0.264	0.510	[0.508]	0.220	0.467	0.487
FedBN	0.310	0.553	0.486	0.226	0.471	0.486
FedProx	[0.262]	[0.507]	**0.513**	[0.211]	[0.459]	**0.517**
FedOpt	**0.245**	**0.492**	0.501	**0.210**	**0.458**	[0.492]
FedDistill	0.276	0.522	0.475	0.233	0.481	0.470
FedRootAvg	0.276	0.522	0.482	0.231	0.477	0.459
FedRootBN	0.272	0.518	0.498	0.263	0.510	0.420
FedRootProx	0.264	0.511	0.481	0.240	0.486	0.447
FedRootOpt	0.285	0.530	0.473	0.240	0.486	0.460
FedRootDistill	0.294	0.539	0.489	0.305	0.550	0.363
	W/ Augmentation					
FedAvg	0.192	0.435	0.547	0.178	[0.420]	**0.586**
FedBN	0.205	0.447	[0.554]	0.188	0.429	0.567
FedProx	0.208	0.451	0.536	**0.173**	**0.412**	[0.581]
FedOpt	0.204	0.450	0.506	[0.177]	[0.420]	0.568
FedDistill	0.221	0.468	0.538	0.205	0.449	0.560
FedRootAvg	0.192	0.433	0.547	0.198	0.440	0.563
FedRootBN	0.197	0.442	**0.558**	0.194	0.439	0.539
FedRootProx	**0.186**	**0.427**	0.541	0.196	0.439	0.552
FedRootOpt	[0.190]	[0.432]	0.528	0.206	0.450	0.549
FedRootDistill	0.225	0.469	0.525	0.222	0.465	0.539

V. Results

A. Federated Learning Benchmark

Table I presents the FL benchmark results on the MANNERS-DB dataset comparing popular FL strategies and their FedRoot-based adaptations for *two* and *ten* clients,

[1]https://flower.dev

TABLE II: Average CPU Usage (s) and GPU Usage (%) per client, per aggregation round for Federated Learning experiments.

Metric	FedAvg	FedRootAvg	FedBN	FedRootBN	FedProx	FedRootProx	FedOpt	FedRootOpt	FedDistill	FedRootDistill
CPU Usage (s) ▼	19.06	**2.55** (▼ 86.6%)	18.4	**2.50** (▼ 86.4%)	18.98	**2.54** (▼ 86.6%)	18.61	**2.46** (▼ 86.7%)	19.46	**2.61** (▼ 86.4%)
GPU Usage (%) ▼	0.63	**0.24** (▼ 61.9%)	0.93	**0.27** (▼ 70.9%)	0.25	**0.15** (▼ 40.0%)	0.69	**0.19** (▼ 72.4%)	0.10	**0.07** (▼ 30.0%)

TABLE III: Average CPU Usage (s) and GPU Usage (%) per client, per aggregation round for Federated Continual Learning experiments.

Metric	FedAvg EWC	FedRoot EWC	FedAvg EWCOnline	FedRoot EWCOnline	FedAvg MAS	FedRoot MAS	FedAvg SI	FedRoot SI	FedAvg NR	FedRoot NR	FedLGR
CPU Usage (s) ▼	24.17	**3.87** (▼ 83.9%)	25.05	**4.97** (▼ 80.2%)	22.41	**5.36** (▼ 76.1%)	20.32	**3.91** (▼ 80.7%)	25.97	**4.05** (▼ 84.4%)	**4.06**
GPU Usage (%) ▼	0.57	**0.04** (▼ 92.9%)	0.13	**0.03** (▼ 76.9%)	0.04	**0.01** (▼ 75.0%)	0.10	**0.05** (▼ 50.0%)	0.06	**0.04** (▼ 33.3%)	**0.02**

both without and with data augmentation, respectively. For without data augmentation experiments with *two* and *ten clients* FedOpt and FedProx emerge as the best performing approaches, on average. The 'general optimisation framework' of FedOpt, with *Adam*-based client and server optimisers are able to efficiently aggregate learning across clients, resulting in the robust performance of the model. FedProx, on the other hand, adds a proximal term $\mu = 0.1$ (similar to [25]) to FedAvg and updates the objective for each client to minimise $F_k(\omega) + \frac{\mu}{2}||\omega - \omega^t||^2$ where F_k is the loss, ω are the local model weights to optimise and ω^t are the global parameters at time-step t. For the proposed FedRoot variants of the compared FL approaches, as task-discriminative *top* weights are kept *strictly local*, we see an overall drop in model performance. This is primarily due to the small size of the MANNERS-DB dataset which does have enough data samples for the FedRoot to optimise model performance, given only feature extracting *root* is aggregated. However, FedRoot offers *sizeable reduction* in CPU and GPU Usage per client, per aggregation round, compared to the original methods, as can be seen in Table II.

Data augmentation has a net positive impact on all models with all approaches reporting better metrics, across evaluations. With the larger amount of data available per client, we see that FedRoot-based methods improve upon their counterparts across all metrics, for the experiments with *two clients*. Similar to without augmentation experiments, using *Adam*-based client and server optimisation as well as adding a proximal term to the model learning objective results in the best performance, however here for FedRootOpt and FedRootProx, respectively. Data available per client still remains relatively low when split across *ten clients*, even when using data augmentation. Thus, despite an overall improvement compared to without augmentation results FedRoot-based approaches are still not able to match the performance of FedProx and FedOpt for *ten clients*, despite offering high resource-efficiency.

B. Federated Continual Learning Benchmark

Table IV presents the FCL results on the MANNERS-DB dataset adapting FedAvg and FedRoot-based weight aggregation strategies to use CL-based learning objectives to mitigate forgetting under incremental settings. Evaluation metrics are reported after training and testing on the data-split depicting the robot operating *within the circle of influence* **(Task 1)** and training on the data-split depicting the robot *in the direction of operation* **(Task 2)** and testing on both the splits. Similar to the FL benchmark, we compare the different FCL strategies for *two* and *ten* clients as well as without and with data augmentation. As *federated averaging* of weights struggles with learning under non-i.i.d data settings [24], extending these methods with CL-based objectives allows us to evaluate how they can objectively contribute towards maintaining model performance when learning *incrementally*.

Our results demonstrate that FedRoot-based approaches are able to outperform their FedAvg-based counterparts across most evaluations. However, similar to FL evaluations, FedRoot-based methods struggle in the scarcity of data, for instance the evaluations across *ten* clients without data augmentation, resulting in much worse loss and RMSE scores. Data augmentation has a net positive impact on model performances, especially for FedRoot-based methods. Furthermore, using FedRoot-based weight aggregation results in a sizeable (up to *84%* for CPU usage and *92%* for GPU usage) reduction in computational expense of running these methods, per client, per aggregation round (see Table III). This can be particularly beneficial for application in resource-constrained devices such as social robots. The proposed FedLGR approach performs the best across all evaluations. This is especially true when data is scarce, that is learning without using data augmentation across *two* and *ten* clients. FedLGR is able to efficiently use pseudo-rehearsal of features to maintain model performance after learning Task 2.

VI. Discussion and Conclusion

Social robots operating in dynamic real-world settings can benefit from federated learning mechanisms where, learning from and adapting towards their unique environmental and data conditions, they can also share their learning with other robots, benefiting from each others' experiences. FL-based approaches enable such a learning paradigm for robots to aggregate model updates across individual agents into a global model, while ensuring that end-user privacy is preserved. However, the resource constraints inhibit such learning settings resulting in most existing methods opting for *centralised* learning, where end-user devices only collect data for a centralised model to be trained, in isolation, for later application *in-the-wild*. The proposed Federated Root (FedRoot) weight aggregation strategy aims to address this very challenge by splitting each client's learning model into

TABLE IV: Federated Continual Learning Results for the MANNERS-DB Dataset for Two (left) and Ten (Right) Clients. **Bold** values denote best while [bracketed] denote second-best values.

	Two Clients						*Ten Clients*					
	After Task 1			After Task 2			After Task 1			After Task 2		
Method	Loss ▼	RMSE ▼	PCC ▲	Loss ▼	RMSE ▼	PCC ▲	Loss ▼	RMSE ▼	PCC ▲	Loss ▼	RMSE ▼	PCC ▲
	W/O Data-augmentation											
FedAvgEWC	0.705	0.826	0.488	0.911	0.944	0.347	0.364	0.591	0.325	0.393	0.617	0.278
FedAvgEWCOnline	0.488	0.685	0.469	0.381	0.608	0.384	[0.328]	[0.566]	0.373	[0.359]	[0.589]	0.321
FedAvgMAS	0.568	0.740	0.570	0.363	0.597	0.357	0.421	0.641	[0.494]	0.464	0.674	0.357
FedAvgSI	0.488	0.686	[0.571]	0.675	0.811	0.343	0.451	0.665	0.323	0.376	0.598	0.306
FedAvgNR	0.403	0.634	0.491	[0.296]	[0.538]	0.333	0.376	0.600	0.447	0.408	0.628	0.341
FedRootEWC	0.433	0.652	0.472	0.998	0.963	0.382	9.320	3.050	0.336	8.950	2.990	0.335
FedRootEWCOnline	0.614	0.778	0.436	0.447	0.659	0.400	9.170	3.030	0.330	9.560	3.090	[0.376]
FedRootMAS	[0.397]	[0.627]	0.429	0.569	0.740	0.422	8.860	2.980	0.363	8.750	2.960	0.369
FedRootSI	0.968	0.984	0.441	0.840	0.893	0.418	9.360	3.060	0.417	9.120	3.020	0.369
FedRootNR	0.762	0.870	0.325	**0.287**	**0.533**	[0.444]	0.446	0.656	0.462	0.482	0.677	0.339
FedLGR	**0.252**	**0.499**	**0.599**	0.465	0.677	**0.534**	**0.299**	**0.539**	**0.515**	**0.320**	**0.553**	**0.421**
	W/ Data-augmentation											
FedAvgEWC	0.669	0.807	0.420	0.348	0.584	0.395	0.397	0.624	0.472	0.433	0.653	0.456
FedAvgEWCOnline	0.495	0.702	0.438	0.629	0.788	0.376	[0.319]	[0.561]	0.365	[0.329]	[0.579]	0.363
FedAvgMAS	0.267	0.514	**0.601**	0.665	0.814	0.394	0.343	0.580	0.480	0.354	0.587	0.328
FedAvgSI	0.280	0.526	[0.546]	[0.328]	[0.568]	0.444	0.471	0.682	0.418	0.435	0.654	0.426
FedAvgNR	0.450	0.669	0.531	0.710	0.839	0.451	0.392	0.618	0.425	0.490	0.698	0.368
FedRootEWC	0.736	0.836	0.529	**0.327**	**0.569**	0.463	0.441	0.648	[0.543]	0.449	0.655	[0.467]
FedRootEWCOnline	0.517	0.712	0.398	0.446	0.661	0.363	0.688	0.805	0.446	0.612	0.761	0.405
FedRootMAS	[0.265]	[0.511]	0.492	0.616	0.782	0.436	0.797	0.851	0.542	1.030	0.930	0.439
FedRootSI	0.388	0.596	0.405	0.735	0.821	[0.478]	0.481	0.688	0.323	0.521	0.708	0.439
FedRootNR	0.781	0.875	0.491	0.368	0.604	0.354	0.539	0.728	0.268	0.409	0.635	0.447
FedLGR	**0.230**	**0.478**	0.531	0.493	0.699	**0.479**	**0.288**	**0.528**	**0.563**	**0.317**	**0.560**	**0.482**

aggregatable feature extraction layers, that is, model *root* and *private* task-relevant *top* layers that learn to predict the social appropriateness of different robot actions. The FL benchmark results (see Table I and Table II) highlight the competitive performance of FedRoot-based approaches with sizeable reductions in CPU and GPU usage for each client.

Furthermore, real-world applications may require robots to learn incrementally, for instance, in the form of developing novel capabilities or applying existing capabilities under novel contextual settings. In this work, we explore the latter where robots need to learn the social appropriateness of different actions depending upon the context in which they are operating, that is, operating *within the circle of influence* or *in a particular direction of operation*. We adapt the proposed FedRootAvg weight aggregation strategy and extend it by adapting popular CL-based learning objectives presenting a novel FCL benchmark on the MANNERS-DB dataset (see Table IV and Table III). In particular, the proposed Federated Latent Generative Replay (FedLGR) approach is seen to outperform other methods across all evaluations. It implements a local *generator* for efficient pseudo-rehearsal of latent features for mitigating forgetting, where the *top* learns task-relevant information. Our work contributes significantly to the fields of Federated Learning and robotics, offering promising avenues for the development of socially intelligent machines capable of learning and adapting in a decentralized and resource-efficient manner.

A. Limitations and Future Work

We present novel FL and FCL benchmarks for learning social appropriateness of high-level robot behaviours in simulated home settings. With our motivating results, we wish to conduct a user study with $2+$ robots operating in different living room settings, sharing their knowledge with each other. However, MANNERS-DB is a relatively small dataset which limits our evaluations into the generalisability of proposed approaches across a large number of clients. In future, we wish to expand the MANNERS-DB dataset to include more scenes as well as different robot embodiments such as Pepper, Nao, PR2 robots, and points-of-view (PoVs) (robot-centric and scene-centric) to extend our evaluations (both FL and FCL), under simulation, to $100+$ clients. This would result in a more generalisable evaluation for large-scale *federated* application for robots in diverse real-world settings. Additionally. it will be beneficial to explore further the scalability of FedRoot and FedLGR in more complex or dynamic environments, especially when data is scarce and data acquisition is complicated, as well as their applicability to a broader range of tasks beyond social appropriateness.

REFERENCES

[1] L. Guerdan and H. Gunes, "Federated Continual Learning for Socially Aware Robotics," in *IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 1522–1529, 2023.

[2] N. Churamani, S. Kalkan, and H. Gunes, "Continual Learning for Affective Robotics: Why, What and How?," in *IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 425–431, IEEE, 2020.

[3] A. Ayub and A. R. Wagner, "What Am I Allowed to Do Here?: Online Learning of Context-Specific Norms by Pepper," in *Social Robotics*, pp. 220–231, Springer International Publishing, 2020.

[4] H. Mahdi, S. A. Akgun, S. Saleh, and K. Dautenhahn, "A survey on the design and evolution of social robots — past, present and future," *Robotics and Autonomous Systems*, vol. 156, p. 104193, 2022.

[5] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *International Conference on Artificial Intelligence and Statistics*, vol. 54 of *Proceedings of Machine Learning Research*, pp. 1273–1282, PMLR, 2017.

[6] A. Imteaj, U. Thakker, S. Wang, J. Li, and M. H. Amini, "A Survey on Federated Learning for Resource-Constrained IoT Devices," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 1–24, 2022.

[7] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on Federated Learning," *Knowledge-Based Systems*, vol. 216, 2021.

[8] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the Convergence of FedAvg on Non-IID Data," in *International Conference on Learning Representations*, 2020.

[9] Y. Xianjia, J. P. Queralta, J. Heikkonen, and T. Westerlund, "Federated Learning in Robotic and Autonomous Systems," *Procedia Computer Science*, vol. 191, pp. 135–142, 2021.

[10] Z. Chen and B. Liu, *Lifelong Machine Learning*. Springer International Publishing, 2018.

[11] R. Hadsell, D. Rao, A. A. Rusu, and R. Pascanu, "Embracing Change: Continual Learning in Deep Neural Networks," *Trends in Cognitive Sciences*, vol. 24, no. 12, pp. 1028–1040, 2020.

[12] J. Yoon, W. Jeong, G. Lee, E. Yang, and S. J. Hwang, "Federated Continual Learning with Weighted Inter-client Transfer," in *International Conference on Machine Learning*, vol. 139, pp. 12073–12086, PMLR, 2021.

[13] J. Tjomsland, S. Kalkan, and H. Gunes, "Mind Your Manners! A Dataset and a Continual Learning Approach for Assessing Social Appropriateness of Robot Actions," *Frontiers in Robotics and AI*, vol. 9, 2022.

[14] B. Zhang and H. Soh, "Large language models as zero-shot human models for human-robot interaction," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7961–7968, 2023.

[15] S. Checker, N. Churamani, and H. Gunes, "Federated Learning of Socially Appropriate Agent Behaviours in Simulated Home Environments," in *Workshop on Lifelong Learning and Personalization in Long-Term Human-Robot Interaction (LEAP-HRI), 19th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2024.

[16] S. Stoychev, N. Churamani, and H. Gunes, "Latent Generative Replay for Resource-Efficient Continual Learning of Facial Expressions," in *17th IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pp. 1–8, 2023.

[17] O. Roesler, E. Bagheri, A. Aly, S. Rossi, and R. Alami, "Socially acceptable robot behavior: Approaches for learning, adaptation and evaluation," *Interaction Studies*, vol. 23, no. Socially Acceptable Robot Behavior, pp. 355–359, 2022.

[18] H. Li, S. Milani, V. Krishnamoorthy, M. Lewis, and K. Sycara, "Perceptions of Domestic Robots' Normative Behavior Across Cultures," in *AAAI/ACM Conference on AI, Ethics, and Society*, AIES '19, p. 345–351, ACM, 2019.

[19] O. Roesler, E. Bagheri, and A. Aly, "Toward understanding the effects of socially aware robot behavior," *Interaction Studies*, vol. 23, no. Socially Acceptable Robot Behavior, pp. 513–552, 2022.

[20] K. Shiarlis, J. Messias, and S. Whiteson, "Acquiring social interaction behaviours for telepresence robots via deep learning from demonstration," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 37–42, 2017.

[21] A. Francis, C. Pérez-d'Arpino, C. Li, F. Xia, A. Alahi, R. Alami, A. Bera, A. Biswas, J. Biswas, R. Chandra, *et al.*, "Principles and guidelines for evaluating social robot navigation algorithms," *arXiv preprint arXiv:2306.16740*, 2023.

[22] Y. Gao, F. Yang, M. Frisk, D. Hemandez, C. Peters, and G. Castellano, "Learning socially appropriate robot approaching behavior toward groups using deep reinforcement learning," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 1–8, IEEE, 2019.

[23] E. McQuillin, N. Churamani, and H. Gunes, "Learning Socially Appropriate Robo-Waiter Behaviours through Real-Time User Feedback," in *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, HRI '22, p. 541–550, IEEE Press, 2022.

[24] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "FedBN: Federated Learning on Non-IID Features via Local Batch Normalization," in *International Conference on Learning Representations*, 2021.

[25] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated Optimization in Heterogeneous Networks," in *Proceedings of Machine Learning and Systems*, vol. 2, pp. 429–450, 2020.

[26] S. J. Reddi, Z. Charles, M. Zaheer, Z. Garrett, K. Rush, J. Konečný, S. Kumar, and H. B. McMahan, "Adaptive Federated Optimization," in *International Conference on Learning Representations*, 2021.

[27] D. Jiang, C. Shan, and Z. Zhang, "Federated Learning Algorithm Based on Knowledge Distillation," pp. 163–167, IEEE, 2020.

[28] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," in *NIPS Deep Learning and Representation Learning Workshop*, 2015.

[29] T. Lesort, V. Lomonaco, A. Stoian, D. Maltoni, D. Filliat, and N. Díaz-Rodríguez, "Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges," *Information Fusion*, vol. 58, pp. 52–68, 2020.

[30] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual Lifelong Learning with Neural Networks: A review," *Neural Networks*, vol. 113, pp. 54–71, 2019.

[31] D. Shenaj, M. Toldo, A. Rigon, and P. Zanuttigh, "Asynchronous Federated Continual Learning," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 5055–5063, 2023.

[32] Y. Chaudhary, P. Rai, M. Schubert, H. Schütze, and P. Gupta, "Federated Continual Learning for Text Classification via Selective Inter-client Transfer," in *Findings of the Association for Computational Linguistics: EMNLP 2022*, pp. 4789–4799, Association for Computational Linguistics, 2022.

[33] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.

[34] J. Schwarz, W. Czarnecki, J. Luketina, A. Grabska-Barwinska, Y. W. Teh, R. Pascanu, and R. Hadsell, "Progress & Compress: A scalable framework for continual learning," in *International Conference on Machine Learning*, vol. 80, pp. 4528–4537, PMLR, 2018.

[35] F. Zenke, B. Poole, and S. Ganguli, "Continual Learning Through Synaptic Intelligence," in *International Conference on Machine Learning - Volume 70*, vol. 70 of *ICML'17*, pp. 3987–3995, JMLR.org, 2017.

[36] R. Aljundi, F. Babiloni, M. Elhoseiny, M. Rohrbach, and T. Tuytelaars, "Memory Aware Synapses: Learning What (not) to Forget," in *European Conference on Computer Vision (ECCV)*, pp. 139–154, 2018.

[37] Y.-C. Hsu, Y.-C. Liu, A. Ramasamy, and Z. Kira, "Re-evaluating Continual Learning Scenarios: A Categorization and Case for Strong Baselines," in *NeurIPS Continual learning Workshop*, 2018.

[38] H. Shin, J. K. Lee, J. Kim, and J. Kim, "Continual Learning with Deep Generative Replay," in *Advances in Neural Information Processing Systems 30*, pp. 2990–2999, Curran Associates, Inc., 2017.

[39] N. Carlini, J. Hayes, M. Nasr, M. Jagielski, V. Sehwag, F. Tramèr, B. Balle, D. Ippolito, and E. Wallace, "Extracting Training Data from Diffusion Models," in *32nd USENIX Security Symposium (USENIX Security 23)*, pp. 5253–5270, USENIX Association, 2023.

[40] E. Guerra, F. Wilhelmi, M. Miozzo, and P. Dini, "The Cost of Training Machine Learning Models Over Distributed Data Sources," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 1111–1126, 2023.

[41] H. Shin, J. K. Lee, J. Kim, and J. Kim, "Continual Learning with Deep Generative Replay," in *Advances in Neural Information Processing Systems 30*, pp. 2990–2999, Curran Associates, Inc., 2017.

[42] L. Pellegrini, G. Graffieti, V. Lomonaco, and D. Maltoni, "Latent Replay for Real-Time Continual Learning," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 10203–10209, 2020.

[43] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.

[44] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*, pp. 37–40, Springer, 2009.

Sādhanā (2024)49:184
https://doi.org/10.1007/s12046-024-02526-8
© Indian Academy of Sciences

Filter based saturated controller design for a class of nonlinear singularly perturbed systems

DIPAK PRASAD[1,*], SUDARSHAN K VALLURU[1] and MADAN MOHAN RAYGURU[2]

[1]Control of Dynamical Systems and Computation Lab, Electrical Engineering Department, Delhi Technological University, New Delhi, Delhi 110042, India
[2]University of Louisville, J.B. Speed School of Engineering, Louisville, KY, USA
e-mail: dipakee15@gmail.com; sudarshan_valluru@dce.ac.in; MMRAYG01@louisville.edu

MS received 28 October 2023; revised 25 January 2024; accepted 20 February 2024

Abstract. This paper describes a novel method to design a saturated controller for a singularly perturbed systems (SPS). These systems frequently exhibit complexity and uncertainty that may have negative impacts on the performance. Further, the SPS are characterised by a substantial time scale difference between fast and slow dynamics. In order to overcome these difficulties, the singular perturbation approach is used to design a nonlinear saturated controller which has a filter and disturbance observer. A strategical approach is presented based on contraction theory to achieve convergence of all closed-loop states of the system. For the analysis and design of these systems, contraction theory offers a rigorous mathematical framework that ensures system trajectories will converge even in the presence of singular perturbation and uncertainties. The suggested approach is validate on pitch angle control of the Twin Rotor MIMO System (TRMS) to replicate the real time situations.

Keywords. Saturated control; singular perturbation; disturbance observer; contraction theory; pitch angle control.

1. Introduction

In the context of control systems,"saturated control"refers to a condition in which a system's control input or output is confined to remain within particular limitations or constraints. This might be due to physical constraints of actuators, safety concerns, or to prevent the system from responding in an unpleasant manner when control signals approach extreme values. When a system is in a saturated mode, the control signal is"clamped"or"limited"to a specific range. This has serious consequences for the system's behaviour and performance. A critical part of control system design is taking actuator constraints into account. Ignoring these boundaries may result poor performance and instability. In [1] a robust saturated controller has been designed for the attitude tracking under bounded disturbances. In this brief a saturated nonlinear controller has been designed for the pitch angle tracking control of TRMS.

Singular perturbation modelling is a method for describing the behaviour of large physical systems composed of multiple slow and rapid subsystems working on different time scales. In singularly perturbed (SP) model the derivative of some states are multiplied by a small parameter which describe the fast and slow phenomena of the system [2]. Slow and fast subsystems of two dynamical systems are connected to form a two-time scale singularly perturbed system [3]. Stochastic filtering, Model order reduction, composite control, optimal control etc utilise singularly perturbed systems [4, 5]. When the perturbation parameter is zero, a singularly perturbed system has a single root for the fast subsystem. Such systems are sometimes referred to as standard models. In contrast, fast systems have several roots or none at all in nonstandard models [6, 7]. Traditionally, controller design and stability analysis for singularly perturbed systems have been successfully accomplished using the quadratic Lyapunov function [8, 9]. To do this, the perturbation parameter is set to zero and the entire system is divided into two reduced-order models. Each reduced system's stability is examined by choosing two suitable quadratic Lyapunov functions. The convex sum of these two functions are called as Composite Lyapunov Function, is then employed to ensure system stability. Depending on the connectivity criteria met by the Lyapunov functions, the resultant stability boundaries are valid for a specific range of the perturbation parameter [10]. Many issues with the well-known SP system have been resolved: composit control design (State and Output feedback), LQG control etc [11–13].

*For correspondence
Published online: 21 May 2024

Many physical systems are influenced by unpredictable and unavoidable environmental impacts, which are referred to as external disturbances and may also change the control variables. Therefore, the need of observers are prerequisite to reduce the impact of external disturbances. Due to presence of environmental disturbances, approximation etc the mathematical model of dynamical systems involve uncertainties. That is why it is necessary to model these uncertainties as ($\delta = 10^3 * tanh(2t)$) while designing a controller. There are many way to eliminate these uncertainties. One of the best method is to estimate the disturbances variables and remove it from the process of control design. For electro-hydraulic systems, a high-gain disturbance observer based backstepping control with position tracking error constraints have been suggested to enhance position tracking performance in the presence of disturbances [14]. Recently, in order to regulate the load following of a pressurized-water reactor (PWR) in presence of disturbances, a high-gain disturbance observer-based double closed-loop sliding mode control (HGDO-DCLSMC) method has been suggested [15].

Recently contraction theory has been proposed for the analysis of stability of the system [16].A contracting dynamical system's paths exponentially converge towards each other irrespective of their initial condition [17, 18].Every trajectory starting inside a region in the state space will converge towards each other and this region is called contracting region [19]. In reference [20] stability of nonlinear singularly perturbed system based on contraction theory has been discussed. While a contracting autonomous system implicitly guaranteed the existence of an equilibrium point, a contraction framework does not always necessitate the presence of an attractor [21]. It is simpler to manage uncertainties in the system model because the exponential convergence behavior is inherently robust to bound disturbances. These fascinating features are used in the investigation of distributed systems [22],observer design [23],networks stability [24], kalman filter [25, 26], synchronisation [27],controller design for mechanical systems [28], backstepping controller synthesis [29] etc. The results of this study are used to stabilise approximate feedback linearisable systems [30]. In this article contraction theory based methodology has been introduced for the convergence of all closed loop states. The contributions of this paper can be summarized as follows:

(1) This work proposes a novel method to design a saturated controller for a singularly perturbed system.
(2) In order to address the system complexity and external disturbances, a singular perturbation approach has been used to design a filter and a disturbance observer.
(3) A strategical approach based on contraction theory has been proposed to achieve convergence of all closed-loop states of the system.
(4) The performance of contraction theory based saturated nonlinear controller has been implemented and validated for pitch control of TRMS (Twin rotor MIMO System).

This article is arranged in the following way:

The system description is provided in section 2. The design of a saturated nonlinear controller is described in section 3. Section 4 and section 5 provide details on convergence analysis and filtered saturated controller with disturbance observer respectively, and section 6 presents the experimental outcomes for pitch angle tracking of TRMS, followed by conclusion and references.

2. System description

We consider the class of nonlinear systems which is in strict feed back form described as:

$$\begin{aligned} \dot{x_1} &= g_1 x_2 + f_1(x_1) \\ \dot{x_2} &= g_2 x_3 + f_2(x_1, x_2) \\ &\vdots \\ \dot{x_n} &= g_n u + f_n(x_1, x_2, \ldots, x_n) \end{aligned} \tag{1}$$

Equation (1) can be represented in the following form:

$$\dot{x_n} = G(x)u + F(x) \tag{2}$$

$f_i(.)$ are sufficiently smooth function and $g_i \neq 0$. The equation (2) can be converted into equation (4) by choosing z_i, α_i and α_{if}: Where,

$$\begin{aligned} & z_i = x_i - \alpha_{if}, i = 1, 2, 3...n \\ & \mu\dot{\alpha}_{if} = -\alpha_{if} + \alpha_i \\ & \alpha_{1f} = \omega_d \\ & \alpha_i = \frac{1}{g_{(i-1)}}[-f_{(i-1)}(x_1, ..x_{(i-1)}) \\ & \qquad - g_{(i-2)}z_{(i-2)} + \dot{\alpha}_{(i-1)f}], i = 2, 3, 4...n \end{aligned} \tag{3}$$

where $\mu \in (0, 1)$, ω_d is desire signal and α_{if} are obtained through the first order filter [31, 32]. The initial conditions are $\alpha_{if}(0) = \alpha_i(0)$, $g_0 = 0$. As we can see in equation (1) and (3) a nonlinear system in strict feedback form and a controller structure have been defined respectively. After applying the controller in the equation (1), equation (4) can be derived as

$$\begin{aligned} \dot{z_1} &= g_1 z_2 \\ \dot{z_i} &= -g_{(i-1)}z_{(i-1)} + g_i z_{(i+1)} \\ \dot{z_n} &= -g_{(n-1)}z_{(n-1)} \end{aligned} \tag{4}$$

This equation (4) can be written in compact form as $\dot{z} = Az$ which can be seen in equation (5).

$$\dot{z} = f(z) = Az \tag{5}$$

$$A = \begin{pmatrix} 0 & g_1 & 0 & & 0 \\ -g_1 & 0 & g_2 & & 0 \\ \ddots & \ddots & \ddots & & \ddots \\ 0 & & 0 & -g_{(n-1)} & 0 \end{pmatrix}$$

where A is a skew symmetric matrix. For a skew-symmetric matrix A, the inequality $A + A^T \leq 0$ always holds true, and the closed loop dynamics described by A is therefore stable even though marginally stable. Additionally, we have added robustifying terms to make the closed loop dynamics uniformly ultimately bounded.

By using equation (3), the equation (1) can be converted into the following form:

$$\dot{z_n} = G(z)u + F(z) \tag{6}$$

Assumption 1 A transformation $\omega = T(x)$ exists for which the equation (2) can be expressed as in the following form:

$$\dot{\omega} = A\omega + B(f(\omega) + h(\omega)u) \tag{7}$$

where A and B are in Brunovsky canonical form. $\omega \in D_\omega \subset \Re^n$. $f(\omega) : D_\omega \to \Re$, $h(\omega) : D_\omega \to \Re$ and $h(\omega) \neq 0$, $\forall \omega \in \Re^n$.

Assumption 2 The function $f(\omega)$ and $h(\omega)$ are locally Lipschitz.

Assumption 3 A smooth saturation function with the following properties ($\sigma(.) : \Re \to \Re$) exists -

(a) $\sigma(0) = 0$ and $s\sigma(s) > 0 \forall s \neq 0$
(b) $\lim_{s\to\infty} \sigma(s) = 1$, $\lim_{s\to-\infty} \sigma(s) = -1$
(c) For a finite set $D_s \subset \Re$, $\frac{\partial\sigma(s)}{\partial s} > 0 \forall s \in D_s$

Assumption 1 can transform the system (1) into a fully linearizable form. Assumption 2 gives information about existence of solution and uniqueness of solution for equation (2). And finally Assumption 3 implies that the saturation function $\sigma(s)$ is continuously differentiable and it will be an odd function such as sigmoid function, hyperbolic tangent function etc.

The aim is find an input u that ensures the closed loop trajectories of the system $\omega(t)$ closely match the states $\omega_d(t)$ of a stable system, as mentioned below:

$$\dot{\omega}_d = A_d\omega_d + B_d r \tag{8}$$

where $\omega_d = [\omega_{d1}, \omega_{d2}, ..., \omega_{dn}]' \in \Re^n$, r represents bounded reference input.

$$A_d = \begin{pmatrix} 0 & 1 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & 1 \\ -a_{d1} & -a_{d2} & & -a_{dn} \end{pmatrix}$$

$$B_d = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ b_d \end{pmatrix}$$

where A_d is Hurwitz for suitable a_{di}, $i = (1, 2, 3...n)$.

3. Controller design

Here, the objective is to solve the tracking problem of non-affine system by using a saturated control rule $u = \beta\sigma(\frac{v}{\beta})$, where v is a new control input.

Error dynamics can be denoted as:

$$\dot{e} = A\omega + BF(\omega, v) - A_d\omega_d - B_d r \tag{9}$$

$$\mu\dot{v} = -sign(j_g)[F(z) + G(z)\sigma(v) - \dot{\omega}_d + \sum_{i=1}^{n} tanh(z_i)] \tag{10}$$

where $e = \omega - \omega_d$ and $F(\omega, v) = f(\omega) + \beta h(\omega)\sigma(\frac{v}{\beta}), h(\omega) > 0$. Equation (7) and Equation (10) can be represented in the following forms:

$$\begin{aligned} \dot{\omega} &= A\omega + BF(\omega, v) \\ \mu\dot{v} &= g(v, z, \omega_d, r) \end{aligned} \tag{11}$$

where

$$g(.) = -sign(j_g)[F(z) + G(z)\sigma(v) - \dot{\omega}_d + \sum_{i=1}^{n} tanh(z_i)] \tag{12}$$

Equation (11) represents the time scale form of the system and the system's dynamics are in a standard singularly perturbed form (The fast dynamics consist of isolated and unique root). This fact can be observed in Theorem 1.

Theorem 1 *The ensuing closed-loop dynamics may be described in the standard singularly perturbed form if the pre-stated assumptions (1-3) continue to hold true and the dynamic controller is chosen as* $u = \beta\sigma(\frac{v}{\beta})$. *In this situation,*

$v = p(z, \omega_d, r)$ of the equation $g(v, z, \omega_d, r) = 0$ is the unique root.

Proof Reference [33, 34] presents findings that prove the existence of a distinct root for a vector field that is partially contracting. The Jacobian matrix of the vector field $g(.)$ with respect to the variable v is designated as:

$$\begin{aligned} j_g &= -\alpha(h(\omega)\frac{\partial(\sigma(v/\beta))}{\partial(v/\beta)}) \\ &= -a_v||((h(\omega)\frac{\partial(\sigma(v/\beta))}{\partial(v/\beta)}))|| \end{aligned} \tag{13}$$

The expression j_g exhibits negative definiteness within a specific bounded reason of D_ω and D_v, as indicated by equation (13). So, the equation $\mu\dot{v} = g(v, z, \omega_d, r)$ will be partially contracting in v and there will be a unique and isolated solution of $g(v, z, \omega_d, r) = 0$ [16, 34]. In other word, we can say that a mapping between the variables z, ω_d, r and control input v exist inside the domain D_ω and D_v. That's why, equation (11) is in the typical form of a singularly perturbed equation.

Corollary 1 *The error dynamics involving the slow reduced system of equations (11) is contracting in e and $(\omega_r - \omega_d) \to 0$.*

Proof Consider the root of the equation $g(v, z, \omega_d, r) = 0$ is denoted by $v = p(z, \omega_d, r)$. Hence, the reduced slow system of equations (11) will be as:

$$\dot{\omega}_r = A\omega + BF(\omega_r, p(z, \omega_d, r)) \tag{14}$$

Using $g(p(.), z, \omega_d, r) = 0$, we can write

$$\begin{aligned} u &= \beta\sigma(\frac{p(.)}{\beta}) \\ &= \frac{1}{G(z)}(-F(z) - \sum_{i=1}^{n} a_{di}\omega_{di} + \sum_{i=1}^{n} b_{di}r \\ &- \sum_{i=1}^{n} tanh(z_i)) \end{aligned} \tag{15}$$

The error dynamics can be expressed as:

$$\dot{e} = A\omega + F(\omega, p(.)) - A_d\omega_d - B_d r = A_e e \tag{16}$$

where the A_e is Hurwitz matrix and it has distinct real eigen values. A positive definite matrix M_{Ae} will exist in such a way that $\overline{A}_e = M_{Ae}A_eM_{Ae}^{-1}$ is negative definite. Here, M_{Ae} is invertible matrix. So, equation (16) is contracting in M_{Ae} and $\omega(t)$ exponentially converge to ω_d i.e. $(\omega_r - \omega_d) \to 0$ and $e(t) \to 0$ [35].

Feasible region: With a limited actuation power any unstable system may not be globally or semi globally stabilized. The tracking problem in this article can be solved inside a certain domain. From equation $u = \beta\sigma(\frac{v}{\beta})$, the dynamic inversion with a limited actuation power β is feasible, provided

$$\begin{aligned} u &= \beta\sigma(\frac{p(.)}{\beta}) \\ &= \frac{1}{G(z)}(-F(z) - \sum_{i=1}^{n} a_{di}\omega_{di} + \sum_{i=1}^{n} b_{di}r \\ &- \sum_{i=1}^{n} tanh(z_i)) \end{aligned} \tag{17}$$

The parameter, denoted as a_{di}, is adjustable, and $||\sigma(.)|| \leq 1$. Consequently, the region of attraction, represented by the finite region $(D_\omega \times D_v)$, can be approximated. Within this region, the vector fields satisfy the following condition.

$$||\frac{1}{h(\omega)}(-f(\omega) + \dot{\omega}_d)|| < \beta \tag{18}$$

4. Convergence analysis

Equations (11) is convergent in nature. This crucial information provides boundedness of closed loop system and represented in the form of lemma.

Lemma 1 *Consider a system $\dot{\omega} = f(\omega)$ which is contracting in a transformation matrix M with rate λ and assume a perturbed system $\dot{\omega}_p = f(\omega_p) + w(\omega_p, \xi)$, ξ: An external signal.*

(1) if $||w(\omega_p, \xi)|| \leq m_0$, $m_0 \in R^+$,then the following equation will be satisfied.

$$||\omega_p(t) - \omega(t)|| \leq \Gamma e^{-\lambda t}||\omega_p(0) - \omega(0)|| + \frac{\Gamma m_0}{\lambda} \tag{19}$$

where Γ represents the supremum of the condition number for M.

(2) Let us consider two positive scalars m_{on}, m_{1n} such that $||w(\omega_p, \xi)|| \leq m_{on} + m_{1n}||\omega_p||$. Also consider that, a forward bounded solution ω_b exists for the system's dynamics, $\dot{\omega} = f(\omega)$, where the inequality $m_{1n} \leq \frac{\lambda}{\Gamma}$, then

$$\dot{\omega}_p = f(\omega_p) + w(\omega_p, \xi) \tag{20}$$

is partially contracting in ω_p and

$$||\omega_p(t) - \omega(t)|| \leq \Gamma e^{-(\lambda - \Gamma m_{1n})t}||\omega_p(0) - \omega(0)|| + m_{2n} \tag{21}$$

bound can be obtained. where, $m_{2n} = \frac{\Gamma(m_{on} + m_{1n}\omega_b)}{\lambda - \Gamma m_{1n}}$.

4.1 *Convergence to Slow Manifold, $p(z, \omega_d, r)$*

Above mentioned controller in equation (10) is given in the form:

$$\mu\dot{v} = g(v, z, \omega_d, r) \tag{22}$$

Let us consider a virtual perturbed system of equation (10) in stretched time scale $\tau = \frac{t}{\mu}$,

$$\frac{dv_{ds}}{d\tau} = g(v_{ds} + p(.), z, \omega_d, r) + p'(v_{ds}, z, \omega_d, r, \mu) \tag{23}$$

where

$$\begin{aligned} p'(v_{ds}, z, \omega_d, r, \mu) &= -\mu \frac{d}{dt}(p(.)) \\ v_{ds} &= v(t) - p(z, \omega_d, r) \end{aligned} \tag{24}$$

$p'(.)$ can be evaluated as:

$$p'(v_{ds}, z, \omega_d, r, \mu) = \mu(\frac{\partial p(.)}{\partial z}\dot{z} + \frac{\partial p(.)}{\partial \omega_d}\dot{\omega_d} + \frac{\partial p(.)}{\partial r}\dot{r}) \tag{25}$$

By using the Lipschitz properties of the function $f(\omega)$, u, and $p(z, \omega_d, r)$ it can be derived that [35]:

$$||p'|| \leq m_{2a}||v_{ds}|| + m_{2b}||\omega_d|| + m_{2c}||\dot{r}|| + m_{2d} \tag{26}$$

where m_{2a}, m_{2b}, m_{2c}, m_{2d} are positive scalars. Let us consider $p'(.) = 0$ then $\frac{dv_{ds}}{d\tau}$ is partially contracting in v_{ds}. Let's assume contraction rate is $\frac{\overline{\lambda_v}}{\mu}$, M_v is transformation matrix and Ω_v is supremum of the condition number for M_v. Reference [36] ensures forward boundedness of trajectories inside the unperturbed system $\frac{dv_{ds}}{d\tau} = g(v_{ds} + p(.), z, \omega_d, r)$. It is stated in Lemma 2.

Lemma 2 *[36]: Let us assume that $\dot{\omega} = f(\omega, t)$, $\omega \in R^n$, is contracting in ω for all $t > 0$. Suppose the contraction matrix is time invariant, finite and also $||f(0, t)|| \leq a < \infty$ for all $t > 0$, then the trajectories $\omega(t)$ remain bounded and the solution $\omega(t)$ will be well defined for all $t > 0$.*

If $v_{ds} = 0$ then $v = p(.)$ which implies that the function $g(v_{ds} + p(.), z, \omega_d, r)$ die out at $v_{ds} = 0$. By using Lemma 2 we can summarize that $v_{ds}(\tau)$ of the unperturbed system $\frac{dv_{ds}}{d\tau} = g(v_{ds} + p(.), z, \omega_d, r)$ are forward bounded for all $t > 0$ and a positive number a_0 exist for which $||v_{ds}(t)|| \leq a_0$ for all $t > 0$. For equation $g(0 + p(.), z, \omega_d, r) = 0$, $v_{ds} = 0$ can act as a particular solution of $\frac{dv_{ds}}{d\tau} = g(v_{ds} + p(.), z, \omega_d, r) + p'(v_{ds}, z, \omega_d, r, \mu)$ in the absence of perturbation term.

There exist a scalar μ_m such that $m_{2a} \leq \frac{\overline{\lambda_v}}{\mu^2\Omega_v}$ for all $\mu < \mu_m$. That is why Lemma 1 can be used in equation $\frac{dv_{ds}}{d\tau} = g(v_{ds} + p(.), z, \omega_d, r) + p'(v_{ds}, z, \omega_d, r, \mu)$ to derive:

$$\begin{aligned} ||v_{ds}(t) - 0|| &\leq ||v(t) - p(z, \omega_d, r)|| \\ &\leq \Omega_v e^{-(\frac{\lambda_v}{\mu})t}||v(0) - p(z(0), \omega_d(0), r)|| + \frac{\mu\Omega_v d_2}{\lambda_v} \end{aligned} \tag{27}$$

where λ_v and d_2 can be evaluated from [36] as $\lambda_v = \overline{\lambda}_v - \mu\Omega_v m_{2a}$ and $d_2 = \mu(m_{2a}v_0 + m_{2c}||r|| + m_{2d})$. By using [36] following bound can be obtained.

$$\lim_{t\to\infty} ||v(t) - p(z, \omega_d, r)|| \leq \mu\Omega_v \frac{d_2}{\lambda_v} \tag{28}$$

4.2 *Convergence analysis of trajectories*

Equation (7) can be written in the following form with extra perturbation terms as:

$$\begin{aligned} \dot{\omega} &= A\omega + BF(\omega, p(z, \omega_d, r)) + BF(\omega, v) \\ &\quad - BF(\omega, p(z, \omega_d, r)) \end{aligned} \tag{29}$$

The first two terms $A\omega + BF(\omega, p(z, \omega_d, r))$ is same as equation (11) and that's why it is partially contracting in ω with transformation matrix M_{Ae}. By employing the Lipschitz properties of the function $F(.)$ it can be determined that there exists a Lipschitz constant denoted as L_1. This constant ensures $||BF(\omega, v) - BF(\omega, p(z, \omega_d, r))|| \leq L_1||v - p(z, \omega_d, r)||$.

The bound of equation (28) is finite inside the region of interest. That is why finite perturbation terms for all $t > 0$ in (29) can be derived. In (29) the nominal unperturbed dynamics is partially contracting, so Lemma 2 can be used to obtain forward boundedness of trajectories $\omega(t)$ $\forall t \geq 0$. By using Lemma 1 on (14) and (29), the following bounded equation can be derived:

$$\lim_{t\to\infty} ||\omega(t) - \omega_r(t)|| \leq \frac{\Omega_{ae}L_1}{\lambda_{ae}}(\frac{\mu\Omega_v d_2}{\lambda_v}) \tag{30}$$

where λ_{ae} is the minimum eigen value of $\overline{A}_e$ and Ω_{ae} is the condition number of M_{Ae}. From the equation (30), bounded steady state tracking error can be derived by using triangle inequalities.

$$\lim_{t\to\infty} ||\omega(t) - \omega_d(t)|| \leq \frac{\Omega_{ae}L_1}{\lambda_{ae}}(\frac{\mu\Omega_v d_2}{\lambda_v}) \tag{31}$$

5. Filtered saturated controller with disturbance observer

A disturbance observer with a high gain is provided to estimate the uncertain dynamics of the systems. The system uncertainties can be eliminated through disturbance observer $\delta_i(.)$ [37].

Table 1. Parametric values of TRMS

TRMS Parameters	Values
I_1 (Moment of inertia of vertical rotor)	$6.8 * 10^{-}2$ Kg m^2
T_{11}(Motor 1 denominator parameter)	1.1
T_{10} (Motor 1 denominator parameter)	1
M_g (Gravity momentum)	0.3 N-m
$B_{1\psi}$ (Friction momentum function parameter)	$6 * 10^{-}3$ N m s/rad
$B_{2\psi}$ (Friction momentum function parameter)	$1 * 10^{-}3$ N m s^2/rad
k_{gy} (Gyroscopic momentum parameter)	0.05 s/rad
a_1 (Static characteristic parameter)	0.0135
b_1 (Static characteristic parameter)	0.0924

$$\begin{aligned}
\dot{\xi}_i &= -k(\xi_i + kx_i + f_i(x_1, x_2, \ldots, x_i) + g_i x_{i+1}) \\
\dot{\xi}_n &= -k(\xi_n + kx_n + f_n(x_1, x_2, \ldots, x_n) + g_n u) \\
\hat{\delta}_i(.) &= \xi_i + kx_i \\
\hat{\delta}_n(.) &= \xi_n + kx_n
\end{aligned} \tag{32}$$

where the scalar $k > 0$ is a large positive constant, $\hat{\delta}_i(.)$ is the estimation of disturbance $\delta_i(.)$, and ξ_i is an intermediate variable. Using observer (32), the dynamics of disturbance estimator ($\hat{\delta}_i$) is given by:

$$\dot{\hat{\delta}}_i = h_i(\hat{\delta}_i, \delta_i) = -k(\hat{\delta}_i - \delta_i). \tag{33}$$

Defining a disturbance error variable $\tilde{\delta}_i = \delta_i(.) - \hat{\delta}_i(.)$, the estimation error dynamics can be written as:

$$\epsilon\dot{\tilde{\delta}}_i(.) = -\tilde{\delta}_i - \epsilon\dot{\delta}_i(.). \tag{34}$$

where $0 \leq \epsilon = 1/k \leq 1$. The boundary layer dynamics resulting from (34) can be written as:

$$\frac{\delta}{\delta\tau}(\tilde{\delta}_i) = \tilde{\delta}_i. \tag{35}$$

5.1 *Conversion to Singularly Perturbed form*

The filter based saturated control law u chosen for system (1) is based on a disturbance observer and filtering.

$$\begin{aligned}
u &= \beta\sigma(\frac{v}{\beta}) \\
\mu\dot{v} &= -sign(j_g)[F(z) + G(z)\sigma(v) - \hat{\delta}_i \\
&\quad - \dot{\omega}_d + \sum_{i=1}^{n} tanh(z_i)] \\
z_i &= x_i - \alpha_{if}, (i = 1, 2.....m) \\
\alpha_{1f} &= \omega_d, \ i \in [2, 3...m] \\
\alpha_i &= \frac{1}{g_{(i-1)}}[-f_{(i-1)}(x_1, ..x_{(i-1)}) \\
&\quad - g_{(i-2)}z_{(i-2)} + \dot{\alpha}_{(i-1)f} - \hat{\delta}_{(i-1)}], i = 2, 3, 4...n.
\end{aligned} \tag{36}$$

The closed-loop system described in equation (1) with the control law given in equation (36) can be represented as follows:

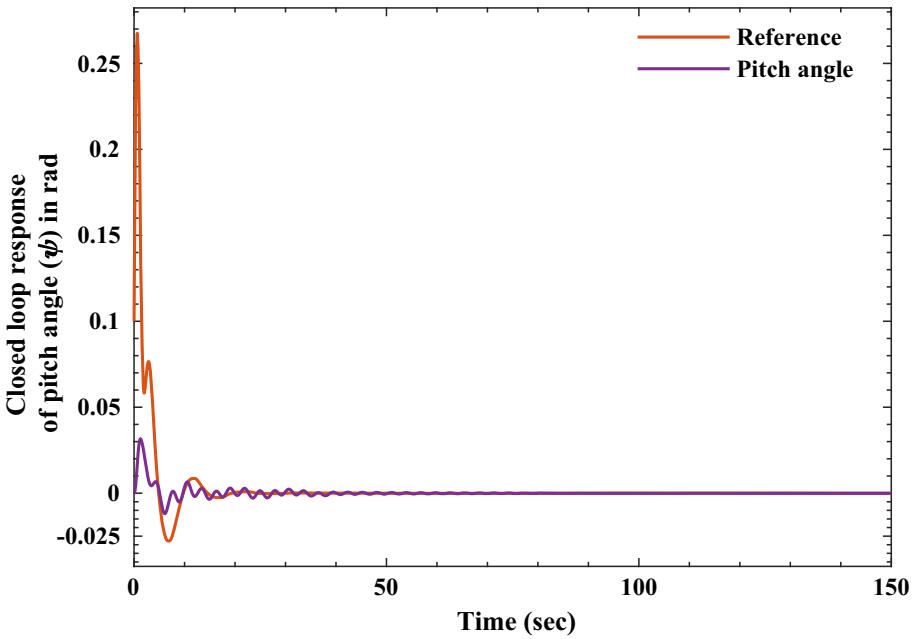


Figure 1. Closed loop response of pitch angle in rad and reference signal.

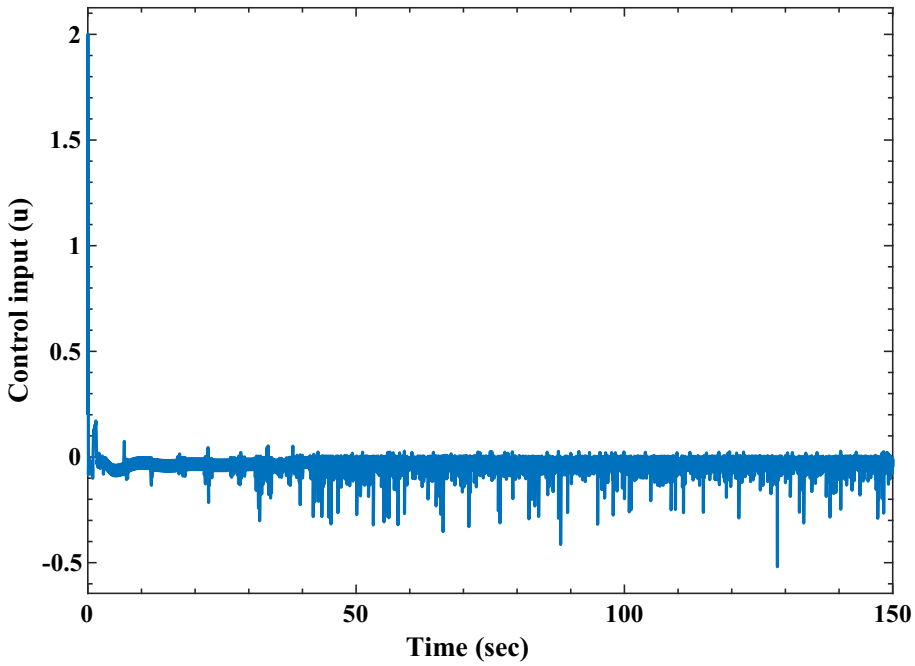


Figure 2. Closed loop control signal.

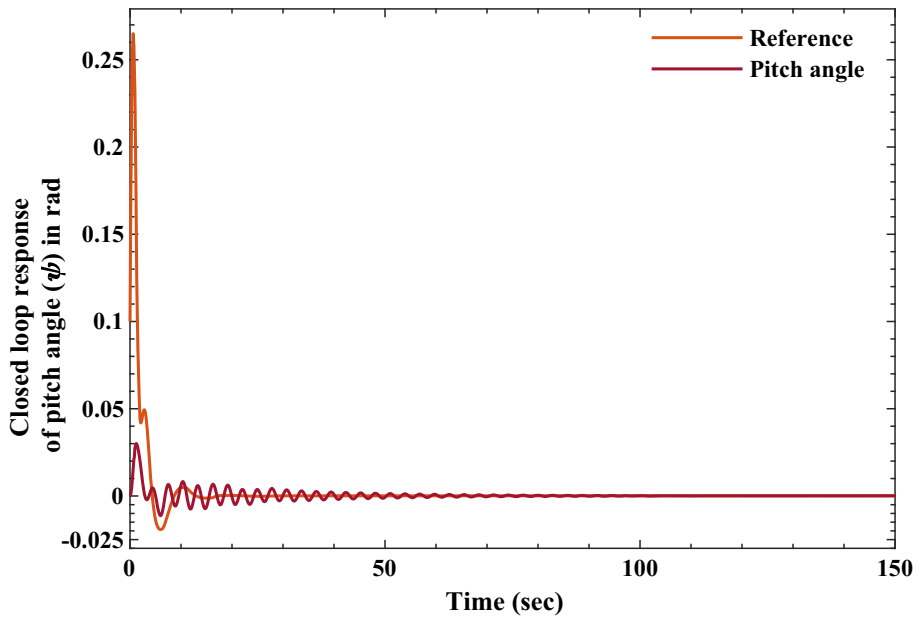


Figure 3. Closed loop response of pitch angle in rad and reference signal.

$$\begin{aligned}
\dot{z}_1 &= g_1 z_2 + g_1 \tilde{\alpha}_{2f} + \tilde{\delta}_1 \\
\dot{z}_i &= g_{(i-1)} z_{i-1} \\
&\quad + g_{(i)} z_{i+1} + g_i \tilde{\alpha}_{(i+1)f} + \tilde{\delta}_i \\
\dot{z}_n &= -g_{(n-1)} z_{n-1} + \tilde{\delta}_{(n)}.
\end{aligned} \tag{37}$$

Rewriting (37) in a compact form as (38),(39) and (40)

$$\dot{z} = f(z, \tilde{\alpha}_f, \tilde{\delta}) \tag{38}$$

$$\mu \dot{\tilde{\alpha}}_f = g(\alpha_f, \alpha_i) = -\tilde{\alpha}_f + \mu \dot{\alpha}_i \tag{39}$$

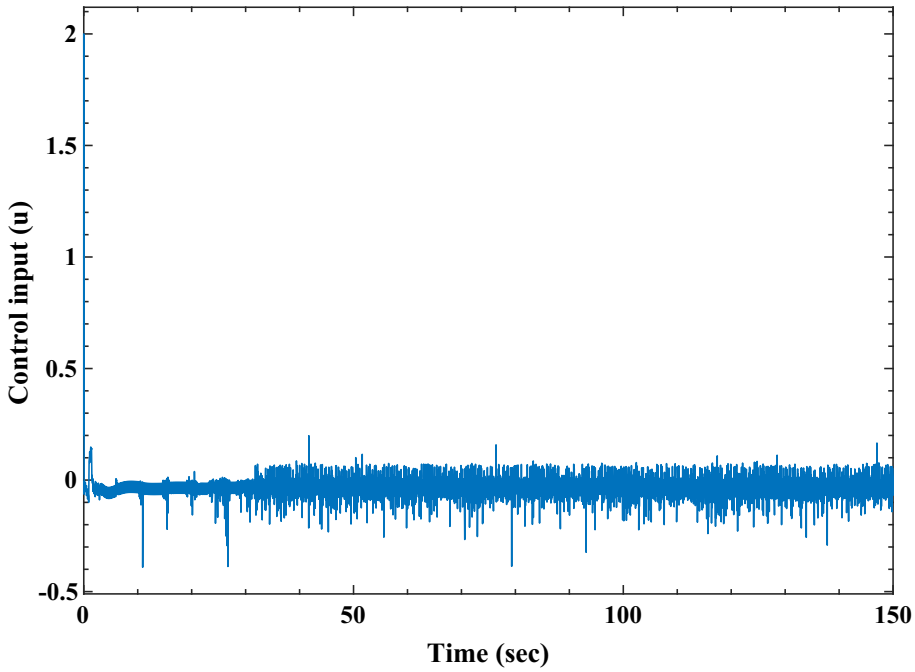


Figure 4. Closed loop control signal

$$\epsilon\dot{\tilde{\delta}} = h(\hat{\delta}, \delta_i) = -\tilde{\delta} + \epsilon\dot{\delta}_i \tag{40}$$

where $\hat{\delta} = [\hat{\delta}_1, \hat{\delta}_2, \ldots, \hat{\delta}_n]$, $\tilde{\delta} = [\tilde{\delta}_1, \tilde{\delta}_2, \ldots, \tilde{\delta}_n]$ $\delta_i = [\delta_1 \delta_2, \ldots, \delta_n]$. In equations (38)–(39), the system dynamics undergo a transformation, resulting in a singularly perturbed system with three distinct time scales. By employing the ratio $\frac{\epsilon}{\mu}$ this system can be further simplified into a two-time scale format. Assuming, without loss of generality, that $\epsilon < \mu$, indicating that the observer dynamics operates at a faster rate as compared to the filter dynamics. A suitable choice for ϵ is $\epsilon = \kappa\mu$, where $\kappa < 1$ is a small positive constant. Express the fast subsystem (39), (40) as:

$$\mu\dot{\tilde{\nu}} = W(\tilde{\nu}, \dot{\nu}, \kappa) = -\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\kappa} \end{bmatrix}\tilde{\nu} + \mu\dot{\nu} \tag{41}$$

where $\tilde{\nu} = [\tilde{\alpha}_f^T, \tilde{\delta}^T]^T$, $\dot{\nu} = [\dot{\alpha}_i^T, \dot{\delta}_i^T]^T$ and $W(.) = [g^T(.), \frac{1}{\kappa} h^T(.)]^T$. By using Equation (41), the dynamics of the closed loop system can be described using two distinct time scales in a singularly perturbed form. Convergence of system trajectories can be analysed by using same reasoning.

6. Experimental outcomes

The TRMS serves as a benchmark in the field of aerodynamics and shares similarities with helicopter dynamics, featuring two inputs and two outputs. This system containing vertical and horizontal rotors which are actuated by two DC motors. The vertical rotor gives vertical take off and landing momentum (pitch angle ψ) and the horizontal rotor gives the steering action (yaw angle ϕ) under strong coupling. The following momentum equations [38] for the vertical take off and landing of TRMS is considered to validate the boundedness and convergence of proposed filter saturated controller:

$$I_1\ddot{\psi} = M_1 - M_{FG} - M_{B\psi} - M_G \tag{42}$$

and

$$\tau_1 = \frac{k_1}{T_{11}s + T_{10}}u \tag{43}$$

where

$M_1 = a_1\tau_1^2 + b_1\tau$, nonlinear static characteristic
$M_{FG} = M_g sin(\psi)$, gravity momentum
$M_{B\psi} = B_{1\psi}\dot{\psi} + B_{2\psi} sign(\dot{\psi})$, friction forces momentum
$M_G = K_{gy}M_1\dot{\phi}cos(\psi)$, gyroscopic momentum

The parametric values of the experimental setup of TRMS [39] is given in table 1. The actuated motor momentum is described by the equation (43). Here, u is bounded control input.

Consider, $x_1 = \psi$, $x_2 = \dot{\psi}$ and $x_3 = \tau_1$.

$$\begin{aligned} \dot{x}_1 &= g_1x_2 + f_1(x_1) \\ \dot{x}_2 &= g_2x_3 + f_2(x_1, x_2) \\ \dot{x}_3 &= g_3u + f_3(x_1, x_2, x_3) \end{aligned} \tag{44}$$

where $g_1 = 1, \quad f_1(x_1) = 0, \quad g_2 = \frac{b_1}{I_1}$, $f_2(x_1, x_2) = -\frac{1}{I_1}[M_g sinx_1 + B_{1\psi}x_2 + B_{2\psi}signx_2], \quad g_3 = \frac{1}{T_{11}}$ and $f_3(x_1, x_2, x_3) = -\frac{T_{10}}{T_{11}}x_3$

The proposed controller was tested and validated on a TRMS under two different test cases. Additionally, in both cases, a nonlinearity term $\delta = 10^3 * tanh(2t)$ has been added to the state x_2.

For the first test case, the reference parameters a_{di} of the stochastic bounded signal has been taken as $a_{d1} = 20$, $a_{d2} = 21$ and $a_{d3} = 42$ with perturbed parameter $\mu = 0.01$. It is observed from the figure 1 the closed loop state response x_1 of the TRMS tracks stochastic bounded reference signal quickly and ensure stability during disturbed environment. The required control signal to track the reference, depicted in figure 2, is designed to achieve quick stability.

The tracking performance of the designed controller has been also tested under second test case for different reference parameters such as $a_{d1} = 25$, $a_{d2} = 26$, $a_{d3} = 40$ with perturbed parameter $\mu = 0.01$. It is observed from the figure 3 the trajectory tracking of the closed loop state response x_1 of TRMS satisfactorily follows the reference signal. The required control signal to track the reference is depicted in figure 4. Finally the results reveals that the nonlinear saturated controller satisfactorily meets control design objectives.

7. Conclusion

In this paper, contraction theory based nonlinear saturated controller has been designed specifically for singular perturbations systems. The effectiveness of the designed controller has been tested on a Twin rotor MIMO System (TRMS) for tracking of pitch angle at two different unknown dynamical scenarios. The obtained outcomes suggest that the controller has the capability to provide reliable and precise control under changing conditions. One of the notable versatile strengths of this novel controller is its boundedness and convergence under all closed loop states. Its effective application and performance analysis on TRMS show that it may be used to a variety of physical systems, such as robotics platforms, inverted pendulums, and single-link manipulators etc. In summary, the designed nonlinear saturated controller exhibits encouraging outcomes when applied to TRMS for pitch angle tracking. Also, It may have great potential to tackle control issues in diverse dynamic systems.

Appendix I. Contraction and Partial Contraction

When any two of the system's trajectories, independent of their starting circumstances, converge towards one another, the system denoted by the $\dot{x} = f(x, t)$ is said to be contracting. If there is a positive definite matrix Θ within an area of the state space, then region is referred to as a contracting region. This matrix Θ is constructed in such a way that its symmetric counterpart $(\dot{\Theta} + \Theta\frac{\partial f(.)}{\partial x})\Theta^{-1}$ is negatively definite inside the region [16]. The minimum eigenvalue of the symmetric component is referred to as the contraction rate, and Θ represents the transformation matrix associated with it. A system $\dot{x} = f(x, y, t)$ is called partial contracting in x if the dynamics is contracting in x for all y [40].

References

[1] Boskovic J D, Li S M and Mehra R K 2004 Robust tracking control design for spacecraft under control input saturation. *Journal of Guidance, Control, and Dynamics* 27(4): 627–633

[2] Daafouz J, Garcia G and Bernussou J 1997 Output feedback control for uncertain singularly perturbed systems. European Control Conference (ECC), 1927–1932

[3] Tikhonov A N 1952 Systems of differential equations containing small parameters in the derivatives. *Mat. Sb. (NS)* 31(73): 575–586

[4] Kokotovic P, Khalil H. K, and O'reilly J 1999 Singular perturbation methods in control: Analysis and design. SIAM

[5] Naidu D S 1988 Singular perturbation methodology in control systems. IET

[6] Khalil H K 1989 Feedback control of nonstandard singularly perturbed systems.American Control Conference, 2172–2177

[7] Narang-Siddarth A 2014 and Valasek J. Nonlinear time scale systems in standard and nonstandard forms, Analysis and control. SIAM

[8] Saberi A and Khalil H 1985 Stabilization and regulation of nonlinear singularly perturbed systems-composite control. *IEEE Transactions on Automatic Control* 30(8): 739–747

[9] Son J-W and Lim J T 2008 Stabilization of approximately feedback linearizable systems using singular perturbation. *IEEE Transactions on Automatic Control* 53(6): 1499–1503

[10] Saberi A and Khalil H 1984 Quadratic-type lyapunov functions for singularly perturbed systems. *IEEE Transactions on Automatic Control* 29(6): 542–550

[11] Calise A J and Moerder D D 1985 Optimal output feedback design of systems with ill-conditioned dynamics. *Automatica* 21(3): 271–276

[12] Haddad A and Kokotovic P 1977 Stochastic control of linear sigularly perturbed systems. *IEEE Transactions on Automatic Control* 22(5): 815–821

[13] Khalil H 1987 Output feedback control of linear two time-scale systems. *IEEE transactions on automatic control* 32(9): 784–792

[14] Won D, Kim W, Shin D and Chung C C 2014 High gain disturbance observer-based backstepping control with output tracking error constraint for electrohydraulic systems. *IEEE transactions on control systems technology* 23(2): 787–795

[15] Hui J and Yuan J 2022 High-gain disturbance observer based double closed-loop sliding mode controller for the load

following of a pressurized-water reactor. *Progress in Nuclear Energy* 150: 10432
[16] Lohmiller W and Slotine J J E 1998 On contraction analysis for non-linear systems. *Automatica* 34(6): 683–696
[17] Sontag E D 2010 Contractive systems with inputs. In: Perspectives in mathematical system theory, control, and signal processing . Springer (pp. 217–228)
[18] Angeli D 2002 A lyapunov approach to incremental stability properties. *IEEE Transactions on Automatic Control* 47(3): 410–421
[19] Aylward E M, Parrilo P A and Slotine J J E 2008 Stability and robustness analysis of nonlinear systems via contraction metrics and sos programming. *Automatica* 44(8): 2163–2170
[20] Rayguru M M and Kar I 2017 Contraction-based stabilisation of nonlinear singularly perturbed systems and application to high gain feedback. *International Journal of Control* 90(8): 1778–1792
[21] Forni F and Sepulchre R 2013 A differential lyapunov framework for contraction analysis. *IEEE transactions on automatic control* 59(3): 614–628
[22] Lohmiller W and Slotine* J J, 2005 Contraction analysis of non-linear distributed systems. *International Journal of Control* 78(9): 678–688
[23] Jouffroy J and Lottin J 2002 On the use of contraction theory for the design of nonlinear observers for ocean vehicles.. Proceedings of the 2002 American Control Conference (IEEE Cat. No. CH37301), 4: 2647–2652
[24] Russo G, Di Bernardo M and Sontag E D 2012 A contraction approach to the hierarchical analysis and design of networked systems. *IEEE Transactions on Automatic Control* 58(5): 1328–1331
[25] Bonnabel S and Slotine J J 2014 A contraction theorybased analysis of the stability of the deterministic extended kalman filter. *IEEE Transactions on Automatic Control* 60(2): 565–569
[26] Jouffroy J and Slotine J J 2004 Methodological remarks on contraction theory. 2004 43rd IEEE Conference on Decision and Control (CDC)(IEEE Cat. No. 04CH37601), 3: 2537–2543
[27] Wang W and Slotine J J E 2005 On partial contraction analysis for coupled nonlinear oscillators. *Biological cybernetics* 92(1): 38–53
[28] Lohmiller W and Slotine J J 2000 Control system design for mechanical systems using contraction theory. *IEEE Transactions on Automatic Control* 45(5): 984–989
[29] Jouffryo J and Lottin J 2002 Integrator backstepping using contraction theory: A brief methodological note. *IFAC Proceedings Volumes* 35(1): 471–475
[30] Rayguru M M and Kar I N 2015 Contraction based stabilization of approximate feedback linearizable systems. *European Control Conference (ECC)* 2015: 587–592
[31] Swaroop D, Hedrick J K, Yip P P and Gerdes J C 2000 Dynamic surface control for a class of nonlinear systems. *IEEE transactions on automatic control* 45(10): 1893–1899
[32] Pan Y and Yu H 2015 Dynamic surface control via singular perturbation analysis. *Automatica* 57: 29–33
[33] Russo G and Di Bernardo M 2009 Contraction theory and master stability function: Linking two approaches to study synchronization of complex networks. *IEEE Transactions on Circuits and Systems II: Express Briefs* 56(2): 177–181
[34] Del Vecchio D and Slotine J J E 2012 A contraction theory approach to singularly perturbed systems. *IEEE Transactions on Automatic Control* 58(3): 752–757
[35] Rayguru M M and Kar I N 2019 High gain observerbased saturated controller for feedback linearizable system. *IEEE Transactions on Circuits and Systems II: Express Briefs* 67(1): 122–126
[36] Pavlov A, Pogromsky A, van de Wouw N and Nijmeijer H 2004 Convergent dynamics, a tribute to boris pavlovich demidovich. *Systems Control Letters* 52(3–4): 257–261
[37] Rayguru M M and Kar I N 2019 Contraction theory approach to disturbance observer based filtered backstepping design. *Journal of Dynamic Systems, Measurement, and Control* 141: 084501
[38] Zeghlache S, Benyettou L, Djerioui A and Ghellab M Z 2020 Twin rotor mimo system experimental validation of robust adaptive fuzzy control against wind effects. *IEEE Systems Journal* 16(1): 409–419
[39] Valluru S K, Singh M, Ayush and Abhiraj D 2020 Design and experimental implementation of multi-loop LQR, PID, and LQG controllers for the trajectory tracking control of twin rotor MIMO system. Intelligent Communication, Control and Devices: Proceedings of ICICCD 2018, Springer Nature, Singapore, 989: 599–608
[40] Bousquet G D and Slotine J J E 2015 A contraction based, singular perturbation approach to near-decomposability in complex systems. arXiv preprint arXiv:1512.08464

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.

https://doi.org/ 10.33472/AFJBS.6.Si2.2024.84-102

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Flavonoids and their Therapeutic Applications in Neurological Disorders

[1] Kanjam Manocha, [2] Navneeta Bharadvaja

[1,2] Delhi Technological University, Delhi, India

Article History

Volume 6,Issue Si2, 2024

Received:22 Feb 2024

Accepted : 13 Mar 2024

doi: 10.33472/AFJBS.6.Si2.2024.84-102

Abstract

Flavonoids are a diverse class of polyphenolic compounds that occur widely in naturally occurring resources like fruits, vegetables, and plant-derived drinks which have potential therapeutic use in the treatment of several life-threatening neurological disorders. This review highlights the role of different flavonoids in various neurodegenerative disorders through their anti-inflammatory, antioxidant, and neuroprotective effects. They seem to have potential stimulatory and inhibitory effects in the mechanism of disease to either stop or delay the progression of various neurological diseases, including Parkinson's, Huntington's, Alzheimer's, and Multiple Sclerosis.

Flavonoids' role in the regulation of oxidative stress, modulation of signaling pathways as well as enhancement in neuroplasticity pose them as a possible neuroprotective agent against a disease in various clinical and models being conducted. Further, their potential to cross the blood-brain barrier and exert neural protective effect make them ideal for therapeutic roles against neurodegenerative disease.

This review highlights the mechanism of various neurodegenerative diseases, the potential of flavonoids to exert neuro-protective effects, their role in modulation of molecular pathways that can be used in combating neurodegenerative diseases, targeted course of action of flavonoids with specificity, and their future potentials in drug discovery and development.

Overall, elucidating the therapeutic potential of flavonoids in neurological disorders holds promise for developing novel therapeutic strategies aimed at improving neurological outcomes and enhancing the way of life for people affected by these debilitating conditions.

Keywords: *Flavonoid, antioxidant, neurological disorder, Parkinson's disorder, anti-inflammatory, Alzheimer's disease*

1. Introduction

Flavonoids being naturally procured secondary metabolites, consisting of polyphenolic compounds that have been observed to show therapeutic effects in cancer, cardiovascular diseases, diabetes, neurological and metabolic disorders. Flavonoids possess potential antimicrobial, antioxidant, anticancer, anti-proliferative, anti-inflammatory, and anti-apoptotic properties, and are extensively studied in lead discovery in drug discovery (Ullah et al., 2020).

They could be classified depending on their chemical structure, degree of saturation, and oxidation of carbon rings and thus further can be differentiated into different subgroups as flavanone, flavanol, flavanonols, flavans, and isoflavonoids (Panche et al., 2016a). These are produced in nature via the phenylpropanoid pathway, and their bioactivity is reliant on their route of absorption and bio-availability. The physicochemical characteristics, such as molecule size, solubility, configuration, lipophilicity, and pKa; influence the absorption of flavonoids(Youdim et al., 2003). Flavonoid glycoside or aglycan are the two types of flavonoids whose structures determine whether they are absorbed in may be in colon or small intestine. While flavonoid glycosides must first be transformed into aglycan to be absorbed, the aglycan portion of flavonoids is absorbed in the small intestine (Dabeek & Marra, 2019).

Our rapidly transforming lifestyles, high-caloric food habits, sedentary lifestyles, and stress have resulted in an increase in many metabolic diseases, cancer, and neurological disorders. Neurodegenerative conditions primarily affect people in their later years of life; a lot of people suffer from conditions such as Parkinson's disorder, Multiple Sclerosis, Huntington's disorder, and Alzheimer's disorder. The accumulation of abnormal proteins, oxidative stress, neuroinflammation, and mitochondrial dysfunction all of which have the potential to cause the degeneration of specific neurons within the brain and are some of the most significant pathophysiological bases of neurodegenerative diseases (Hamsalakshmi et al., 2022). Synaptic and cognitive dysfunction have an intrinsic relationship to major neurological disorders such as Parkinson's and Alzheimer's, which deteriorate with the loss of these neural cells and the signals that they generate.

The key causes of neurodegenerative diseases are oxidative stress and neuroinflammation. Numerous factors, including abnormal protein accumulation, disruptions in the peroxidation and imbalance in polyunsaturated fatty acid levels, and increased Ca2+ transmission across neurons, are contributors to oxidative stress. Increased oxidative stress triggers the activated interaction of numerous biochemical pathways that cause lipids, proteins, and DNA to undergo oxidative damage, which eventually results in neuronal cell death and neurodegeneration.

This overview provides of the natural flavonoids' resources, structural characteristics, processes, prospects, and potential future therapy approaches for neurological disorders. Employing a range of mechanisms that include molecular pathways and specific molecular targets linked to the pathophysiology of neurological disorders, including oxidative stress, inflammation, vascular remodeling, genetics, ionic channels, and cell proliferation, which with developments in this field would lead in drug discovery. The comprehensive mode of mechanism of several flavonoids were being highlighted against various life-threatening neurological disorders. Flavonoids' interactions with oxidative stress pathways, inflammatory mediators and several other biochemical compounds addressed its potential in having a neuroprotective effect, which either could be harnessed in drug discovery or facilitate the drug in medical treatment or side effects prevention. Further, it has been compared the selective and specific target of different flavonoids in the mechanism of a disease, and thus having a therapeutic action against the disease. The inhibitory and stimulatory effect of different flavonoids were being compared to each other, which would be crucial in knowing their course

of action. This study withstands with such comparisons that may be crucial in target selection, identification in drug discovery and development.

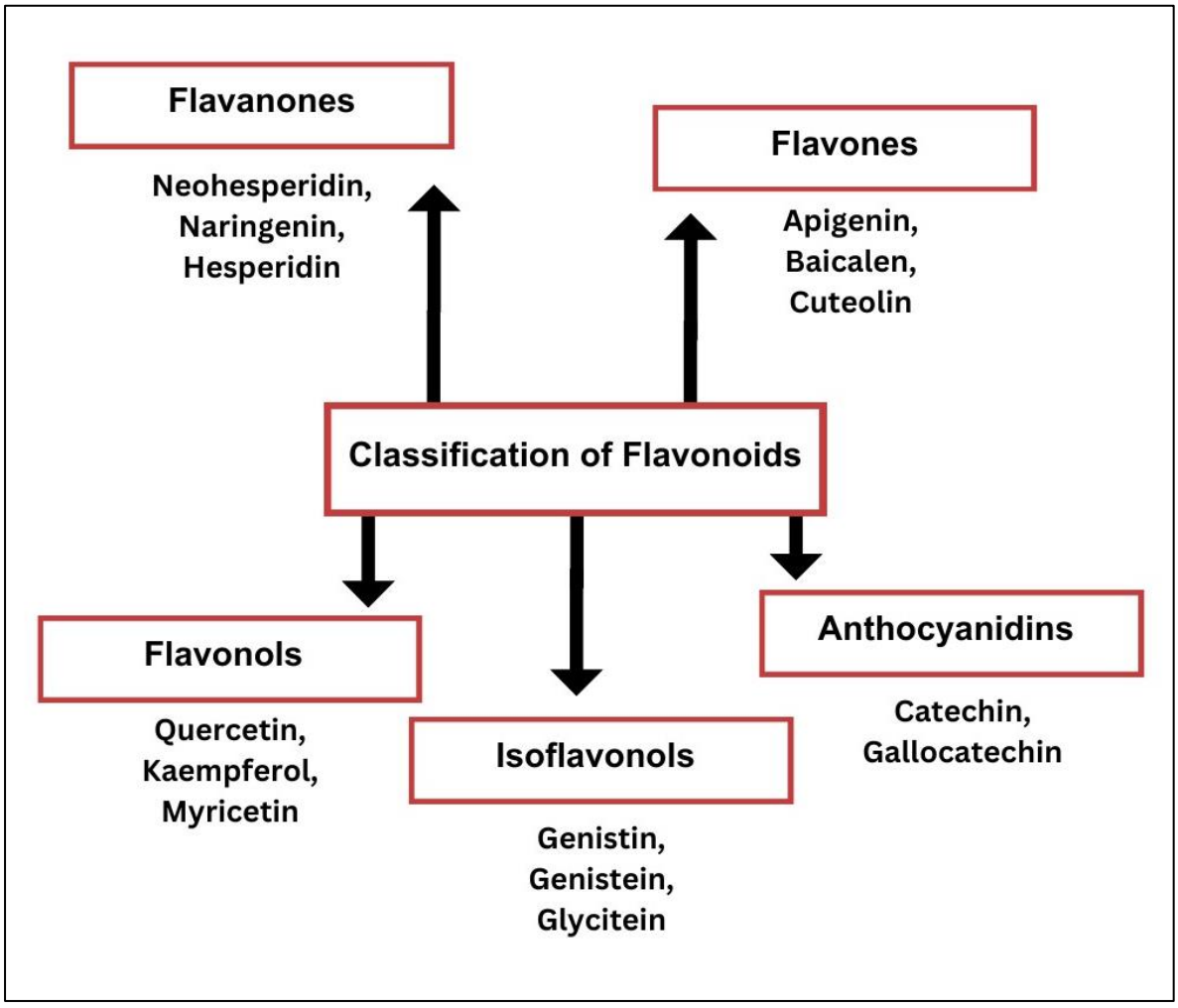


Figure 1: Classification of Flavonoids

2. Biochemical activities of Flavonoids

Plants possess substantial quantities of flavonoids, a sub class of polyphenolic compounds, especially in fruits, vegetables, grains, herbs, and beverages like tea and wine. Different characteristics of Flavonoids are:

2.1 Antioxidant Properties:

One of the most well-known characteristics of flavonoids is their potent antioxidant activity, that they scavenge free radicals and reactive oxygen species (ROS), thereby protecting cells from oxidative damage and reducing the risk of various chronic diseases, including cardiovascular diseases, cancer, and neurodegenerative disorders (Lobo et al., 2010).

2.2 Anti-inflammatory Activity:

Flavonoids possess anti-inflammatory properties, which contribute to their therapeutic potential in mitigating inflammation-associated diseases as they modulate inflammatory

pathways by inhibiting the production of pro-inflammatory cytokines, enzymes (such as cyclooxygenases and lipoxygenases), and signaling molecules involved in the inflammatory response (Rakha et al., 2022).

2.3 Neuroprotective Effects:

Flavonoids have been shown to have neuroprotective benefits in an array of neurological disorders experimental models as they protect neurons from damage caused by oxidative stress, inhibit neurological inflammation, promote neuronal survival, and enhance neuroplasticity these properties make flavonoids promising candidates for the prevention and treatment of neurodegenerative diseases like Alzheimer's and Parkinson's (Ayaz et al., 2019).

2.4 Cardioprotective Effects:

Several flavonoids have been linked to improvements for cardiovascular health. They improve endothelial function, reduce blood pressure, inhibit platelet aggregation, and lower LDL cholesterol levels, thus reducing the risk of cardiovascular diseases such as coronary artery disease and stroke (Ciumărnean et al., 2020).

2.5 Antimicrobial Activity:

Certain flavonoids exhibit antimicrobial properties, inhibiting the growth of bacteria, fungi, viruses, and parasites as they can be used as natural alternatives to conventional antimicrobial agents and may help combat microbial infections and antibiotic resistance (Vaou et al., 2021).

2.6 Potential Anti-cancer Properties:

The potential anti-cancer properties of flavonoids have sparked interest in the scientific community as they exert anti-proliferative, pro-apoptotic, and anti-angiogenic effects on cancer cells, inhibiting tumor growth metastasis, and may also be used to enhance the efficacy of conventional cancer therapies and reduce their side effects (Kopustinskiene et al., 2020).

S.No	Common Name	Plant	Class	Bioactive Compound	References
1	Apple	Malus pumila	Flavanols	Quercetin	(H. Khan et al., 2020)
2	Berries	Vaccinium corymbosum	Flavanols	Myricetin	(Zhu et al., 2020)
3	Broccoli	Brassica oleracea	Flavanols	Kaempferol	(Alam et al., 2020)
4	Cassava	Manihot esculenta	Flavanols	Kaempferol	(Baião et al., 2017)
5	Cassava	Manihot esculenta	Flavanols	Quercetin	(Uarrota et al., 2016)
6	Cassava	Manihot esculenta	Flavanols	Rutin	(Buschmann, 2000)
7	Cauliflower	Brassica oleracea	Flavanols	Kaempferol	(HERRMANN, 1976)

8	Cocoa nibs	Theobroma cacao	Flavones	Apigenin	(Häkkinen et al., 1999)
9	Grapes	Vitis vinifera	Flavanones	Hesperidin	(Egert & Rimbach, 2011)
10	Lemon	Citrus limon	Flavanones	Hesperidin	(Benavente-García & Castillo, 2008)
11	Mango	Mangifera indica	Flavanols	Quercetin	(H. Khan et al., 2020)
12	Onion	Allium cepa	Flavanols	Fisetin	(Sahu et al., 2014)
13	Orange	Citrus sinensis	Flavanones	Hesperidin	(Panche et al., 2016b)
14	Persimmons	Diospyros kaki	Flavanols	Fisetin	(Sahu et al., 2014)
15	Pomegranate	Punica granatum	Flavanols	Quercetin	(H. Khan et al., 2020)
16	Soybean	Glycine max	Isoflavones	Genistein	(Malenčić et al., 2012)
17	Soybean	Glycine max	Isoflavones	Glycitein	(Malenčić et al., 2012)
18	Strawberry	Fragaria ananassa	Flavanols	Kaempferol	(Alam et al., 2020)
19	Tea	Camellia sinensis	Flavanols	Myricetin	(Zhu et al., 2020)
20	Tomato	Solanum lycopersicum	Flavanols	Quercetin	(Panche et al., 2016b)

Table 1: Different Flavonoids and its sources

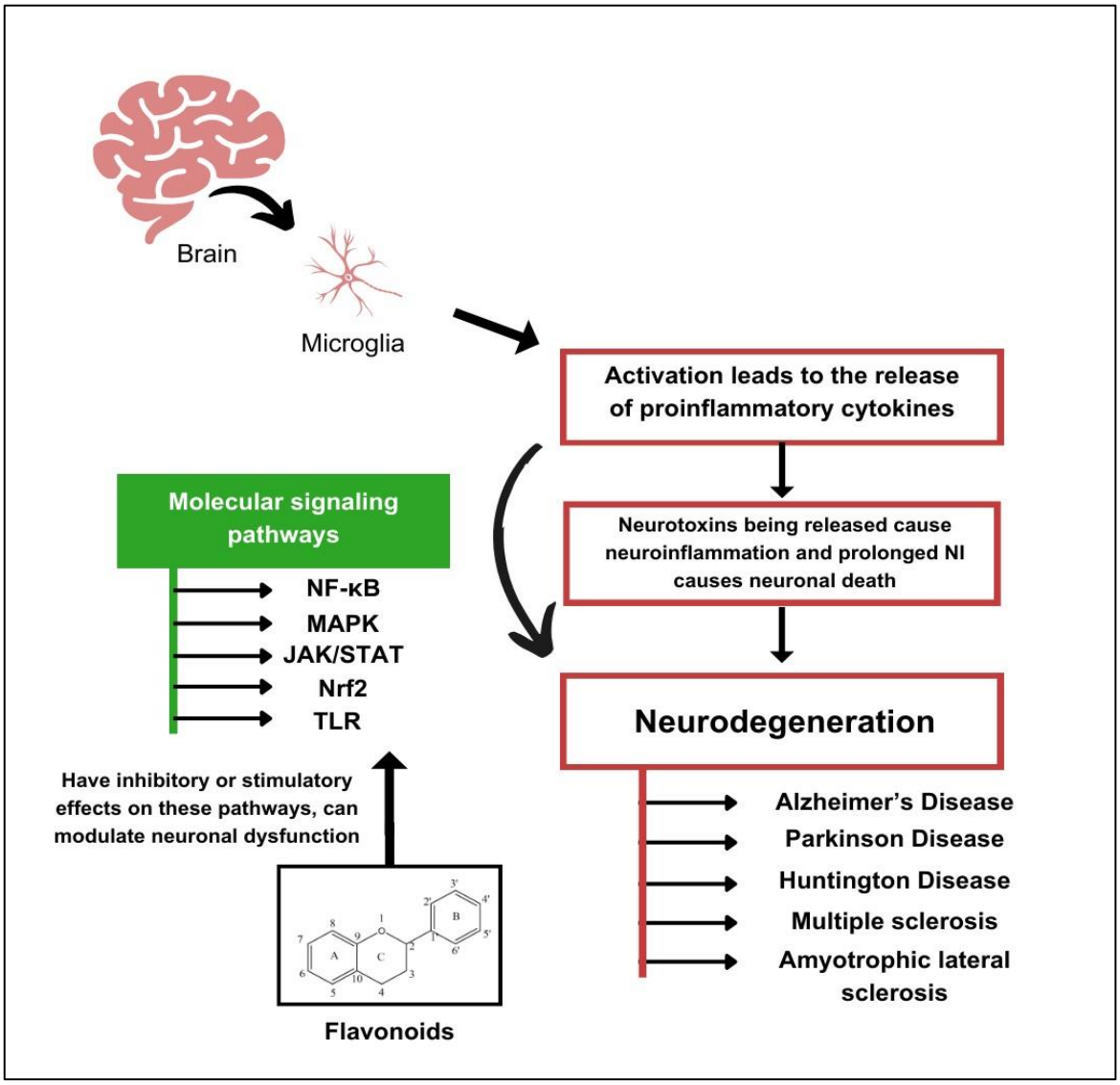


Figure 2: Schematic overview of Neurodegeneration and flavonoid actions

3. Neurological disorders and their current mechanisms

The Electron Transport Chain is usually characterized by producing reactive oxygen species in the bodies of human. Essential metabolic processes like protein phosphorylation, apoptosis, and activation of various transcriptional factors are being carried by Electron Transport Chain, but it is also responsible for lipid peroxidation that causes cell membrane damage. The potential difference exists in membrane of cells, with negative charges on the inner side and positive on the outer, the damage caused by Reactive oxygen species causes alteration in the osmotic pressure, and membrane potential thus causing cell death. The human defense system uses different mechanisms and enzymes to battle endogenous elevated Reactive oxygen species (Brunetti et al., 2013).

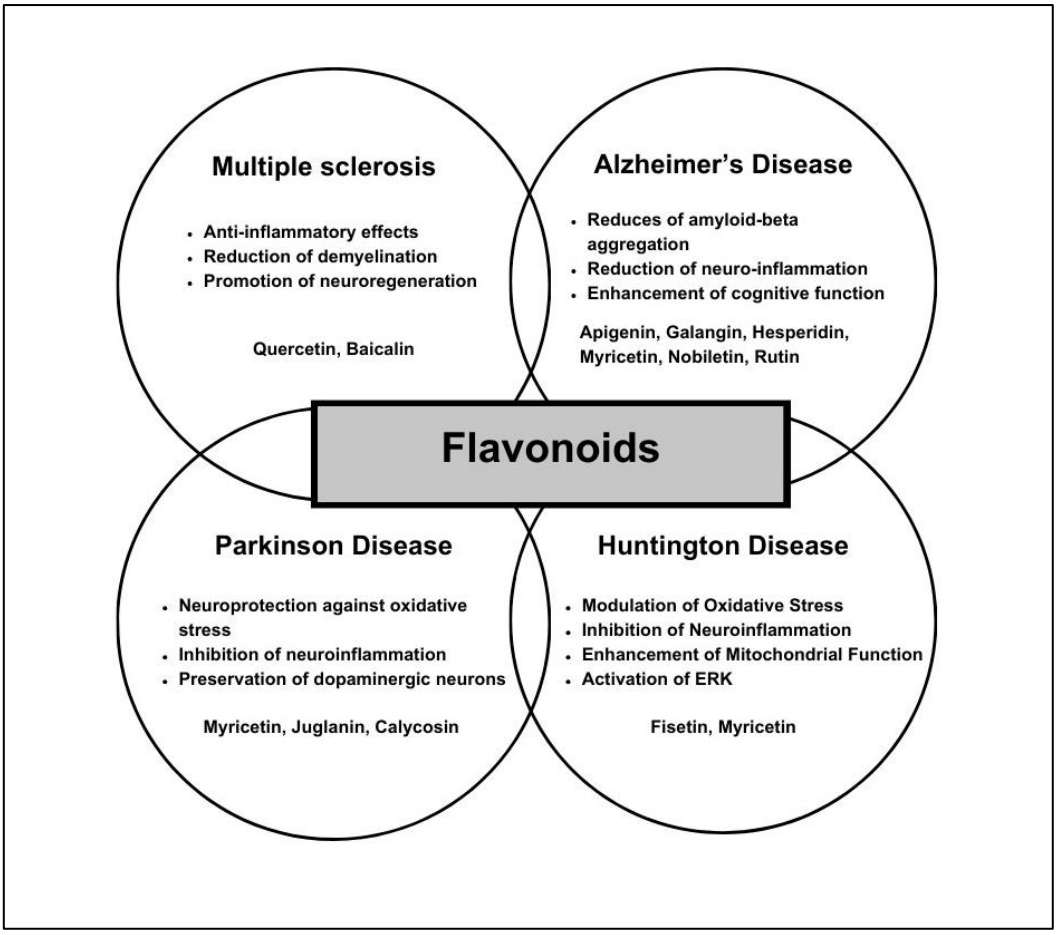


Figure 3: Main mechanism of actions of flavonoids as a neuromodulator

Neuronal interaction with microglia, astrocytes, and other Immunol signaling molecules plays a crucial role in neuroimmunol activation. Microglia are the central nervous system immune cells that usually perform sensing, housekeeping, and protective functions (Mitoma & Manto, 2019; Tchessalova et al., 2018). Microglia activation is an important factor of NI which promotes the release of pro-inflammatory cytokines resulting in progressive neuronal death (Qin et al., 2023). Neurotoxins released from microglia contrary to adverse neuroimmune function led to neurological inflammation and neuronal death and are considered to be the main cause of neurodegenerative diseases (W. Zhang et al., 2023).Neuroprotective therapies continue to be challenging in practice because neurodegeneration is caused by a variety of pathophysiological mechanisms. Thus, of this moment, there is no known cure for neurodegenerative disorders; the only available treatment focuses on symptom management.

Flavonoids appear to be potential leads, particularly for the treatment of neurological disorders. Several key molecular signaling pathways such as MAPK, CREB, Nrf2, NF-kB, JAK/STAT, and TLR are modulated by flavonoids. Inhibitory or stimulatory effects on these pathways by flavonoids significantly affect cellular functions by changing the phosphorylation state of target molecules and by modifying gene expressions (Mansuri et al., 2014; Pérez-Cano & Castell, 2016). These compounds also can modulate neuronal dysfunction through the activation of various metabolic pathways and are found to delay or prevent the onset of neurodegenerative diseases and provide neuroprotection (Makkar et al., 2020).

The antioxidant potential of flavonoids is associated with the molecular structure, and more precisely, with the location and total number of the –OH groups, the conjugation and resonance effects, the surrounding environment that modifies the thermodynamically favored antioxidant site, and the antioxidant mechanism for a compound (Zheng et al., 2019).

3.1 Alzheimer's disease

Somewhere five million individuals globally struggle with Alzheimer's disease, which makes it the most prevalent type of dementia. In Alzheimer there is the buildup of abnormal protein, increased production of amyloid beta production, and tau tangles which are responsible for causing impairment in synaptic loss, neuronal dysfunction and further leading to cognitive and memory dysfunction. The accumulation of amyloid beta leads to elevation of oxidative stress, neuroinflammation, and neurodegeneration (Williamson et al., 2018). Currently, there is no effective treatment available for the treatment and prevention of AD (Yiannopoulou & Papageorgiou, 2013). Nrf2 is a transcription factor binding to the antioxidant response element which controls oxidative response and regulates anti-inflammatory, antioxidant, and detoxifying genes (Ma, 2013). Nrf2 malfunction and its alternative localzation are common in most Neurodegenerative disorders including Amyotrophic lateral sclerosis, Alzheimer, and Parkinson. Nrf2 activation prevents our body from harmful stress by inhibiting inflammation, upregulating antioxidant defense pathways, and maintaining homeostasis (Villavicencio Tejo & Quintanilla, 2021). It was observed that Nrf2 was responsible for an array of critical roles to eliminate reactive oxygen species, preservation of redox homeostasis, and regulation of peripheral inflammatory responses thus has emerged as a promising therapeutic target and its seen flavonoids express their neuroprotective effects by inhibiting glutamate receptors and activating the Nrf2 (Moratilla-Rivera et al., 2023). However, some of the flavonoids have key potential for preventive and therapeutic effects.

3.1.1 Quercetin

Quercetin, a flavanol, possesses some pharmacological effects and has therapeutic effects. In various in vivo and in vitro models, it has been found that it protects neural cells by reducing oxidative stress and neurological inflammation. Some of the key properties that have a therapeutic effect include inhibition of amyloid beta aggregation, inhibition of tau phosphorylation, and restoration of acetylcholine levels through the inhibition of hydrolysis of acetylcholine by AChE enzyme (H. Khan et al., 2020).

3.1.2 Myricetin

Myricetin exhibits multifaceted neuroprotective effects by reducing the formation of amyloid beta 42 formations by blocking the conformational change, increasing the enzymatic activity of alpha-secretase and reducing that of beta-secretase countering the AD progression, suppressing neuroinflammation by inhibiting the activation of microglia and pro-inflammatory cytokines and mitigating oxidative stress (Taheri et al., 2020).

3.2 Parkinson’s Disorder

Parkinson's Disorder is the most common neurodegenerative disorder after Alzheimer’s disease and occurs due to the loss of dopaminergic neurons in the substantia nigra resulting in a decrease in the production of dopamine in the striatum and, subsequently, motor dysfunction (de Andrade Teles et al., 2018). The exact cause of Parkinson’s disease is unknown, but it

occurs due to the accumulation of abnormal proteins, oxidative stress, environmental toxins, and accumulation of alpha-synuclein (Dias et al., 2013; Jo et al., 2019). During neuroinflammatory diseases including Multiple sclerosis and Parkinson, it has been observed that JAK/STAT behaviour differently. JAK/STAT pathway initiates Dysregulation which generates oncological, metabolic, neurological, or other types of diseases to create disorders. The abnormalities are being regulated by JAK/STAT pathway inhibitors during the treatment of neurological disorders (Yan et al., 2018). Flavonoids possess potential in reduction of the phosphorylation of TYK2, STATs ; reduction in the translocation of nuclear STAT1 and STAT3; inhibition of COX-2 NO, iNOS, and proinflammatory cytokines such as TNF-alpha, interleukin 6 and 1beta in macrophages. Reactive oxygen species synthesis delivers the antioxidant properties of Flavonoids (S.-M. Qi et al., 2018).

3.2.1 Myricetin

Myricetin has shown some of the key neuroprotective effects that could be crucial against Parkinson's disorder are through inhibiting the activation of microglia and other proinflammatory cytokines, suppressing intracellular production of reactive oxygen species, inhibiting the formation of alpha-synuclein and destabilizing it, and re-establishing the mitochondrial transmembrane potential (Taheri et al., 2020).

3.2.2 Calycosin

Calycosin has been shown to have crucial neuroprotective effects against Parkinson's disorder by preserving and enhancing mitochondrial biogenesis, reducing oxidative stress-induced damage, and modulating and inhibiting TLR/NF-kB, and MAPK pathway (Yang et al., 2019).

3.3 Huntington’s Disease

Another neurodegenerative illness that causes involuntary movements, cognitive impairments , psychiatric disturbances, and dementia is Huntington's disease (HD). Genetically it is associated with the expansion of cytosine adenine guanine trinucleotide repeats in the Huntingtin gene (Hung et al., 2016; Velusamy et al., 2017). There is no treatment available for the management of HD (Frank, 2014).

3.3.1 Myricetin

Myricetin has been shown to interact with RNA, especially the CAG motif to inhibit the translation and destabilize the aggregates of mutant huntingtin protein, potentially reducing their cytotoxicity (E. Khan et al., 2018).

3.4 Multiple Sclerosis

In multiple sclerosis, chronic inflammation contributes to demyelination and neuronal damage. The activation of NF-kB plays a major role in many inflammatory diseases such as MS (Tak & Firestein, 2001). NF-kB is involved during functionating of the brain especially NDs (O’Neill & Kaltschmidt, 1997). Flavonoids play a significant role in providing the anti-inflammatory and antioxidant effect by hindering the NF-kB signaling pathway (Jones et al., 2012). Thus, NF-kB pathway targeting offers a novel treatment method and potential for neurodegenerative disorders.

S .No	Flavonoid	Disease	Action	References
1	*Apigenin*	Alzheimer disease	Inhibits the MAPK Pathway	(Zhao et al., 2013)
2	*Galangin*	Alzheimer disease	Docking studies revealed interaction with BChE active site forms H bonds and thus inhibits it	(Katalinić et al., 2010)
3	*Hesperidin*	Alzheimer disease	Reduces the increased levels of Tumour necrosis factor-alpha and Interleukin-1beta levels	(Ciftci et al., n.d.)
4	*Icariside II*	Alzheimer disease	Increase the antiapoptotic activity by the MAPK pathway	(He et al., 2018)
5	*Myricetin*	Alzheimer disease	Reduced the formation of amyloid b 42 formation by blocking the conformation change	(Fiori et al., 2012)
6	*Nobiletin*	Alzheimer disease	Reducing the level of soluble A1–40, reactive oxygen species, NF-kB ,and MAPKs	(G. Qi et al., 2019)
7	*Rutin*	Alzheimer disease	Decrease the amyloid beta induced ROS levels	(Moghbeli nejad et al., 2014)
8	*Quercetin*	Alzheimer disease	inhibition of amyloid beta aggregation and tau phosphorylation	(H. Khan et al., 2020)
9	*Fisetin*	Amyotrophic lateral sclerosis	Reduced ROS levels and activation of ERK	(Wang et al., 2018)
10	*Myricetin*	CNS	minimizes the accumulation of various abnormal proteins and gets rid of different harmful proteins connected to neurological disorders. stimulates Hsp70 molecular chaperone's physiological activities and lowers misfolded protein levels	(Joshi et al., 2019)
11	*Fisetin*	Huntington Disease	Activation of ERK	(Maher et al., 2011)
12	*Myricetin*	Huntington Disease	Interaction with RNA, especially CAG motif, and thus decrement in the huntingtin protein translation and	(E. Khan et al., 2018)

			sequestration that reduces cytotoxicity in HD.	
13	*Galangin*	Inflammation and free radicals	Decreased levels of pro-inflammatory cytokines, reduced levels of Nrf2 and HO-1, ERK and p38 pathways	(Y.-C. Huang et al., 2017)
14	*Quercetin*	Multiple sclerosis	Inhibits interleukin 12; STAT 3,4; TYK2	(Muthian & Bright, 2004)
15	*Baicalin*	Multiple sclerosis	Inhibits Th1,17; JAK/STAT; NF-kB	(Liu et al., 2014)
16	*Fisetin*	Neuroinflammation	Reduced the interleukin 6, NF-kB, TNF-alpha, NF-kB levels	(Chen et al., 2018)
17	*Nobiletin*	Neuroinflammation	Decrease in pro-inflammatory cytokine	(Cui et al., 2010)
18	*Quercetin*	Neuroinflammation	Inhibits the MAPK Pathway, NF-kB, NO	(Sun et al., 2015)
19	*Myricetin*	Parkinson Disease	Suppresses intracellular ROS production, re-establishes mitochondrial trans-membrane potential and also impedes MKK4 and JNK activation	(K. Zhang et al., 2011)
20	*Myricetin*	Parkinson Disease	Reduces the activation of microglia (neuroinflammation), and expression of pro-inflammatory mediators.	(B. Huang et al., 2018)
21	*Juglanin*	Parkinson Disease	Inhibits TLR4/NF-kB pathway	(F.-X. Zhang & Xu, 2018)
22	*Calycosin*	Parkinson Disease	Inhibits MAPK pathway	(Yang et al., 2019)

Table 2: Different Flavonoids and its therapeutic actions

4. Discussion and Future Directives

Flavonoid's ability to reach the target site (i.e. brain) in effective concentrations and exert its neuroprotective effects is hindered due to factors like low solubility, fast metabolic rate, and limited absorption in the gastrointestinal tract. Adsorption of flavonoids and delivery to specific target sites has been a challenge and needs innovative novel strategies to overcome. Delivery

systems such as nanoparticles, liposomes, or microemulsions could be used to deliver at the target site. These novel delivery systems possess the potential to improve the stability and solubility of flavonoids, improving in absorption and bioavailability of flavonoids having neuroprotective effects to cross the blood-brain barrier and reach the target site.

Complex structures and diversity in flavonoids pose challenges in understanding precise mechanisms of action and determining in optimal dosage to have therapeutic effects and metabolic activity. Flavonoids' distinct structural properties, bioactivities, and mechanisms make it difficult to find specific flavonoids that would modulate, stimulate, and interact with the abnormal proteins at the target site and cause a neuroprotective effect. This specificity and selective limitations could be overcome by conducting structure-activity relationship studies to know the peculiarities of flavonoids for their neuroprotective effects. Advanced computational tools and molecular docking technologies must be deployed in predicting the flavonoids' interactions with biological targets and pathways, these could further guide drug development efforts.

Flavonoid interactions with other medications or bio compounds are less known and could be affecting their efficacy or causing adverse effects. These potential side effects and interactions with other bio compounds are crucial to know as would have deteriorative effects on individuals having multiple diseases or some underlying health conditions. Preclinical and clinical studies could be conducted to find out the potential interaction of flavonoids with medications. Devising specialized systems to monitor the adverse effects related to flavonoid therapeutic action, to ensure timely detection and antidote for the side effects.

To summarize, flavonoids show immense promise in treating Neurological Disorders, but several limitations warrant further exploration. Firstly, leveraging tools in drug discovery, such as computer-aided design and high-throughput screening, can enhance understanding of flavonoids' safety, biochemical activity, and targeted mechanisms of action. Secondly, investigating the modifications in structure of these compounds is crucial for optimizing their efficacy. Lastly, conducting scientific clinical trials and comprehensive evaluations of potential side effects are essential to determine both the specificity and safety of flavonoids in treating Neurological Disorders. By continuing to advance our understanding, overcoming barriers, and translating research findings into clinical practice, we can harness flavonoids’ full potential in promoting brain health and combating neurological diseases.

5. Reference

Alam, W., Khan, H., Shah, M. A., Cauli, O., & Saso, L. (2020). Kaempferol as a Dietary Anti-Inflammatory Agent: Current Therapeutic Standing. *Molecules*, *25*(18). https://doi.org/10.3390/molecules25184073

Ayaz, M., Sadiq, A., Junaid, M., Ullah, F., Ovais, M., Ullah, I., Ahmed, J., & Shahid, M. (2019). Flavonoids as Prospective Neuroprotectants and Their Therapeutic Propensity in Aging Associated Neurological Disorders. *Frontiers in Aging Neuroscience*, *11*, 155. https://doi.org/10.3389/fnagi.2019.00155

Baião, D., Defreitas, C., Gomes, L., Silva, D., Correa, A. C. N. T. F., Pereira, P., Delaguila, E., & Paschoalin, V. M. F. (2017). Polyphenols from root, Tubercles and grains cropped

in brazil: Chemical and nutritional characterization and their effects on human health and diseases. *Nutrients*, *9*(9). https://doi.org/10.3390/nu9091044

Benavente-García, O., & Castillo, J. (2008). Update on Uses and Properties of Citrus Flavonoids: New Findings in Anticancer, Cardiovascular, and Anti-inflammatory Activity. *Journal of Agricultural and Food Chemistry*, *56*(15), 6185–6205. https://doi.org/10.1021/jf8006568

Brunetti, C., Di Ferdinando, M., Fini, A., Pollastri, S., & Tattini, M. (2013). Flavonoids as antioxidants and developmental regulators: relative significance in plants and humans. *International Journal of Molecular Sciences*, *14*(2), 3540–3555. https://doi.org/10.3390/ijms14023540

Buschmann, H. (2000). Accumulation of Hydroxycoumarins During Post-harvest Deterioration of Tuberous Roots of Cassava (Manihot esculenta Crantz). *Annals of Botany*, *86*(6), 1153–1160. https://doi.org/10.1006/anbo.2000.1285

Chen, C., Yao, L., Cui, J., & Liu, B. (2018). Fisetin Protects against Intracerebral Hemorrhage-Induced Neuroinflammation in Aged Mice. *Cerebrovascular Diseases (Basel, Switzerland)*, *45*(3–4), 154–161. https://doi.org/10.1159/000488117

Ciftci, O., Ozcan, C., Kamisli, • Ozden, Cetin, • Aslı, Nese Basak, •, & Aytac, B. (n.d.). *Hesperidin, a Citrus Flavonoid, Has the Ameliorative Effects Against Experimental Autoimmune Encephalomyelitis (EAE) in a C57BL/J6 Mouse Model*. https://doi.org/10.1007/s11064-015-1571-8

Ciumărnean, L., Milaciu, M. V., Runcan, O., Vesa, Ștefan C., Răchișan, A. L., Negrean, V., Perné, M.-G., Donca, V. I., Alexescu, T.-G., Para, I., & Dogaru, G. (2020). The Effects of Flavonoids in Cardiovascular Diseases. *Molecules (Basel, Switzerland)*, *25*(18). https://doi.org/10.3390/molecules25184320

Cui, Y., Wu, J., Jung, S.-C., Park, D.-B., Maeng, Y.-H., Hong, J. Y., Kim, S.-J., Lee, S.-R., Kim, S.-J., Kim, S. J., & Eun, S.-Y. (2010). Anti-neuroinflammatory activity of nobiletin on suppression of microglial activation. *Biological & Pharmaceutical Bulletin*, *33*(11), 1814–1821. https://doi.org/10.1248/bpb.33.1814

Dabeek, W. M., & Marra, M. V. (2019). Dietary Quercetin and Kaempferol: Bioavailability and Potential Cardiovascular-Related Bioactivity in Humans. *Nutrients*, *11*(10). https://doi.org/10.3390/nu11102288

de Andrade Teles, R. B., Diniz, T. C., Costa Pinto, T. C., de Oliveira Júnior, R. G., Gama E Silva, M., de Lavor, É. M., Fernandes, A. W. C., de Oliveira, A. P., de Almeida Ribeiro, F. P. R., da Silva, A. A. M., Cavalcante, T. C. F., Quintans Júnior, L. J., & da Silva Almeida, J. R. G. (2018). Flavonoids as Therapeutic Agents in Alzheimer's and Parkinson's Diseases: A Systematic Review of Preclinical Evidences. *Oxidative Medicine and Cellular Longevity*, *2018*, 7043213. https://doi.org/10.1155/2018/7043213

Dias, V., Junn, E., & Mouradian, M. M. (2013). The role of oxidative stress in Parkinson's disease. *Journal of Parkinson's Disease*, *3*(4), 461–491. https://doi.org/10.3233/JPD-130230

Egert, S., & Rimbach, G. (2011). Which sources of flavonoids: Complex diets or dietary supplements? In *Advances in Nutrition* (Vol. 2, Issue 1, pp. 8–14). https://doi.org/10.3945/an.110.000026

Fiori, J., Naldi, M., Bartolini, M., & Andrisano, V. (2012). Disclosure of a fundamental clue for the elucidation of the myricetin mechanism of action as amyloid aggregation inhibitor by mass spectrometry. *Electrophoresis*, *33*(22), 3380–3386. https://doi.org/10.1002/elps.201200186

Frank, S. (2014). Treatment of Huntington's disease. *Neurotherapeutics : The Journal of the American Society for Experimental NeuroTherapeutics*, *11*(1), 153–160. https://doi.org/10.1007/s13311-013-0244-z

Häkkinen, S. H., Kärenlampi, S. O., Heinonen, I. M., Mykkänen, H. M., & Törrönen, A. R. (1999). Content of the Flavonols Quercetin, Myricetin, and Kaempferol in 25 Edible Berries. *Journal of Agricultural and Food Chemistry*, *47*(6), 2274–2279. https://doi.org/10.1021/jf9811065

Hamsalakshmi, Alex, A. M., Arehally Marappa, M., Joghee, S., & Chidambaram, S. B. (2022). Therapeutic benefits of flavonoids against neuroinflammation: a systematic review. In *Inflammopharmacology* (Vol. 30, Issue 1, pp. 111–136). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s10787-021-00895-8

He, L., Deng, Y., Gao, J., Zeng, L., & Gong, Q. (2018). Icariside II ameliorates ibotenic acid-induced cognitive impairment and apoptotic response via modulation of MAPK pathway in rats. *Phytomedicine : International Journal of Phytotherapy and Phytopharmacology*, *41*, 74–81. https://doi.org/10.1016/j.phymed.2018.01.025

HERRMANN, K. (1976). Flavonols and flavones in food plants: a review†. *International Journal of Food Science & Technology*, *11*(5), 433–448. https://doi.org/10.1111/j.1365-2621.1976.tb00743.x

Huang, B., Liu, J., Ma, D., Chen, G., Wang, W., & Fu, S. (2018). Myricetin prevents dopaminergic neurons from undergoing neuroinflammation-mediated degeneration in a lipopolysaccharide-induced Parkinson's disease model. *Journal of Functional Foods*, *45*, 452–461. https://doi.org/10.1016/j.jff.2018.04.018

Huang, Y.-C., Tsai, M.-S., Hsieh, P.-C., Shih, J.-H., Wang, T.-S., Wang, Y.-C., Lin, T.-H., & Wang, S.-H. (2017). Galangin ameliorates cisplatin-induced nephrotoxicity by attenuating oxidative stress, inflammation and cell death in mice through inhibition of ERK and NF-kappaB signaling. *Toxicology and Applied Pharmacology*, *329*, 128–139. https://doi.org/10.1016/j.taap.2017.05.034

Hung, K.-C., Huang, H.-J., Wang, Y.-T., & Lin, A. M.-Y. (2016). Baicalein attenuates α-synuclein aggregation, inflammasome activation and autophagy in the MPP+-treated

nigrostriatal dopaminergic system in vivo. *Journal of Ethnopharmacology*, *194*, 522–529. https://doi.org/10.1016/j.jep.2016.10.040

Jo, M. G., Ikram, M., Jo, M. H., Yoo, L., Chung, K. C., Nah, S.-Y., Hwang, H., Rhim, H., & Kim, M. O. (2019). Gintonin Mitigates MPTP-Induced Loss of Nigrostriatal Dopaminergic Neurons and Accumulation of α-Synuclein via the Nrf2/HO-1 Pathway. *Molecular Neurobiology*, *56*(1), 39–55. https://doi.org/10.1007/s12035-018-1020-1

Jones, Q. R. D., Warford, J., Rupasinghe, H. P. V., & Robertson, G. S. (2012). Target-based selection of flavonoids for neurodegenerative disorders. *Trends in Pharmacological Sciences*, *33*(11), 602–610. https://doi.org/10.1016/j.tips.2012.08.002

Joshi, V., Mishra, R., Upadhyay, A., Amanullah, A., Poluri, K. M., Singh, S., Kumar, A., & Mishra, A. (2019). Polyphenolic flavonoid (Myricetin) upregulated proteasomal degradation mechanisms: Eliminates neurodegenerative proteins aggregation. *Journal of Cellular Physiology*, *234*(11), 20900–20914. https://doi.org/10.1002/jcp.28695

Katalinić, M., Rusak, G., Domaćinović Barović, J., Sinko, G., Jelić, D., Antolović, R., & Kovarik, Z. (2010). Structural aspects of flavonoids as inhibitors of human butyrylcholinesterase. *European Journal of Medicinal Chemistry*, *45*(1), 186–192. https://doi.org/10.1016/j.ejmech.2009.09.041

Khan, E., Tawani, A., Mishra, S. K., Verma, A. K., Upadhyay, A., Kumar, M., Sandhir, R., Mishra, A., & Kumar, A. (2018). Myricetin Reduces Toxic Level of CAG Repeats RNA in Huntington's Disease (HD) and Spino Cerebellar Ataxia (SCAs). *ACS Chemical Biology*, *13*(1), 180–188. https://doi.org/10.1021/acschembio.7b00699

Khan, H., Ullah, H., Aschner, M., Cheang, W. S., & Akkol, E. K. (2020). Neuroprotective effects of quercetin in alzheimer's disease. In *Biomolecules* (Vol. 10, Issue 1). MDPI AG. https://doi.org/10.3390/biom10010059

Kopustinskiene, D. M., Jakstas, V., Savickas, A., & Bernatoniene, J. (2020). Flavonoids as Anticancer Agents. *Nutrients*, *12*(2). https://doi.org/10.3390/nu12020457

Liu, Y., Holdbrooks, A. T., De Sarno, P., Rowse, A. L., Yanagisawa, L. L., McFarland, B. C., Harrington, L. E., Raman, C., Sabbaj, S., Benveniste, E. N., & Qin, H. (2014). Therapeutic efficacy of suppressing the Jak/STAT pathway in multiple models of experimental autoimmune encephalomyelitis. *Journal of Immunology (Baltimore, Md. : 1950)*, *192*(1), 59–72. https://doi.org/10.4049/jimmunol.1301513

Lobo, V., Patil, A., Phatak, A., & Chandra, N. (2010). Free radicals, antioxidants and functional foods: Impact on human health. *Pharmacognosy Reviews*, *4*(8), 118–126. https://doi.org/10.4103/0973-7847.70902

Ma, Q. (2013). Role of nrf2 in oxidative stress and toxicity. *Annual Review of Pharmacology and Toxicology*, *53*, 401–426. https://doi.org/10.1146/annurev-pharmtox-011112-140320

Maher, P., Dargusch, R., Bodai, L., Gerard, P. E., Purcell, J. M., & Marsh, J. L. (2011). ERK activation by the polyphenols fisetin and resveratrol provides neuroprotection in

multiple models of Huntington's disease. *Human Molecular Genetics*, *20*(2), 261–270. https://doi.org/10.1093/hmg/ddq460

Makkar, R., Behl, T., Bungau, S., Zengin, G., Mehta, V., Kumar, A., Uddin, M. S., Ashraf, G. M., Abdel-Daim, M. M., Arora, S., & Oancea, R. (2020). Nutraceuticals in Neurological Disorders. *International Journal of Molecular Sciences*, *21*(12). https://doi.org/10.3390/ijms21124424

Malenčić, D., Cvejić, J., & Miladinović, J. (2012). Polyphenol content and antioxidant properties of colored soybean seeds from central Europe. *Journal of Medicinal Food*, *15*(1), 89–95. https://doi.org/10.1089/jmf.2010.0329

Mansuri, M. L., Parihar, P., Solanki, I., & Parihar, M. S. (2014). Flavonoids in modulation of cell survival signalling pathways. *Genes & Nutrition*, *9*(3), 400. https://doi.org/10.1007/s12263-014-0400-z

Mitoma, H., & Manto, M. (Eds.). (2019). *Neuroimmune Diseases*. Springer International Publishing. https://doi.org/10.1007/978-3-030-19515-1

Moghbelinejad, S., Nassiri-Asl, M., Farivar, T. N., Abbasi, E., Sheikhi, M., Taghiloo, M., Farsad, F., Samimi, A., & Hajiali, F. (2014). Rutin activates the MAPK pathway and BDNF gene expression on beta-amyloid induced neurotoxicity in rats. *Toxicology Letters*, *224*(1), 108–113. https://doi.org/10.1016/j.toxlet.2013.10.010

Moratilla-Rivera, I., Sánchez, M., Valdés-González, J. A., & Gómez-Serranillos, M. P. (2023). Natural Products as Modulators of Nrf2 Signaling Pathway in Neuroprotection. *International Journal of Molecular Sciences*, *24*(4). https://doi.org/10.3390/ijms24043748

Muthian, G., & Bright, J. J. (2004). Quercetin, a flavonoid phytoestrogen, ameliorates experimental allergic encephalomyelitis by blocking IL-12 signaling through JAK-STAT pathway in T lymphocyte. *Journal of Clinical Immunology*, *24*(5), 542–552. https://doi.org/10.1023/B:JOCI.0000040925.55682.a5

O'Neill, L. A., & Kaltschmidt, C. (1997). NF-kappa B: a crucial transcription factor for glial and neuronal cell function. *Trends in Neurosciences*, *20*(6), 252–258. https://doi.org/10.1016/s0166-2236(96)01035-1

Panche, A. N., Diwan, A. D., & Chandra, S. R. (2016a). Flavonoids: an overview. *Journal of Nutritional Science*, *5*, e47. https://doi.org/10.1017/jns.2016.41

Panche, A. N., Diwan, A. D., & Chandra, S. R. (2016b). Flavonoids: An overview. In *Journal of Nutritional Science* (Vol. 5). Cambridge University Press. https://doi.org/10.1017/jns.2016.41

Pérez-Cano, F. J., & Castell, M. (2016). Flavonoids, Inflammation and Immune System. *Nutrients*, *8*(10). https://doi.org/10.3390/nu8100659

Qi, G., Mi, Y., Fan, R., Li, R., Liu, Z., & Liu, X. (2019). Nobiletin Protects against Systemic Inflammation-Stimulated Memory Impairment via MAPK and NF-κB Signaling

Pathways. *Journal of Agricultural and Food Chemistry*, *67*(18), 5122–5134. https://doi.org/10.1021/acs.jafc.9b00133

Qi, S.-M., Li, Q., Jiang, Q., Qi, Z.-L., & Zhang, Y. (2018). [Chrysin inhibits lipopolysaccharide-induced inflammatory responses of macrophages via JAK-STATs signaling pathway]. *Nan Fang Yi Ke Da Xue Xue Bao = Journal of Southern Medical University*, *38*(3), 243–250. https://doi.org/10.3969/j.issn.1673-4254.2018.03.02

Qin, J., Ma, Z., Chen, X., & Shu, S. (2023). Microglia activation in central nervous system disorders: A review of recent mechanistic investigations and development efforts. *Frontiers in Neurology*, *14*, 1103416. https://doi.org/10.3389/fneur.2023.1103416

Rakha, A., Umar, N., Rabail, R., Butt, M. S., Kieliszek, M., Hassoun, A., & Aadil, R. M. (2022). Anti-inflammatory and anti-allergic potential of dietary flavonoids: A review. *Biomedicine & Pharmacotherapy*, *156*, 113945. https://doi.org/10.1016/J.BIOPHA.2022.113945

Sahu, B. D., Kalvala, A. K., Koneru, M., Mahesh Kumar, J., Kuncha, M., Rachamalla, S. S., & Sistla, R. (2014). Ameliorative Effect of Fisetin on Cisplatin-Induced Nephrotoxicity in Rats via Modulation of NF-κB Activation and Antioxidant Defence. *PLoS ONE*, *9*(9), e105070. https://doi.org/10.1371/journal.pone.0105070

Sun, G. Y., Chen, Z., Jasmer, K. J., Chuang, D. Y., Gu, Z., Hannink, M., & Simonyi, A. (2015). Quercetin Attenuates Inflammatory Responses in BV-2 Microglial Cells: Role of MAPKs on the Nrf2 Pathway and Induction of Heme Oxygenase-1. *PloS One*, *10*(10), e0141509. https://doi.org/10.1371/journal.pone.0141509

Taheri, Y., Suleria, H. A. R., Martins, N., Sytar, O., Beyatli, A., Yeskaliyeva, B., Seitimova, G., Salehi, B., Semwal, P., Painuli, S., Kumar, A., Azzini, E., Martorell, M., Setzer, W. N., Maroyi, A., & Sharifi-Rad, J. (2020). Myricetin bioactive effects: Moving from preclinical evidence to potential clinical applications. In *BMC Complementary Medicine and Therapies* (Vol. 20, Issue 1). BioMed Central. https://doi.org/10.1186/s12906-020-03033-z

Tak, P. P., & Firestein, G. S. (2001). NF-kappaB: a key role in inflammatory diseases. *The Journal of Clinical Investigation*, *107*(1), 7–11. https://doi.org/10.1172/JCI11830

Tchessalova, D., Posillico, C. K., & Tronson, N. C. (2018). Neuroimmune Activation Drives Multiple Brain States. *Frontiers in Systems Neuroscience*, *12*, 39. https://doi.org/10.3389/fnsys.2018.00039

Uarrota, V. G., Nunes, E. da C., Peruch, L. A. M., Neubert, E. de O., Coelho, B., Moresco, R., Domínguez, M. G., Sánchez, T., Meléndez, J. L. L., Dufour, D., Ceballos, H., Becerra Lopez-Lavalle, L. A., Hershey, C., Rocha, M., & Maraschin, M. (2016). Toward better understanding of postharvest deterioration: biochemical changes in stored cassava (Manihot esculenta Crantz) roots. *Food Science & Nutrition*, *4*(3), 409–422. https://doi.org/10.1002/fsn3.303

Ullah, A., Munir, S., Badshah, S. L., Khan, N., Ghani, L., Poulson, B. G., Emwas, A. H., & Jaremko, M. (2020). Important flavonoids and their role as a therapeutic agent. In *Molecules* (Vol. 25, Issue 22). MDPI AG. https://doi.org/10.3390/molecules25225243

Vaou, N., Stavropoulou, E., Voidarou, C., Tsigalou, C., & Bezirtzoglou, E. (2021). Towards Advances in Medicinal Plant Antimicrobial Activity: A Review Study on Challenges and Future Perspectives. *Microorganisms*, *9*(10). https://doi.org/10.3390/microorganisms9102041

Velusamy, T., Panneerselvam, A. S., Purushottam, M., Anusuyadevi, M., Pal, P. K., Jain, S., Essa, M. M., Guillemin, G. J., & Kandasamy, M. (2017). Protective Effect of Antioxidants on Neuronal Dysfunction and Plasticity in Huntington's Disease. *Oxidative Medicine and Cellular Longevity*, *2017*, 3279061. https://doi.org/10.1155/2017/3279061

Villavicencio Tejo, F., & Quintanilla, R. A. (2021). Contribution of the Nrf2 Pathway on Oxidative Damage and Mitochondrial Failure in Parkinson and Alzheimer's Disease. *Antioxidants (Basel, Switzerland)*, *10*(7). https://doi.org/10.3390/antiox10071069

Wang, T. H., Wang, S. Y., Wang, X. D., Jiang, H. Q., Yang, Y. Q., Wang, Y., Cheng, J. L., Zhang, C. T., Liang, W. W., & Feng, H. L. (2018). Fisetin Exerts Antioxidant and Neuroprotective Effects in Multiple Mutant hSOD1 Models of Amyotrophic Lateral Sclerosis by Activating ERK. *Neuroscience*, *379*, 152–166. https://doi.org/10.1016/j.neuroscience.2018.03.008

Williamson, G., Kay, C. D., & Crozier, A. (2018). The Bioavailability, Transport, and Bioactivity of Dietary Flavonoids: A Review from a Historical Perspective. *Comprehensive Reviews in Food Science and Food Safety*, *17*(5), 1054–1112. https://doi.org/10.1111/1541-4337.12351

Yan, Z., Gibson, S. A., Buckley, J. A., Qin, H., & Benveniste, E. N. (2018). Role of the JAK/STAT signaling pathway in regulation of innate immunity in neuroinflammatory diseases. *Clinical Immunology (Orlando, Fla.)*, *189*, 4–13. https://doi.org/10.1016/j.clim.2016.09.014

Yang, J., Jia, M., Zhang, X., & Wang, P. (2019). Calycosin attenuates MPTP-induced Parkinson's disease by suppressing the activation of TLR/NF-κB and MAPK pathways. *Phytotherapy Research : PTR*, *33*(2), 309–318. https://doi.org/10.1002/ptr.6221

Yiannopoulou, K. G., & Papageorgiou, S. G. (2013). Current and future treatments for Alzheimer's disease. *Therapeutic Advances in Neurological Disorders*, *6*(1), 19–33. https://doi.org/10.1177/1756285612461679

Youdim, K. A., Dobbie, M. S., Kuhnle, G., Proteggente, A. R., Abbott, N. J., & Rice-Evans, C. (2003). Interaction between flavonoids and the blood-brain barrier: in vitro studies. *Journal of Neurochemistry*, *85*(1), 180–192. https://doi.org/10.1046/j.1471-4159.2003.01652.x

Zhang, F.-X., & Xu, R.-S. (2018). Juglanin ameliorates LPS-induced neuroinflammation in animal models of Parkinson's disease and cell culture via inactivating TLR4/NF-κB

pathway. *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie*, *97*, 1011–1019. https://doi.org/10.1016/j.biopha.2017.08.132

Zhang, K., Ma, Z., Wang, J., Xie, A., & Xie, J. (2011). Myricetin attenuated MPP(+)-induced cytotoxicity by anti-oxidation and inhibition of MKK4 and JNK activation in MES23.5 cells. *Neuropharmacology*, *61*(1–2), 329–335. https://doi.org/10.1016/j.neuropharm.2011.04.021

Zhang, W., Xiao, D., Mao, Q., & Xia, H. (2023). Role of neuroinflammation in neurodegeneration development. *Signal Transduction and Targeted Therapy*, *8*(1), 267. https://doi.org/10.1038/s41392-023-01486-5

Zhao, L., Wang, J.-L., Wang, Y.-R., & Fa, X.-Z. (2013). Apigenin attenuates copper-mediated β-amyloid neurotoxicity through antioxidation, mitochondrion protection and MAPK signal inactivation in an AD cell model. *Brain Research*, *1492*, 33–45. https://doi.org/10.1016/j.brainres.2012.11.019

Zheng, Y.-Z., Deng, G., Guo, R., Fu, Z.-M., & Chen, D.-F. (2019). The influence of the H5⋯OC4 intramolecular hydrogen-bond (IHB) on the antioxidative activity of flavonoid. *Phytochemistry*, *160*, 19–24. https://doi.org/10.1016/j.phytochem.2019.01.011

Zhu, M. L., Zhang, P. M., Jiang, M., Yu, S. W., & Wang, L. (2020). Myricetin induces apoptosis and autophagy by inhibiting PI3K/Akt/mTOR signalling in human colon cancer cells. *BMC Complementary Medicine and Therapies*, *20*(1), 209. https://doi.org/10.1186/s12906-020-02965-w

Journal of Polymers and the Environment
https://doi.org/10.1007/s10924-024-03317-2

ORIGINAL PAPER

Fructose-Based Non-Isocyanate Polyurethane/Poly (Sodium Acrylate) Hydrogels: Design, Synthesis and Environmental Applications

Pooja Singh[1] · **Raminder Kaur**[1]

Accepted: 8 May 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

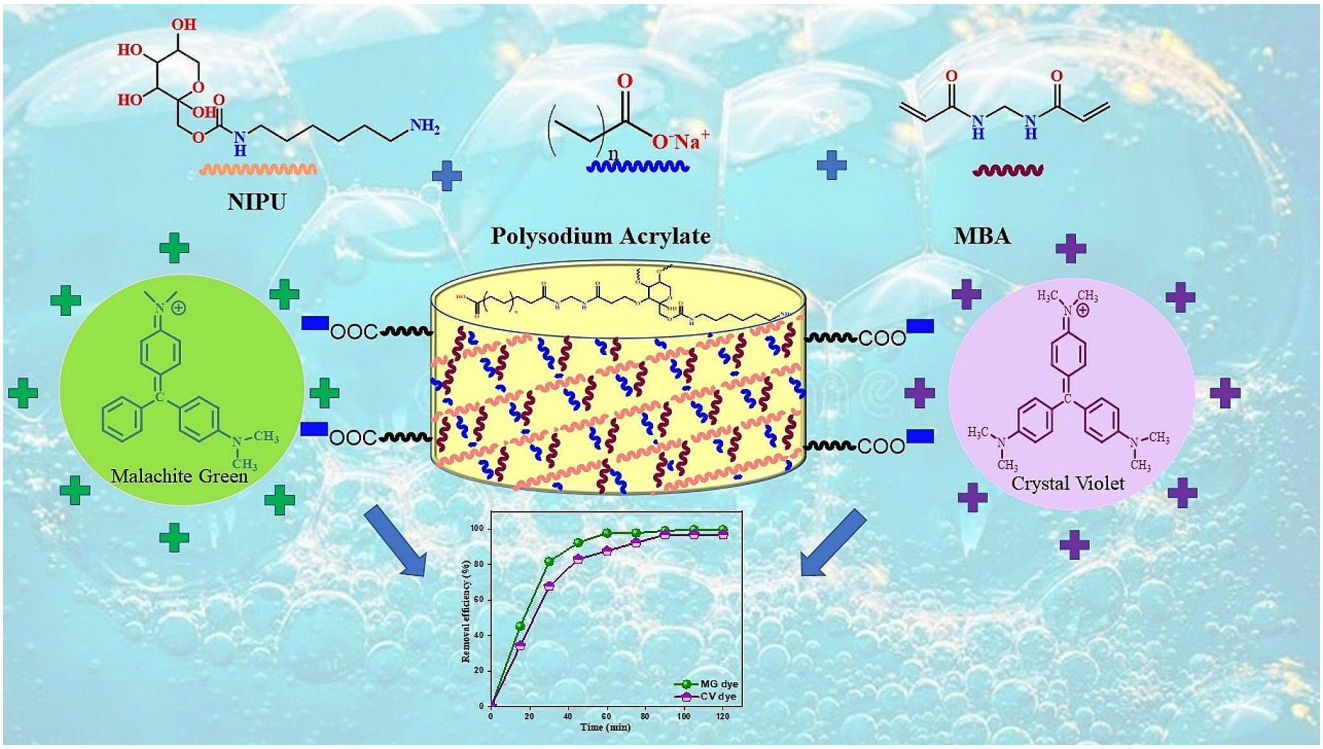
Conventional methods for synthesising polyurethane hydrogels encompass toxic isocyanates and organic solvents, limiting their eco-friendliness and ease of synthesis. In response, this study introduces an innovative approach to synthesising fructose-based non-isocyanate polyurethane (NIPU) hydrogel (FNHG), eliminating the need for isocyanates. Initially, fructose-based NIPU (FNPU) was synthesised using dimethyl carbonate and hexamethylene diamine under mild reaction conditions, paving the way for a greener polyurethane variant. Subsequently, a free radical polymerization technique was employed in an aqueous medium. This process allowed for the integration of poly(sodium acrylate), and N, N-methylene bisacrylamide, leading to to the development of FNHG. Remarkably short gelation time of just 30 min at 60 °C was achieved, signifying a significant advancement in the synthesis process. The synthesized NIPU-based hydrogels exhibited outstanding efficiency in the removal of crystal violet (CV) and malachite green (MG) dyes from aqueous media. With an impressive removal efficiency of 96.87% for CV and an astounding 99.8% for MG, these hydrogels demonstrated high effectiveness in remediation efforts. The study's novelty lies in both the synthesis methodology, utilising FNPU, and the exceptional efficiency exhibited by these hydrogels in eliminating diverse dyes from contaminated water. Furthermore, the structure of FNPU was confirmed using FTIR and ^{1}H NMR spectroscopy, adding robustness to our findings. This research not only presents a solution to the limitations of traditional polyurethane synthesis but also demonstrates the potential of fructose-based NIPU hydrogels (FNHG) as eco-friendly and efficient agents for water purification.

✉ Raminder Kaur
raminderkaur@dce.ac.in; raminderkaur141983@gmail.com

[1] Discipline of Polymer Science and Chemical Technology, Department of Applied Chemistry, Delhi Technological University (Formerly Delhi College of Engineering), Delhi 110042, India

Published online: 28 May 2024

Graphical Abstract



Keywords Isocyanate · NIPU · Fructose · Hydrogel · Water Remediation

Introduction

PUs are versatile polymers with exceptional properties, making them ideal for use as engineering materials in various industrial applications ranging from lightweight construction materials, coatings in marine and aerospace industries, rollers, gaskets, seals, conveyor belts, flexible or rigid cellular foams, insulating components for energy conservation, and even as ideal candidate for biomedical industries due to their biocompatibility. However, the synthesis of PUs involves toxic isocyanates, which pose a risk to human health and the environment. Isocyanate exposure can cause health issues like dermatitis and asthma [1]. Some isocyanates have been linked to cancer, mutation, and reproductive harm [2]. In addition, their rapid hydrolysis is a challenge in humid environments. Subsequently, the one-pot one-step synthesis of PU hydrogels in water is challenging due to the rapid hydrolysis of isocyanates [3]. They are usually synthesized in bulk or organic solvents, followed by swelling in water [4]. Consequently, the development of alternative and sustainable approaches to polyurethane synthesis has become a significant area of research.

Recently, the development of non-isocyanate polyurethanes (NIPU) has gained significant consideration as a sustainable alternative to conventionally synthesized PU. NIPUs can be easily fabricated from renewable resources and do not require the use of toxic isocyanates, making them environmentally-friendly and safer for production as well as for application [5]. One of the most promising alternative approaches for the isocyanate-free synthesis of PUs proceeds via the formation of polyhydroxyurethanes via the reaction of hydroxyl group-containing precursors with cyclic carbonates, and then reacting the resulting product with diamines [6]. Nearly all these reports included the reaction of synthetic cyclic or dicyclic carbonates with aliphatic diamines (via aminolysis), but only a handful of these studies reported the use of bio-based carbonates and diamines because of the complexities in the synthesis process and poor results [7]. Unfortunately, there were still issues observed, such as harsh reaction conditions such as high temperature and the need for a catalyst, along with the use of a highly toxic reagent (epichlorohydrin) and high-pressure conditions [8]. Potential solutions to this hindrance in the synthesis of bio-based cyclic carbonates are the conversion of epoxidized vegetable oils or the utilization of glycerin carbonate intermediates. An even better strategy for the synthesis of NIPU is the use of simpler, non-cyclic dimethyl carbonate for the initial stage of carbonation, since

it is neither irritating nor mutagenic and avoids the need for ancillary steps of carbonate cyclisation [9]. NIPU materials synthesised by these sustainable routes find application in the coatings and polymer industries, similar to PU. Thermosetting coatings, UV-stable coatings, monolithic flooring, etc. could all be applied using NIPU technologies [10]. NIPUs offer distinct advantages, such as decreased toxicity and environmental impact compared to traditional PUs, yet their potential in environmental applications remains largely unexplored [11, 12]. This gap in the literature highlights the need for focused investigations into the environmental applications of NIPUs, which represent a promising area for multidisciplinary research and the formation of sustainable materials with reduced ecological impact. Among the renewable precursors explored for NIPU synthesis, fructose can emerge as a promising building block due to its abundance, biodegradability, and wide availability.

A significant portion of the population faces challenges such as malnutrition, illnesses, and, in extreme cases, fatalities, primarily due to the unavailability of safe drinking water. Water contamination stands as a pressing global challenge, posing risks to ecosystems, public health, and the environment [13–15]. The multifaceted properties of polyurethane serve as the foundation for its effectiveness in eliminating a broad range of pollutants, including oil, heavy metals, nitrates, phenols, phthalates, and dyes [16–19]. Engineered with high surface area and porosity, it effectively adsorbs pollutants such as heavy metals and dyes. Its chemical stability and robust mechanical strength ensure sustained performance when exposed to different pollutants, while the ability to customize surface chemistry enhances its affinity for specific contaminants like nitrates and phthalates [20]. These characteristics are enhanced when incorporated into hydrogels, attributed to a physically well-defined three-dimensional porous architectures and an increased surface area-to-volume ratio, thus, leading to an improved adsorption capacity [21–23]. The ability to construct a flexible network of polymer chains due to the hydrophilic nature of hydrogel adsorbents enables solutes to readily permeate the network and form stable complexes with the functional groups.

Limited research is available on the utilization of bio-based PU hydrogels for environmental applications. The first polyhydroxyurethane (PHU) hydrogels were based on a two-step process involving first the preparation of PHU in the bulk by polyaddition of a hydrophilic poly(ethylene glycol) dicyclic carbonate (PEG-diCC) to a diamine and a crosslinker; followed by their swelling in water [24]. Another example of PHU hydrogels prepared in water consisted of the aminolysis of PEG-diCC with polyethyleneimine at room temperature [25]. Fructose offers a renewable, eco-friendly source for polymer synthesis, reducing environmental impact and enabling the development of sustainable, biocompatible materials. This facile study aimed to synthesize fructose-based NIPU by using dimethyl carbonate and hexamethylene diamine solution as precursors employing mild reaction conditions, and to study their application in the fabrication of hydrogels, having potential application in dye removal. The hydrogels were prepared (with poly(sodium acrylate)e and N, N-methylene bisacrylamide) through a free radical polymerization method in an aqueous medium. In the process of hydrogel synthesis, sodium acrylate was synthesized by reacting acrylic acid (AA) with sodium hydroxide (NaOH). By harnessing the potential of fructose-based NIPUs and their hydrogels, this research offers valuable insights into the development of effective and eco-friendly solutions for dye removal and highlights the significance of NIPUs in addressing water pollution challenges.

Materials and Methods

Materials

Non-isocyanate polyurethane was synthesized by employing commercially available hexamethylene diamine (HMDA), which was obtained from Sigma Aldrich. D-Fructose, potassium persulphate (KPS), dimethyl carbonate (DMC), crystal violet dye (CV), and malachite green dye (MG) were purchased from Thermo Fisher Scientific. Acrylic acid (AA) and sodium hydroxide (NaOH) were acquired from CDH Pvt. Ltd whereas N, N'-methylene bis (acrylamide) (MBA) was procured from Merck. All chemicals and solvents used in the analytical measurement trials were of analytical grade and utilized exactly as received.

Preparation of the Fructose-Based NIPU (FNPU)

In a round bottom flask, 9.99 g of fructose ($C_6H_{12}O_6$), 6.71 g of dimethyl carbonate (CH_3OCOCH_3), and 8.33 g of water (H_2O) were added. The reaction mixture was then subjected to reflux at 80°C for 60 min. Subsequently, 1 g of a 70% solution of hexamethyl diamine (HMDA) was introduced into the reaction mixture, followed by further heating at 90°C for another 60 min. The resulting solution, upon cooling to room temperature, yielded the desired product, non-isocyanate polyurethane (NIPU), labelled as FNPU. A schematic representation and detailed reaction mechanism of the synthesis of NIPU are shown in Figs. 1 and 2, respectively.

NaOH was used to neutralise AA (2 mL), resulting in sodium acrylate (SA), which was subsequently polymerized to produce the hydrogel. Subsequently, the aqueous

solutions of each FNPU and SA were mixed and 0.05 g of MBA, and 0.05 g of KPS were added, with continuous stirring. The resulting solution was then transferred into a test tube and heated for 30 minutes in a water bath at a controlled temperature of 60 °C. The gelation process resulted in the formation of the FNHG, which was carefully removed and then further divided into tiny pellets. The hydrogel pellets were subsequently dried in an oven for 48 h at 50 °C, to remove the moisture. A schematic representation of the synthesis of NIPU-based hydrogel (FNHG) is shown in Fig. 1.

Swelling Studies

The investigation involved an examination of the swelling behaviour exhibited by synthesized hydrogels via the gravimetric approach at consistent one-hour intervals spanning 24 h in distilled water. In the initial phase, desiccated hydrogel discs were immersed in distilled water. Any surplus fluid was blotted away using filter paper [26]. Subsequently, the swollen hydrogels were extracted after a specified interval, their mass was carefully recorded and subsequently re-immersed into the distilled water. The determination of the swelling (%) was accomplished by employing the following formula (Eq. 1) [26].

$$\text{Swelling}\,(\%) = ((W_e - W_d)/W_d) \times 100 \quad (1)$$

where W_e is the weight of the swollen hydrogel at time t and W_d is the initial weight of the dried hydrogel.

pH Dependent Equilibrium Swelling Percent

Pre-weighed dried samples of the hydrogels were immersed in distilled water and buffer solutions of pH 2,4, 9 and 12 respectively. After a defined period of 24 h (equilibrium state), the hydrogel samples were carefully removed and their final weights were recorded. The equilibrium swelling percent (ES %) of the hydrogel samples was calculated using the Eq. (2) [27]. This calculation allows us to determine the extent of swelling exhibited by the hydrogel samples under specific pH conditions. A higher ES (%) value indicates greater swelling of the hydrogel, while a lower value suggests minimal swelling. By measuring ES (%) at different pH levels, pH sensitivity of the hydrogels can be determined.

$$\text{ES}\,(\%) = ((W_t - W_i)/W_i) \times 100 \quad (2)$$

- W_t is the weight of the hydrogel sample after swelling,
- W_i is the initial dry weight of the hydrogel sample.

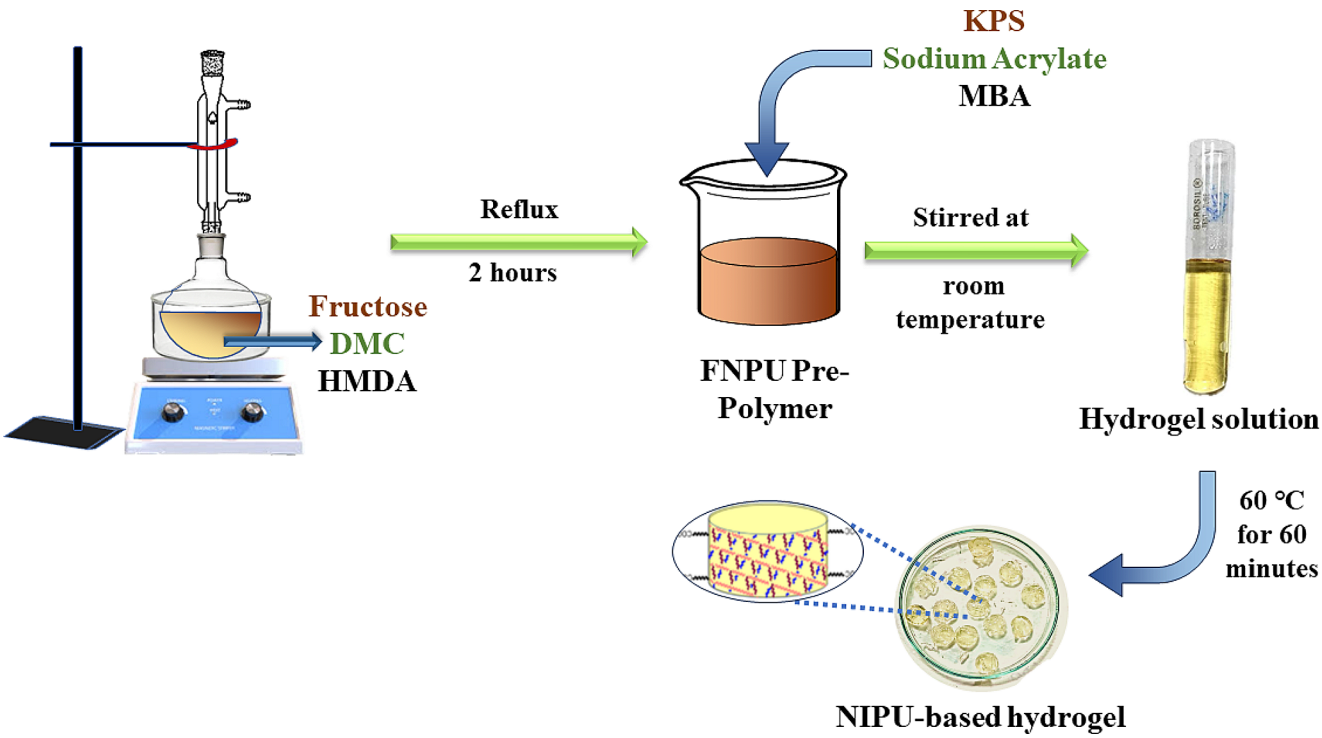


Fig. 1 Schematic diagram for the synthesis of NIPU-based hydrogel (Color figure online)

Swelling Kinetics

In order to understand the mechanism of water diffusion into the hydrogel samples, kinetic swelling studies were conducted using both the Kosmeyer-Peppas model (Eq. 3) and the Higuchi model (Eq. 4). Kosmeyer-Peppas model is particularly useful for hydrogels undergoing Fickian or non-Fickian diffusion, which is commonly observed in polymeric networks with complex structures. By fitting experimental swelling data to the Korsmeyer-Peppas model, we can determine the swelling exponent (*n*), which provides the information about the mechanism of the swelling. On contrary, the Higuchi model provides insights into the kinetics of swelling by describing the swelling as diffusion mechanism. This model helps us to understand how quickly a hydrogel reaches the equilibrium swelling, under specific conditions. The most appropriate and well-fitting model for explaining the drug release mechanism is the one that shows an R^2 value approximately equal to 1.

$$\mathrm{Log}\ (\mathrm{Swelling}\,\%) = \mathrm{Log\,K} + \mathrm{n\,Log\,t} \tag{3}$$

$$\mathrm{Swelling}\% = \mathrm{K.t}^{1/2} \tag{4}$$

where is K is rate constant and t is time.Sol-gel Analysis

The synthesised hydrogel's uncrosslinked and crosslinked fraction was determined via sol-gel analysis. A hydrogel's insoluble cross-linked portion is called a gel, while its soluble uncross-linked portion is implied by the term "sol". A sol–gel study was performed to examine the uncrosslinked portion of FNHG and control Polysodium acrylate (PSA) hydrogel. Weighed (Mi) dried hydrogels were immersed in distilled water for 24 h at room temperature. The hydrogel discs were then dried in an oven at 45 °C until a constant weight (Me) was achieved. The following equations were used to calculate the sol and gel fractions [26]:

$$\mathrm{Sol\ Fraction}\ (\%) = \frac{M_i - M\mathrm{e}}{M_e} \times 100 \tag{5}$$

$$\mathrm{Gel\ Fraction}\ (\%) = 100 - \mathrm{Sol\ Fraction} \tag{6}$$

Dye Sorption by FNHG

Batch studies were undertaken to assess the efficacy of the synthesized hydrogel FNHG in removing malachite green (MG) and crystal violet (CV) dyes at room temperature to ensure uniform experimental conditions. In each study, 0.05 g of the testing FNHG sample was put into 25 mL aqueous solution with 20 ppm of CV and MG dye solutions at room temperature for 60 min, until equilibrium was reached. The main objective of this study is to assess the dye sequestration efficiency of the synthesized hydrogel FNHG. The progression of the sorption phenomenon was precisely monitored using a UV-vis spectrophotometer at a wavelength of 617 and 582 nm for MG and CV dye, respectively. In order to assess the removal efficiency of the dyes and to ascertain the dye elimination capacity of the prepared hydrogel, the following mathematical formulation was employed:

$$\mathrm{Removal\,efficiency}\ (\%) = \frac{\mathrm{C_o} - \mathrm{C_e}}{\mathrm{C_o}}\ 100\% \tag{7}$$

where C_o refers to the initial dye concentration and C_e refers to the equilibrium dye concentration.

Instrumentation

Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy (ATR-FTIR)

Examination of samples was conducted using an attenuated total reflectance (ATR) Miracle diamond crystal and a Perkin Elmer Frontier ATR-FTIR. The clamp mechanism was firmly fastened when the sample powders were positioned on the diamond eye (1.8 mm) of the ATR apparatus. For the purpose of capturing the spectra in the wave number range of 600 to 4000 cm^{-1}, each sample was scanned 32 times at a resolution of 4 cm^{-1}.

Nuclear Magnetic Resonance (NMR)

The structural analysis of FNPU was performed by employing ^{1}HNMR spectroscopy using Jeol (Model: JNM-ECZ 400 S) NMR spectrometer (at 100 MHz). Deuterated dimethyl sulfoxide was employed as a solvent to dissolve the sample for analysis.

Thermogravimetric Analysis (TGA)

The thermal stability of the NIPU and hydrogel was evaluated by thermogravimetric analysis (TGA; Perkin Elmer TGA 4000). As a purge gas, dinitrogen was used. The thermal properties in the range of 50 to 250 °C were examined at a rate of 20 °C min^{-1} in order to determine the thermal parameters of transition.

Powder X-ray Diffraction (PXRD)

PXRD was employed using a high-resolution Bruker diffractometer (D8 Discover), in conjunction with a point detector (scintillation counter), to analyze the morphology and structure of the developed samples. Cu K radiation was received via a Gobel mirror that covers the range of 2α, 5–80°, with a scan rate of 1.0 s/step and step size 0.02°, at 298 K.

Scanning Electron Microscopy (SEM)

The morphology and internal structure of the hydrogel were examined using a scanning electron microscope (Jeol JSM-6610LVmodel).

Results and Discussions

Mechanistic Pathway Followed for the Synthesis of FNPU

During the synthesis of fructose-based NIPU (FNPU), the fructose molecules act as the starting material for the formation of the polyurethane backbone. Non-cyclic dimethyl carbonate is utilised for the initial stage of carbonation [2], and then nucleophilic substitution is used to carboxymethylate the existing hydroxy groups in fructose [28]. Afterwards, the adduct is allowed to react with hexamethylene diamine to produce NIPU at ambient pressure as shown in Fig. 2.

Mechanistic Pathway Followed for the Synthesis of FNHG

The synthesized FNPU was subsequently employed as a precursor for hydrogel fabrication with SA as a co-monomer, MBA as a crosslinker, and KPS as an initiator. The various chemical interactions and reactions that occur during the synthesis process are presented in Fig. 3.

KPS serves as an initiator for the free radical polymerization of SA during the hydrogel formation. KPS dissociates in water, generating sulfate radicals ($SO_4^{\bullet -}$) through thermal decomposition. These sulfate radicals initiate the polymerization of SA monomers, leading to the formation of long polymer chains of poly(sodium acrylate) (PSA) that further crosslink to create the hydrogel matrix. PSA contains a carboxylate group (–COO–) that reacts with the hydroxyl groups present in the fructose-based NIPU backbone. This reaction involves a nucleophilic addition of the carboxylate group to the hydroxyl group, leading to the formation of ester linkages (–COO–) between the PSA and the NIPU backbone. In addition to the ester linkages formed between PSA and the NIPU backbone, MBA serves as a crosslinker attributing to the presence of two acrylamide groups that

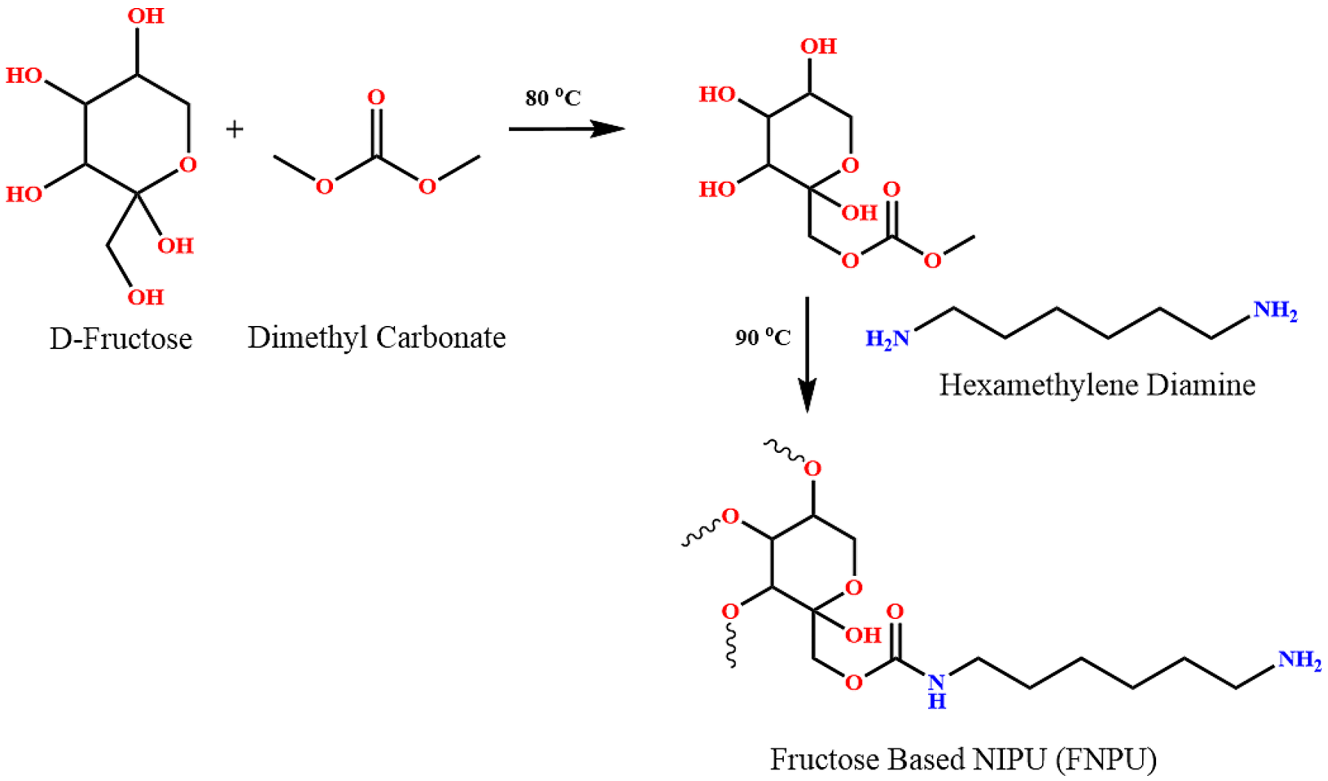


Fig. 2 Proposed mechanism for the synthesis of fructose-based NIPU (FNPU) (Color figure online)

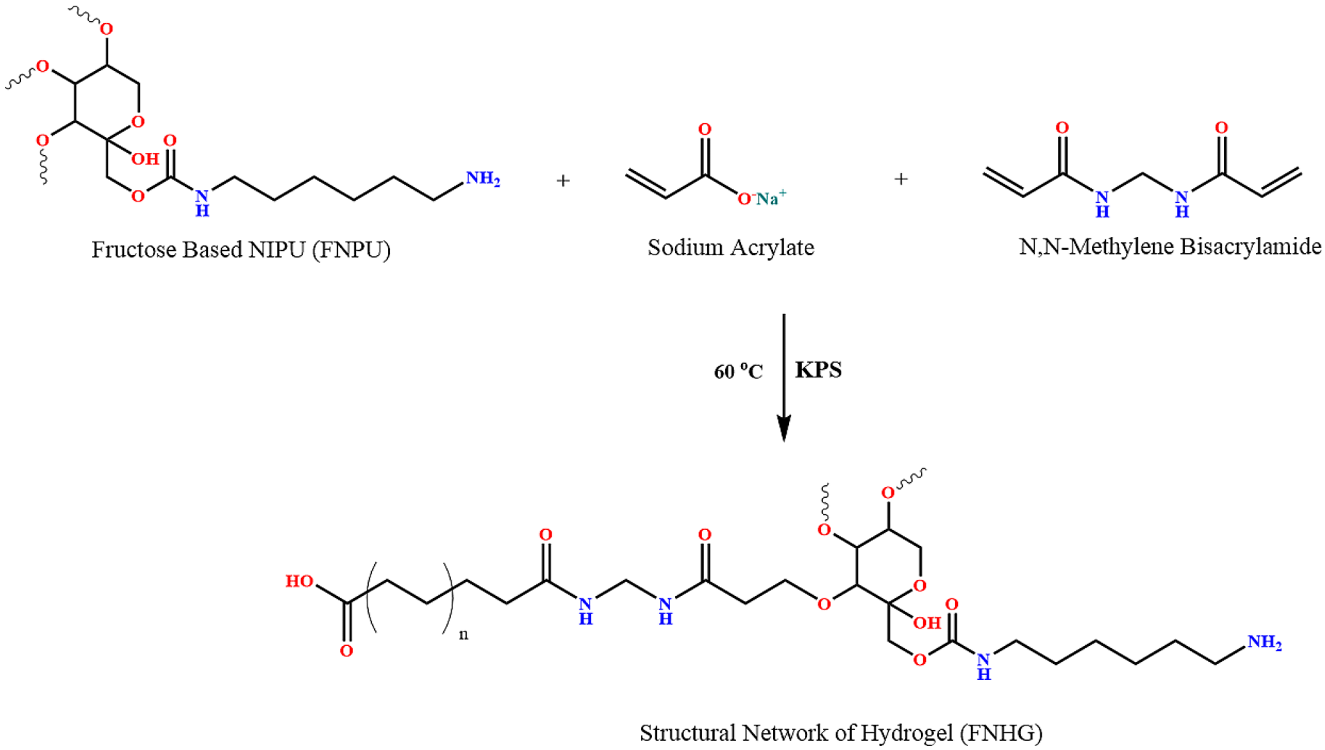


Fig. 3 Proposed mechanism for the synthesis of fructose-based NIPU hydrogels (FNHG) (Color figure online)

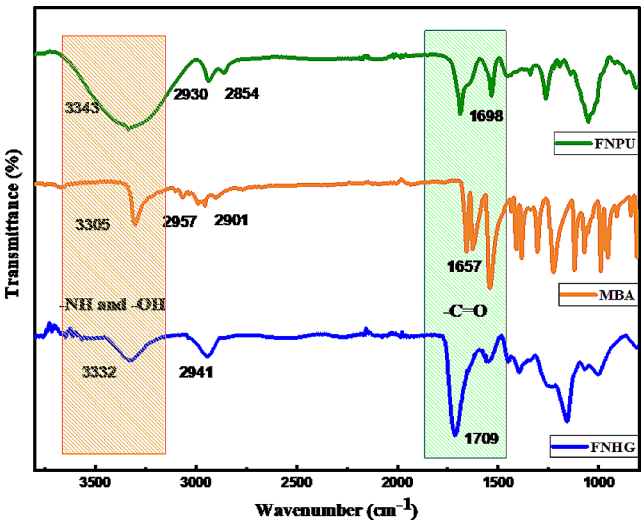


Fig. 4 FTIR spectra of FNPU, MBA, and FNHG (Color figure online)

react with the carboxylate groups present in SA. The cross-linking reaction involves the development of amide linkages (–CONH–) between the acrylamide groups of MBA and the carboxylate groups of PSA. This crosslinking reaction reinforces the three-dimensional network structure of the hydrogels, enhancing its mechanical integrity and stability.

Characterization of FNPU and FNHG

In order to study the functional group modifications occurring during the formation of FNPU, FTIR analysis was done (Fig. 4). In the case of FNPU, the broad peak at 3332 cm^{-1} is indicative of extensive hydrogen bonding, attributed to the stretching frequencies of both hydroxyl (-OH) and amine (-NH) groups. Bands around 2930–2854 cm^{-1} are associated with C-H stretching vibration in -CH_2- groups. The band around 1698 cm^{-1} confirms the presence of urethane-derived carbonyl bands of a complex nature [9]. The characteristic absorption peak at 1530 cm^{-1} correlates to bending of urethane-derived C-N-H and is typically observed in all polyurethanes [1, 9, 11, 12]. Another band at 1180–980 cm^{-1} corresponds to ether C-O-C (carbonate backbone) and alcohol C-O stretching vibrations. In contrast, the FTIR spectrum of MBA reveals a characteristic peak associated with the C-N bond at 1311 cm^{-1}. However, upon cross-linking to form FNHG, a notable shift of this peak to 1376 cm^{-1} is observed, indicating a reduction in conjugation as a consequence of the cross-linking process [26]. As a result, it can be established that the cross-linking of the FNHG has been successfully achieved. Moreover, the presence of carboxyl and urethane groups within the hydrogel matrix is evidenced by bands at 1709 cm^{-1} [9]. Additionally, the C-O ester groups of the network structure are associated with the unique absorption band at 1248 cm^{-1} [1,

10–12, 29, 30]. Thus, the FTIR spectra of FNPU, MBA, and FNHG confirm the successful synthesis of fructose-based NIPU/ PSA employing an isocyanate-free process at atmospheric pressure.

To further confirm the structure of FNPU, ^{1}HNMR studies were conducted. After dissolving a predetermined amount of FNPU (about 20 mg) in deuterated DMSO, the sample solution was transferred into an NMR tube. Once the ^{1}H NMR data acquisition was completed, characteristics peaks of urethane linkage and other entities were determined. In the NMR spectra presented in Fig. 5, peaks are observed at around 2.5 ppm and 3.4 ppm, which can be attributed to deuterated dimethyl sulfoxide (DMSO) and its water. In the NMR spectra of the synthesized compound (FNPU), several distinct peaks provide valuable insights into its molecular structure. Firstly, the peak at δ 8.12 ppm corresponds to the urethane NH group, confirming the successful formation of the urethane bond critical for enhancing mechanical strength and stability. Furthermore, Small peaks around δ 5.44 ppm are likely attributed to the anomeric proton of the glucopyranose unit. Additionally, the presence of hydroxyl (OH) groups is indicated by a visible hump around δ 3.80–3.89 ppm, contributing to the broad signal in this region. Hydroxyl groups play crucial roles in the chemistry of the synthesized compound, facilitating hydrogen bonding interactions. Furthermore, the peak at δ 3.56 ppm corresponds to the methylene group adjacent to the urethane moiety, likely belonging to the polymer backbone while peaks in the range of δ 3–4 can be attributed to the protons of the glucopyranose ring. Its chemical environment may differ from other methylene groups due to its proximity to the urethane linkage. Lastly, a cluster of peaks within the δ range of 1.21–1.49 ppm arises from the protons of methylene groups in the carbon chain of hexamethylene diamine, a common building block in polymer synthesis. The distinct chemical shifts of these methylene protons facilitate their differentiation from other methylene groups in the compound [31].

In the PXRD pattern presented in Fig. 6, the presence of a dominant peak at 20° implies that the synthesized material is segmented PU (NIPU in this case) and the broadness of the peak signifies that FNPU has a chemical structure in accordance with the previously reported studies [16, 17]. This corroborates that fructose-based NIPU has been successfully synthesized using an isocyanate-free approach at atmospheric pressure. The amorphous nature of FNPU is reduced in the process of formation of FNHG; this could be due to the introduction of other functional groups (such as acrylate entities and amide) and an increase in the number of hydroxyl groups, which may induce an ordered structure [32].

TGA was carried out to determine the thermal stability of the synthesized FNPU (Fig. 7 (a)) and FNHG (Fig. 7 (b)), in nitrogen (N_2) atmosphere across temperatures ranging from 25 to 700 °C. The TGA curve displays mass loss as a two-step process, as presented in Fig. 7 (a). The initial weight loss step between 104 and 224 °C involves the decomposition of

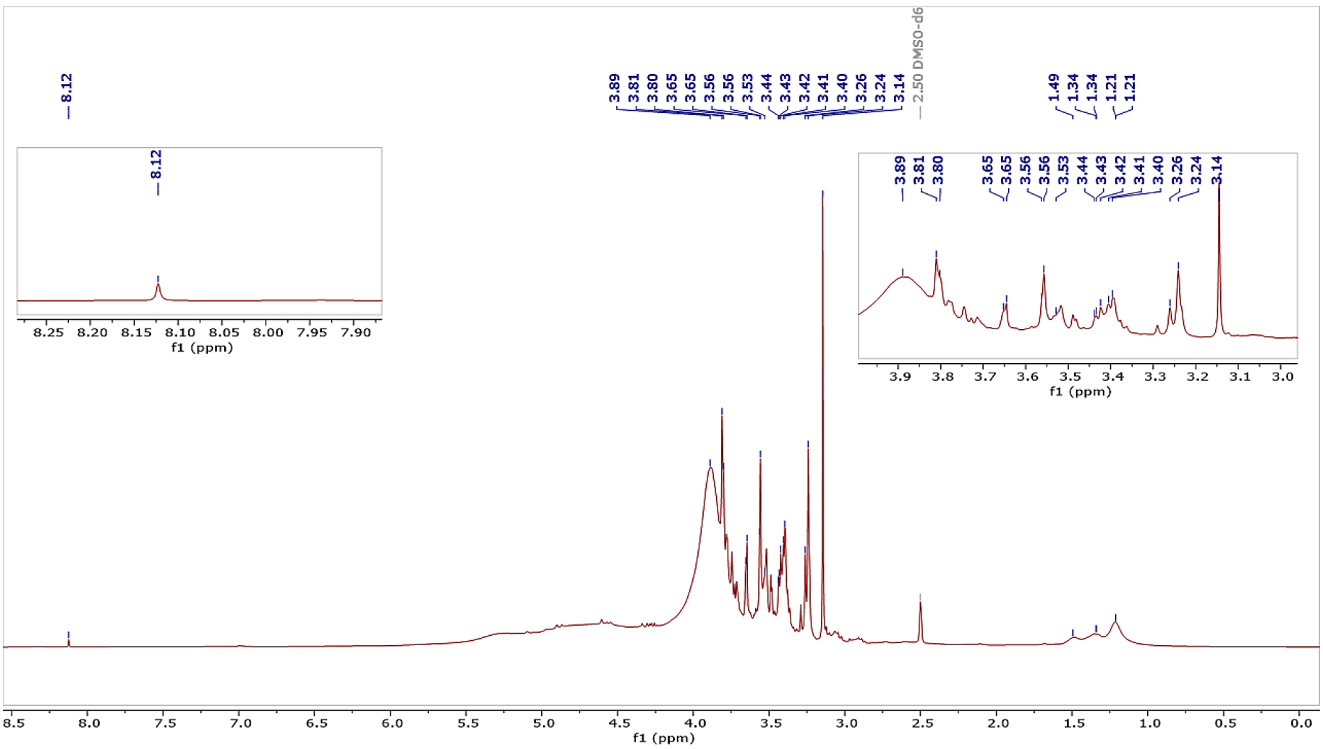


Fig. 5 ^{1}HNMR Spectra of the FNPU (Color figure online)

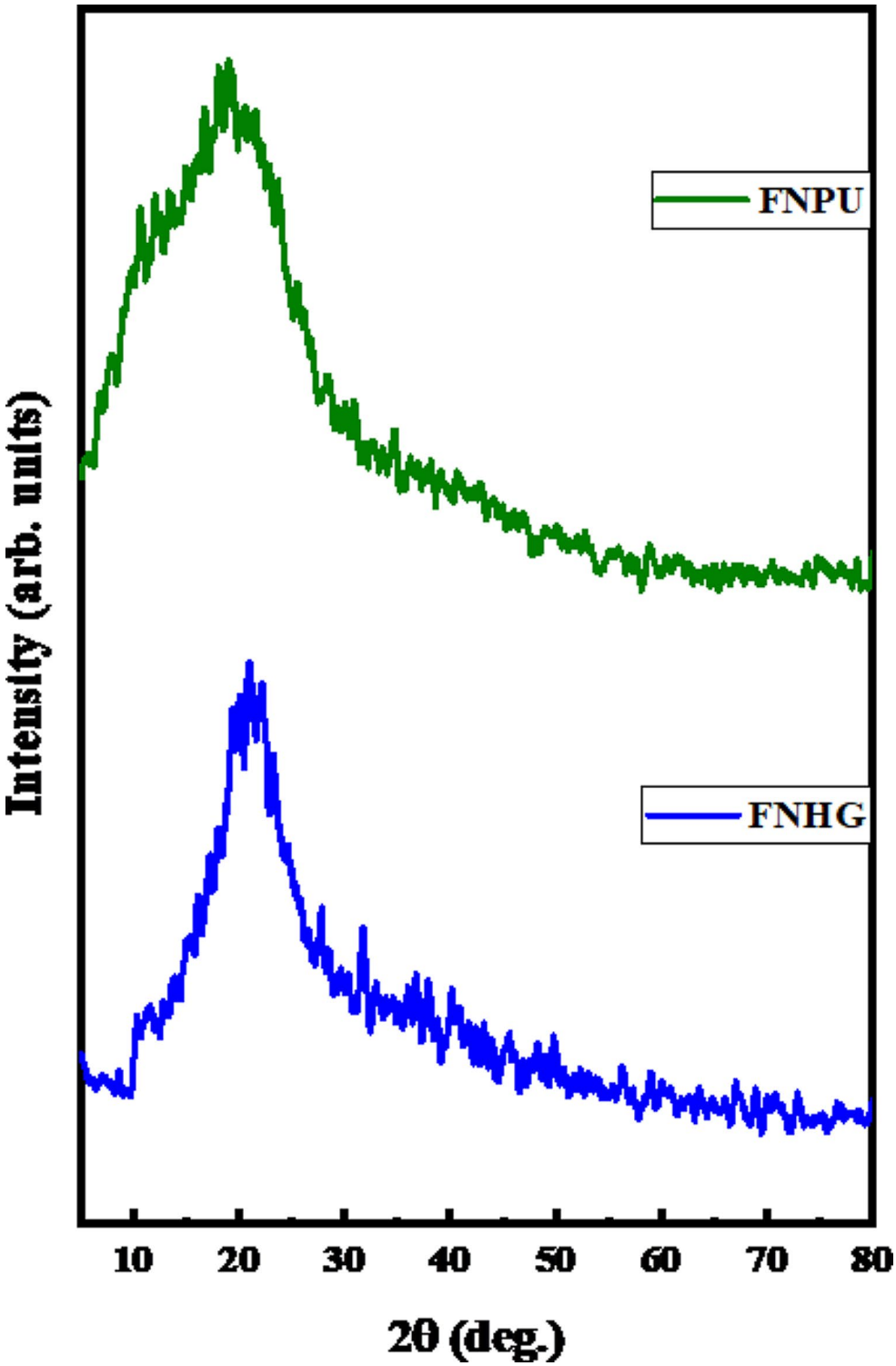


Fig. 6 PXRD pattern of FNPU and FNHG (Color figure online)

hexamine as well as the discharge of absorbed water molecules and evaporation of monomers/oligomers generated during the course of the reaction below 200 °C [24, 25]. The second weight loss step, as observed between 333 and 508 °C is a result of the collapse of the polymer backbone due to the cleavage of the urethane bond. The third and final weight loss step occurs above 500 °C and involves more than 80% mass loss, which can be credited to the possible breakdown of C–C bonds and degradation of the remnants of the pyrolysis by-products from the prior two steps.

Thermal degradation of FNHG hydrogel follows four stages of decomposition as shown in Fig. 7 (b). The first

stage occurs from 74 to 164 °C with weight loss of 1.09% indicating the vaporization of water molecules in FNHG hydrogel. The second stage indicates the loss of carboxyl and hydroxyl groups from the polymeric backbone with weight loss of 41.76% from 200 to 307 °C in FNHG hydrogel. In the third weight loss step, a weight loss of 21.92% from 333 °C to 412 °C was observed and can be attributed to the breakdown of cross-links present in FNPU hydrogel. Finally, the last weight loss of 8.03% from 412 to 508 °C can be attributed to the complete degradation of FNHG hydrogel, respectively. The FNHG hydrogel showed an enhanced thermal stability as compared to that of FNPU, thus, indicating that MBA results in the formation of a chemically cross-linked PSA and NIPU network.

The ability of FNHG to swell by retaining water in its sponge-like structure was demonstrated by a swelling study. The FNHG shows the maximum swelling of 53.20% in distilled water, as observed in Fig. 8 (a). The ES % for FNHG was evaluated in distilled water and in pH 2,4,10, and 12, as presented in Fig. 8 (b). The equilibrium swelling percent was observed maximum at pH 12 and minimum at pH 2. The ionisation characteristics of the synthesised hydrogel can be used to explain its pH-dependent swelling behaviour. At higher pH of 10 and 12, the ES % was found to have greater swelling, these values indicate that the carboxylate groups in poly(sodium acrylate) deprotonate, which increases ionisation and the electrostatic repulsion between polymer chains. The hydrogel network structure loosens as a result of this repulsion, allowing for increased water absorption and percent equilibrium swelling (PES) values. In contrast, the carboxylate groups are primarily protonated at low pH of 2 and 4 values, which results in less ionisation and reduced electrostatic repulsion between polymer chains. Overall, the ionisation state of carboxylate groups controls the pH-dependent swelling behaviour of the hydrogel. This eventually affects the hydrogel's capacity to swell by modifying the electrostatic interactions and hydrophilicity of the polymer chains [33].

The hydrogel samples showed behaviour that was in line with first-order kinetics and the Korsmeyer-Peppas model, according to the examination of the swelling kinetics data (Fig. 8c). A moderate match of the data to the Higuchi kinetics model is suggested by the derived R^2 value of 0.73582. Furthermore, the Korsmeyer-Peppas model—which takes n to be the diffusion constant and k to be the kinetic constant—was used to examine the hydrogel samples' swelling behaviour. The obtained values of R^2 (0.98019) and n (0.64034) show that the data fit the Korsmeyer-Peppas model well, implying that non-Fickian diffusion kinetics govern the hydrogel samples' swelling behaviour. It suggests that the swelling kinetics of the hydrogel may not be entirely controlled by molecular diffusion, but rather other processes like polymer chain relaxation might also be involved [26, 27].

The control (PSA hydrogel) and FNHG have an observed sol fraction of roughly 22.12% and 19.02% and a gel fraction of roughly 77.88% and 80.98% respectively. These results can be ascribed to the crosslinking density. The addition of NIPU improves the gel fraction by adding more crosslinks to the polymer matrix. These crosslinks aid in the development of a three-dimensional network, giving the hydrogel its structural stability. Low molecular weight species, oligomers, or unreacted monomers in the hydrogel formulation could be the cause of the observed sol fraction.

Figure 9 illustrates the hydrogel's surface morphology before the dye removal experiments. The SEM micrograph

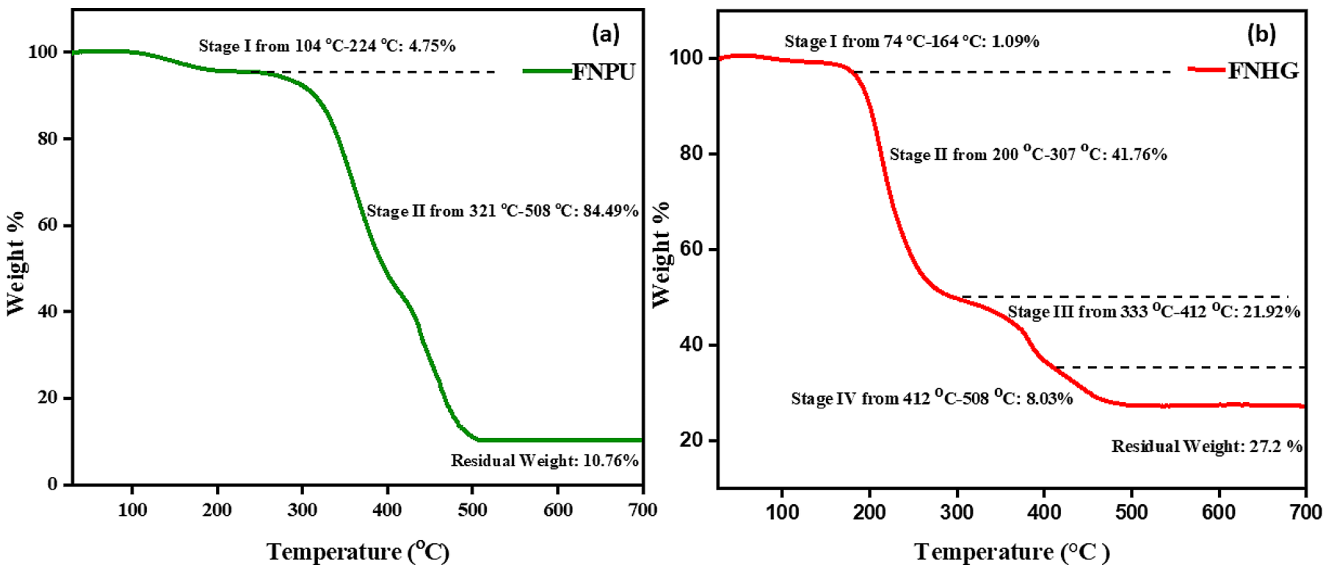


Fig. 7 TGA thermogram of (**a**) FNPU and (**b**) FNHG (Color figure online)

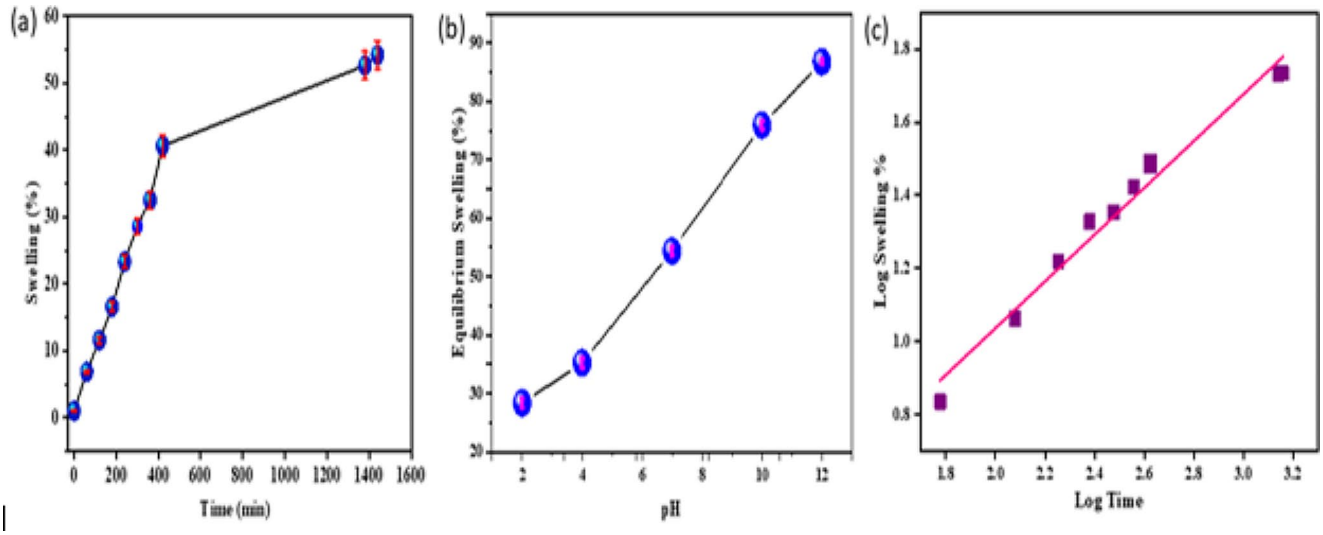


Fig. 8 (**a**) Rate of swelling of FNHG concerning time (**b**) Equilibrium swelling per cent at different pH and (**c**) Korsmeyer-Peppas model for swelling (Color figure online)

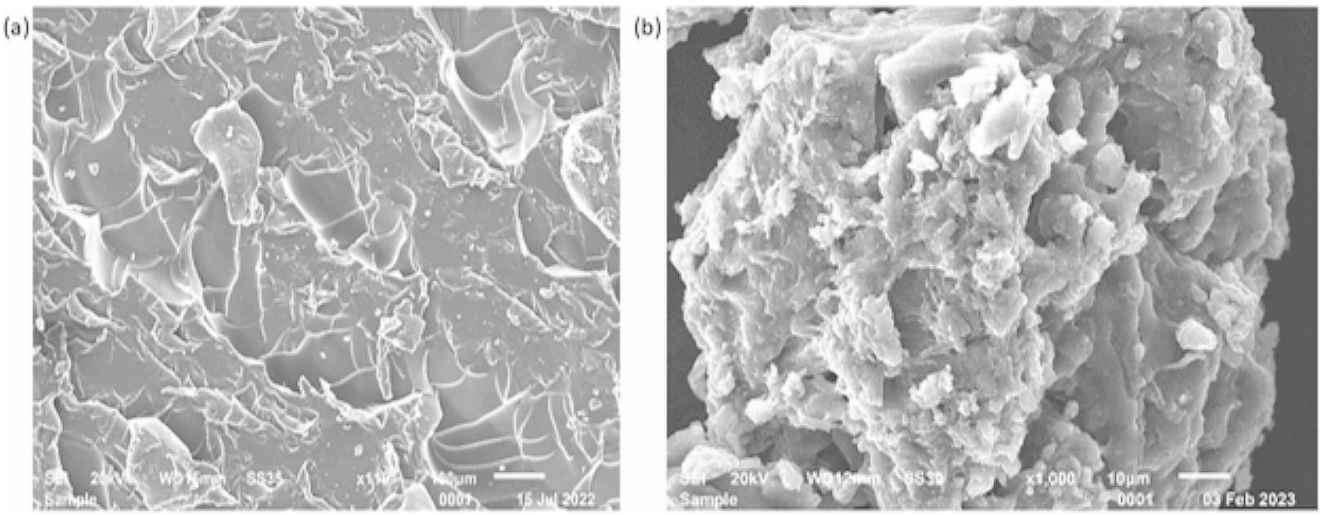


Fig. 9 SEM image of (**a**) PSA hydrogel and (**b**) FNHG

displays that FNHG hydrogel exhibits a characteristic macroporous three-dimensional structure. A closer inspection reveals that the FNHG hydrogel has a surface that is substantially rough and contains scattered pores in comparison to the controlled hydrogel (PSA hydrogel). The surface roughness of the FNHG has significantly increased with addition of NIPU due to irregularities or changes in the surface topography of the hydrogel, as seen in Fig. 9 (a) and (b) respectively. Changes in the crosslinking density and the incorporation of additional polymer chains, as a result of the integration of NIPU, are two reasons for this increase in surface roughness. The hydrogel's overall texture may be impacted by the surface imperfections resulting from the NIPU-polysodium acrylate interactions [34]. This feature of the hydrogel's surface morphology is beneficial for the effective collection and retention of dye molecules, which is a critical characteristic for its application in dye removal processes. The efficient removal of the dye molecules is greatly aided by this complex surface roughness. The existence of the pores and the uneven surface texture increases the surface area available for interactions between the dye molecules and the polymeric network [17]. As a result, this structural characteristic improves the hydrogel's ability to arrest molecules, which makes it suitable for removing dye molecules from the surrounding environment [35].

CV and MG Dye Sequestration by FNHG

The removal efficiency of the FNHG hydrogel was assessed for the sequestration of CV and MG dyes under optimum

conditions (adsorbent = 0.05 g, adsorbate solution = 25 mL, concentration of CV and MG dyes = 20 mg/L, and contact time = 60 min). During the process, a pattern was observed, as seen in Fig. 10, wherein the absorbance values for both dyes show a steady fall over time, with their levels stabilizing after 60 min. Within the first 15 min of the interaction of the dye with the synthesized hydrogel, there is an abrupt and noticeable increase in dye uptake. The initial high concentrations of the dye solutions along with the availability of readily accessible adsorption sites on the hydrogel surface are responsible for this phenomenon. The degree of removal, however, steadily slows down as time passes and the system gets closer to equilibrium, eventually achieving equilibrium at 60 min. The decline in the concentration of CV and MG solution and the gradual occupation of previously available empty sites on the sorbent material are two concurrent processes that can be attributed to this slowing down of the dye removal rate.

The removal efficiency of controlled hydrogel was found to be around 86.84% and 90.8% for CV and MG dye, respectively whereas that for FNHG it was found to be 96.87% and 99.80% for CV dye and MG dye respectively. The removal efficiency of FNHG hydrogel (Fig. 11a) is significantly higher than that of the control PSA hydrogel (Fig. 11b). These enhanced removal efficiencies of dyes via FNHG highlight the expediency of the synthesised hydrogel as an efficient material for the uptake of dyes from wastewater.

Possible Interactions Between the Synthesized Hydrogel and Cationic Dyes

It has been reported that organic and inorganic contaminants found in wastewater can be effectively removed by polyurethane foams [17–20]. However, several studies have also demonstrated that modified polyurethane-based adsorbents exhibit enhanced adsorption attributes, for example, surface area, pore size, and functional groups. These modification can be achieved through various methods, such as coating, grafting, and crosslinking; to introduce specific chemical moieties that can interact with target pollutants [36]. In this study, NIPU has been modified via crosslinking with PSA. Depending on the structure and charge of the pollutants, removal can be accomplished through a variety of interactions between the synthesized hydrogel FNHG and dye molecules, including electrostatic interactions, ion exchange, and hydrophobic interactions, etc. CV and MG dyes are water soluble and exhibit a positive charge when dissolved in aqueous solutions [37]. Negatively charged polymer entities present on the surface of hydrogel, owing to the inclusion of PSA into the matrix may have primarily facilitated the electrostatic attraction of positively charged cationic dye molecules. Leite et al. studied the inclusion of sodium dodecyl sulfate (SDS) into the PU matrix which enabled the establishment of a hydrophobic ionic pairing between the cationic dye and the anionic surfactant, displaying a strong affinity for PU [38]. Therefore, it can be suggested that PSA might be acting as a carrier to adsorb CV and MG dye molecules into the polymeric network similar to SDS. Despite the predominance of electrostatic interactions, some weaker

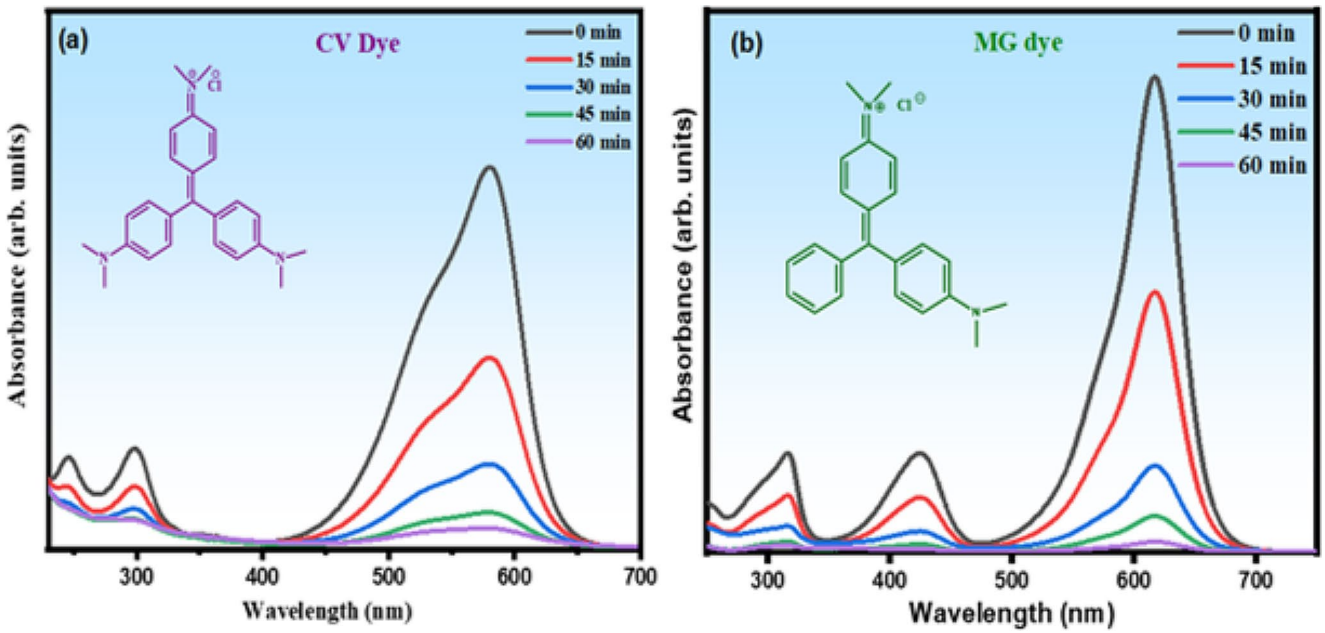


Fig. 10 Absorbance spectra of CV (**a**) and MG (**b**) dyes (Color figure online)

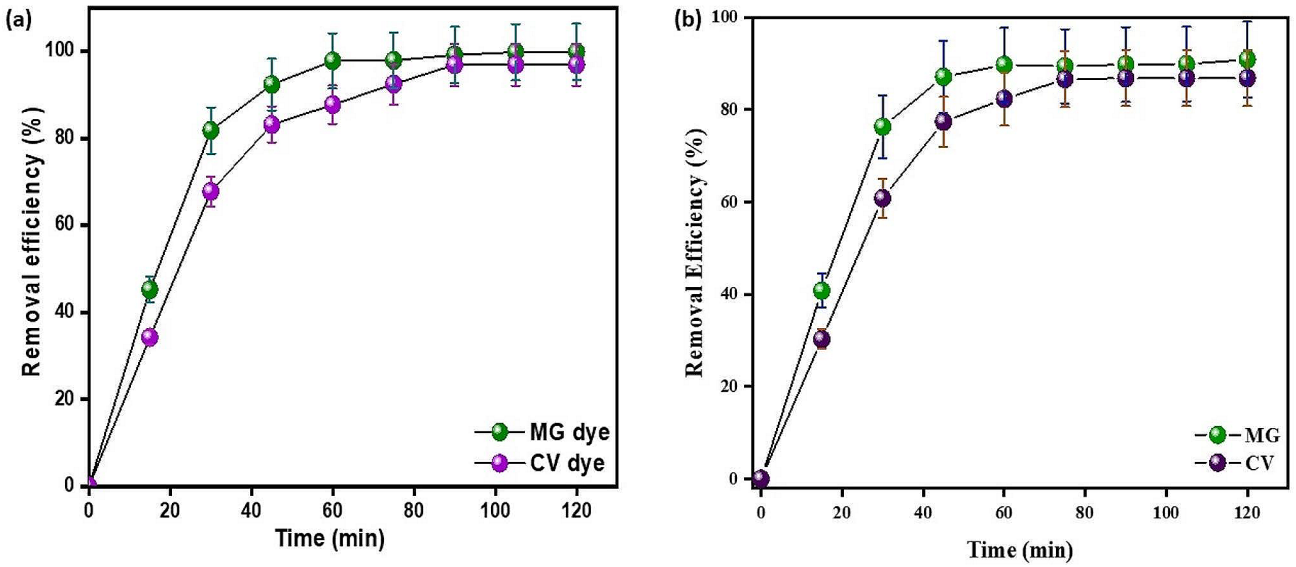


Fig. 11 Removal efficiency of CV and MG dyes (**a**) FNHG and (**b**) Control PSA Hydrogel (Color figure online)

hydrogen bonding may also occur amongst the amine groups of dye molecules and hydroxyl groups of the synthesized hydrogel [27, 39]. The interactions between CV and MG dyes with PSA polymer led to their attachment and probable adsorption onto the polymer surface. This property of PSA polymer makes it a useful carrier in the NIPU matrix for applications related to dye sequestration and material formulation.

Conclusions

The present work demonstrated the synthesis of fructose-based NIPU and then its hydrogels via a facile and novel approach. The adopted method results in the formation of hydrogel in a more ecologically friendly manner by avoiding harsh reaction conditions and by reducing gelation time to 30 minutes. The fundamental structural framework of the NIPU backbone is fructose, which was selected as a sustainable and easily accessible building element. The acquired results from FTIR and ^{1}H NMR spectroscopy confirm the formation of urethane linkages throughout the polymer matrix. The hydrogel's three-dimensional network structure is strengthened by the addition of the crosslinker MBA. Furthermore, PSA serves as a comonomer in the hydrogel formulation, facilitating the introduction of various functional groups. TGA results revealed that this structural reinforcement not only improves the hydrogel's thermal stability but also provides the mechanical resilience required for its practical applications. SEM morphology displays the presence of a rough surface providing additional surface area. The study also demonstrated that the fabricated hydrogel can act as a potential material for the sequestration of cation dyes. The NIPU-based hydrogel exhibited high removal efficiency towards CV and MG dyes i.e., 96.87% and 99.8% respectively. The novel fructose-based NIPU hydrogels have enormous potential as an effective material for dye removal from aqueous solutions.

Acknowledgements The authors are thankful to the Department of Applied Chemistry, Delhi Technological University for its infrastructural facilities.

Author Contributions P.S. conceptualized the study, conducted investigations, devised the methodology, and drafted the original manuscript. R.K. provided supervision, contributed to manuscript editing, and critically reviewed the content. All authors participated in reviewing and approving the final manuscript.

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

Declarations

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

1. Singh P, Kaur R (2023) One pot synthesis of bio-based porous isocyanate-free polyurethane materials. Mater Lett 331:133433. https://doi.org/10.1016/j.matlet.2022.133433
2. Niesiobędzka J, Datta J (2023) Challenges and recent advances in bio-based isocyanate production. Green Chem 25:2482–2504. https://doi.org/10.1039/d2gc04644j
3. Polo Fonseca L, Trinca RB, Felisberti MI (2018) Amphiphilic polyurethane hydrogels as smart carriers for acidic hydrophobic drugs. Int J Pharm 546:106–114. https://doi.org/10.1016/j.ijpharm.2018.05.034
4. Bourguignon M, Thomassin JM, Grignard B et al (2021) Water-Borne Isocyanate-Free polyurethane hydrogels with adaptable functionality and behavior. Macromol Rapid Commun 42:1–13. https://doi.org/10.1002/marc.202000482
5. Filippova OV, Maksimkin AV, Dayyoub T et al (2023) Sustainable elastomers for actuators: Green Synthetic approaches and Material properties. Polym (Basel) 15:1–32. https://doi.org/10.3390/polym15122755
6. Bizet B, Bizet B, Grau É et al (2020) Water-based non-isocyanate polyurethane-ureas (NIPUUs). Polym Chem 11:3786–3799. https://doi.org/10.1039/d0py00427h
7. Kaur R, Singh P, Tanwar S et al (2022) Assessment of Bio-based polyurethanes: perspective on applications and Bio-degradation. Macromol 2:284–314. https://doi.org/10.3390/macromol2030019
8. Malik M, Kaur R (2018) Synthesis of NIPU by the carbonation of canola oil using highly efficient 5,10,15-tris(pentafluorophenyl) corrolato-manganese(III) complex as novel catalyst. Polym Adv Technol 29:1078–1085. https://doi.org/10.1002/pat.4219
9. Xi X, Pizzi A, Gerardin C, Du G (2019) Glucose-biobased non-isocyanate polyurethane rigid foams. J Renew Mater 7:301–312. https://doi.org/10.32604/jrm.2019.04174
10. Dechent SE, Kleij AW, Luinstra GA (2020) Fully bio-derived CO_2 polymers for non-isocyanate based polyurethane synthesis. Green Chem 22:969–978. https://doi.org/10.1039/c9gc03488a
11. Singh P, Kaur R (2022) Sustainable xylose-based non-isocyanate polyurethane foams with Remarkable Fire-Retardant properties. J Polym Environ 31(3):1–11. https://doi.org/10.1007/s10924-022-02638-4
12. He X, Xu X, Wan Q et al (2019) Solvent- and Catalyst-free synthesis, hybridization and characterization of Biobased Non-isocyanate polyurethane (NIPU). https://doi.org/10.3390/polym11061026
13. Pathak J, Singh P (2023) Adsorptive Removal of Congo Red using organically modified zinc–copper–nickel Ternary Metal Hydroxide: kinetics, isotherms and Adsorption studies. J Polym Environ 31:327–344. https://doi.org/10.1007/s10924-022-02612-0
14. Kumar L, Sharma JG, Kaur R (2022) Catalytic Performance of Cow-Dung Sludge in Water Treatment Mitigation and Conversion of Ammonia Nitrogen into Nitrate. Sustainability 14(4):2183. https://doi.org/10.3390/su14042183
15. Duhan M, Kaur R (2021) Nano-Structured Polyaniline as a potential adsorbent for Methylene Blue Dye removal from Effluent. J Compos Sci 5(1):7. https://doi.org/10.3390/jcs5010007(MDPI)
16. Gandara M, Mulinari DR, Monticeli FM, Capri MR (2020) Sugarcane bagasse fibers Reinforced in polyurethane for Sorption of Vegetal Oil. J Nat Fibers 00:1–12. https://doi.org/10.1080/15440478.2019.1710653
17. Qin J, Qiu F, Rong X et al (2015) Adsorption behavior of crystal violet from aqueous solutions with chitosan-graphite oxide modified polyurethane as an adsorbent. J Appl Polym Sci 132:1–10. https://doi.org/10.1002/app.41828
18. Sultan M (2017) Polyurethane for removal of organic dyes from textile wastewater. Environ Chem Lett 15:347–366. https://doi.org/10.1007/s10311-016-0597-8
19. Song X, Wang Y, Zhou L et al (2021) Halloysite nanotubes stabilized polyurethane foam carbon coupled with iron oxide for high-efficient and fast treatment of arsenic(III/V) wastewater. Chem Eng Res Des 165:298–307. https://doi.org/10.1016/j.cherd.2020.11.001
20. Barroso-Solares S, Merillas B, Cimavilla-Román P et al (2020) Enhanced nitrates-polluted water remediation by polyurethane/sepiolite cellular nanocomposites. J Clean Prod 254:120038. https://doi.org/10.1016/j.jclepro.2020.120038
21. Qi X, Li Z, Shen L et al (2019) Highly efficient dye decontamination via microbial salecan polysaccharide-based gels. Carbohydr Polym 219:1–11. https://doi.org/10.1016/j.carbpol.2019.05.021
22. Qi X, Wei W, Su T et al (2018) Fabrication of a new polysaccharide-based adsorbent for water purification. Carbohydr Polym 195:368–377. https://doi.org/10.1016/j.carbpol.2018.04.112
23. Qi X, Wu L, Su T et al (2018) Polysaccharide-based cationic hydrogels for dye adsorption. Colloids Surf B Biointerfaces 170:364–372. https://doi.org/10.1016/j.colsurfb.2018.06.036
24. Gennen S, Grignard B, Thomassin JM et al (2016) Polyhydroxyurethane hydrogels: synthesis and characterizations. Eur Polym J 84:849–862. https://doi.org/10.1016/j.eurpolymj.2016.07.013
25. Bourguignon M, Thomassin JM, Grignard B et al (2019) Fast and facile one-pot one-step preparation of nonisocyanate polyurethane hydrogels in water at room temperature. ACS Sustain Chem Eng 7:12601–12610. https://doi.org/10.1021/acssuschemeng.9b02624
26. Meena P, Singh P, Warkar SG (2023) Development and assessment of carboxymethyl tamarind kernel gum-based pH-responsive hydrogel for release of diclofenac sodium. Eur Polym J 197:112340. https://doi.org/10.1016/j.eurpolymj.2023.112340
27. Bekiari V, Lianos P (2010) Poly (sodium acrylate) hydrogels as potential pH-sensitive sorbents for the removal of model organic and inorganic pollutants from water. Glob Nest J 12:262–269. https://doi.org/10.30955/gnj.000720
28. Fiorani G, Perosa A, Selva M (2018) Dimethyl carbonate: a versatile reagent for a sustainable valorization of renewables. Green Chem 20:288–322. https://doi.org/10.1039/c7gc02118f
29. Ghasri M, Bouhendi H, Kabiri K et al (2019) Superabsorbent polymers achieved by surface cross linking of poly(sodium acrylate) using microwave method. Iran Polym J (English Ed 28:539–548. https://doi.org/10.1007/s13726-019-00722-6
30. Blattmann H, Fleischer M, Bähr M, Mülhaupt R (2014) Isocyanate- and Phosgene-Free routes to Polyfunctional Cyclic carbonates and Green polyurethanes by fixation of Carbon Dioxide. Macromol Rapid Commun 35:1238–1254. https://doi.org/10.1002/marc.201400209
31. Eltayeb M, Li SX, Okoye PU, Wang S (2021) Carbodiimide-assisted synthesis of high purity Bis(cyclic carbonate) under Atmospheric conditions for Preparation of Non-isocyanate polyurethane. J Polym Environ 29:1880–1893. https://doi.org/10.1007/s10924-020-01996-1
32. Das M, Yadav M, Shukla F et al (2020) Facile design of a dextran derived polyurethane hydrogel and metallopolymer: a sustainable approach for elimination of organic dyes and reduction of nitrophenols. New J Chem 44:19122–19134. https://doi.org/10.1039/d0nj01871f
33. Bajpai SK, Swarnkar MP (2017) Controlled release of anti-diabetic drug metformin hydrochloride from cellulose/PEG/poly (SA) ter-polymeric hydrogels. J Macromol Sci Part Pure Appl Chem 54:186–193. https://doi.org/10.1080/10601325.2017.1265406

34. Cui L, Yao Y, Yim EKF (2021) The effects of surface topography modification on hydrogel properties. APL Bioeng 5:31509. https://doi.org/10.1063/5.0046076
35. Salama A, Shukry N, El-Sakhawy M (2015) Carboxymethyl cellulose-g-poly(2-(dimethylamino) ethyl methacrylate) hydrogel as adsorbent for dye removal. Int J Biol Macromol 73:72–75. https://doi.org/10.1016/j.ijbiomac.2014.11.002
36. Selvasembian R, Gwenzi W, Chaukura N, Mthembu S (2021) Recent advances in the polyurethane-based adsorbents for the decontamination of hazardous wastewater pollutants. J Hazard Mater 417:125960. https://doi.org/10.1016/j.jhazmat.2021.125960
37. George G, Saravanakumar MP (2018) Facile synthesis of carbon-coated layered double hydroxide and its comparative characterisation with Zn–Al LDH: application on crystal violet and malachite green dye adsorption—isotherm, kinetics and Box-Behnken design. Environ Sci Pollut Res 25:30236–30254. https://doi.org/10.1007/s11356-018-3001-3
38. Leite BT, Robaina NF, Gustavo L (2012) Removal of Malachite Green from Aqueous Medium employing polyurethane foam as Adsorbent and Sodium Dodecylsulfate as Carrier. 1303–1313. https://doi.org/10.1007/s11270-011-0946-y
39. Pandey B, Singh P (2023) Statistical Optimization of Process Parameters for Ultrafast Uptake of Anionic Azo Dyes by efficient sorbent: Zn/Cu layered double hydroxide. Appl Organomet Chem 37:1–18. https://doi.org/10.1002/aoc.7072

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

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

arXiv:2405.11029v1 [cs.LG] 17 May 2024

Generative Artificial Intelligence: A Systematic Review and Applications

Sandeep Singh Sengar[1*], Affan Bin Hasan[1], Sanjay Kumar[2], Fiona Carroll [1]

[1]Cardiff School of Technologies, Cardiff Metropolitan University, Cardiff, CF5 2YB, United Kingdom.
[2]Department of Computer Science and Engineering, Delhi Technological University, New Delhi, 110042, India.

*Corresponding author(s). E-mail(s): SSSengar@cardiffmet.ac.uk, Phone: +44-7780390651;
Contributing authors: ahasan@cardiffmet.ac.uk; sanjay.kumar@dtu.ac.in; fcarroll@cardiffmet.ac.uk;

Abstract

In recent years, the study of artificial intelligence (AI) has undergone a paradigm shift. This has been propelled by the groundbreaking capabilities of generative models both in supervised and unsupervised learning scenarios. Generative AI has shown state-of-the-art performance in solving perplexing real-world conundrums in fields such as image translation, medical diagnostics, textual imagery fusion, natural language processing, and beyond. This paper documents the systematic review and analysis of recent advancements and techniques in Generative AI with a detailed discussion of their applications including application-specific models. Indeed, the major impact that generative AI has made to date, has been in language generation with the development of large language models, in the field of image translation and several other interdisciplinary applications of generative AI. Moreover, the primary contribution of this paper lies in its coherent synthesis of the latest advancements in these areas, seamlessly weaving together contemporary breakthroughs in the field. Particularly, how it shares an exploration of the future trajectory for generative AI. In conclusion, the paper ends with a discussion of Responsible AI principles, and the necessary ethical considerations for the sustainability and growth of these generative models.

Keywords: Generative Artificial Intelligence, Generative Adversarial Networks, Diffusion, Segmentation, Variational Autoencoder, Transformers.

1 Introduction

The recent advancement in Artificial Intelligence has been mainly the result of Generative Artificial Intelligence (often referred to as Generative AI or GenAI) being introduced. Generative AI encompasses artificial intelligence systems with the ability to create text, images, or various forms of media through the utilization of generative models. These models acquire an understanding of the underlying patterns and structures within their training data, subsequently producing fresh data that share similar traits and characteristics. The motivation of this systematic review is to gather, evaluate, and synthesize existing research on a GenAI. This paper presents a systematic review that highlights key applications and variations of the architecture of Generative Artificial Intelligence models and their performance. We conducted this review to (a) understand the state-of-the-art generative AI techniques including summarizing key methodologies, algorithms, and findings across a range of studies (b) systematically review a large body of literature, which includes emerging trends, common challenges, and recurring patterns in the development and application of generative AI techniques (c) compare and contrast different generative AI approaches, such as Autoencoders, Generative Adversarial Networks, Transformers, and Diffusion models (d) explore successful applications of generative AI such as image translation, video synthesis and generation, natural language processing, knowledge graph generation, etc. (e) identify ethical challenges and propose solutions for responsible AI development.

In this research, we outline the most recent research and advancement in the field of Generative Artificial Intelligence. It details the approach used to navigate and analyze cutting-edge developments, ensuring a comprehensive and insightful review of the current landscape in Generative AI. The following criteria were applied for searching the used research papers.

Time Period: This paper presents a comprehensive overview of the advancements and applications of Generative AI, focusing on significant developments between 2018 and 2023. Additionally, it offers a concise historical perspective, tracing the evolution of foundational models from 2012 to 2018, which laid the groundwork for the current state of Generative AI techniques. This historical context enriches the understanding of the field's rapid progression and its burgeoning applications.

Keywords: This paper employs a targeted keyword search strategy, incorporating specific terms such as 'Generative Adversarial Networks', 'Transformers', 'Variational Autoencoders', and 'Diffusion Models'. This approach also includes searching for advancements in 'image translation', 'video synthesis', and various applications of Generative AI in 'natural language processing' and 'knowledge graph generation'. This methodology ensures a focused and comprehensive review of the latest developments in the field of Generative AI.

Databases: The work primarily sources relevant literature from Google Scholar, focusing on the specified timeframe. It selectively includes research that showcases advancements in generative models. This criteria ensures the inclusion of studies where developed models were rigorously tested on well-recognized datasets, and where results are communicated effectively and clearly. This approach guarantees that the paper presents a detailed and credible overview of significant developments in the field of Generative AI.

Inclusion Criteria: This work exclusively incorporates peer-reviewed papers, conference, and journal papers that are written in English. It emphasizes studies that highlight either significant advancements or innovative applications in the realm of Generative AI, ensuring that the focus remains on cutting-edge and impactful developments within this field.
Exclusion Criteria: This paper meticulously filters its sources, excluding non-peer-reviewed materials, papers not written in English, and studies that fall outside the 2012-2023 timeframe. Additionally, it deliberately omits any papers that do not directly contribute to the advancement or understanding of Generative AI, ensuring a focused and relevant academic discourse.
Evaluation Criteria: In each subsection evaluating the advancements in Generative AI techniques, the paper compares the performance of various models using standardized datasets commonly cited in the field. This comparison focuses on how different state-of-the-art models perform on these datasets, providing a clear and consistent basis for assessing the progress and effectiveness of these techniques in their respective domains.

The following is the summary of the major contributions of our work-

- **Paradigm Shift in Artificial Intelligence:** The paper discusses the paradigm shift in artificial intelligence and highlights the significant impact of generative models in the field of machine learning.
- **Historical Context:** The paper includes a section that gives a straightforward overview of how key AI models have developed from 2012 to 2018, helping to better understand how the field has grown and changed over time.
- **Real-World Uses of Generative AI:** The paper describes how Generative AI is used in different areas like image translation, diagnosing medical conditions, combining text and images, processing natural language, etc.
- **Systematic Review of Generative AI:** The work provides a comprehensive review and analysis of recent advancements in Generative AI, focusing on techniques and applications, including application-specific models. We have also provided information on relevant datasets for each used application.
- **Impact on Language and Image Translation:** The paper discusses the major impact of generative AI in language generation with large language models and in the field of image translation.
- **Responsible AI Principles:** The paper ends with a discussion on Responsible AI principles and ethical considerations necessary for the sustainability and growth of generative models.

After this introductory section, Sections 2 and 3 review the basic early architecture of Generative adversarial Networks and their variants. Section 4 deeply explores the recent applications and the advancements in Generative AI application-specific techniques. Section 5 provides the Challenges and opportunities of Generative AI. Lastly, Section 6 concludes the work and highlights the future directions of generative AI.

2 What is Generative Artificial Intelligence?

As discussed Generative Artificial Intelligence refers to artificial intelligence systems with the capability to create text, images, or other forms of media through the utilization of generative models. These models acquire an understanding of patterns and structures within their training data, subsequently generating novel data with akin characteristics. Generative Artificial Intelligence encompasses various types, each tailored for specific tasks or forms of media generation. The following are some of the more well-known types: Generative Adversarial Networks (GANs) [71], Transformer-based Models (TRMs) [133], Variational Autoencoders (VAEs) [62], and Diffusion models (DMs) [68], to name a few. The following sections will discuss these in more detail.

2.1 Generative Adversarial Networks

A generative adversarial network (GAN) is a class of machine learning framework and a prominent framework for approaching generative AI. The aspect that is novel in this generative adversarial network set-up is that it does not depend upon heavily annotated training data. Moreover, the architecture that it affords is quite unique from the conventional Deep Neural Networks [27]. Indeed, it consists of two major components named *Generator* and *Discriminator*. The main operation of the generator is to keep on generating the fake data using the noise while the purpose of the discriminator is to distinguish whether the generated image is real or fake. The discriminator is trained using the real images of the domain that the generator is trying to synthetically produce and the discriminator's sole purpose is to identify whether the output produced by the generator is fake or not. The overall system is based on the zero-sum game dynamics, the winner will remain unchanged and the loser model each time has to modify its parameters, it will keep on doing this until the discriminator is unable to detect whether the generator output is fake or not [44]. The sole purpose of this is to build a powerful generator model that generates synthetic data that looks real.

Fig. 1 demonstrates how the generator and discriminator work together. The generator aims to deceive the discriminator by providing the synthetically generated image with the objective that it is proven real. The discriminator discerns between genuine and counterfeit images and generates the output signal. This output signal then goes to both the generator and discriminator, allowing the generator to produce better synthetic output. And, in case the discriminator fails to prove the image is fake, it also uses the signal to change its weights to give better predictions. In this entire architecture, it is important to note that only the discriminator has access to the real image, synthetic image, and its own signal output while the generator only learns from the output signal of the discriminator [27].

During the initial stages of development, Generative Adversarial Network (GAN)-based models encountered significant challenges in their training process. These difficulties primarily revolved around issues like training divergence and model collapse [87]. Training divergence refers to situations where the GAN's generator and discriminator fail to achieve a stable equilibrium during training, leading to oscillations and unreliable model outputs. This problem results in inconsistent and sub-optimal generation performance, hindering the GAN's ability to produce high-quality samples. On

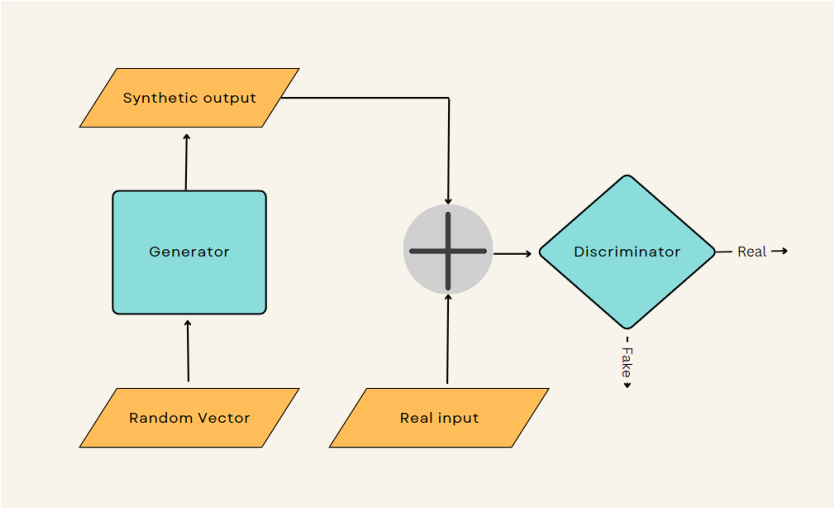


Fig. 1 GAN- Generator and discriminator working

the other hand, model collapse occurs when the GAN's generator produces limited and repetitive outputs, ignoring a large portion of the data distribution. This phenomenon causes the generator to focus on a small subset of data points, resulting in a lack of diversity and novelty in the generated samples. Addressing these challenges has been a main focus in the advancement of GAN-based models, with numerous research efforts aimed at improving stability, convergence, and diversity during the training process. As a result, substantial progress has been made, leading to the development of more robust and effective GAN architectures, which have significantly enhanced the performance and applicability of these generative models.

2.2 Transformers

Further Generative AI techniques called *Transformers* were introduced by Vaswani et al. [134]. This breakthrough architecture laid the foundation for various tasks, including machine translation and language generation, and it continues to influence subsequent neural network designs. The paper's emphasis on attention mechanisms highlighted their pivotal role in sequence-to-sequence tasks, advancing the state of the art. Transformers use both the self-attention and Multi-Head Attention mechanisms to learn the dependencies between the objects regardless of the distance between them and to learn the different relations and patterns between the input respectively. Often in Natural Language Processing, these methods are combined with positional encoding added to the input sequence to make the transformers keep track of the position of a specific word in an input sequence. Transformers are commonly used to build Generative AI Models such as Generative Pre-trained Transformers (GPT) models which are capable of generating coherent and contextually relevant text [109]. Bidirectional Encoder Representations from Transformers (BERT) and Open AI GPT are based on transformers.

2.3 Variational Autoencoders

Another model in the field of generative AI is Variational Autoencoders (VAEs), introduced by Kingma et al. [65]. As the name suggests the VAEs consist of an encoder and a decoder. The purpose of an encoder is to encode the given input in a lower dimension called *latent space* and the decoder decodes that latent output of the encoder into its original input shape. During this whole process, variation is introduced to the latent space by using the standard Gaussian distribution. The main goal is to achieve the output with a similar mean and variance as the given input after the introduction of the variance. This provides a structured way to learn meaningful representations of data and then generate new samples from that data distribution.

2.4 Diffusion Models

The Diffusion Models have been designed to improve the performance of the Simple Generative Adversarial Network, the technique was introduced by Salimans et al [115]. At a later stage, Kingma et al [64] introduced a variant of the diffusion model called Inverse Autoregressive Flow (IAF) as a building block for generative models. IAF is a type of normalizing flow. This is a type of generative model that aims to learn complex probability distributions by transforming a simple base distribution into the target distribution through a series of invertible transformations.

3 Evolution of Generative AI Models: A Look at Earlier Variants

3.1 Earlier GAN Variants

In the early stages of the introduction of Generative Artificial Intelligence models the major issue that researchers were facing was the convergence problem of generative models [87]. To avoid this problem, different approaches were adopted by the researchers to make the GAN more stable (e.g., by understanding the behavior of GAN training). In detail, Mescheder et al. [88] explained the analysis of local convergence and stability properties during the training of GAN. This involves an examination of the eigenvalues of the Jacobian matrix associated with the gradient vector field. Specifically, when the equilibrium point is characterized by solely negative real-part eigenvalues in the Jacobian, GAN training demonstrates local convergence, specifically when utilizing relatively small learning rates. However, the situation changes when the eigenvalues of the Jacobian are situated on the imaginary axis. In such cases, the local convergence of GAN training is generally compromised. Also, it is important to note that if the eigenvalues are in proximity and not directly on the imaginary axis, the training algorithm may necessitate exceedingly small learning rates to achieve convergence [88].

The study by Mescheder et al. [88] identified instances of eigenvalues near the imaginary axis in practical scenarios. This observation does not definitively address whether such proximity to the imaginary axis is a prevalent phenomenon. Furthermore, it does not conclusively establish whether these eigenvalues are the fundamental cause

behind the training instabilities that practitioners commonly encounter in their GAN training endeavors. Moreover, Nagarajan et al. [93] contributed a partial response to this query. They demonstrated that in the context of absolutely continuous data and generator distributions, these findings establish that GANs exhibit local convergence for sufficiently small learning rates. However, to emphasize, this assertion relies on the premise of absolute continuity.

Goodfellow et al. [44] presented the basic GAN architecture, other researchers also advanced more variants with some architectural differences. However, due to the potentially limiting overlap between real and generated data distributions, the Jensen-Shannon divergence presented in the objective function can become a constant value. It is this phenomenon that contributes to the challenge of the vanishing gradient, hindering effective training of GANs when employing gradient descent methods. To address the vanishing gradient problem, the Wasserstein GAN (W-GAN) was introduced [6], using the EarthMover distance instead of the Jensen-Shannon divergence to compare real and generated data distributions. W-GAN employs a critic function 'f' with a Lipschitz constraint as its discriminator, significantly improving GAN training stability. However, W-GAN may still face issues like suboptimal sample generation and occasional convergence problems in specific cases. In order to restrict the discriminative capacity of the discriminator, an alternative approach has also been introduced by [105] in the form of Loss-Sensitive GAN (LS-GAN). Both W-GAN and LS-GAN retain the fundamental GAN architecture.

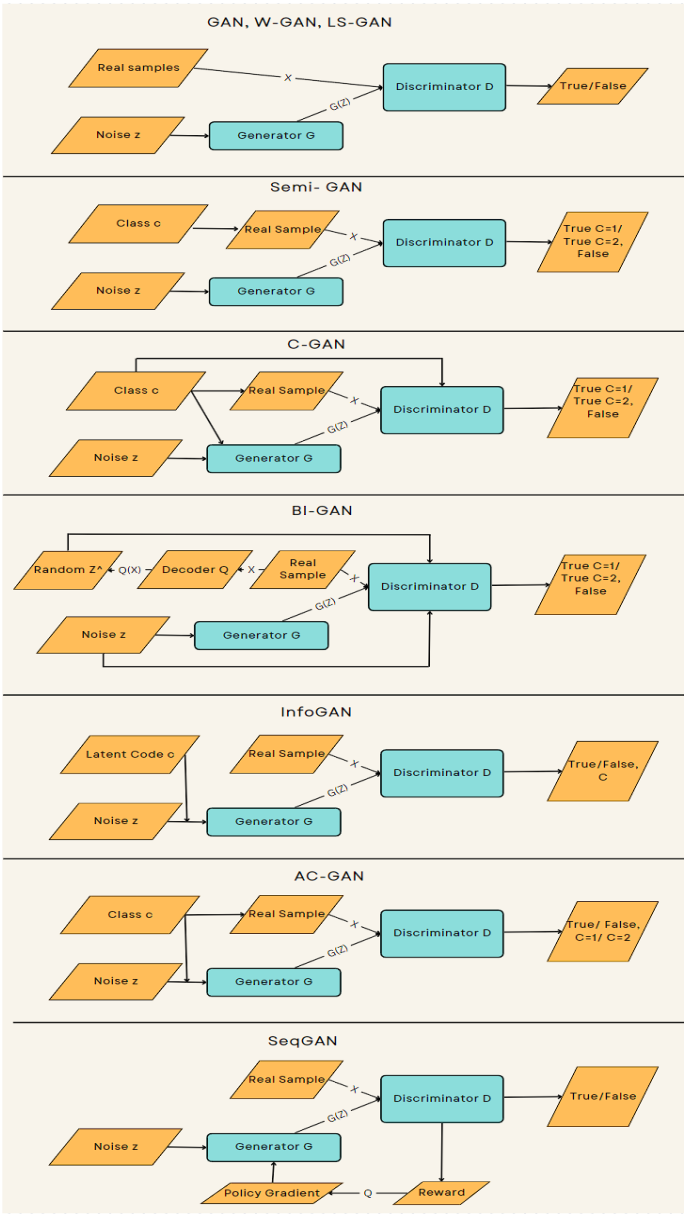


Fig. 2 Basic GANS Variants

Moreover, Qi et al. [97] introduced the Semi-GAN model, which involves the incorporation of real data labels into the discriminator's training process. Additionally, semi-GAN is an approach involving the integration of auxiliary information 'y' into both the generator 'G,' the discriminator 'D,' and the real data 'x' for the discriminator [91]. This auxiliary information can encompass labels or other supplementary data.

In the context of conventional GANs, the primary objective revolves around acquiring a generative model capable of mapping latent variable distributions to intricate real data distributions. Expanding upon this concept, Donahue et al. [33] propose Bidirectional GANs (BiGANs) to facilitate the mapping of real data to the latent variable space, thereby enabling feature learning. BiGANs extend the fundamental GAN structure by incorporating an additional decoder 'Q,' which facilitates the transformation of real data 'x' into the latent space. Consequently, this modification transforms the optimization problem into the form $\min_{G,Q} \max_D f(D, Q, G)$.

Chen et al. [21] introduce InfoGAN to capture mutual information between a subset of latent variables and observed data.

In InfoGAN, the correlation is quantified using $I(c; G(z, c))$, where c is a latent code and $G(z, c)$ is the generated output. The objective function is:

$$\min_G \max_D \{f(D, G) - \lambda I(c; G(z, c))\}$$

Here, $f(D, G)$ includes the adversarial term and penalty term $\lambda I(c; G(z, c))$. The goal is to minimize the generator's loss while maximizing the discriminator's loss with respect to mutual information.

Due to the challenge of computing $p(c|x)$, a lower bound estimation is used through variational information maximization.

Odena et al. [98] introduce the auxiliary classifier GAN (AC-GAN) approach tailored for semi-supervised synthesis. Their formulated objective function comprises two integral components: the logarithmic likelihood related to the accurate data source and the corresponding accurate class. The essence of AC-GAN lies in its capacity to seamlessly integrate label information into the generator and to adapt the discriminator's objective function accordingly. This integration yields noticeable enhancements in the generative and discriminative capabilities of the GAN framework.

In another context, Yu et al. [150] introduced SeqGAN, a pioneering framework for sequence generation using GANs. It extends GANs to handle discrete token sequences, treating the generator as a stochastic policy in reinforcement learning. SeqGAN employs policy gradient-based mechanisms to enhance sequence generation by effectively propagating errors from the discriminator. These advancements build upon the foundational work of GANs [44].

3.2 Earlier Transformer Variants

The concept of Transformers was introduced by Vaswani et al. [134]. It was a revolutionary step in the field of generative AI specifically in natural language processing and generating synthetic content. The basic concept of the Transformers was introduced in [125] by Sutskever et al as a sequence modelling technique. The basic technique of pretraining transformers was introduced and used as a state-of-the-art technique by [106]. This was used in answering different queries and also used as a chatbot to give results that are competitive and accurate. Indeed, these early Transformer Developments paved the way for State-of-the-Art NLP Chatbots.

3.3 Earlier Variational Autoencoder Variants

Variational Autoencoders [72] is one of the oldest techniques of unsupervised learning and generative modeling. The foundational work of Variational Autoencoders was done by Kingma at el in [66]. The Variational Autoencoders model combines the probabilistic modeling with the basics of Autoencoders. This concept not only learns the properties of the latent space but they also learn the probabilistic distribution of it, which gives them the ability to generate new synthetic data samples. Early variants of autoencoders also include denoising autoencoders which use denoising techniques whilst trained locally to get rid of corrupted versions of their inputs [135]. Moreover, [84] used the auto-encoders coupled with the convolutional network to solve the image recognitional problems. Without a doubt, the pioneering work of [66] has paved the way for numerous subsequent developments and applications of VAEs in a wide range of domains, including image generation, natural language processing, and more.

3.4 Earlier Diffusion Model Variants

Diffusion-based models employ a sequential diffusion process to iteratively transform simple data distributions into complex, high-dimensional ones. The Non-Linear independent component estimation (NICE) introduced the concept of invertible transformations as a foundation for generative artificial intelligence [31]. This was followed by Real NVP (Real Non-Volume Preserving), which expanded the capabilities by incorporating neural networks into the transformation process [32]. Additionally, Glow (Generative Latent Optimization) extended these ideas to high-resolution image generation, highlighting the potential of diffusion-based models in computer vision [63]. Furthermore, Diffusion Probabilistic Models (DPMs) leveraged the diffusion process to model the likelihood of data samples, making it an essential contribution to the development of diffusion-based generative models. Continuous-time flows (CTFs) diffusion models ventured into continuous-time modelling using stochastic differential equations [45]. These earlier works have laid the foundation for an exciting and rapidly evolving field of generative modelling using diffusion-based techniques.

4 Advancements in Generative AI and Their Diverse Applications

4.1 Generative AI for Image Translation

Image translation [44] is becoming a rapidly growing technology, particularly within the realm of medical applications. This innovation holds remarkable potential, not only in terms of cost-saving implications related to equipment usage but also in the facilitation of informed medical decisions.

The performance of generative AI models, particularly in the subfield of image translation, is typically assessed using specialized datasets. Among these, two notable datasets stand out: ImageNet [147], ClebA [82], and in the field of medical science: MIMIC [54], BRATS [86], FastMRI [129] and ChestX-ray [139] Each of these datasets is uniquely designed to challenge and evaluate the models' abilities to accurately and

effectively translate images, providing a comprehensive benchmark for their performance capabilities. These datasets are freely accessible to researchers for testing their models, under certain conditions. Users must properly cite the source of the dataset in their work. For datasets containing medical data, researchers are required to sign a Data Use Agreement. This agreement sets forth guidelines on the appropriate usage and security of the data and strictly prohibits any attempts to identify individual patients. This ensures that while fostering innovation and research, the datasets are used ethically and responsibly.

The utilization of AI-driven image translation yields images that are not only more polished and precise but also empower medical professionals with a heightened ability to discern critical information [146]. Yan et al. proposed a GANs-based model that uses the *Swin Transformers* in the Generator. The Swin Transformer represents a notable stride forward in the evolution of architecture. Its most remarkable enhancement entails the replacement of the conventional multiple self-attention (MSA) modules with an innovative shift window-based module while keeping the remaining layers largely unchanged. This transformer-based generator allows for the production of the output content which is the same as source images and the same information required by the target image. They tested the model using the BraTs2018 [86] and FastMRI [118] datasets. The Swin-based Transformers method attains its highest level of performance in the specific task of converting T1 mode to T2 mode images using the clinical brain MRI dataset. Moreover, they conducted evaluations using the unpaired *BraTs2018* dataset (see the results depicted in the Figure 3). These highlight that the innovative MMTrans approach stands out as the leader in terms of translation performance.

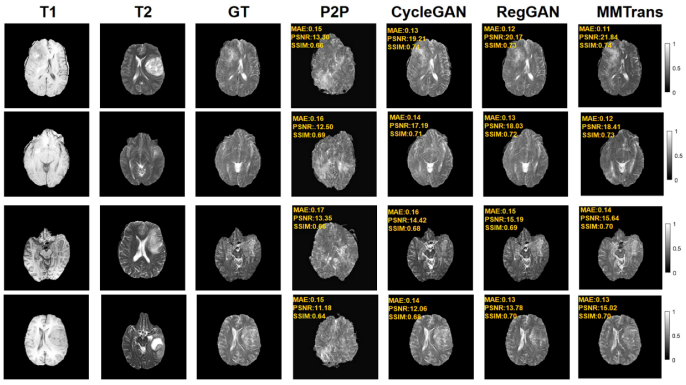


Fig. 3 qualitative outcomes obtained from various translation techniques employed to generate T2 images from T1 images within the unpaired BraTs2018 dataset, Source: [146]

Indeed, Figure 3 shows that the introduced MMTrans method performs better in translating the image when compared to other methods such as Pix2Pix [52], CycleGAN [159] and RegGAN [114]. The Image of MMTrans is closest to the Ground Truth (GT).

Furthermore, Dar et al. [28] used Conditional GANs to solve the problem of Image translation in MRI. This variant of GANs was introduced for image-image translation with the conditional label given to both generator and discriminator to instruct about what they have to forge and to predict real or fake respectively [52]. In fact, there are two types of GAN variants present for the purpose of image translation, Pix2Pix GAN is a conditional GAN [52], where the generator takes both an input image and a target condition as input and then generates an output image that adheres to the specified condition. However, for these types of GANs pixel-aligned images are required which is quite difficult to acquire [146]. Moreover, unpaired GANs do not require corresponding pairs of images for training. Instead, they focus on learning the mapping between two domains by using cycle consistency as a constraint [131].

Indeed, unpaired image translation presents a significant challenge. The objective is to ensure that translating an image from one domain to another should not compromise its fundamental characteristics (e.g., allowing for a seamless reversion back to its initial state)[159]. To address this issue, the Cyclic Generative Adversarial Network (Cyclic GAN) was developed [159]. Torbunov et al. [131] developed a Vision Transformer based GAN (UVCGAN). This works on the principle of cyclic GANs and without sacrificing the image regeneration capabilities gives better results than previous simple cyclic models. The evaluation of image-to-image translation performance commonly employs two widely accepted metrics, namely Frechet Inception Distance (FID) [46] and Kernel Inception Distance (KID) [13]. These metrics quantify the similarity between the translated images and those within the target domain, with a lower score indicative of higher similarity. UVCGAN model's superior performance is evident across most image-to-image translation tasks, as illustrated in Table 1. Operating similarly to a CycleGAN-like model, their approach consistently produces translated images that exhibit strong correlations with the input images, capturing essential aspects like hair color and facial orientations (as exemplified in Figure 4).

The high-quality result produced by UVCGAN in Figure 4 is paramount for enhancing scientific simulations. It can be observed that translations generated by ACL-GAN and Council-GAN tend to overly emphasize features that aren't pivotal for achieving the intended translation, such as non-essential attributes like hair color, background color, and length. Even in some cases, Council-GAN changed the background.

Table 1 FID and KID scores of UVCGAN and other models. Lower is better, Source:[131]

Model	Selfie to Anime		Anime to Selfie	
	FID	KID (×100)	FID	KID (×100)
ACL-GAN	99.3	3.22 ± 0.26	128.6	3.49 ± 0.33
Council-GAN	91.9	2.74 ± 0.26	126.0	2.57 ± 0.32
CycleGAN	92.1	2.72 ± 0.29	127.5	2.52 ± 0.34
U-GAT-IT	95.8	2.74 ± 0.31	108.8	1.48 ± 0.34
UVCGAN	79.0	1.35 ± 0.20	122.8	2.33 ± 0.38
Model	Male to Female		Female to Male	
	FID	KID (×100)	FID	KID (×100)
ACL-GAN	9.4	0.58 ± 0.06	19.1	1.38 ± 0.09
Council-GAN	10.4	0.74 ± 0.08	24.1	1.79 ± 0.10
CycleGAN	15.2	1.29 ± 0.11	22.2	1.74 ± 0.11
U-GAT-IT	24.1	2.20 ± 0.12	15.5	0.94 ± 0.07
UVCGAN	9.6	0.68 ± 0.07	13.9	0.91 ± 0.08
Model	Remove Glasses		Add Glasses	
	FID	KID (×100)	FID	KID (×100)
ACL-GAN	16.7	0.70 ± 0.06	20.1	1.35 ± 0.14
Council-GAN	37.2	3.67 ± 0.22	19.5	1.33 ± 0.13
CycleGAN	24.2	1.87 ± 0.17	19.8	1.36 ± 0.12
U-GAT-IT	23.3	1.69 ± 0.14	19.0	1.08 ± 0.10
UVCGAN	14.4	0.68 ± 0.10	13.6	0.60 ± 0.08

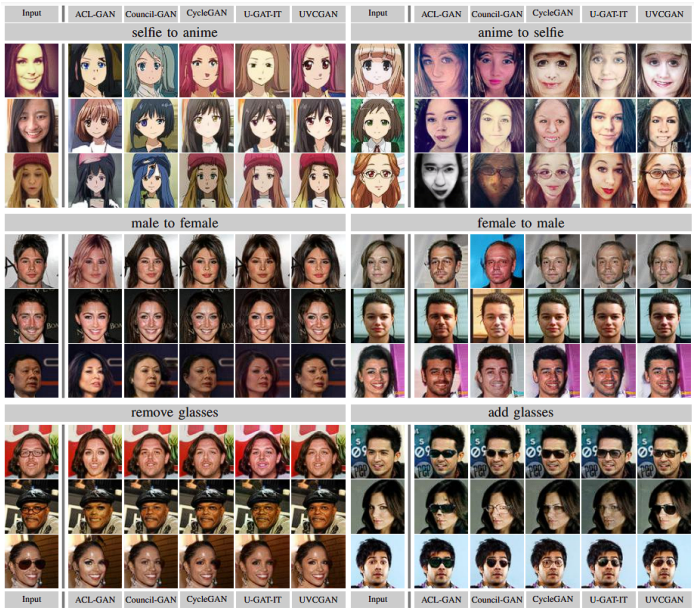


Fig. 4 unpaired UVCGAN vs Others image-to-image translation, Source: [131]

Another application under the umbrella of image translation is the Synthetic Aperture Radar (SAR) image translation [155]. Both the techniques of paired and unpaired GANs are used for this application. Indeed, the SAR-generated images are

not that visually clear but they can be captured at any time so it's a better technique when compared to optical imaging. Particularly, if it is combined with the GANs image translation methods to convert the SAR-generated image into an optical high-resolution clear image [155]. Different GANs methods of unpaired image translation such as CycleGAN [159], NICE GANs [50], Attn-CycleGAN [79] and paired GANs such as Pix2Pix [52], and Bicycle GANs [160] are used for Satellite image translation. In Wei et al. [141] the researchers presented a new technique using the generative AI that is specifically designed for the translation of unpaired SAR images to optical images. In detail, they introduced an approach known as Cross-Fusion Reasoning and Wavelet Decomposition GAN (CFRWD-GAN) [141]. The primary objective of CFRWD-GAN is twofold: to effectively retain structural intricacies and elevate the quality of high-frequency band details. This is achieved through a unique framework that integrates cross-fusion reasoning (CFR) structure, adept at preserving both high-resolution, fine-grained features and low-resolution semantic attributes throughout the entire process of feature reasoning. Additionally, to address speckle noise inherent in SAR images, the method employs discrete wavelet decomposition (WD), enabling the translation of high-frequency components. Through the convergence of these techniques, CFRWD-GAN demonstrates its capability to significantly enhance the translation process for unpaired image-to-image scenarios. The model was evaluated using Root Mean Squared Error (RMSE) [53, 117], structural similarity index (SSIM) [73, 140], peak signal-to-noise ratio (PSNR) [120, 128], learned perceptual image patch similarity (LPIPS) [119, 152] and produced a better result than the other state of art models.

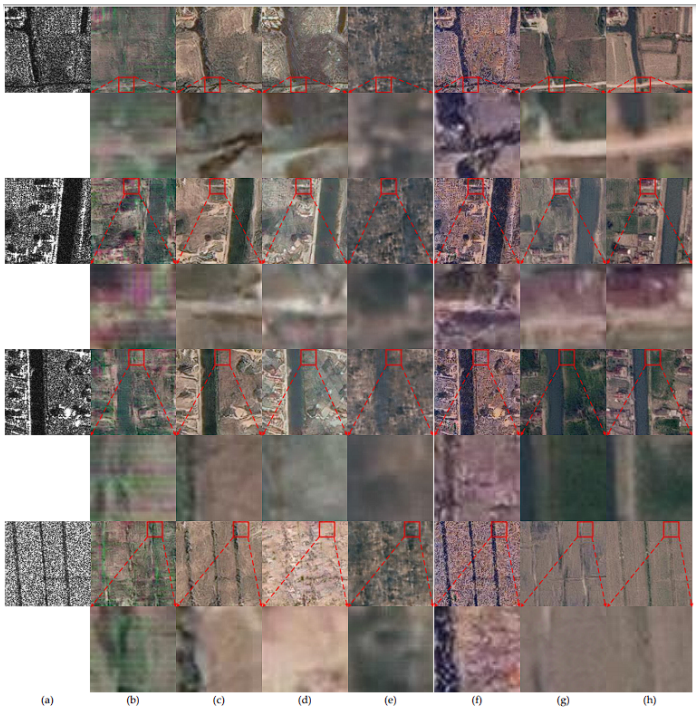


Fig. 5 Highlighted with red boxes and magnified for emphasis, the images are presented in the following order: (a) SAR images, (b) Pix2Pix, (c) CycleGAN, (d) S-cycle-GAN, (e) NICE-GAN, (f) GANILLA, (g) CFRWD-GAN, and (h) ground truth optical images., Source: [141]

The other technique for image generation and translation is Variational Autoencoders developed by Kingma et al. [66]. Furthermore, Zhu et al. [161] compared the generative capabilities of Conditional Variation Autoencoders and also compared it with the other generative techniques for image generation and translation. The main idea of this work was to find the algorithm that performs the best in balancing the diversity and realism in the generated data. The best-performing model in their work was the Bicycle GANs.

Table 2 Generative AI and Its Applications in Image Translation

Domain	Methods	References
Medical-MRI	MM-Transformers, Cyclic-GAN, Pix2Pix GAN, VAE	[1, 5, 17, 28, 30, 144, 146]
Satellite Image Translation	Cyclic-GAN, Pix2Pix GAN, NICE-GAN, Attn-CycleGAN, PSGAN	[80, 101, 155]
Facial Expression Editing	VAE, UPGPT	[22, 39, 149]
Style Transfer	VAE, GANs, DD-GAN	[9, 126, 162]
Text-to-Image Translation	TextControlGAN	[70]
Image Upscaling	GIGA-GAN	[58]

VAE was also used for the molecule generations, generating the 3-Dimensional synthetic molecule structure [151]. Jumper et al. [56] developed an architecture of a generative algorithm specifically for predicting the structure of molecules and proteins. This is the best approach till now in generative AI techniques for generating and predicting molecular architectures.

4.2 Generative AI for Video Synthesis and Generation

Generative AI has transformative applications in the field of video and animation, enabling the creation of visually stunning and dynamic content. The evaluation of video generative models often involves a set of widely recognized datasets: Voxcleb [94], HDTF [154], ClebV [142], Kinetics [59], UCF101 [123], and for specifically audio performance testing VCTK Corpus[132] and LibriSpeech [100]. These datasets are publicly available to researchers, with the stipulation that any use of these resources must include proper citation of the source.

In Hong et al. [49] the researchers introduced a GANs variant called depth-aware GAN. This model provided strong competition to the state of art models and the problem of replacing the face in a video. In detail the dataset on which they tested the built model and compared the results was called Voxcleb [94] and ClebV [142]. The model produced better results than the present state-of-the-art models for achieving talking head video face generation.

Indeed, the replacement of a face in a talking head video is the most predominant application in the video generation problem. Many advancements have been made in generative Artificial intelligence to master this application. Hong et al. [48] state that all this model needs is a target video and a 2-D picture with good pixels and facial features and it will translate the video expression features into the static picture supplied. The *DaGAN++* framework proposed by [48] comprises three key components: (a) an uncertainty-aware face depth learning network that reconstructs detailed 3D facial geometry from self-supervised face videos, without requiring camera parameters or explicit 3D annotations (b) geometry-guided facial keypoint detection, which employs the facial depth network to estimate depth maps, is used alongside RGB images for accurate facial keypoint estimation (c) a geometry-enhanced multi-layer generation process that incorporates learned motion fields, occlusion maps, and facial geometry into each layer of image generation through cross-modal geometry-guided attention. This comprehensive approach enables the synthesis of images enriched with geometry-related attributes derived from facial videos.

It is evident from Figure 6 that DaGAN++ exhibits superior capabilities. Notably, DaGAN++ excels in capturing expression-related facial movements within the driving frame, with heightened accuracy observed in regions such as the eyes and mouth. This performance enhancement can be attributed to the precise facial geometry estimation, which greatly contributes to the refinement of expression-related facial motions. Furthermore, Min et al. [90] introduce StyleTalker, an innovative audio-driven talking head generative model designed to synthesize a talking person's video using a single reference image. It features highly accurate lip synchronization, realistic head poses, and natural eye blinks synchronized to the provided audio. To achieve this, they leverage a pre-trained image generator and an image encoder to estimate latent codes for the



Fig. 6 DaGAN++ vs other state of art model on HDTF dataset [154], Source: [48]

talking head video that align faithfully with the given audio input. This achievement stems from the integration of novel components which includes a contrastive lip-sync discriminator. This ensures precise lip synchronization, a conditional sequential variational autoencoder that captures a motion space disentangled from lip movements. In more detail, it enables independent manipulation of motions and lip movements while preserving identity. In addition, it affords an auto-regressive prior enhanced with normalizing flow, facilitating the acquisition of a complex audio-to-motion multi-modal latent space. With these components in place, StyleTalker has the capacity to produce talking head videos, both in a motion-controllable manner when another motion source video is available. Also, entirely driven by audio inputs, wherein it infers real motions from the provided audio.



Fig. 7 StyleTalker vs Other models, Source: [90]

The qualitative evaluation of audio-driven talking head generation performance on VoxCeleb2 dataset in Figure 7 reveals distinct differences. In the first row (marked by a yellow box), frames corresponding to the provided audio are displayed. Conversely, the single image input (highlighted in a red box) represents a reference image of the desired target identity. When observing the generated video frames produced by the StyleTalker in comparison to those generated by other audio-driven generation

models [103, 138, 157], a notable distinction becomes evident. StyleTalker consistently produces talking head videos of exceptional quality, skillfully preserving the distinctive identity of the intended.

Advancing further, Li et al. [76] introduced Multiscale Vision Transformers (MViTv2) as a comprehensive architectural framework suitable for tasks encompassing image and video classification, along with object detection. Within this study, the researchers introduce an enhanced version of MViT, featuring decomposed relative positional embeddings and residual pooling connections. By deploying this upgraded architecture across five different scales, we meticulously assess its performance in scenarios such as ImageNet classification, COCO object detection, and Kinetics video recognition. Remarkably, MViTv2 surpasses previous benchmarks in terms of effectiveness and provided video classification accuracy of 86.1% on the Kinetics-400 dataset as shown in Table 3.

Table 3 Comparative analysis with other Models on the Kinetics-400 [59] dataset, Source: [76]

Model	**Top-1**	**Top-5**	**FLOPs×views**	**Param**
SlowFast 16×8 +NL [37]	79.8	93.9	234×3×10	59.9
X3D-XL [36]	79.1	93.9	48.4×3×10	11.0
MoViNet-A6 [69]	81.5	95.3	386×1×1	31.4
MViTv1, 16×4 [35]	78.4	93.5	70.3×1×5	36.6
MViTv1, 32×3 [35]	80.2	94.4	170×1×5	36.6
MViTv2-S, 16×4 [76]	**81.0**	**94.6**	64×1×5	34.5
MViTv2-B, 32×3 [76]	**82.9**	**95.7**	225×1×5	51.2
ViT-B-VTN in 21k [96]	78.6	93.7	4218×1×1	114.0
ViT-B-TimeSformer [12] in 21k	80.7	94.7	2380×3×1	121.4
ViT-L-ViViT [7] in 21k	81.3	94.7	3992×3×4	310.8
Swin-L$^+$ in 21k [83]	84.9	96.7	2107×5×10	200.0
MViTv2-L$^+$, 40×3, in 21k [76]	**86.1**	**97.0**	2828×3×5	217.6

In the context of the Kinetics-400 dataset [59], Table 3 presents a comparison between MViTv2 and previous methodologies, encompassing both state-of-the-art Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). Upon training from the ground up, MViTv2-S and MViTv2-B models exhibit top-1 accuracy of 81.0% and 82.9% respectively, surpassing their MViTv1 [35] counterparts by +2.6% and +2.7%. Notably, earlier ViT-based models necessitate substantial pre-training on the IN-21K dataset to achieve peak accuracy on Kinetics-400 as shown in Table 3 (see last 5 rows). In contrast, MViTv2 achieves an exceptional top-1 accuracy of 86.1% when fine-tuning, the MViTv2-L model with a large spatio-temporal input size of 40×3 (time $\times$ space^2).

The rest of the reviewed applications of Generative Artificial Intelligence in the field of Video Generation are given in Table 4

Table 4 Generative AI and Its Applications in Video Generation

Domain	Methods	References
Face swapping videos	Depth Awareness GANs, DaGAN++	[48, 49]
Video/Image Classification	MViTv2	[76]
Audio-Based Facial Expression video Translation	Styletalker Sequential VAE	[90]
Simulation in Metaverses	Multi-task DT offloading model	[144]
ECG Synthesis to Improve Deep ECG Classification	SimGAN	[43]
3D human motion prediction	HP-GAN	[11]

4.3 Generative AI for Natural Language Processing

Generative AI models have demonstrated remarkable achievements across a spectrum of Natural Language Processing tasks. These encompass language comprehension, logical reasoning, and text generation.

In the domain of natural language processing, specific datasets have become standard benchmarks for evaluating state-of-the-art models in various tasks. For Named Entity Recognition, the CoNLL-2003 dataset [116] is frequently utilized. In the area of text summarization, two prominent datasets are DUC 2002 [85] and QMSUM [156]. Additionally, for Natural Language Inference (NLI) tasks across multiple languages, the XNLI dataset [24] serves as a crucial resource. All of these datasets are publicly available, providing researchers with essential tools to advance and assess the capabilities of their models in these specific NLP tasks.

Presently, a significant query the AI community poses revolves around the extent and confines of these model's capabilities [2]. Ahuja et al. [2] raise the question that most of the large language models are made and tested on only the English language. Therefore these researchers took the state-of-the-art models and trained them on other languages on certain available datasets and compared their question-answering and classificational accuracies [2]. Generative AI even has found its way into education[19, 34]. In the midst of this dynamic backdrop that challenges conventional modes of thinking, recent research endeavors investigating the implications of generative AI within the educational landscape yield valuable insights. Notably, these studies shed light on the opportunities and obstacles arising from the integration of generative AI. For instance, in a recent article [130] highlights the need for a new and creative way of teaching that effectively incorporates the progress brought by AI. They mentioned the significance of cultivating an ethical and personalized chatbot solution while augmenting digital proficiency to fully harness the manifold benefits of AI. Furthermore, the researchers advocate for the incorporation of AI literacy as an essential technological skill for navigating the complexities of the 21st century.

In the current landscape, Bozkurt [14] advocates for a significant reevaluation of the roles played by human educators and AI within the educational realm. They assert that the emergence of AI offers a unique juncture to redefine these roles. Especially, as

AI possesses the capacity to assume an increasing array of educational tasks that were traditionally the exclusive domain of human educators. This perspective underscores the importance of adopting a forward-thinking outlook that reconsiders the contributions of both technology and human educators to the educational process. Bozkurt [14] also emphasizes that generative AI's arrival presents a propitious opportunity to redefine these roles further.

The researchers further delve into the opportunities and challenges ushered in by the advent of generative AI. Generative AI, they elaborate, provides a diverse range of opportunities. These encompass personalized learning, fostering inclusive curriculum provision, and enhancing collaboration and cooperation throughout educational processes. Also, they can cater to automated assessment benefits, ensuring improved accessibility, optimizing efficiency in terms of time and effort, cultivating language skills, and enabling the round-the-clock availability of these technologies.

Generative AI can also enhance synthetic data generation, with the use of Transformers and GANs. It is clear that the understanding of data is better and now the AI is able to generate synthetic data even in the health field. For example, Frid-Adar et al. [40] used GANs to generate synthetic data for the liver lesion classification problem. The variant they used was called DCGAN. In another problem, Wang et al. [137] used the SSIM embedded-cycle GAN to count the number of people in the crowd. The dataset that they used for this problem was fully synthetically generated. It had all the weather conditions covered which made the model perform the best compared to the state-of-the-art models. Another technique of generative AI that is among the state-of-the-art techniques is BERT (Bidirectional Encoder representations from Transformers) which was introduced by the Google AI team [29]. It represents a significant advancement in pre-training techniques for NLP tasks. BERT is based on the transformer architecture and is designed to capture contextual information from both the left and right sides of a word in a sentence, hence the term 'bidirectional'. In Bert's pre-training, the model learns to predict missing words in a sentence by training on a large corpus of text. This helps BERT develop a deep understanding of syntax, semantics, and context. After pre-training, the model is fine-tuned on specific downstream tasks, such as sentiment analysis, question answering, and named entity recognition, using task-specific labelled data.

Table 5 shows the BERT model tested with the CoNLL-2003 [116] dataset for the task focusing on named entity recognition. BERTLarge demonstrates strong competitiveness with state-of-the-art techniques. The most successful approach involves concatenating token representations from the uppermost four hidden layers of the pre-trained Transformer. Remarkably, this approach lags by only 0.3 F1 behind the performance achieved by fine-tuning the complete model. This finding underscores the effectiveness of BERT for both fine-tuning and feature-based methodologies.

Another model related to Generative Natural language processing is ELMo [55]. This stands for 'Embeddings from Language Models'. ELMo utilizes a bidirectional LSTM (Long Short-Term Memory) network [47] for contextual word embeddings. In fact, ELMo embeddings have been shown to be effective in improving the performance of various NLP tasks, including sentiment analysis, question answering, and named entity recognition. The ability to capture context-specific information makes ELMo

Table 5 BERT vs others for Named Entity Recognition task on the CoNLL-2003 [116] dataset (comparison in terms of F1 Score on Validation DataSet (Dev) and Testing Dataset (Test), Source:[29]

System	Dev F1	Test F1
ELMo [55]	95.7	92.2
CVT [23]	-	92.6
CSE [3]	-	93.1
Fine-tuning approach		
BERTLARGE	96.6	92.8
BERTBASE	96.4	92.4
Feature-based approach (BERTBASE)		
Embeddings	91.0	-
Second-to-Last Hidden	95.6	-
Last Hidden	94.9	-
Weighted Sum Last Four Hidden	95.9	-
Concat Last Four Hidden	96.1	-
Weighted Sum All 12 Layers	95.5	-

embeddings particularly useful for tasks where word meanings can vary based on the surrounding context.

Another application of generative AI models used in natural language processing is malware classification. In particular, machine language malware classification is a big concern that can be solved by using generative AI. For example, Kale et al. [57] used the Bert and ELMo to train the embeddings of the models to classify the malware and the results provided remarkable improvements.

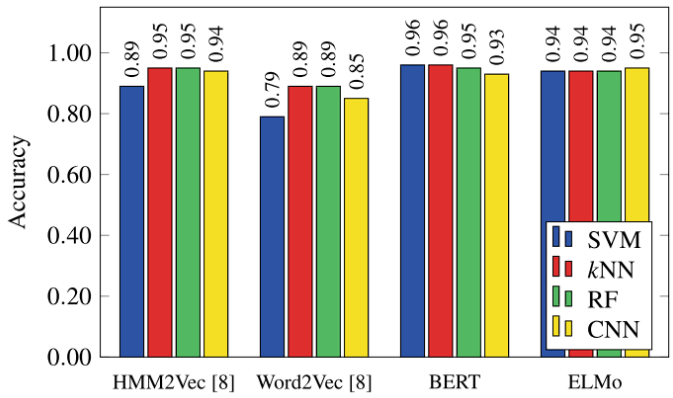


Fig. 8 HMM2Vec, Word2Vec, BERT, and ELMo based classification techniques, Source: [57]

Additionally, Figure 8 provides a summarized depiction of the optimal accuracies achieved by the classification methodologies based on HMM2Vec[20], Word2Vec[89],

BERT[29], and ELMo[102]. The graph illustrates that BERT-SVM and BERT-kNN secured the highest performance with a notable accuracy of 96%. In close pursuit HMM2Vec-kNN, HMM2Vec-RF, BERT-RF, and ELMo-CNN, achieved an accuracy of 95%. Slightly trailing behind, HMM2Vec-CNN, as well as all ELMo-based techniques, demonstrated an accuracy of 94%. Interestingly, the Word2Vec embeddings consistently yielded accuracies below 90% across all four classifiers.

A further application of Generative artificial intelligence is the text-summarization. Joshi et al. [112] introduced *Ranksum* which is an innovative technique designed for extractive text summarization of individual documents. This method hinges on the fusion of four distinct multi-dimensional sentence features, namely topic information, semantic content, significant keywords, and position. By independently acquiring sentence saliency rankings for each feature in an unsupervised manner, Ranksum subsequently amalgamates these scores through weighted fusion, yielding a comprehensive ranking of sentence significance. It is important to note that these scores are generated in a completely unsupervised manner.

Topic ranking is established through the application of probabilistic topic models, while semantic content is captured using sentence embeddings. Sentence embeddings are generated using Siamese networks to craft abstractive sentence representations, followed by a novel strategy to organize them based on their relative importance. To identify significant keywords and their associated sentence rankings within the document, a graph-based approach is employed. Additionally, a mechanism to gauge sentence novelty is formulated, relying on bigrams, trigrams, and sentence embeddings. This eliminates redundant sentences from the summary.

Table 6 Comparative analysis of RankSum with state-of-the-art algorithms conducted on the DUC 2002 [85] and QMSUM [156] dataset, Source: [112, 148]

Method	ROUGE-1	ROUGE-2	ROUGE-L
LEAD	43.6	21.0	40.2
ILP	45.4	21.3	42.8
NN-SE	47.4	23.0	–
SummaRuNNer	47.4	24.0	14.7
Egraph+coh	47.9	23.8	–
Tgraph+coh	48.1	24.3	–
URANK	48.5	21.5	–
SummCoder	51.7	27.5	44.6
HSSAS	**52.1**	24.5	**48.8**
CoRank	52.6	25.8	–
Rank-emb	49.9	24.8	45.6
Rank-topic	51.4	25.9	47.2
Rank-keyword	52.0	26.3	48.6
RankSum	**53.2**	27.9	**49.3**
PGNet on QMSUM	31.52	8.69	27.63
BART on QMSUM	32.18	8.48	28.56
HMNet on QMSUM	**36.06**	**11.36**	**31.27**
ChatGPT on QMSUM	**36.83**	**12.78**	**24.23**

Table 6 showcases the performance outcomes of the novel RankSum framework in comparison to other state-of-the-art algorithms on the DUC 2002 dataset, evaluated through ROUGE metrics. ROUGE metrics are typically used in the field of machine translation, text summarization, and other tasks where the quality of generated text needs to be evaluated automatically. The metrics involve comparing n-grams (sequences of n words) between the generated and reference texts. Common versions of ROUGE metrics include ROUGE-N (which considers overlapping n-grams), ROUGE-L (which focuses on the longest common subsequence) [78]. Ranksum achieves notable ROUGE-1, ROUGE-2, and ROUGE-L scores of 53.2, 27.9, and 49.3, respectively. Impressively, it outperforms all recent methods examined for this extractive text summarization dataset. Notably, this approach surpasses the highly accurate summarization systems, HSSAS [4] and Co-Rank, with a substantial margin of 0.6, 0.8, and 0.5 for ROUGE-1, ROUGE-2, and ROUGE-L scores respectively [112]. Additionally, the outcomes of PGnet [136], BART[75], HMNet[143], and ChatGPT[15] were assessed using the QMSUM dataset[156]. It is important to note that the current pinnacle of meeting summarization models, HMNet, achieves the most impressive performance in terms of ROUGE-L. This might be attributed to its cross-domain pretraining approach, which imparts HMNet with a heightened familiarity with the style of meeting transcripts [158]. However, it's worth highlighting that in the case of ROUGE-1 and ROUGE-2 metrics, ChatGPT emerges as the leader. ChatGPT excels due to extensive training on diverse data, enabling superior relationship comprehension in one to two grams, and boosting metric performance.

Model	en	ar	bg	de	el	es	fr	hi	ru	sw	th	tr	ur	vi	zh	avg
Fine-tuned Baselines	80.8	64.3	68.0	70.0	65.3	73.5	73.4	58.9	67.8	49.7	54.1	60.9	57.2	69.3	67.8	65.4
Prompt-Based Baselines	67.5	60.7	46.5	54.0	47.4	61.2	61.4	56.8	53.3	50.4	43.8	42.7	50.0	61.0	56.7	54.2
Open AI Models	76.2	59.0	63.5	67.3	65.1	70.3	67.7	55.5	62.5	56.3	54.0	62.6	49.1	60.9	62.1	62.1
mBERT	80.8	64.3	68.0	70.0	65.3	73.5	73.4	58.9	67.8	49.7	54.1	60.9	57.2	69.3	67.8	65.4
mT5-Base	84.7	73.3	78.6	77.4	77.1	80.3	79.1	70.8	77.1	69.4	73.2	72.8	68.3	74.2	74.1	75.4
XLM-R Large	88.7	77.2	83.0	82.5	80.8	83.7	82.2	75.6	79.1	71.2	77.4	78.0	71.7	79.3	78.2	79.2
TuLRv6 - XXL	93.3	89.0	90.6	90.0	90.2	91.1	90.7	86.2	89.2	85.5	87.5	88.4	82.7	89.0	88.4	88.8
gpt-3.5-turbo	76.2	59.0	63.5	67.3	65.1	70.3	67.7	55.5	62.5	56.3	54.0	62.6	49.1	60.9	62.1	62.1
gpt-3.5-turbo (TT)	76.2	62.7	67.3	69.4	67.2	69.6	69.0	59.9	63.7	55.8	59.6	63.8	54.0	63.9	62.6	64.3
text-davinci-003	79.5	52.2	61.8	65.8	59.7	71.0	65.7	47.6	62.2	50.2	51.1	57.9	50.0	56.4	58.0	59.3
text-davinci-003 (TT)	79.5	65.1	70.8	71.7	69.3	72.2	71.8	63.3	67.3	57.3	62.0	67.6	55.1	66.9	65.8	67.1
gpt-4-32k	84.9	73.1	77.3	78.8	79.0	78.8	79.5	72.0	74.3	70.9	68.8	76.3	68.1	74.3	74.6	75.4

Table 7 Performance comparison among different models on all languages within the XNLI dataset, Source:[2].

Ahuja et al. [2] tested the performance of the state-of-the-art models on multilingual XNLI dataset [24] data. The results are given in Table 7, TuLRv6 - XXL achieves the highest average accuracy across all languages (88.8%). It performs exceptionally well in most languages, with accuracy scores consistently above 85%, XLM-R Large is the model that comes in second place with an average accuracy of 79.2%. While not quite as high as TuLRv6, it still maintains a strong performance across all languages and demonstrates its multilingual capabilities. With an average accuracy of 75.4%, mT5-Base takes the third spot. GPT-4-32k achieves an average accuracy of 75.4%. It exhibits consistent performance across languages, demonstrating its effectiveness in handling multilingual tasks.

Table 8: Generative AI Applications in Natural Language Processing: Major Papers and Descriptions

Paper	Field	Description	Citation
"Attention Is All You Need"	NLP / Machine Translation	Introduces the Transformer model using self-attention mechanisms for various NLP tasks, revolutionizing sequence-to-sequence models	[134]
"BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension"	NLP / Text Generation	Introduces BART, a sequence-to-sequence model pre-trained using denoising autoencoders, capable of various NLP tasks	[75]
"CTRL: A Conditional Transformer Language Model"	Language Generation	Proposes a generative model (CTRL) that can condition its output on specific attributes, enabling fine-grained control over text generation	[61]
"T5: Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer"	NLP / Transfer Learning	Presents T5, a model that casts all NLP tasks as a text-to-text problem, achieving state-of-the-art results across diverse tasks	[111]
"GPT-2: Language Models are Unsupervised Multitask Learners"	Language Generation	Describes the GPT-2 model, a large-scale generative model that demonstrates impressive text generation capabilities across a range of tasks	[110]
"LayoutLM: Pre-training of Text and Layout for Document Image Understanding"	Document Analysis	Presents LayoutLM, a model that pre-trains on document images with associated text, improving document understanding tasks	[145]
"ERNIE: Enhanced Language Representation with Informative Entities"	NLP / Knowledge Enhancement	Introduces ERNIE, a model that enhances language representations by incorporating knowledge from knowledge bases	[153]
"DALL·E: Creating Images from Text"	NLP / Image Generation	Introduces DALL·E, a generative model capable of generating images from textual descriptions	[113]

Continued on next page

Table 8 – continued from previous page

Paper	Field	Description	Citation
"CLIP: Connecting Text and Images for Supervised Learning"	NLP / Vision	Proposes CLIP, a model that learns to understand images and text jointly, achieving impressive results in cross-modal tasks	[108]
"WebGPT: Browser-assisted question-answering with human feedback"	NLP, Human feedback	Introduces an approach that leverages a text-based web-browsing environment, enabling the model to access and navigate online resources. The methodology is structured in a manner that aligns with human capabilities, thus facilitating model training through imitation learning.	[95]
"GPT-4 Technical Report"	NLP/ Text Generation	GPT-4, a groundbreaking advancement, is introduced as a multimodal model with the ability to process both image and text inputs while generating text-based outputs. GPT-4's accomplishments encompass passing a simulated bar exam with a score that ranks within the top 10% of test takers.	[99]
"Let's Verify Step by Step"	Mathematical Reasoning	The introduced approach, centered around a process-supervised model, achieves a commendable success rate of 78% when addressing problems sourced from a representative subset of the MATH test set.	[77]

The landscape of Natural Language Processing (NLP) has witnessed remarkable advancements in recent years, driven primarily by the innovative applications of generative AI. The Table 8 highlights a selection of influential papers that showcase the evolution and impact of generative AI techniques within the NLP domain. These advancements have led to groundbreaking developments in various subfields of NLP, transforming the way we process and understand human language.

As evidenced by the papers presented, recent years have seen generative AI techniques reshape NLP in profound ways. These advancements not only enhance the quality and diversity of text generation but also enable more sophisticated control, cross-modal understanding, and knowledge integration. As the field continues to

evolve, it is likely that further innovations in generative AI will continue to drive NLP's progress, unlocking new frontiers of language understanding and generation across a wide array of applications.

4.4 Generative AI for Knowledge Graph Generation

Researchers and practitioners have leveraged the power of generative AI to enhance the creation and refinement of knowledge graphs—a structured representation of relationships between entities. This section explores the burgeoning landscape of generative AI applications within knowledge graph generation, highlighting pioneering research papers and their contributions to this evolving field. For evaluating models in the field of knowledge graph generation, a variety of datasets are employed, with Dbpedia [74], Cora Dataset [16] and Googles Knowledge graph [122] being among the most commonly used. However, the scope of available datasets extends beyond just these. These datasets are freely accessible, offering researchers and developers an invaluable resource to test and refine the capabilities of their knowledge graph generation models. A knowledge graph is a structured representation of information that captures relationships between entities and concepts. It goes beyond traditional databases by not only storing data but also organizing it in a way that highlights connections and context. Knowledge graphs, first introduced by Google in 2012 [124], are designed to model real-world relationships, making them a powerful tool for representing and querying complex information. Cai and Wang [18] introduced KBGAN, an innovative adversarial learning framework designed to enhance the performance of various existing knowledge graph embedding models. This approach is not dependent on the specific structures of the generator and discriminator, allowing for the incorporation of a wide range of knowledge graph embedding models as fundamental components. This enables KBGAN to significantly enhance the training dynamics and performance of existing knowledge graph embedding models. Liu et al. [81] introduce K-BERT, a novel approach that empowers language representation with knowledge graphs, enabling the incorporation of commonsense and domain-specific knowledge. The K-BERT methodology comprises two fundamental steps. Initially, knowledge from a knowledge graph (KG) is seamlessly integrated into a sentence, rendering it a knowledge-rich sentence tree. Subsequently, the utilization of soft-position and visible matrix techniques serves to regulate the extent of knowledge integration, thereby preventing any deviation from the original sentence meaning.

Despite the challenges presented by handling heterogeneous entity spans (HES) and keyphrases not seen in training (KN), the investigation yields promising outcomes across a spectrum of twelve open-domain and specific-domain natural language processing (NLP) tasks. Empirical evidence underscores the considerable efficacy of knowledge graphs, particularly in tasks that are driven by domain-specific knowledge. Moreover, K-BERT's compatibility with the model parameters of BERT offers a seamless integration of knowledge enhancement within a well-established framework.

Link prediction is a fundamental task involving the prediction of missing facts within a knowledge graph using available information. In this context, Balazevic et al. [10] introduce TuckER, a linear model that employs Tucker decomposition of the binary tensor representation of knowledge graph triples. Despite its straightforward

nature, TuckER demonstrates remarkable efficacy. It surpasses previous state-of-the-art models on widely recognized link prediction datasets, solidifying its position as a potent baseline for more sophisticated models in this domain. Moreover, Zeb et al. [121] have introduced ComplexGCN, an innovative graph convolutional network that leverages standard graph convolutional architecture(GCN) [67] to learn complex embeddings. Within the ComplexGCN framework, both node and relation features are projected into complex space through the use of learnable weights associated with neighboring nodes at each convolutional layer. To maintain the integrity of initial embedding information in the final node embeddings, a residual connection between the input and output of the convolutional stack is implemented. These researchers [121]

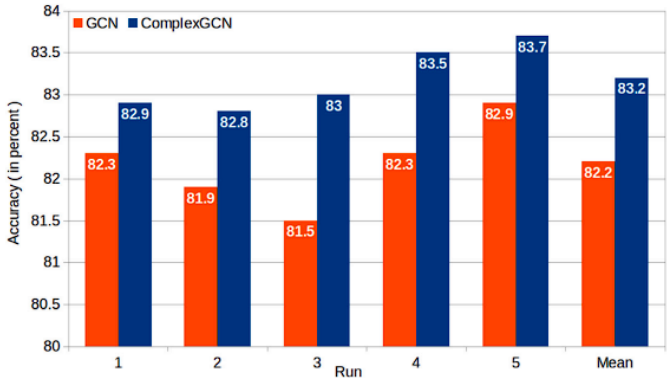


Fig. 9 GCN vs CGCN on node completion task- Cora Dataset, Source: [121]

conducted an evaluation of ComplexGCN's performance on the node classification task using the Cora dataset. This consisted of 2708 nodes categorized into 7 classes and 5429 edges representing citation links between documents. Both the standard GCN and ComplexGCN were trained on the Cora dataset for 200 epochs, with the objective of minimizing the cross-entropy loss. The training process was repeated 5 times for each model, and the results were averaged and reported in Figure 9. In terms of accuracy percentage, ComplexGCN exhibited improved performance compared to the standard GCN, achieving a 1% increase in mean accuracy.

The presented selection of papers highlights the innovative strides researchers have made in leveraging generative AI to enhance knowledge graph generation. From employing adversarial learning for improved embeddings to bridging the gap between unstructured text and structured knowledge, these papers showcase the multifaceted nature of the advancements.

4.5 Interdisciplinary Applications of Generative AI

The groundbreaking capabilities of generating synthetic data and content generation have given Generative AI the ability to be applicable in interdisciplinary fields. Further recent applications are discussed in this section. The synthetic data generation capability of generative AI is useful in mechanical fault detections. Gao et al. [42]

highlight a fault detection method that combined Finite Element Method (FEM)[26] simulations and Generative Adversarial Network (GAN) [44] to address two primary challenges. Firstly, it aimed to fill the gaps in fault samples by leveraging FEM simulations. Secondly, it sought to enhance fault detection accuracy by utilizing GAN to generate a substantial number of synthetic fault samples. Initially, FEM was employed to generate simulation signals of specific lengths to complete missing fault samples. Subsequently, GAN was utilized to create additional fault samples based on FEM simulations, resulting in a more comprehensive dataset. Finally, classifiers such as Support Vector Machines (SVM)[25], Extreme Learning Machines (ELM)[51], Decision Trees (DTree)[107], and others were employed to detect faults in cases where the faults were previously unknown.

Table 9 Comparison of Classification accuracy with and without GANs sampling, Source: [42]

Description	**FEM & AI**	**FEM, GAN & AI**
No of fault Samples	360	360
No of Synthetic Fault Samples	–	3240
Accuracy SVM (%)	84.44	86.11
Accuracy ELM (%)	89.72	91.67
Accuracy Decision Tree (%)	91.11	97.5
Average Accuracy (%)	88.42	91.76

In Table 9 the result clearly identifies the use of Generative AI to generate the synthetic samples which helps the machine learning models to achieve greater accuracy. In Feng et al. [38] a traffic generation model is developed using the Generative AI, named TrafficGen. The model outperforms the previous state-of-art techniques such as SeneGen [127] in generating the synthetic traffic scenarios. The introduced models were also able to generate the trajectory of the generated traffic and synthetic snapshots. This enabled the creation of numerous fresh traffic scenarios and the enhancement of the ones that already exist.

Music generation is also gaining popularity in the field of Generative AI, different methods such as Variational Auto Encoders, Transformers and Recurrent Neural Networks are being used in generating synthetic music, and different new approaches of music generation [60]. Another application is the Handwriting generation [41]. This proposed HiGAN+ which presented the capability to generate a wide range of authentic handwritten texts while being guided by arbitrary textual content and distinct calligraphic styles. These styles are separated from reference images or randomly drawn from a prior normal distribution. Traditional style transfer methods, which rely on pixel-level mappings, may not be suitable for HiGAN+, hence they introduce the contextual loss to notably enhance the stylistic consistency of generated images. The model performed very well in generating readable handwriting samples. In the field of software engineering, generative AI is introduced to help write better code and solve errors in the code, debugging and even writing the documentation of the work (e.g., Copilot) [92].

Emerging technologies in the realm of generative AI aim to simplify human life, yet they also underscore the imperative for responsible AI development. These innovations should prioritize ethical considerations, ensuring that their generated content aligns with societal values. The Contemporary efforts in the generative AI method development are increasingly conscious of these ethical dimensions. In a recent review, Pudari and Ernst [104] delved into Copilot and emphasized that generative AI, while valuable, won't supplant humans in the field of software engineering. This is because it struggles to grasp intricate software design principles and identify coding issues, known as 'code smells'. Instead, its role primarily revolves around aiding developers in crafting more efficient code. Researchers [8] highlight responsible AI as a comprehensive concept, mandating systematic adoption of AI principles. Besides explainability, it emphasizes fairness, accountability, and privacy in real-world AI model implementations, especially in scenarios involving sensitive information and regulatory demands for data privacy.

Apart from these discussed papers, a multitude of groundbreaking advancements are unfolding within the domain of generative AI. It is evident that in this rapidly advancing research, the previously discussed papers represent just a prominent subset of the recent developments in this field.

5 Challenges and opportunities of Generative AI

There are various domains for which we can discuss both the challenges and opportunities of Generative AI. Let's start with challenges and their proposed solutions followed by opportunities-

5.1 Challenges and their proposed solutions

Ethical Concerns:
Challenge: GenAI can be used for malicious purposes, such as the creation of deepfakes for identity theft or misinformation.
Solution: Establishing ethical governance structures, guidelines, and regulations to guide the responsible development and deployment of GenAI.
Security Concerns:
Challenge: There might be vulnerabilities in generative models that could be exploited for adversarial attacks.
Solution: The development of security measures to protect generative models from manipulation and continuous research into adversarial robustness.
Bias and Fairness:
Challenge: Generative models may amplify and perpetuate biases in the training data, leading to discriminatory and unfair outputs.
Solution: Extensive research and implementation of methods to detect and mitigate bias in training data, as well as encouraging inclusivity and diversity in datasets.
Data Privacy:
Challenge: Generative models trained on large datasets may inadvertently remember sensitive information, posing privacy risks.

Solution: Adherence to data protection regulations and implementation of privacy-preserving approaches to protect personal privacy.
Interpretability:
Challenge: Mostly, it is difficult to understand the decision-making process of generative algorithms due to their 'black boxes' nature.
Solution: Research and development of explainable AI approaches to improve interpretability and transparency and, permitting users to understand algorithm outputs.

5.2 Opportunities

Human-AI Collaboration: Collaborative work between GenAI and humans can lead to innovative solutions in design, problem-solving, and creativity.
Creative Expression: Generative AI facilitates innovative creative expressions, such as music, generative art, and literature.
Content Generation: Applications in content creation, for example- image synthesis, text generation, and video creation enhance productivity in numerous industries.
Education and Training: Generative models can be employed for simulating scenarios for training, creating interactive educational materials, and enhancing learning experiences.
Personalization: Generative algorithms can be used for personalized recommendations in entertainment, e-commerce, and other user-oriented domains.
Innovative Design: GenAI can assist in creating optimized and innovative designs in industries like architecture and product design.
Scientific Discovery: GenAI contributes to scientific investigation by simulating complex systems, predicting outcomes, and generating hypotheses.
Healthcare Applications: GenAI advances personalized medicine, drug discovery, medical imaging, and healthcare system.
Balancing the potential benefits of generative AI with the demand for responsible development and deployment is essential. Ethical considerations, transparency, and ongoing research will play major roles in maximizing the positive impact of generative AI while minimizing risks.

6 Conclusion and Future Direction

This paper offers a comprehensive systematic literature review of recent advancements in the field of generative AI. Specifically, it thoroughly explores key algorithms within the realm of Generative AI, including Diffusion Models, Transformer-based models, Generative Adversarial Networks, Variational Autoencoders, and their advancements tailored to specific applications.

Within the paper, we discuss advanced methodologies developed by various researchers, representing the current state-of-the-art achievements in the field of generative AI. A primary focus of generative AI's impact is evident in the domains of NLP and Video Translation, where advanced models have emerged with the capacity to tackle a wide array of human-centric challenges. These include tasks like question answering, code generation, language translation, image transformation, and more interdisciplinary applications. The paper highlights the recent achievements made

in these areas, shedding light on the cutting-edge advancements achieved through generative AI techniques.

Moreover, it seems evident that the future direction of generative AI will be a transformative journey. One critical avenue of exploration involves the continuous evolution of AI architectures, aiming to create models that surpass current machine and human capabilities. Additionally, the ethical dimension of AI is set to gain even more prominence, with research and development focusing on ensuring responsible AI generation, minimizing biases, and aligning with evolving ethical standards. Indeed, interdisciplinary collaborations will flourish, as generative AI is applied to complex challenges in fields like healthcare, climate science, and education, amplifying its real-world impact.

No doubt, the synergy between humans and AI will deepen, emphasizing AI's role as a collaborative partner across various domains. Advancements in NLP will persist, with an emphasis on question-answering, multilingual translation, and code generation. The domain of image, video, and multimedia processing will witness expansion, with generative AI contributing to content creation, enhancement, and interpretation. As we journey into this new and exciting future, it is also clear that we need to remain committed to responsible AI development and ethical considerations in parallel to developing these more advanced generative AI methods.

Statements and Declarations

Funding

This research received partial financial support from the Wales Innovation Network and Global Wales Small Grant Fund, Grant Number: GW-230433/414 (1.3.1b).

Conflict of interest/Competing interests

All authors declare that they have no conflict of interest.

Ethics approval

Not Required.

Data availability statement

This research paper is a systematic review that does not utilize any specific dataset for analysis. Instead, it focuses on collating and evaluating existing literature to provide a comprehensive overview of the field.

References

[1] Ahmad, B., Sun, J., You, Q., Palade, V., and Mao, Z. (2022). Brain tumor classification using a combination of variational autoencoders and generative adversarial networks. *Biomedicines*, 10(2):223.

[2] Ahuja, K., Diddee, H., Hada, R., Ochieng, M., Ramesh, K., Jain, P., Nambi, A., Ganu, T., Segal, S., Axmed, M., Bali, K., and Sitaram, S. (2023). Mega: Multilingual evaluation of generative ai.

[3] Akbik, A., Blythe, D., and Vollgraf, R. (2018). Contextual string embeddings for sequence labeling. In *Proceedings of the 27th international conference on computational linguistics*, pages 1638–1649.

[4] Al-Sabahi, K., Zuping, Z., and Nadher, M. (2018). A hierarchical structured self-attentive model for extractive document summarization (hssas). *IEEE Access*, 6:24205–24212.

[5] Ali, H., Biswas, M. R., Mohsen, F., Shah, U., Alamgir, A., Mousa, O., and Shah, Z. (2022). The role of generative adversarial networks in brain mri: a scoping review. *Insights into imaging*, 13(1):98.

[6] Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein gan.

[7] Arnab, A., Dehghani, M., Heigold, G., Sun, C., Lučić, M., and Schmid, C. (2021). Vivit: A video vision transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6836–6846.

[8] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115.

[9] Atapour-Abarghouei, A. and Breckon, T. P. (2018). Real-time monocular depth estimation using synthetic data with domain adaptation via image style transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[10] Balazevic, I., Allen, C., and Hospedales, T. (2019). TuckER: Tensor factorization for knowledge graph completion. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics.

[11] Barsoum, E., Kender, J., and Liu, Z. (2018). Hp-gan: Probabilistic 3d human motion prediction via gan. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[12] Bertasius, G., Wang, H., and Torresani, L. (2021). Is space-time attention all you need for video understanding? In *ICML*, volume 2, page 4.

[13] Bińkowski, M., Sutherland, D. J., Arbel, M., and Gretton, A. (2018). Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*.

[14] Bozkurt, A. (2023). Generative artificial intelligence (ai) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*, 18(1).

[15] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

[16] Cabanes, C., Grouazel, A., von Schuckmann, K., Hamon, M., Turpin, V., Coatanoan, C., Guinehut, S., Boone, C., Ferry, N., Reverdin, G., et al. (2012). The cora dataset: validation and diagnostics of ocean temperature and salinity in situ measurements. *Ocean Science Discussions*, 9(2):1273–1312.

[17] Cabreza, J. N., Solano, G. A., Ojeda, S. A., and Munar, V. (2022). Anomaly detection for alzheimer's disease in brain mris via unsupervised generative adversarial learning. In *2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 1–5.

[18] Cai, L. and Wang, W. Y. (2018). Kbgan: Adversarial learning for knowledge graph embeddings.

[19] Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., and Sun, L. (2023). A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt.

[20] Chandak, A., Lee, W., and Stamp, M. (2021). A comparison of word2vec, hmm2vec, and pca2vec for malware classification.

[21] Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., and Abbeel, P. (2016). Infogan: Interpretable representation learning by information maximizing generative adversarial nets.

[22] Cheong, S. Y., Mustafa, A., and Gilbert, A. (2023). Upgpt: Universal diffusion model for person image generation, editing and pose transfer.

[23] Clark, K., Luong, M.-T., Manning, C. D., and Le, Q. V. (2018). Semi-supervised sequence modeling with cross-view training. *arXiv preprint arXiv:1809.08370*.

[24] Conneau, A., Lample, G., Rinott, R., Williams, A., Bowman, S. R., Schwenk, H., and Stoyanov, V. (2018). Xnli: Evaluating cross-lingual sentence representations. *arXiv preprint arXiv:1809.05053*.

[25] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20:273–297.

[26] Courant, R. (1943). Variational methods for the solution of problems of equilibrium and vibrations.

[27] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65.

[28] Dar, S. U. H., Yurt, M., Karacan, L., Erdem, A., Erdem, E., and Çukur, T. (2018). Image synthesis in multi-contrast mri with conditional generative adversarial networks.

[29] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[30] Dimitriadis, A., Trivizakis, E., Papanikolaou, N., Tsiknakis, M., and Marias, K. (2022). Enhancing cancer differentiation with synthetic mri examinations via generative models: a systematic review. *Insights into Imaging*, 13(1):188.

[31] Dinh, L., Krueger, D., and Bengio, Y. (2015). Nice: Non-linear independent components estimation.

[32] Dinh, L., Sohl-Dickstein, J., and Bengio, S. (2017). Density estimation using real nvp.

[33] Donahue, J., Krähenbühl, P., and Darrell, T. (2016). Adversarial feature learning. *arXiv preprint arXiv:1605.09782*.

[34] Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., Carter, L., Chowdhury, S., Crick, T., Cunningham, S. W., Davies, G. H., Davison, R. M., Dé, R., Dennehy, D., Duan, Y., Dubey, R., Dwivedi, R., Edwards, J. S., Flavián, C., Gauld, R., Grover, V., Hu, M.-C., Janssen, M., Jones, P., Junglas, I., Khorana, S., Kraus, S., Larsen, K. R., Latreille, P., Laumer, S., Malik, F. T., Mardani, A., Mariani, M., Mithas, S., Mogaji, E., Nord, J. H., O'Connor, S., Okumus, F., Pagani, M., Pandey, N., Papagiannidis, S., Pappas, I. O., Pathak, N., Pries-Heje, J., Raman, R., Rana, N. P., Rehm, S.-V., Ribeiro-Navarrete, S., Richter, A., Rowe, F., Sarker, S., Stahl, B. C., Tiwari, M. K., van der Aalst, W., Venkatesh, V., Viglia, G., Wade, M., Walton, P., Wirtz, J., and Wright, R. (2023). Opinion paper: "so what if chatgpt wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy. *International Journal of Information Management*, 71:102642.

[35] Fan, H., Xiong, B., Mangalam, K., Li, Y., Yan, Z., and Malik, J. (2021). Christoph feichtenhofer. multiscale vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6824–6835.

[36] Feichtenhofer, C. (2020). X3d: Expanding architectures for efficient video recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 203–213.

[37] Feichtenhofer, C., Fan, H., Malik, J., and He, K. (2019). Slowfast networks for video recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6202–6211.

[38] Feng, L., Li, Q., Peng, Z., Tan, S., and Zhou, B. (2023). Trafficgen: Learning to generate diverse and realistic traffic scenarios. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3567–3575.

[39] Fontanini, T., Ferrari, C., Bertozzi, M., and Prati, A. (2023). Automatic generation of semantic parts for face image synthesis.

[40] Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., and Greenspan, H. (2018). Synthetic data augmentation using gan for improved liver lesion classification.

[41] Gan, J., Wang, W., Leng, J., and Gao, X. (2022). Higan+: Handwriting imitation gan with disentangled representations. *ACM Trans. Graph.*, 42(1).

[42] Gao, Y., Liu, X., and Xiang, J. (2020). Fem simulation-based generative adversarial networks to detect bearing faults. *IEEE Transactions on Industrial Informatics*, 16(7):4961–4971.

[43] Golany, T., Radinsky, K., and Freedman, D. (2020). SimGANs: Simulator-based generative adversarial networks for ECG synthesis to improve deep ECG classification. In III, H. D. and Singh, A., editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3597–3606. PMLR.

[44] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial networks.

[45] Grathwohl, W., Chen, R. T., Bettencourt, J., Sutskever, I., and Duvenaud, D. (2018). Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*.

[46] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

[47] Hochreiter, S. and Schmidhuber, J. (1996). Lstm can solve hard long time lag problems. *Advances in neural information processing systems*, 9.

[48] Hong, F.-T., Shen, L., and Xu, D. (2023). Dagan++: Depth-aware generative adversarial network for talking head video generation.

[49] Hong, F.-T., Zhang, L., Shen, L., and Xu, D. (2022). Depth-aware generative adversarial network for talking head video generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3397–3406.

[50] Hoyez, H., Schockaert, C., Rambach, J., Mirbach, B., and Stricker, D. (2022). Unsupervised image-to-image translation: A review. *Sensors*, 22(21).

[51] Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2004). Extreme learning machine: a new learning scheme of feedforward neural networks. In *2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541)*, volume 2, pages 985–990. Ieee.

[52] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2018). Image-to-image translation with conditional adversarial networks.

[53] Jain, V., Sengar, S. S., and Ronickom, J. F. A. (2023). Age-specific diagnostic classification of asd using deep learning approaches. *Studies in Health Technology and Informatics*, 309:267–271.

[54] Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L.-w. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., and Mark, R. G. (2016). Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9.

[55] Joshi, V., Peters, M., and Hopkins, M. (2018). Extending a parser to distant domains using a few dozen partially annotated examples.

[56] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589.

[57] Kale, A. S., Pandya, V., Di Troia, F., and Stamp, M. (2023). Malware classification with word2vec, hmm2vec, bert, and elmo. *Journal of Computer Virology and Hacking Techniques*, 19(1):1–16.

[58] Kang, M., Zhu, J.-Y., Zhang, R., Park, J., Shechtman, E., Paris, S., and Park, T. (2023). Scaling up gans for text-to-image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10124–10134.

[59] Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., Viola, F., Green, T., Back, T., Natsev, P., Suleyman, M., and Zisserman, A. (2017). The kinetics human action video dataset.

[60] Keerti, G., Vaishnavi, A., Mukherjee, P., Vidya, A. S., Sreenithya, G. S., and Nayab, D. (2022). Attentional networks for music generation. *Multimedia Tools and Applications*, 81(4):5179–5189.

[61] Keskar, N. S., McCann, B., Varshney, L. R., Xiong, C., and Socher, R. (2019). Ctrl: A conditional transformer language model for controllable generation.

[62] Khamparia, A., Gupta, D., Rodrigues, J. J., and de Albuquerque, V. H. C. (2021). Dcavn: Cervical cancer prediction and classification using deep convolutional and variational autoencoder network. *Multimedia Tools and Applications*, 80:30399–30415.

[63] Kingma, D. P. and Dhariwal, P. (2018). Glow: Generative flow with invertible 1x1 convolutions. *Advances in neural information processing systems*, 31.

[64] Kingma, D. P., Salimans, T., Jozefowicz, R., Chen, X., Sutskever, I., and Welling, M. (2016). Improved variational inference with inverse autoregressive flow. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

[65] Kingma, D. P. and Welling, M. (2013a). Auto-encoding variational bayes.

[66] Kingma, D. P. and Welling, M. (2013b). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

[67] Kipf, T. N. and Welling, M. (2017). Semi-supervised classification with graph convolutional networks.

[68] Kollem, S., Reddy, K. R., and Rao, D. S. (2023). A novel diffusivity function-based image denoising for mri medical images. *Multimedia Tools and Applications*, 82(21):32057–32089.

[69] Kondratyuk, D., Yuan, L., Li, Y., Zhang, L., Tan, M., Brown, M., and Gong, B. (2021). Movinets: Mobile video networks for efficient video recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16020–16030.

[70] Ku, H. and Lee, M. (2023). Textcontrolgan: Text-to-image synthesis with controllable generative adversarial networks. *Applied Sciences*, 13(8):5098.

[71] Kumar, L. and Singh, D. K. (2023). A comprehensive survey on generative adversarial networks used for synthesizing multimedia content. *Multimedia Tools and Applications*, 82(26):40585–40624.

[72] Kumar, S., Mallik, A., and Sengar, S. S. (2023). Community detection in complex networks using stacked autoencoders and crow search algorithm. *The Journal of Supercomputing*, 79(3):3329–3356.

[73] Lakshmi, P. B., Reddy, V. D., Ghosh, S., and Sengar, S. S. (2023). Classification of autism spectrum disorder based on brain image data using deep neural networks. In *International Conference on Frontiers of Intelligent Computing: Theory and Applications*, pages 209–218. Springer.

[74] Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., Van Kleef, P., Auer, S., et al. (2015). Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic web*, 6(2):167–195.

[75] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., and Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension.

[76] Li, Y., Wu, C.-Y., Fan, H., Mangalam, K., Xiong, B., Malik, J., and Feichtenhofer, C. (2022). Mvitv2: Improved multiscale vision transformers for classification and detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4804–4814.

[77] Lightman, H., Kosaraju, V., Burda, Y., Edwards, H., Baker, B., Lee, T., Leike, J., Schulman, J., Sutskever, I., and Cobbe, K. (2023). Let's verify step by step.

[78] Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

[79] Lin, Y., Wang, Y., Li, Y., Gao, Y., Wang, Z., and Khan, L. (2021). Attention-based spatial guidance for image-to-image translation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 816–825.

[80] Liu, Q., Zhou, H., Xu, Q., Liu, X., and Wang, Y. (2020). Psgan: A generative adversarial network for remote sensing image pan-sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12):10227–10242.

[81] Liu, W., Zhou, P., Zhao, Z., Wang, Z., Ju, Q., Deng, H., and Wang, P. (2019). K-bert: Enabling language representation with knowledge graph.

[82] Liu, Z., Luo, P., Wang, X., and Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*.

[83] Liu, Z., Ning, J., Cao, Y., Wei, Y., Zhang, Z., Lin, S., and Hu, H. (2022). Video swin transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3202–3211.

[84] Masci, J., Meier, U., Cireşan, D., and Schmidhuber, J. (2011). Stacked convolutional auto-encoders for hierarchical feature extraction. In *Artificial Neural*

Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part I 21, pages 52–59. Springer.

[85] McKeown, K., Barzilay, R., Blair-Goldensohn, S., Evans, D., Hatzivassiloglou, V., Klavans, J., Nenkova, A., Schiffman, B., and Sigelman, S. (2002). The columbia multi-document summarizer for duc 2002. In *Workshop on Automatic Summarization*, pages 1–8.

[86] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al. (2014). The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024.

[87] Mescheder, L., Geiger, A., and Nowozin, S. (2018a). Which training methods for gans do actually converge?

[88] Mescheder, L., Nowozin, S., and Geiger, A. (2018b). The numerics of gans.

[89] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

[90] Min, D., Song, M., and Hwang, S. J. (2022). Styletalker: One-shot style-based audio-driven talking head video generation.

[91] Mirza, M. and Osindero, S. (2014). Conditional generative adversarial nets.

[92] Moradi Dakhel, A., Majdinasab, V., Nikanjam, A., Khomh, F., Desmarais, M. C., and Jiang, Z. M. J. (2023). Github copilot ai pair programmer: Asset or liability? *Journal of Systems and Software*, 203:111734.

[93] Nagarajan, V. and Kolter, J. Z. (2017). Gradient descent gan optimization is locally stable. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

[94] Nagrani, A., Chung, J. S., and Zisserman, A. (2017). VoxCeleb: A large-scale speaker identification dataset. In *Interspeech 2017*. ISCA.

[95] Nakano, R., Hilton, J., Balaji, S., Wu, J., Ouyang, L., Kim, C., Hesse, C., Jain, S., Kosaraju, V., Saunders, W., Jiang, X., Cobbe, K., Eloundou, T., Krueger, G., Button, K., Knight, M., Chess, B., and Schulman, J. (2022). Webgpt: Browser-assisted question-answering with human feedback.

[96] Neimark, D., Bar, O., Zohar, M., and Asselmann, D. (2021). Video transformer network.

[97] Odena, A. (2016). Semi-supervised learning with generative adversarial networks.

[98] Odena, A., Olah, C., and Shlens, J. (2017). Conditional image synthesis with auxiliary classifier gans.

[99] OpenAI (2023). Gpt-4 technical report.

[100] Panayotov, V., Chen, G., Povey, D., and Khudanpur, S. (2015). Librispeech: An asr corpus based on public domain audio books. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5206–5210.

[101] Paola, Z. L., Jesús, L. S., Christian, A. H., and Sonia, R. U. (2023). Correction of banding errors in satellite images with generative adversarial networks (gan). *IEEE Access*.

[102] Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations.

[103] Prajwal, K., Mukhopadhyay, R., Namboodiri, V. P., and Jawahar, C. (2020). A lip sync expert is all you need for speech to lip generation in the wild. In *Proceedings of the 28th ACM international conference on multimedia*, pages 484–492.

[104] Pudari, R. and Ernst, N. A. (2023). From copilot to pilot: Towards ai supported software development.

[105] Qi, G.-J. (2018). Loss-sensitive generative adversarial networks on lipschitz densities.

[106] Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., and Huang, X. (2020). Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, 63(10):1872–1897.

[107] Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1:81–106.

[108] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. (2021). Learning transferable visual models from natural language supervision.

[109] Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al. (2018). Improving language understanding by generative pre-training.

[110] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

[111] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551.

[112] Rani, R. and Lobiyal, D. (2021). An extractive text summarization approach using tagged-lda based topic modeling. *Multimedia tools and applications*, 80:3275–3305.

[113] Reddy, M. D. M., Basha, M. S. M., Hari, M. M. C., and Penchalaiah, M. N. (2021). Dall-e: Creating images from text. *UGC Care Group I Journal*, 8(14):71–75.

[114] Rezagholiradeh, M. and Haidar, M. A. (2018). Reg-gan: Semi-supervised learning based on generative adversarial networks for regression. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 2806–2810. IEEE.

[115] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., and Chen, X. (2016). Improved techniques for training gans. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

[116] Sang, E. F. and De Meulder, F. (2003). Introduction to the conll-2003 shared task: Language-independent named entity recognition. *arXiv preprint cs/0306050*.

[117] Sengar, S. S. and Kumar, S. (2022). Content-based secure image retrieval in an untrusted third-party environment. In *International Conference on Frontiers of Intelligent Computing: Theory and Applications*, pages 287–297. Springer.

[118] Sengar, S. S., Meulengracht, C., Boesen, M. P., Overgaard, A. F., Gudbergsen, H., Nybing, J. D., Perslev, M., and Dam, E. B. (2023). Multi-planar 3d knee mri segmentation via unet inspired architectures. *International Journal of Imaging Systems and Technology*, 33(3):985–998.

[119] Sengar, S. S. and Mukhopadhyay, S. (2016). Moving object tracking using laplacian-dct based perceptual hash. In *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pages 2345–2349. IEEE.

[120] Sengar, S. S. and Mukhopadhyay, S. (2020). Motion segmentation-based surveillance video compression using adaptive particle swarm optimization. *Neural Computing and Applications*, 32(15):11443–11457.

[121] Shi, X., Lv, F., Seng, D., Zhang, J., Chen, J., and Xing, B. (2021). Visualizing and understanding graph convolutional network. *Multimedia Tools and Applications*, 80:8355–8375.

[122] Singhal, A. (2012). Introducing the knowledge graph: Things, not strings,.

[123] Soomro, K., Zamir, A. R., and Shah, M. (2012). Ucf101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402*.

[124] Steiner, T., Verborgh, R., Troncy, R., Gabarro, J., and Van de Walle, R. (2012). Adding realtime coverage to the google knowledge graph. In *11th International Semantic Web Conference (ISWC 2012)*, volume 914, pages 65–68. Citeseer.

[125] Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

[126] Tahir, R., Cheng, K., Memon, B. A., and Liu, Q. (2022). A diverse domain generative adversarial network for style transfer on face photographs.

[127] Tan, S., Wong, K., Wang, S., Manivasagam, S., Ren, M., and Urtasun, R. (2021). Scenegen: Learning to generate realistic traffic scenes. In *Proceedings - 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021*, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 892–901. IEEE Computer Society. Funding Information: Work done at Uber ATG. Publisher Copyright: © 2021 IEEE; 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021 ; Conference date: 19-06-2021 Through 25-06-2021.

[128] Tanchenko, A. (2014). Visual-psnr measure of image quality. *Journal of Visual Communication and Image Representation*, 25(5):874–878.

[129] Tibrewala, R., Dutt, T., Tong, A., Ginocchio, L., Keerthivasan, M. B., Baete, S. H., Chopra, S., Lui, Y. W., Sodickson, D. K., Chandarana, H., and Johnson, P. M. (2023). Fastmri prostate: A publicly available, biparametric mri dataset to advance machine learning for prostate cancer imaging.

[130] Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., and Agyemang, B. (2023). What if the devil is my guardian angel: Chatgpt as a case study of using chatbots in education. *Smart Learning Environments*, 10(1):15.

[131] Torbunov, D., Huang, Y., Yu, H., Huang, J., Yoo, S., Lin, M., Viren, B., and Ren, Y. (2023). Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 702–712.

[132] van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio.

[133] Vasanthi, P. and Mohan, L. (2023). Multi-head-self-attention based yolov5x-transformer for multi-scale object detection. *Multimedia Tools and Applications*, pages 1–27.

[134] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

[135] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., and Bottou, L. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(12).

[136] Wang, P., Zhang, C., Qi, F., Liu, S., Zhang, X., Lyu, P., Han, J., Liu, J., Ding, E., and Shi, G. (2021a). Pgnet: Real-time arbitrarily-shaped text spotting with point gathering network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 2782–2790.

[137] Wang, Q., Gao, J., Lin, W., and Yuan, Y. (2019). Learning from synthetic data for crowd counting in the wild.

[138] Wang, S., Li, L., Ding, Y., Fan, C., and Yu, X. (2021b). Audio2head: Audio-driven one-shot talking-head generation with natural head motion. *arXiv preprint arXiv:2107.09293*.

[139] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., and Summers, R. M. (2017). Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106.

[140] Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

[141] Wei, J., Zou, H., Sun, L., Cao, X., He, S., Liu, S., and Zhang, Y. (2023). Cfrwd-gan for sar-to-optical image translation. *Remote Sensing*, 15(10):2547.

[142] Wu, W., Zhang, Y., Li, C., Qian, C., and Loy, C. C. (2018). Reenactgan: Learning to reenact faces via boundary transfer.

[143] Xiao, S., Duan, L., Xie, G., Li, R., Chen, Z., Deng, G., and Nummenmaa, J. (2021). Hmnet: Hybrid matching network for few-shot link prediction. In *International Conference on Database Systems for Advanced Applications*, pages 307–322. Springer.

[144] Xu, I. R., Van Booven, D. J., Goberdhan, S., Breto, A., Porto, J., Alhusseini, M., Algohary, A., Stoyanova, R., Punnen, S., Mahne, A., et al. (2023). Generative adversarial networks can create high quality artificial prostate cancer magnetic

resonance images. *Journal of Personalized Medicine*, 13(3):547.

[145] Xu, Y., Li, M., Cui, L., Huang, S., Wei, F., and Zhou, M. (2020). Layoutlm: Pre-training of text and layout for document image understanding. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1192–1200.

[146] Yan, S., Wang, C., Chen, W., and Lyu, J. (2022). Swin transformer-based gan for multi-modal medical image translation. *Frontiers in Oncology*, 12:942511.

[147] Yang, K., Yau, J., Fei-Fei, L., Deng, J., and Russakovsky, O. (2022). A study of face obfuscation in imagenet. In *International Conference on Machine Learning (ICML)*.

[148] Yang, X., Li, Y., Zhang, X., Chen, H., and Cheng, W. (2023). Exploring the limits of chatgpt for query or aspect-based text summarization.

[149] Yeh, R., Liu, Z., Goldman, D. B., and Agarwala, A. (2016). Semantic facial expression editing using autoencoded flow.

[150] Yu, L., Zhang, W., Wang, J., and Yu, Y. (2017). Seqgan: Sequence generative adversarial nets with policy gradient.

[151] Zeng, X., Wang, F., Luo, Y., Kang, S.-g., Tang, J., Lightstone, F. C., Fang, E. F., Cornell, W., Nussinov, R., and Cheng, F. (2022). Deep generative molecular design reshapes drug discovery. *Cell Reports Medicine*.

[152] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595.

[153] Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., and Liu, Q. (2019). Ernie: Enhanced language representation with informative entities. *arXiv preprint arXiv:1905.07129*.

[154] Zhang, Z., Li, L., Ding, Y., and Fan, C. (2021). Flow-guided one-shot talking face generation with a high-resolution audio-visual dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3661–3670.

[155] Zhao, Y., Celik, T., Liu, N., and Li, H.-C. (2022). A comparative analysis of gan-based methods for sar-to-optical image translation. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5.

[156] Zhong, M., Yin, D., Yu, T., Zaidi, A., Mutuma, M., Jha, R., Awadallah, A. H., Celikyilmaz, A., Liu, Y., Qiu, X., and Radev, D. (2021). QMSum: A new benchmark for query-based multi-domain meeting summarization. In *Proceedings of the 2021*

Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 5905–5921, Online. Association for Computational Linguistics.

[157] Zhou, Y., Han, X., Shechtman, E., Echevarria, J., Kalogerakis, E., and Li, D. (2020). Makelttalk: speaker-aware talking-head animation. *ACM Transactions On Graphics (TOG)*, 39(6):1–15.

[158] Zhu, C., Xu, R., Zeng, M., and Huang, X. (2020). A hierarchical network for abstractive meeting summarization with cross-domain pretraining. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 194–203, Online. Association for Computational Linguistics.

[159] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017a). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251.

[160] Zhu, J.-Y., Zhang, R., Pathak, D., Darrell, T., Efros, A. A., Wang, O., and Shechtman, E. (2017b). Toward multimodal image-to-image translation. *Advances in neural information processing systems*, 30.

[161] Zhu, J.-Y., Zhang, R., Pathak, D., Darrell, T., Efros, A. A., Wang, O., and Shechtman, E. (2017c). Toward multimodal image-to-image translation. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

[162] Zuo, Z., Zhao, L., Lian, S., Chen, H., Wang, Z., Li, A., Xing, W., and Lu, D. (2022). Style fader generative adversarial networks for style degree controllable artistic style transfer. In *Proc. Int. Joint Conf. on Artif. Intell.(IJCAI)*, pages 5002–5009.

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.

Graph-based Semi-supervised Learning for 3D Pose Estimation

Ravi Kant Sharma
ravikantsharma635@gmail.com
Delhi College of Engineering: Delhi Technological University https://orcid.org/0009-0003-8321-031X
Raghav Kumar Jha
Delhi College of Engineering: Delhi Technological University
Shweta Meena
Delhi College of Engineering: Delhi Technological University

Research Article

Keywords: Detectron2, Feed Forward Neural Network, GNN, LSTM, Multi-Layer Perceptron, WMPJPE

Posted Date: May 27th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4232631/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

The challenge of estimating 3D pose has been thoroughly explored and researched in computer vision due to its broad range of applications. However, due to complex structures, occlusion, frame rates, varying sizes and resolutions, this problem is highly challenging. This paper demonstrates the ability of graph neural networks (GNNs) and long short-term memory (LSTM) for 2D to 3D pose estimation in a sequence of frames. LSTM is used as a feature extractor. The bone length and joint angle play a crucial role in pose estimation. Our model uses a GNN to analyze the relationships between nearby joints and their angles and then predicts the final 3D joint positions using a multi-layer perceptron (MLP) model. A Semi-supervised learning and frame-dropping strategy approach is employed to enhance the accuracy of our model. Our model outperforms various latest advancements models, achieving an accuracy of 6.5 mm in joint localization on the HumanEva-1 and Human3.6m datasets.

1 Introduction

In the research on 3D human pose estimation from 2D pose estimation in videos, emphasis is placed on comprehending human posture and limb articulation for computer vision tasks such as augmented and virtual reality, human activity recognition, and sports analysis [1]. The use of a 2D representation of human position, deemed simpler to estimate, is common, although occlusion and foreshortening in 2D positions may introduce ambiguity. Depth information inherent in 3D poses is leveraged to eliminate uncertainty and enhance performance[2]. The challenge lies in the inherent non-uniqueness of mapping a 2D pose to 3D, as multiple projections onto a 2D plane are possible [3]. Acquiring datasets for 3D pose is challenging due to the expense and complexity of motion capture setups, unlike 2D pose datasets where key points can be manually identified [4].

Over time, various methods have been employed to address the challenge of 3D pose estimation[5]. Earlier techniques focused on extracting backdrop-independent features from photos and mapping them into 3D human poses. Recent approaches prioritize training deep convolutional neural networks (CNN) using end-to-end training for direct 3D pose estimation from input frames [6].

Two significant contributions are presented in the paper. The utilization of LSTM and GNN models for 3D pose estimation from 2D pose sequences is demonstrated [7]. A comparative analysis with other approaches, including FCN, CNN, and RNN-based methods, is conducted in terms of parameters, training accuracy, and FLOPs. Additionally, the effectiveness of a dropping frames strategy in improving model accuracy is showcased.

Furthermore, a semi-supervised learning strategy is proposed, utilizing 2D key points from a video sequence to gradually generate corresponding 3D key points [8]. The research illustrates how limited training data does not constrain the model's ability to generalize.

2 Related Works

2.1 End-to-End 3D Pose Estimation

The conversion of pixel intensities from a single picture to 3D postures is facilitated by end-to-end pose estimation algorithms. It becomes necessary for the pattern to be retained and the 3D stance deduced by deep neural networks (DNNs) [6]. Poor performance and limited generalization capabilities are observed in these networks due to the extensive requirement of heavily labeled training data and the inherent difficulty in learning. This limitation of 3D pose estimation is acknowledged [7].

2.2 3D Pose from 2D Key Points

2D-to-3D lifting techniques are employed to leverage 2D pose estimation derived from input frames or video frames. Commonly utilized for 2D pose estimation are MMPose, OpenPose, and Detectron2. Functioning as an intermediary for 3D pose estimation, 2D pose estimators serve as inputs for the deep neural network (DNN) responsible for predicting 3D key points [9]. A simple yet effective fully connected residual network for generating 3D key points from 2D key points locations was introduced by Martinez et al., utilizing only a single frame. Rather than relying on a single frame for predicting 3D pose, the utilization of videos is advocated, offering temporal information that contributes to accuracy improvement and enhances model robustness. Temporal convolutions were employed by Pavllo et al. to predict 3D pose from 2D joints using multiple frames simultaneously [10]. Some models incorporated bone length and bone directions in the loss functions to ensure temporal consistency across different frames [11]. Time-series-based models, as proposed by Hossain and Little, employed a Long Short-Term Memory (LSTM) network to predict from previous temporal information between input frames [12].

2.3 Semi-supervised Learning

The In multitask networks for action recognition and joint 2D and 3D pose estimation, supervised learning is employed. Training for 3D pose estimation involves the use of various multi-view recordings, but obtaining such recordings for unsupervised learning proves challenging [1]. The differentiation between realistic and unrealistic poses generated by the model is accomplished by GANs models, and 3D poses are discriminated by the discriminator through random projection back to 2D points [4]. Additionally, depth estimation can be utilized for the projection of different key points based on depth estimation, such as determining if the “left hand is behind the right hand [8].”

3 Preliminaries

Graph Neural Networks (GNNs) are characterized as a class of deep learning models explicitly crafted to operate on graph-structured data, wherein graphs comprise nodes (referred to as vertices) interconnected by edges (referred to as links or edges) [7]. The intricate relationships and dependencies

among the nodes in a graph are intended to be captured and modeled by GNNs. Traditional deep learning models, including convolutional neural networks (CNNs) or recurrent neural networks (RNNs), are found to be inadequately suited for processing graph data [13]. To address this limitation, GNNs consider both the local and global structure of the graph, engaging in the propagation of information between nodes and updating node representations based on the features of their neighbors [9].

The fundamental building block of a GNN is a message-passing scheme, where information is exchanged between nodes and aggregated for the purpose of updating node representations. The message-passing operation is described through the following steps:

1. *Initialization:* An initial feature vector is assigned to each of the 17 nodes in a frame *[14]*.
2. *Message Passing:* Information exchange occurs between neighboring nodes. Each node, in turn, aggregates and transforms the messages received from its neighbors to update its representation *[15]*. This step is typically performed iteratively across multiple layers to capture higher-order dependencies *[16]*.
3. *Readout:* Following the message passing step, the final node representations find application in downstream tasks, such as node classification, link prediction, or graph-level prediction *[17]*.

4 Methodology

We have described poses with exact key points opposed to the heatmap methodology used by various detectors. This allows us to use LSTM as a feature extractor for our GNN. Pose can be considered as a relational data, where one joint relates to other in a certain way which can be modelled using the GNN [18]. The poses are shifted to a common origin point and are rescaled to the same size which makes our model independent of the subject size and position in the frame. The temporal information is exploited using the LSTM network [17]. Though FCNs suffer from vanishing gradient problem, this problem can be eliminated by using residual connections in the network. These models fail to efficiently learn temporal information as already proved in various studies [19].

4.1 Model Architecture

The model architecture as shown in figure 1, is simple and consists of 3 main blocks:

1. *Feature Extraction Module:* Joint positions can be considered as a type of spatial-temporal graph. Each joint can be considered as a node which represents the current trajectory of that particular joint. As already proven, raw text or data is not rich in features and hence we require a feature extractor module [19]. The input $x \in R^{L\times17\times2}$ consists of sequence of frames, each frame has 17 (x, y) coordinates for each joint of the body. All the 17 joints are shown in figure 1. The trajectory of each joint is passed to separate LSTM models. Opposed to GRUs, LSTMs don't suffer from vanishing gradient problem [20].
2. *Relational Extraction Module:* Relational Modelling Network: The joint representations generated from the feature extractor are fed into to relational modelling module [21]. It consists of a GNN which models relationships between nearby joints.

4.2 Semi-supervised Training

In semi-supervised training, our approach capitalizes on the limited availability of labeled data by incorporating a strategy that blends supervised and unsupervised learning. The joint representations obtained from the relational modeling module are utilized to train the model on labeled data, facilitating the prediction of corresponding 3D poses from 2D key points [4]. Simultaneously, the model is exposed to unlabeled data, enabling it to discern and exploit underlying patterns and relationships present in the data without explicit annotations. This semi-supervised strategy enhances the model's ability to generalize to diverse scenarios and adapt to variations in poses that may not be fully covered by the labeled dataset [7]. Through this dual-learning paradigm, our model achieves a more robust performance, effectively leveraging both labeled and unlabeled data to enhance its overall capability in 3D pose estimation from 2D key points [23].

5 Experimental Setup

5.1 Dataset

Two datasets, Human3.6m and HumanEva-I, have been utilized for training the model. Both datasets encompass images along with corresponding 2D and 3D poses [6].

The Human3.6m dataset comprises 3.6 million images featuring associated 3D human poses. Eleven actors (comprising 6 males and 5 females) performed 17

diverse activities, including smoking, sitting, walking, using the phone, taking photographs, discussing, standing, etc [13]. Figure 3 displays selected samples from the dataset, illustrating images with corresponding 3D poses represented in stick form. Captured by four calibrated cameras operating at a 50 Hertz refresh rate [8], this dataset provides accurate 3D posed joint positions and angles using high-speed capture cameras. It involves labeling 24 body parts with precise background removal, making it a widely used dataset for pose estimation and prediction tasks.

HumanEva-I encompasses seven calibrated video sequences (four grayscale and three color), temporally synchronized with 3D body poses captured by a motion capture system [3]. The database incorporates four subjects engaged in six common actions, such as walking, jogging, gesturing, etc [2]. Participants are provided with error metrics to calculate errors in 2D and 3D poses. Figure 4 showcases selected samples from the dataset, illustrating images with corresponding 3D poses represented in box form [5].

End-to-end pose estimation involves the conversion of pixel intensities from a single picture to 3D postures [9]. This process necessitates the retention of patterns and deduction of the 3D stance by deep neural networks (DNNs) [12]. Due to the substantial requirement for heavily labeled training data and the inherent challenges in learning, these networks exhibit poor performance and limited generalization capabilities. This limitation is inherent in 3D pose estimation [4].

5.2 Loss Function

The model uses 4 loss functions:

1. Joint Loss (MPJPE Loss): This is measures of Euclidean distance between the predicted and ground truth 3D joint positions and is used in semi-supervised learning [21]. Given a set of points J with predicted positions P and ground truth positions G, the MPJPE loss is often calculated is:

$$E_{JointLoss} = \frac{1}{|J|}\sum_{j=1}^{J} || P_j - G_j ||$$

1

2. Weighted Joint Loss (WMPJPE Loss): As the subject moves away from the frame, its trajectory gets worse and it gets difficult to track its trajectory hence the WMPJPE Loss gives less weight to further subjects. This allows us to assign different importance to different joints [7].

$$E_{WeightedJointLoss} = \frac{1}{\sum_{j=1}^{J} w_j}\sum_{j=1}^{J} w_j * || P_j - G_j ||$$

2

3. Bone Angle Loss: Measure the absolute difference in lengths between predicted and ground truth bones. As the human skeleton has limited range of angular motion of joints, hence bone angle loss is used to neglect the 3D pose predictions having bone joint angle greater than maximum limit of human body [11]. Apart from this, bone angle loss combined with LSTM provides better angular accuracy to physical and dynamic joints [10]. If $\varnothing$ represents the set of bones angles, A is the set of predicted bone vectors, and B is the set of ground truth bone vectors, the bongle angle loss formulated as:

$$E_{BoneAngleLoss} = \frac{1}{|\varnothing|}\sum_{\theta \in \varnothing} \cos^{-1}\left(\frac{A_\theta * B_\theta}{||A_\theta||_2 * ||B_\theta||_2}\right)$$

3

4. Bone Length Loss: Adding a soft constraint to align the mean bone lengths of unlabeled batch subjects with those of the labeled batch is a beneficial approach for self-supervision in the context of regressing the 3D trajectory of a person [6], [13]. This method involves training everything jointly, incorporating an elastic restriction to align the average bone lengths of unlabeled predictions with the labeled counterparts [5]. If L represents the set of bones length, $L_{\varnothing p}$ is the set of predicted bone lengths, and $L_{\varnothing g}$ is the set of ground truth bone lengths, the bone length loss is:

$$E_{BondLengthLoss} = \frac{1}{|L|}\sum_{l \in L} |L_p - L_g|$$

4

$$E_{Total} = E_{JointLoss} + E_{WeightedJointLoss} + E_{BoneAngleLoss} + E_{BoneLengthLoss}$$

5

6 Results

Table 3 displays the results for both evaluation techniques for our model. The model does not depend on additional data like other techniques, and has a lowest value for average error in comparison to all other methods. Our model outperforms the earlier best result [21] under procedure 1 (Table 3) by an average of 6 mm, and an error reduction by 11%. Notably, whereas our model does not employ ground-truth boxes, [2] unlike others .

For enhancing our results on the missing frames, During training we black out frames randomly in the given dataset which helps our model to understand how much it needs to rely on past frames information [24]. Our model significantly predicts right output also for the missing frames.

Table 1. Reconstruction error is evaluated on 3.6 million human dataset, with every 64th frame used for evaluation

	Photo	Greet	Walk D.	Smoke	Pose	Wait	Dir.	Disc.	Phone	SitD.	Eat	WalkT.	Sit	Walk	Purch.	Avg
Hossain & Little [16] ECCV' 18	72.6	55.2	62.4	59.0	53.0	57.3	59.0	50.7	63.1	80.9	57.2	49.6	66.1	46.6	51.7	58.2
Pavlakos et al. [40] CVPR' 18	65.3	52.0	60.9	56.6	49.9	52.9	48.5	54.4	59.4	71.1	54.4	47.8	65.8	52.9	52.9	56.2
Martinez et al. [34] ICCV' 17	78.4	59.0	65.1	62.3	55.2	59.1	51.8	56.2	69.5	94.6	58.1	52.4	74.0	49.5	58.1	62.9
Sun et al. [50] ICCV' 17	67.2	54.3	61.6	61.5	53.1	53.4	52.8	54.8	61.8	86.7	54.2	53.4	71.7	53.4	53.6	59.1
Yang et al. [56] CVPR' 18	65.4	57.0	69.2	57.4	49.8	58.4	51.5	58.9	62.1	85.2	50.4	47.7	69.2	58.4	52.7	58.6
Lee et al. [18] ECCV' 18	75.0	52.6	**43.6**	54.1	50.2	55.6	**40.2**	49.2	50.1	73.9	47.8	43.3	**55.8**	55.6	**43.0**	52.8
Tekin et al. [52] ICCV' 17	78.3	61.2	74.3	69.3	63.1	70.3	54.2	61.4	79.4	107.3	60.2	63.2	70.1	70.3	81.6	69.7
Fang et al. [10] AAAI' 18	73.3	57.1	62.7	60.3	53.4	57.7	50.1	49.2	66.6	88.6	57.0	50.6	72.8	57.7	55.7	60.4
Ours, 243 frames	**55.0**	**45.2**	49.1	**46.9**	**44.9**	**44.5**	45.2	**46.3**	**48.1**	**65.3**	**43.5**	**34.0**	57.1	**33.2**	44.0	**46.8**

Next, we assess how the 2D key point detector affected the outcome. Table 3 details the model's accuracy using GT 2D poses, for the Detectron2, hourglass-network predictions, and CPN. The performance of Mask R-CNN and CPN significantly outperforms the stacked hourglass network [24]. This is brought on by a stronger feature combination and better heatmap resolution and a more varied dataset used for pre-training. Using 2D ground truth poses as training data, our method enhances results by 8.33 mm and approaches proposed by Lee et al. which are based on LSTM by 1.2mm for method 1. Thus, advancements are not attributable to a better 2D detector [1].

The smoothness of predictions is crucial for the videos which are not measured by the absolute position errors. This is evaluated using joint velocity errors termed as joint loss, which are the first derivative of joint loss of 3D posture sequences [20]. The Joint loss error produced by the single-frame is reduced by our semi-supervised approach on the transformer model by an average of 15%, resulting in noticeably smoother poses, as seen in Table 1.

Table 2 represents results on the Human-EVA dataset and the Mask R-CNN detector as a 2D key point detector for the testing purpose. Table 3 represents our model and also performs better for the small dataset [24].

Table 2. Analysis of the HumanEva-1 dataset error

Model	Parameters	FLOPs
Hossain et al. [21]	16.86M	34.0M
Our Method	1.14M	17.59M

In terms of complexity, Table 2 contrasts our model performance with the LSTM model [2]. We show the model parameters together with an estimation of FLOPs required for prediction of one frame at a time.

7 Conclusion

In conclusion, this has provided an overview of 3D human pose estimation from 2D pose estimation[14] in videos, highlighting the challenges, techniques, and proposed solutions in this field [10]. However, there are several avenues for future research and development in this area.

Firstly, improving the accuracy and robustness of 3D pose estimation [18] algorithms remains a crucial direction. While deep learning-based models have shown promising results, there is still room for enhancing their performance, especially in challenging scenarios such as occlusions, complex poses [19], and cluttered backgrounds. Additionally, exploring novel architectures and fusion strategies that can effectively leverage both spatial and temporal information could lead to further improvements [6].

Secondly, addressing the issue of limited labeled 3D pose datasets is essential [11]. The development of methods that can leverage unlabeled or weakly labeled data to enhance the training process holds great potential. Semi-supervised and unsupervised learning techniques could be explored to make 5 better use of the vast amount of unlabeled video data available [16]. Furthermore, the integration of multimodal data sources, such as depth sensors, RGB-D cameras, and wearable devices, could offer richer information for more accurate and detailed 3D pose estimation. Leveraging additional cues, such as scene context, object interactions, or human-object relationships, could further enhance the understanding of human poses in complex environments [7], [23].

Table 3
Analysis of the effect 2D detector on the expected results

Method	P1	Method	P1
Lee et al. [19] (GT)	37.9	Our Method (D Ff)	50.9
Martinez et al. [1] (SH Ff)	63.0	Our Method (SH Fr)	53.3
Hossain & Little (GT)	42.7	Our Method (D Ff)	54.7
Martinez et at. [1] (GT)	44.9	Our Method (D Ff)	39.3
Martinez et at. [1] (SH PT)	67.8	Our Method (D Ff)	57.1
Our Method	51.9	Our Method (D Ff)	48.9

Lastly, deploying 3D pose estimation in real-time applications, such as human-computer interaction, augmented reality, and robotics, is an exciting area for future exploration [6]. Developing efficient and lightweight models that can run in real-time on resource-constrained devices would enable the integration of 3D pose estimation into a wide range of practical applications [12].

Declarations

Conflict of interest

The authors have no relevant financial or non-financial interests to disclose. The authors have no conflicts of interest to decalre that are relevant to the content of this article.

Research involving human participants and/or animals

This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent

Not applicable.

References

1. Martinez J, Hossain R, Romero J, Little JJ (2017) A simple yet effective baseline for 3d human pose estimation. In: IEEE International Conference on Computer Vision, pp. 2640–2649
2. Sun X, Xiao B, Wei F, Liang S, Wei Y (2017) Compositional Human Pose Regression. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2602–2611
3. Pavlakos G, Zhou X, Derpanis KG, Daniilidis K (2016) Coarse-to-Fine Volumetric Prediction for Single-Image 3D Human Pose. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7025–7034
4. Zhou K, Han X, Jiang N, Jia K, Lu J (2022) HEMlets PoSh: Learning Part-Centric Heatmap Triplets for 3D Human Pose and Shape Estimation. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 06, pp. 3000–3014
5. Xia H, Xiao M (2020) 3D Human Pose Estimation With Generative Adversarial Networks. In: IEEE Access, Institute of Electrical and Electronics Engineers (IEEE), vol. 8, pp. 206198–206206
6. Lonini L, Moon Y, Embry K, Cotton RJ, McKenzie K, Jenz S, Jayaraman A (2022) Video-Based Pose Estimation for Gait Analysis in Stroke Survivors during Clinical Assessments: A Proof-of-Concept Study. Digital Biomarkers, vol 6. S. Karger AG, pp 9–18. 1
7. Li S, Ke L, Pratama K, Tai Y-W, Tang C-K, Cheng K-T (2020) Cascaded Deep Monocular 3D Human Pose Estimation With Evolutionary Training Data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition
8. Cai Y, Ge L, Liu J, Cai J, Cham T-J, Yuan J, Thalmann NM (2019) Exploiting Spatial-Temporal Relationships for 3D Pose Estimation via Graph Convolutional Networks. In: Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV), pp. 2272–2281

9. Cheng Y, Yang B, Wang B, Wending Y, Tan R (2019) Occlusion-Aware Networks for 3D Human Pose Estimation in Video. In: Proceedings of IEEE/CVF International Conference Computer Vision (ICCV), pp. 723–732
10. Dabral R, Mundhada A, Kusupati U, Afaque S, Sharma A, Jain A (2018) Learning 3D Human Pose from Structure and Motion. Computer Vision – ECCV 2018. Springer International Publishing, pp 679–696
11. Pavllo D, Feichtenhofer C, Grangier D, Auli M (2018) 3D human pose estimation in video with temporal convolutions and semi-supervised training. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7745–7754
12. Hossain MRI, Little JJ (2018) Exploiting Temporal Information for 3D Human Pose Estimation. Computer Vision – ECCV 2018. Springer International Publishing, pp 69–86
13. Ji X, Fang Q, Dong J, Shuai Q, Jiang W, Zhou X (2020) A survey on monocular 3D human pose estimation. Virtual Reality & Intelligent Hardware. In: Elsevier BV, vol. 2, no. 6, pp. 471–500
14. Tekin B, Marquez-Neila P, Salzmann M, Fua P (2017) Learning to Fuse 2D and 3D Image Cues for Monocular Body Pose Estimation. In: IEEE International Conference on Computer Vision
15. Lee K, Lee I, Lee S (2018) Propagating LSTM: 3D Pose Estimation Based on Joint Interdependency. Computer Vision – ECCV 2018. Springer International Publishing, pp 123–141
16. Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, Wang L, Li C, Sun M (2020) Graph neural networks: A review of methods and applications. In: AI Open, Elsevier BV, vol. 1, pp. 57–81
17. Chao X, Ge Z, Leung H (2023) Video2mesh: 3D human pose and shape recovery by a temporal convolutional transformer network. In: IET Computer Vision, vol. 17, no. 4. Institution of Engineering and Technology (IET), pp. 379–388
18. Yuan J, Yu H, Cao M, Xu M, Xie J, Wang C (2021) Semi-Supervised and Self-Supervised Classification with Multi-View Graph Neural Networks. In: Proceedings of the 30th ACM International Conference on Information and Knowledge Management
19. Ma N, Wu Z, Cheung Y, Guo Y, Gao Y, Li J, Jiang B (2022) A Survey of Human Action Recognition and Posture Prediction. Tsinghua Science and Technology, vol 27. no. 6. Tsinghua University, pp 973–1001
20. Park S, Hwang J, Kwak N (2016) 3D Human Pose Estimation Using Convolutional Neural Networks with 2D Pose Information. Lecture Notes in Computer Science. Springer International Publishing, pp 156–169
21. Lin J, Li S, Qin H, Wang H, Cui N, Jiang Q, Jian H, Wang G (2023) Overview of 3D Human Pose Estimation. In: Computer Modeling in Engineering and Sciences, vol. 134, no. 3. Computers, Materials and Continua (Tech Science Press), pp. 1621–1651
22. Zhang S, Hu X (2023) The Latest Progress in Human Pose Estimation. In: 15th International Conference on Computer Research and Development (ICCRD), IEEE
23. Wang J, Tan S, Zhen X, Xu S, Zheng F, He Z, Shao L (2021) Deep 3D human pose estimation: A review. Computer Vision and Image Understanding, vol 210. Elsevier BV, p 103225
24. Zhang D, Wu Y, Guo M, Chen Y (2021) In: Electronics (ed) Deep Learning Methods for 3D Human Pose Estimation under Different Supervision Paradigms: A Survey, vol 10. MDPI AG, p 18

Figures

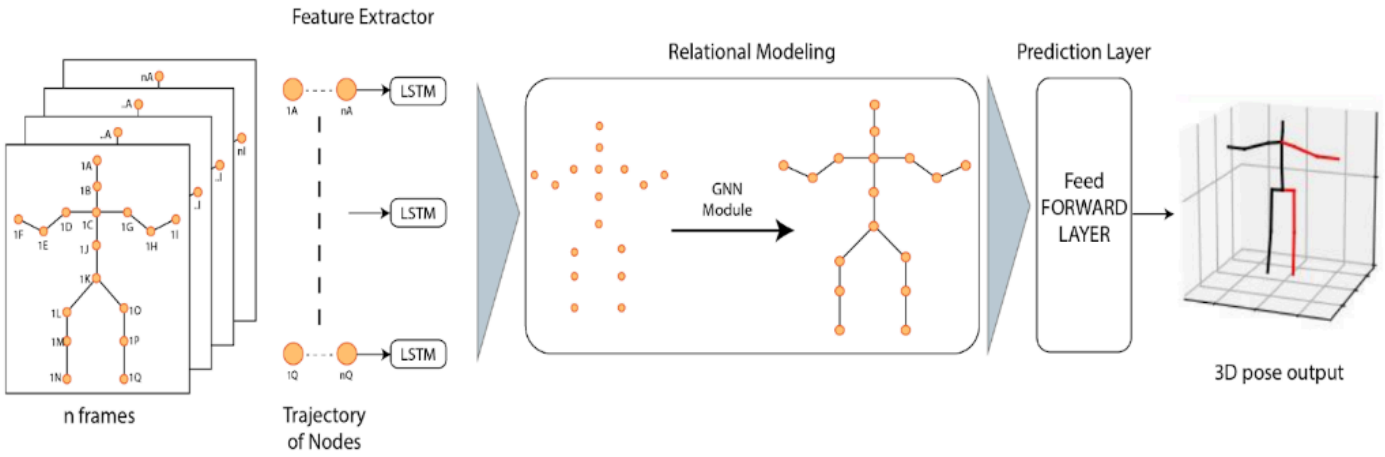


Figure 1

Semi-supervised Learning. A regression is done in the unlabelled 2D pose trajectory to predict the 3D pose constrained by the bone length and bone joint angle

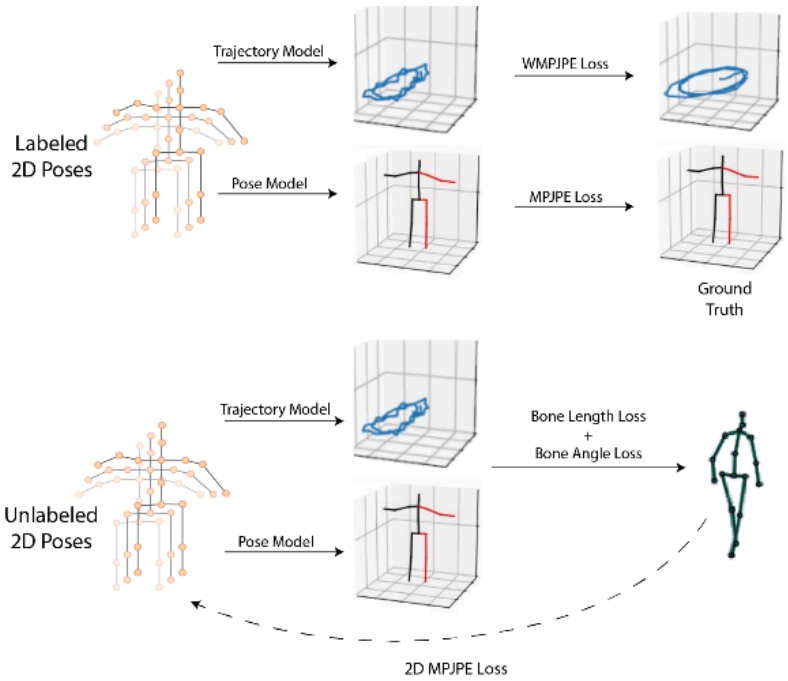


Figure 2

Semi-supervised Learning. A regression is done in the unlabelled 2D pose trajectory to predict the 3D pose constrained by the bone length and bone joint angle.

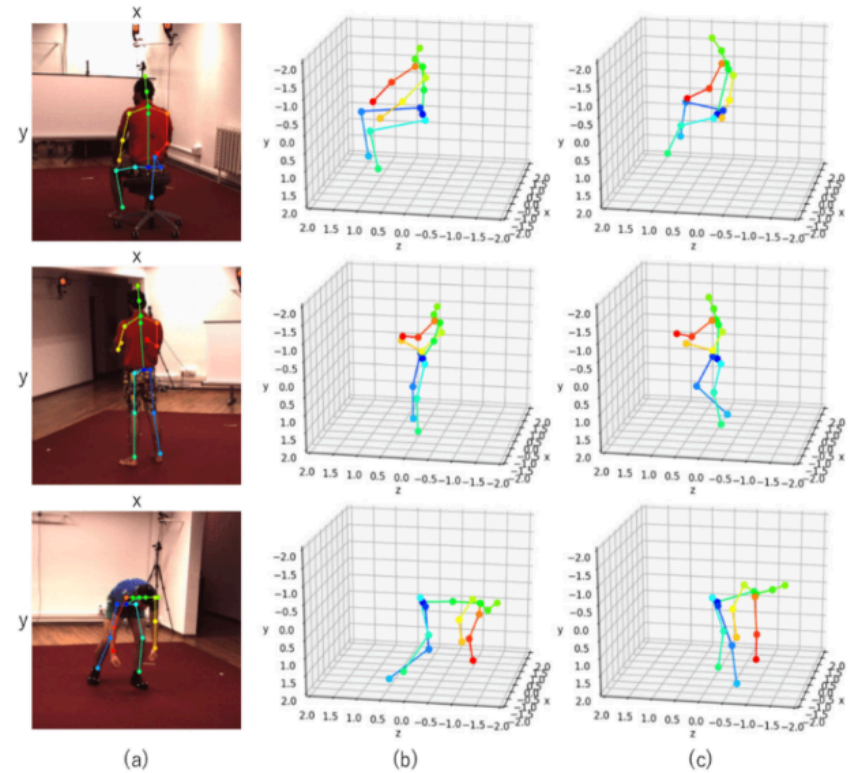


Figure 3

Some labeled samples from the Human3.6m dataset with their predicted 3D poses (middle) and ground truth (right).

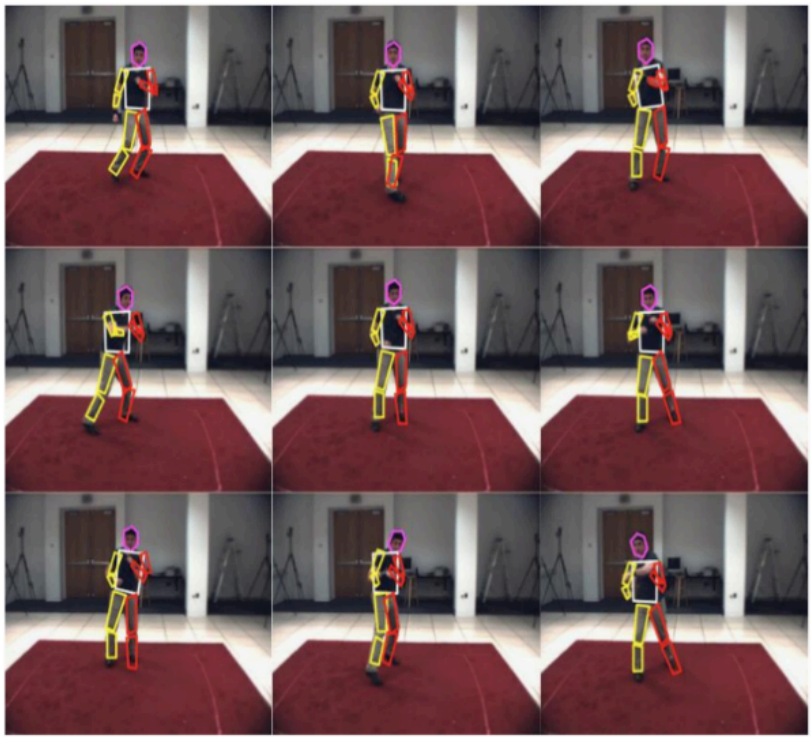


Figure 4

Some labeled samples from the HumanEva-1 dataset.

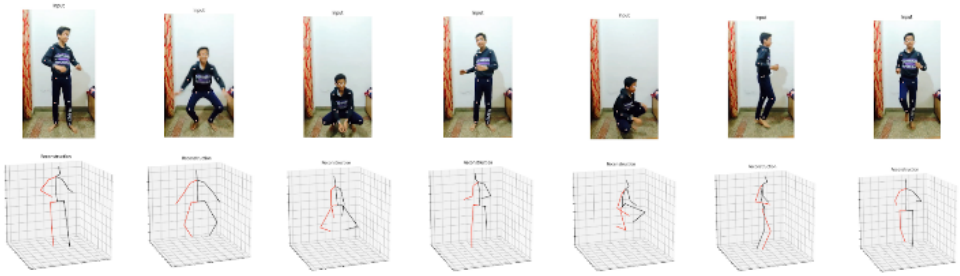


Figure 5

Predicted 3D poses on custom video. The images on top are the input to the model and the bottom images are their predicted 3D poses.

Check for updates

Received: 12 October 2023 | Accepted: 11 April 2024

DOI: 10.1111/ajsp.12621

REGULAR ARTICLE

Asian Journal of Social Psychology

Gratitude research: Review and future agenda using bibliometric analysis of the studies published in the last 20 years

Naval Garg

Delhi Technological University, New Delhi, India

Correspondence
Naval Garg, Delhi Technological University, New Delhi, India.
Email: naval.garg@dtu.ac.in

Abstract

Although there is an exponential rise in the literature on gratitude research, there is no attempt to systematically identify the historical development and recent trends in gratitude research. Gratitude has been defined diversely as an emotion, an attitude, a moral virtue, a habit, a personality trait or a coping response. As an emotion, people experience gratitude when they receive something valuable from someone else. As a trait, it is defined as a tendency to recognise and respond to goodness in others. The present study uses bibliometric analysis to review gratitude publications from the last 20 years, that is from 2001 to September 2023. WoS (Web of Science) identified 2982 publications referring to gratitude in title, abstract or keywords. It comprised 2860 articles and 122 review papers. The extracted data were analysed and visualised with the help of two analytical tools, that is the WoS analysis and VOSviewer (version 1.6.16). This study elicits the number of publications and citations from 2001 to September 2023, and most cited publications, and the most influential authors, articles, publishers, universities and countries as performance analysis. Furthermore, collaboration among countries, keyword co-occurrence and recent trend analysis are employed through science mapping. The results reveal that the major research areas of gratitude research are psychology, social sciences, business economics, psychiatry and public environmental and occupational health. The keywords co-occurrence suggest five major research clusters: evolutionary studies of gratitude, gratitude and health, gratitude and positive psychology, gratitude among children and adolescent and mediating and moderating studies of gratitude. Also, the analysis of recent 5-year studies highlights a clear trend of scientific explorations of gratitude against earlier trends of articles on philosophical and religious connotations of gratitude.

KEYWORDS
bibliometric, gratitude, VOSviewer

1 | INTRODUCTION

Gratitude has been considered a revered virtue in nearly all major religions and cultures of the world. The very notion of expressing gratitude to God through prayers and rituals permeates the teaching of the holy book 'The Bible' (Emmons & Kneezel, 2005). Gratitude is one of the pivotal moral values of Islam (Ali et al., 2019). Sixty-nine verses in the Holy Quran glorify gratitude or one of its cognates. In addition, the five rnas (auspicious duties) conceptualisation of Hinduism offer a foundation for gratitude towards ancestors, parents, friends, teachers, knowledge, God, nature and the environment (Garg, 2023). Hinduism offers one of the most comprehensive conceptualisations of gratitude, wherein Hindus worship almost everything (animals, mountains, rivers, sun, moon, soil, season, etc.) to express their gratitude for others

© 2024 Asian Association of Social Psychology and John Wiley & Sons Australia, Ltd.

(Garg, 2017). Apart from these three largest religions, other religions like Buddhism, Jainism, Judaism, etc., also favour gratitude for individual, group and societal well-being.

Although appreciated and encouraged traditionally, gratitude has captured the attention of modern scientists in the last 20 years or so. Traditionally, gratitude was a subject matter for philosophers and religious leaders. Still, a few early psychologists and sociologists, like Simmel (1950) and Heider (1958) tried to conceptualise gratitude as a construct that could be measured and studied scientifically. The fundamental scientific exploration of gratitude started with a series of articles by Robert Emmons, Michael McCullough and colleagues during the early 2000s (McCullough et al., 2001). There has been no looking back since then, and gratitude has been widely researched in various disciplines like psychology, psychiatry, business and economics, health care, clinical and therapeutic studies, etc. Today, a large body of research delineates various definitions, concepts, theories, forms, dimensions, interventions and benefits of gratitude.

The existing literature on gratitude could be understood through the following themes: (a) what is gratitude; (b) forms and dimensions of gratitude; (c) individual, social and cultural factors linked to gratitude; (d) individual and social benefits of gratitude; (e) gratitude interventions; and (f) gratitude scales. Regarding the definition of gratitude, there is no all-agreed definition and conceptualisation of gratitude (Wood et al., 2010). Gratitude has been operationalised diversely as 'emotion, an attitude, a moral virtue, a habit, a personality trait or a coping response' (Emmons & McCullough, 2003, p. 337). Rosenberg (1998) suggested that gratitude could be understood at three levels: emotion, mood and trait. As an emotion, people experience gratitude when they receive something valuable from someone else (Watkins, 2014). As an affective trait, gratitude is a 'general tendency to recognize and respond with grateful emotions to the roles of other people's benevolence in the positive experiences and outcomes that one obtains' (McCullough et al., 2002, p. 112). With the emergence of positive psychology 2.0, researchers are contemplating a new conceptualisation of gratitude: existential gratitude, which is the tendency to remain grateful or thankful even in challenging or difficult situations (Jans-Beken & Wong, 2019).

Although there is a rapid growth in the literature on gratitude research, there has been no attempt to systematically identify the historical development and recent trends in gratitude research. To the best of our knowledge and explorations, no previous researcher has reviewed and analysed the trends within the rapidly expanding domain of scientific knowledge of gratitude. Although a few scholars conducted meta-analysis and systematic reviews of gratitude studies (please see Boggiss et al., 2020; Cregg & Cheavens, 2020; Davis et al., 2016; Dickens, 2017; Renshaw & Steeves, 2016), the bibliometric analysis of gratitude researcher is still unknown. In their meta-analysis, Cregg and Cheavens (2020) investigated the impact of gratitude interventions on symptoms of anxiety and depression. In another study, Boggiss et al. (2020) analysed 19 articles to systematically analyse the impact of gratitude interventions on physical health and health behaviour. Based on 38 studies, Dickens (2017) conducted a meta-analysis of gratitude interventions. Similarly, Davis et al. (2016) and Renshaw and Steeves (2016) also conducted meta-analyses of the available literature on gratitude interventions and their impact on various specific variables. They listed down different gratitude interventions like counting your blessings, daily gratitude listing, gratitude collages, gratitude letters and visits, and introspection. There are several cultural-specific gratitude interventions also. For instance, Japanese people often express gratitude through short poems, which are referred to as 'Haiku'. Similarly, Indians express their gratefulness through different festivals and functions. For instance, Indians celebrate a 15-day festival known as '*shrad'* to articulate their gratitude for their ancestors and forefathers. Contemporary studies are exploring the beneficial effects of technology-based gratitude interventions such as telehealth services, counselling and psychotherapy platforms. Moreover, researchers are also investigating the impacts of diverse web-based interventions like daily gratitude writeups on social media, blogging, and gratitude journaling apps. It is argued that meta-analysis explains '(1) the overall strength and direction of effects or relationships, and (2) the across-study variance in the distribution of effect-size estimates and the factors that explain this variance' (Aguinis et al., 2010, p. 310). Further, meta-analyses are restricted to the examination of the direction and strength of effects and relationships among variables (Donthu et al., 2021). A systematic literature review uses systematic procedures to acquire, arrange and assess available literature. They are restricted to niche research areas like domain-based reviews, method-based reviews or intervention-based reviews.

In contrast, bibliometric analysis is appreciated and recommended for three obvious reasons. First, bibliometric analysis is not confined to the exploration of any specific domain, interventions or association between two variables. It extensively explores different facets of a phenomenon through a comprehensive evaluation of its research impact. Second, it is highly appreciated for its objectivity and ability to handle large data and files using mathematical and statistical methods (Donthu et al., 2021; Singh et al., 2020). Objectivity and a logic-driven approach differentiate bibliometric analysis from meta-analysis and systematic literature reviews. Third, it offers significant indicators of the growth and development of a field of knowledge, including

1467839x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

growth of scholastic publications, prominent scholars, journals, institutions or countries. Accordingly, the present study uses bibliometric analysis to offer an extensive understanding of publications, citations, and authors, including their institutions and countries. It highlights how gratitude research has unfolded over the last two decades. This study also reports how gratitude research organically develops and disseminates across countries. It evaluates historical academic developments, emerging and active research areas and trends in the field of gratitude. 'It enables us to unpack the evolutionary nuances of a specific field, while shedding light on the emerging areas in that field' (Donthu et al., 2021, p. 285). As suggested earlier, the scientific studies on gratitude started in 2001 with efforts of early researchers like Robert Emmons, Michael McCullough and colleagues; thus, this study performs a bibliometric analysis of gratitude research over a period of 2001 to September 2023. It tends to answer the following research questions.

RQ1. What are the current publication and citation trends of gratitude research?

RQ2. Which are the most influential authors, articles, journals, publishers, and countries publishing extensively in the domain of gratitude?

RQ3. What are the different research areas that carry out research in the gratitude domain?

RQ4. Which themes of gratitude are contemporary researchers exploring, and which topics have scope for future research?

To answer these questions, performance analysis and science mapping (two categories of bibliometric analysis) are performed. Performance analysis highlights the most impactful and influential articles, authors, countries and journals. It also reveals trends in publications and articles. Thus, performance analysis helps in addressing the first two research questions of the study. In addition, science mapping is used to underline the network of publications based on the attributes and characteristics of the article (Donthu et al., 2021; Zupic & Čater, 2014). In the present study, two bibliometric techniques of science mapping are used. First, co-citation analysis examines the associations among cited publications, which offers an insight into the knowledge base of gratitude research accumulated over the years (Dabic et al., 2013). It addresses the third research question by revealing the major research areas that focus on exploring different facets of gratitude. Second, keyword co-occurrence analysis explores the relationship between author keywords. It addresses the fourth research questions by highlighting major research clusters and recent trends in the area of research.

2 | METHOD

2.1 | Data source, search strategy and data retrieval

The web of Science (WoS) database was used to extract data for bibliometric analysis. It has a vast collection of over 24,000 journals across 254 subject disciplines. The availability of data about publication, citation, author, affiliation, country, etc., makes WoS suitable for bibliometric analysis. Previously, scores of researchers, including Huang et al. (2022), Donthu et al. (2021) and Ohlan et al. (2022), have also relied on the WoS database for bibliometric studies. Other databases like Google scholar, PubMed, PsycInfo and Scopus were also considered but WoS database was selected for its more relevance in the present context. Google scholar does not provide a dataset in such a form suited for bibliometric analysis. The PubMed dataset predominantly comprises research in life sciences and biomedical topics. And the PsycInfo database consists of works in the field of psychology. However, gratitude studies are distributed in psychology, management, sciences, psychiatry, etc. Thus, the WoS database was preferred over Google scholar, PubMed and PsycInfo databases. Also, the WoS database was preferred over Scopus for two obvious reasons. First, WoS is an older database than Scopus. Second, WoS has 68 data field tags whereas Scopus offers 43 data field tags. Thus, WoS provides comparatively more avenues for data analysis (Singh et al., 2021). Accordingly, the WoS database was used for the present study. The word 'gratitude' was searched on WoS in the title, abstract or keywords of the publications for the year from 2002 to September 2023. Only documents of the article and review types were considered, and other document types like books, chapters, letters, etc were discarded. It resulted in a collection of 3227 publications, which were considered for analysis. The abstracts of all these publications were manually screened to check for any casual use of word gratitude in title, keywords or abstract. It leads to removal of 249 articles from the dataset. In addition, articles published in 2001 were manually included in the study. It is realised that there were four publications in the year 2001. Thus, finally 2982 publications were considered for this study. WoS allows the export of data in files with only 500 documents; thus, six files were exported.

2.2 | Data analysis and visualisation

The extracted data were analysed and visualised with the help of two analytical tools, that is the WoS analysis and VOSviewer (version 1.6.16). The WoS provides

1467839x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

information about the most prominent authors, countries, affiliations, journals and publication years. Apart from the information provided by WoS, the VOSviewer provides visualisation facilities for country collaboration, co-authorship and keyword co-occurrence. Both performance analysis and science mapping techniques are used to analyse the exported 2982 articles. This study elicits the number of publications and citations from 2001 to September 2023, the most cited publications, and the most influential authors, articles, publishers, universities and countries as performance analysis. Also, major research areas of gratitude research are explored. Furthermore, collaboration among countries, keyword co-occurrence and recent trend analysis are employed through science mapping. Donthu et al. (2021) recommend that performance analysis and co-citation analysis reveal the knowledge foundation (past knowledge), and keyword co-occurrence analysis reveals the knowledge production (present trend of knowledge development). Accordingly, this study employs both methods to dive deep into the ocean of knowledge in gratitude research.

3 | RESULTS

3.1 | Number of publications over 2001–September 2023

WoS identified 2982 publications referring to gratitude in title, abstract or keywords. It comprised 2860 articles and 122 review papers. The first article in the period between 2001 and September 2023 was by McCullough et al.'s (2001), 'Is gratitude a moral affect?' The article conceptualises gratitude as a moral affect. It suggests three morally relevant functions of gratitude: moral barometer, moral motivator and moral reinforcer. The number of publications remained less than 100 and 200 per year until 2012 and 2016, respectively. After that, there was an exponential growth in the number of publications, reaching 432 in 2022 and subsequently 292 until September 2023. The last 5 years, that is 2018–2022, recorded 1638 publications, which were 54.92% of total publications since 2001. The number of articles has increased from 205 in 2018 to 432 in 2022; thereby, publications in gratitude showed a growth of more than 100% in the last 5 years. Figure 1 provides the year-wise growth in publication on gratitude. It clearly reflects a spurt in publication. The trend is quite obvious, as earlier studies primarily focused on conceptualisation and theoretical development. And banked upon these developments, later researchers looked upon gratitude to solve societal problems. Similarly, earlier studies conceptualised gratitude as emotion, mood, trait, etc. and developed scales and interventions to measure and augment gratitude. And, in recent years, scholars have relied on previously accumulated knowledge to explore the utility of gratitude in different fields like physical and mental health, professional organisations, societal challenges, etc.

3.2 | Most cited publications

Table 1 reports the top 10 cited publications in the field of gratitude. The most cited publication is Tangney et al.'s (2007) 'Moral emotions and moral behavior'. It is published in the *Annual Review of Psychology* and has been cited 1712 times (107 citations per year). It was a review paper that examined gratitude as a moral emotion. The most cited empirical research examined the relationship between gratitude and subjective well-being with the help of an experimental research design (Emmons & McCullough, 2003). It was the first empirical manuscript to use the 'gratitude journal' methodology. Another

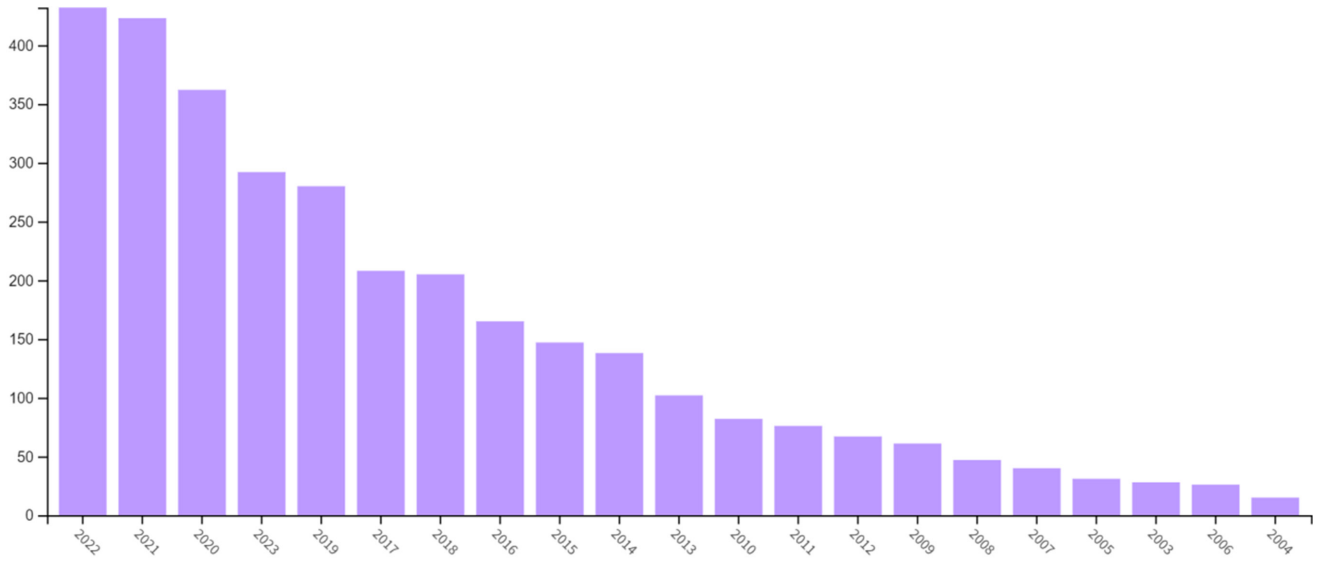


FIGURE 1 Publication trends from 2001 to 2022. *Source*: WoS database.

14678390, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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 Top 10 most cited publications on gratitude during 2001–September 2023.

S. No	Title	Authors	Journal	Total citations (citation per year)
1	Moral emotions and moral behaviour	Tangney et al. (2007)	*Annual Review of Psychology*	1712 (107)
2	Counting blessings versus burdens: An experimental investigation of gratitude and subjective well-being in daily life	Emmons and McCullough (2003)	*Journal of Personality and Social Psychology*	1551 (77)
3	The grateful disposition: A conceptual and empirical topography	McCullough et al. (2002)	*Journal of Personality and Social Psychology*	1542 (73)
4	What good are positive emotions in crises? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001	Fredrickson et al. (2003)	*Journal of Personality and Social Psychology*	1511 (75)
5	Gratitude and well-being: A review and theoretical integration	Wood et al. (2010)	*Clinical Psychology Review*	934 (72)
6	Strengths of character and well-being	Park et al. (2004)	*Journal of Personality and Social Psychology*	761 (40)
7	Gratitude and prosocial behavior	Bartlett and DeSteno (2006)	*Psychological Science*	715 (39)
8	Feeling and believing: The influence of emotion on trust	Dunn and Schweitzer (2005)	*Journal of Personality and Social Psychology*	704 (42)
9	Witnessing excellence in action: the 'other-praising' emotions of elevation, gratitude, and admiration	Algoe and Haidt (2009)	*Journal of Positive Psychology*	671 (56)
10	The nature of feelings: evolutionary and neurobiological origins	Damasio and Carvalho (2013)	*Nature Reviews Neuroscience*	589 (59)

Source: WoS database.

top-cited publication investigates the role of positive emotions, including gratitude, in crisis management regarding the terrorist attacks on the United States on 11 September 2001 (Fredrickson et al., 2003). In another most-cited empirical article, Algoe and Haidt (2009) differentiated between elevation, gratitude and admiration. They suggested that gratitude motivates trust-worthy relationships with benefactors, and admiration leads to self-improvement. Interestingly, 4 of the top 10 cited publications on gratitude were published in the *Journal of Personality and Social Psychology*. In addition, all 10 top-cited articles on gratitude were published in journals related to psychology. It reflects a clear dominance of psychology-related explorations of gratitude. Also, all 10 most-cited articles and reviews were published in or before 2013. Among all types of articles published after 2013, the most cited publication focused on the multidimensional approach to well-being using the PREMA (positive emotions, engagement, relationships, meaning and accomplishment) framework (Kern et al., 2014).

3.3 | Most prolific journals and publisher

There are 1053 journals with at least one publication on gratitude. Also, there are only 20 journals with 10 or more publications. The *Journal of Positive Psychology* has published the highest number of papers on gratitude (136 publications and 4260 citations) followed by *Frontiers in Psychology* (111 publications and 1405 citations). The *Journal of Personality and Social Psychology* has only 18 publications but has the highest citation of 5027. It is interesting to note that seven out of the top 10 are psychology journals (except *Journal of Religion and Health*, *International Journal of Environmental Health and Public Health*, and *PLoS One*). It suggests a clear dominance of gratitude studies in the domain of psychology. This observation corresponds with the findings of Table 1. Another prominent area of gratitude research is 'religion and spirituality', with two journals related to religion featuring in the list. These two journals are *Journal of Religion and Health*, and *Religion*. It highlights the significance of gratitude in major religions of the world. The list also reveals gratitude studies in social areas, including social psychology, social behaviour and social indicators. Also, only one journal, *Journal of Business Research*, directly linked to business and management is there in the list. It clearly reflects the scope of future researchers. As far as publishers are concerned, Taylor and Francis (471 publications) have published the maximum number of papers on gratitude, followed by Elsevier (450 publications) and Springer Nature (443

1467839x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

publications). Top rated journals published by Taylor and Francis are *Journal of Positive Psychology* and *Cognition and Emotion*. Tables 2 and 3 provide lists of journals and publishers with the highest number of documents on gratitude.

3.4 | Most prolific authors, universities and countries

Table 4 illustrates the list of authors with 10 or more publications on gratitude, along with their affiliation and country. Todd B. Kashdan from George Mason University (USA) has contributed a maximum number (25 publications) of papers. The author's notable contribution includes exploring the associations between gratitude and suicidal ideation, depressive symptoms, social anxiety and subjective well-being. Other top contributors in the field of gratitude are Sonja Lyubomirsky and Sara B. Algoe, with 18 publications each. Sonja Lyubomirsk majorly examined the role of gratitude in sustaining positive emotions, self-improvement, positive change and humility. Sara B. Algoe's significant contribution to the literature on gratitude is the 'find-remind-and-bind theory of gratitude'. The theory states that gratitude helps find and remind goodness in others and in relationships. And, thus, it binds one with goodness and social relations. Another notable contributor, although with 11 publications, is Philip Watkins. His notable contributions include the development and validation of the 'Gratitude Resentment and Appreciation Scale' (GRAT) and GRAT-16 short scale. His book titled 'Gratitude and the Good Life: Toward a Psychology of Appreciation' has also received much attention and appreciation throughout the world. The list also features an eminent scholar, Robert A. Emmons. However, it is disheartening to observe the non-inclusion of a few stalwarts like Michael E. McCullough. The probable reasons for such exclusion could be the non-indexing of journals in WoS, articles prior to 2001, and publications excluded from the scope of the present study like book, book chapters and proceeding articles. Out of 20, 11 top contributors are from the United States, followed by Taiwan with two publications. Table 5 reports the top 25 universities and countries in terms of the number of publications of gratitude. The top contributing country is the United States, with 1509 documents comprising 50.6% of the total publications considered in this study. Other top contributing countries are China (357 publications), England (273 publications), Australia (181 publications) and Canada (187 publications). It is also observed that citations per 1 million population are highest for Switzerland (168.89 citations per 10 million population), followed by New Zealand (118.10 citations per 10 million population), and the United States (117.10 citations per 10 million population). Also, the University of Michigan, the University of North Carolina and the University of Pennsylvania

TABLE 2 Journals with 15 or more publications on gratitude during 2001–September 2023 (WoS).

S.No	Journal	No. of publications	Citations	IF[a]
1	*Journal of Positive Psychology*	136	4260	4.30
2	*Frontiers in Psychology*	111	1405	4.62
3	*Journal of Happiness Studies*	77	2604	3.80
4	*Personality and Individual Differences*	57	1925	6.80
5	*Current Psychology*	53	337	2.80
6	*Religion*	44	315	1.00
7	*International Journal of Environmental Research and Public Health*	40	252	NA
8	*Emotion*	38	2959	4.20
9	*Journal of Religion and Health*	30	297	2.80
10	*PLoS One*	25	257	
11	*Cognition and Emotion*	24	1384	2.60
12	*Journal of Business Research*	21	548	11.30
13	*Journal of Personality and Social Psychology*	18	5027	7.60
14	*Personality and Social Psychology Bulletin*	16	703	4.00

[a]Impact Factor as per Journal Citation Reports-Clarivate Analytics (2022).
Source: WoS database.

TABLE 3 Top 10 publishers (number of documents).

S.No	Publisher	No. of publications
1	Taylor and Francis	471
2	Elsevier	450
3	Springer Nature	443
4	Wiley	345
5	Sage	330
6	Frontiers media	148
7	MDPI	110
8	Amer Psychological Association	103
9	Cambridge University Press	96
10	Lippincott Williams & Wilkins	96

Source: WoS database.

14678398, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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 Authors with 10 or more publications on gratitude.

S.No	Authors	Affiliation	Country	Document	Citation
1	Todd B. Kashdan	George Mason University	USA	25	1999
2	Sara B Algoe	University of North Carolina	USA	22	2082
3	Sonja Lyubomirsky	University of California, Riverside	USA	22	1922
4	Arndt Buessing	Witten Herdecke University	Germany	20	256
5	Lung Hung Chen	National Taiwan Sport University	Taiwan	17	437
6	Robert H. Pietrzak	Yale University	USA	17	476
7	Alex M. Wood	Leeds Trinity University	UK	16	3016
8	Chih-che Lin	National Taipei Univ. of Technology	Taiwan	16	278
9	Jeff C. Huffman	Harvard Medical School	USA	15	493
10	Willibald Ruch	University of Zurich	Switzerland	15	933
11	Jonathan R. H. Tudge	University of North Carolina	USA	14	208
12	Feng Kong	Shaanxi Normal University	China	13	342
13	Jeffrey J. Froh	Hofstra University	USA	12	2666
14	Frank D. Fincham	University of Florida	USA	12	1023
15	Christopher M. Celano	Harvard Medical School	USA	12	424
16	Lourdes Rey	University of Malaga	Spain	12	130
17	Emily Impett	University of Toronto	Canada	12	397
18	Naval Garg	Delhi Technological University	India	12	47
19	Neal Krause	University of Michigan	USA	12	329
20	Robert A. Emmons	University of California, Davis	USA	11	1071

Source: WoS database.

TABLE 5 Top 20 universities and countries/regions (number of documents).

Countries/regions	Document	Citation	Citation/ Population	Universities	Document	Citation
1. USA	1509	46,043	117.10	1. University of Michigan	52	4342
2. China	357	4498	1.99	2. University of North Carolina	46	2422
3. England	273	7807	95.76	3. University of Pennsylvania	41	3425
4. Canada	187	2919	95.11	4. Harvard University	39	1324
5. Australia	181	3344	56.76	5. Harvard Medical School	36	454
6. Germany	104	1543	15.33	6. George Mason University	36	4119
7. Spain	81	1188	19.57	7. Baylor University	35	975
8. Taiwan	75	1147	2.63	8. Duke University	34	1094
9. South korea	75	925	35.88	9. University of California, San Diego	33	590
10. Netherlands	72	1611	72.83	10. Beijing Normal University	31	494
11.Japan	70	427	11.65	11. Yale University	31	1411
12. Italy	67	117	15.51	12. University of Toronto	30	704
13. Poland	58	605	168.89	13. Massachusetts General Hospital	30	488
14. Switzerland	57	1709	47.43	14. Florida State University	30	1312
15. Sweden	51	614	6.39	15. University of Minnesota	28	576
16. India	48	265	11.65	16. University of Melbourne	28	635
17. Israel	48	550	54.36	17. Northwestern university	27	524
18. Singapore	45	783	4.35	18. University of California, Riverside	27	2058
19. Norway	41	810	2.45	19. University of California, Berkeley	26	1990
20. New Zealand	39	620	118.10	20. National University of Singapore	26	551

Note: World population as per world bank/countrymeters.info, Citation/population (I millions).
Source: WoS database.

1467839x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

are the top three contributing universities. The top five universities are from the United States.

3.5 | Research areas

Table 6 enlists major research areas of gratitude. Most research on gratitude has been carried out in psychology (1276 publications, 42.79% of total publications of 2982 from 2001 to September 2023). Other vital research areas include social sciences (271 publications, 9.08%), business economics (219 publications, 7.35%), psychiatry (204 publications, 6.91%) and public environmental and occupational health (207 publications, 6.98%). It suggests that although the majority of gratitude studies belong to psychology, there are gratitude researchers in other areas like social science, management, religion, philosophy, etc. too. Table 6 also reports publications in different research areas in two time frames: 2001–September 2023 and 2017–2023. It shows a decline in the research on gratitude (percentage-wise) in psychology, religion and philosophy, social science, education, and nursing from 2001–16 to 2017–23. During the same period, some research areas highlight a promising increase in research on gratitude. These areas include psychiatry, health care services, public environmental and occupational health and business economics. It suggests that contemporary researchers seem more interested in examining health and organisational utilities of gratitude. A closer look at research areas of publications of recent years suggests an increase in gratitude research in other research areas too.

TABLE 6 Top 10 research areas.

Research areas	2002–2023	%	2017–2023	%
Psychology	1276	42.79	881	40.44
Social Sciences	271	9.08	178	8.17
Business economics	219	7.35	160	7.34
Public Environmental and Occupational Health	207	6.91	157	7.20
Psychiatry	204	6.98	165	7.57
Religion	195	6.53	114	5.23
Nursing	161	5.39	116	5.34
Education Educational Research	120	4.02	90	4.13
Health Care Sciences	115	3.85	92	4.30
Philosophy	88	2.95	42	1.92

Note: %, Number of papers published during a period * 100/Number of paper published.
Source: WoS database.

3.6 | Collaboration among co-authors and countries

Figure 2 shows the networks among the top 151 authors with five or more publications from 2001 to September 2023 on gratitude. Seven author clusters were delineated with the help of VOSviewer. The top three authors with the highest linkages are Jeff C. Huffman (total link strength=35), Christopher M. Celano (total link strength=34) and Eleanor E. Beale (total link strength=27). Also, Figure 3 highlights the network among the top 38 countries with 10 or more published documents on gratitude. Again, seven country clusters were delineated with the help of VOSviewer. The top five linking countries are the United States (total link strength=336), China (total link strength=238), Australia (total link strength=94), Canada (total link strength=77) and Germany (total link

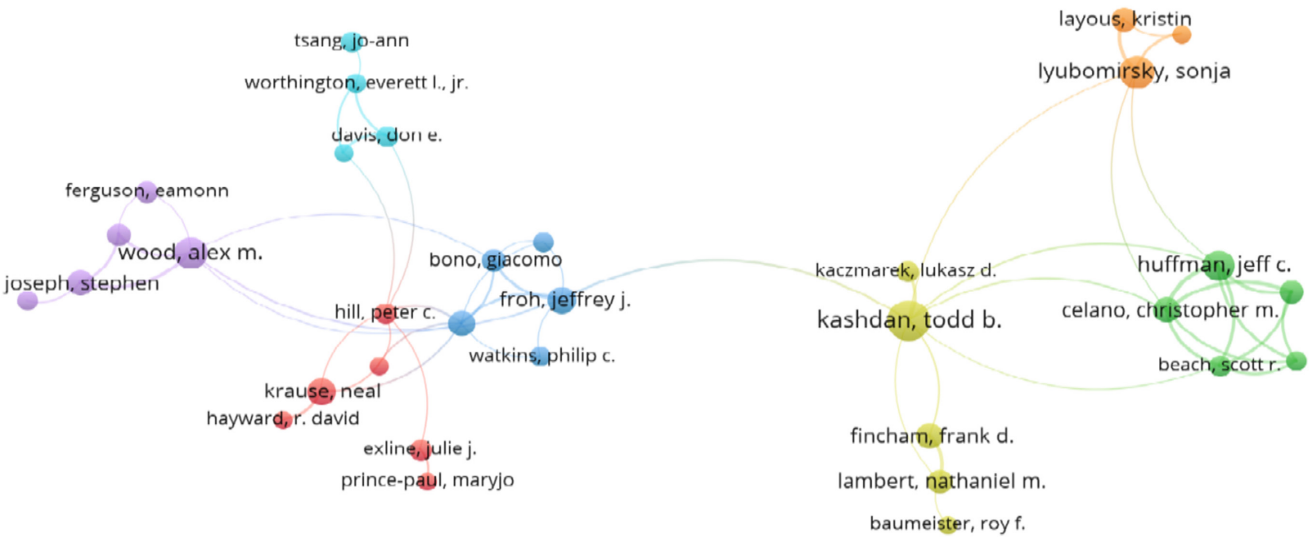


FIGURE 2 Co-authorship network among authors with five or more publications. *Source*: Delineated from VOSviewer.

14678388, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

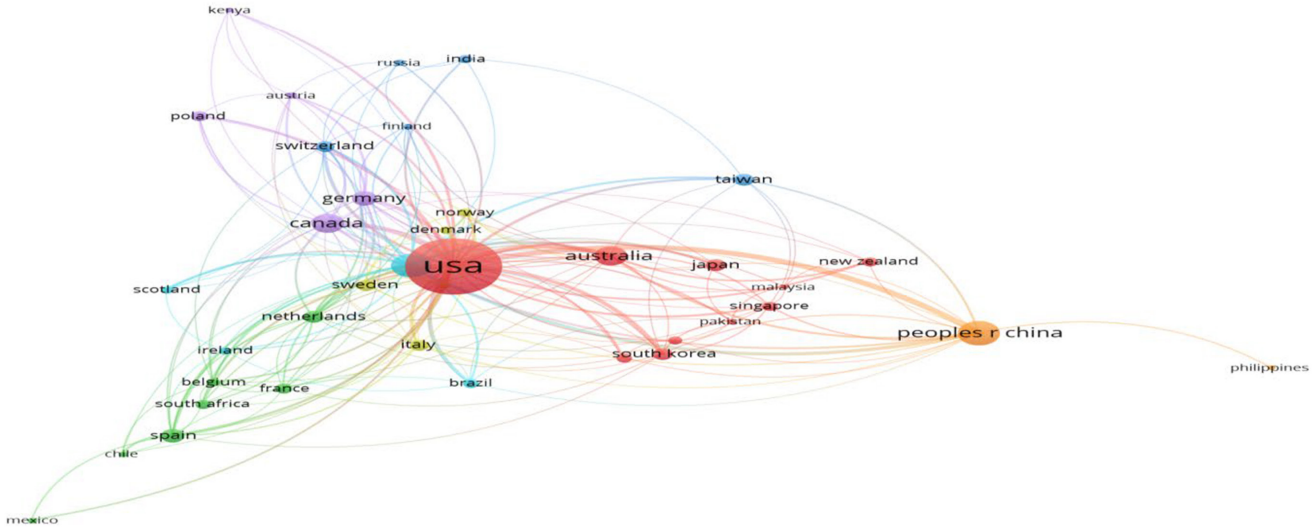


FIGURE 3 Co-authorship network among the top 50 countries with 10 or more publications. *Source*: Delineated from VOSviewer.

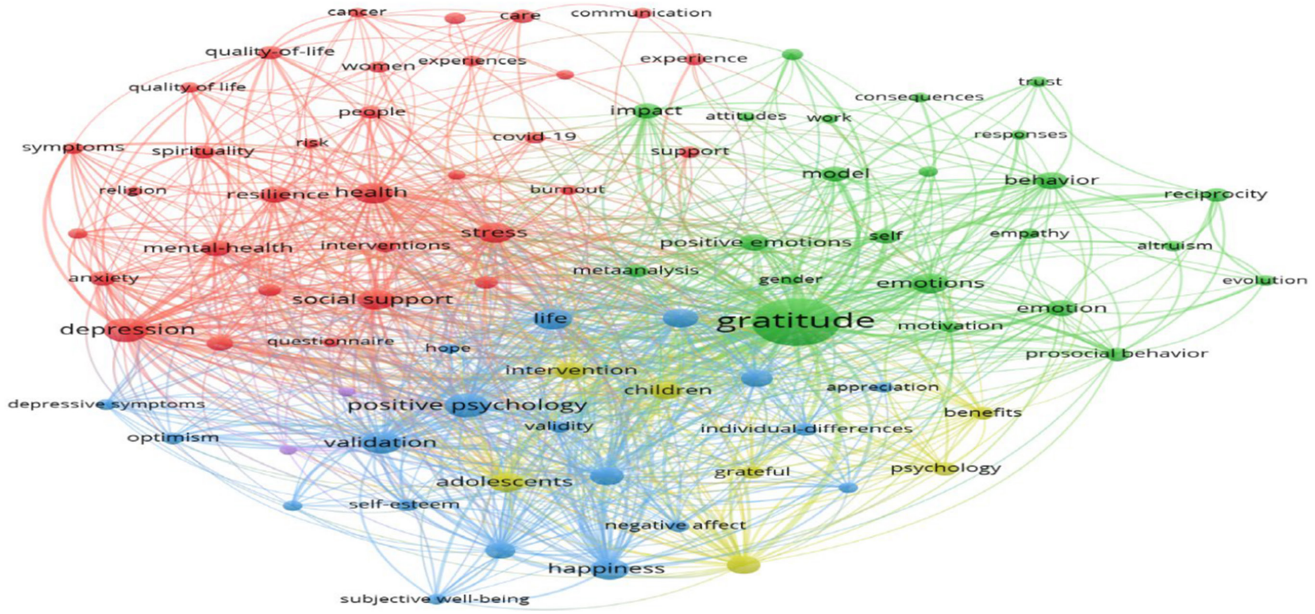


FIGURE 4 Keywords co-occurrence clusters in gratitude literature. *Note*: Five clusters are shown in different colours. The size of circles indicates the total link strength of keyword.

strength = 64). Among the two countries, the United States and China are the most networked, with link strength of 45. It is followed by the United States and Australia (link strength = 38) and the USA and Canada (link strength = 34). The most networked cluster has countries like the United States, Australia, Japan, Malaysia, New Zealand, etc.

3.7 | Keywords co-occurrence

VOSviewer was used to visualise keywords co-occurrence in the publications on gratitude. Figure 4 shows the following five keyword clusters.

3.7.1 | Cluster-1: Evolutionary studies of gratitude (in green colour)

The most prominent cluster (in green) includes keywords like positive emotion, prosocial behaviour, behaviour, attitudes, consequences, altruism, attitude, trait, work, etc. This cluster seems to inscribed early studies that conceptualised gratitude as emotion, mood, trait, attitude, habit, etc. A few notable contributors are Emmons and McCullough (2003), McCullough et al. (2002), Watkins (2014), Fredrickson (2004), Adler and Fagley (2004), Wood, Joseph, and Maltby (2009), Wood, Joseph, Lloyd, et al. (2009), etc. These studies

1467839x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

conceptualise gratitude as an emotion (feeling that arises when one acknowledges that someone else does something good for them), a mood (more frequent and pervasive grateful emotions), a trait (tendency to recognise goodness in one's life) and an attitude (acknowledging favourable and positive aspects of life). The first cluster also discusses general and work-related consequences/impacts of gratitude. Gratitude leads to lower materialism (Polak & McCullough, 2006), higher life-satisfaction (Lambert et al., 2009), counteracts hedonic adaptation (Lyubomirsky et al., 2005), broadens one's outlook (Fredickson, 2004), develops psychological and social capital (Fredickson, 2004) and cancels out negative effects (Garland et al., 2010). Work-related benefits of gratitude include higher job performance (Grant & Wrzesniewski, 2010), job satisfaction (Garg, 2017), organisational citizenship behaviour (Andersson et al., 2007), lower work stress (Saxena et al., 2020), less chances of workplace burnout (Chan, 2011), improved leadership (Garg et al., 2020) and greater work-family enrichment.

3.7.2 | Cluster-2: Gratitude and health (in red colour)

It comprises keywords like health, stress, anxiety, mental health, burnout, covid-19, depression, cancer, symptoms, care, etc. This cluster captures vital studies that investigate the linkage between gratitude and health. Various researchers, including, Krause et al. (2014), Hill et al. (2013), and Wood, Joseph, and Maltby (2009), Wood, Joseph, Lloyd, et al. (2009) reported a positive linkage between gratitude and better physical health. Emmons and McCullough (2003) and Jackowska et al. (2016) observed better sleep among participants who kept a gratitude journal. Similarly, notable scholars like Wood et al. (2008), Froh et al. (2009), Petrocchi and Couyoumdjian (2015) and Disabato et al. (2016) reported that grateful people experience lower levels of stress, depression and anxiety. Several other studies have revealed a link between different gratitude interventions and decreased mental stress and anxiety. Studies reported decreased depression owing to the three/five good things activity (Seligman et al., 2005), gratitude journals (Kerr et al., 2014) and gratitude letters (Toepfer et al., 2011).

3.7.3 | Cluster-3: Gratitude and positive psychology (in blue colour)

The third cluster includes positive psychology, validation, happiness, appreciation, subjective well-being, self-esteem, measurement, optimism, hope, life satisfaction, etc. This cluster majorly focuses on two major subthemes: (a) development and validation of measures of gratitude; (b) relationship of gratitude with psychological constructs like hope, happiness, satisfaction, resilience, optimism, etc. It includes studies that propagates major gratitude measures like the 'Gratitude Questionnaire-6 (GQ-6)' (McCullough et al., 2002), 'Gratitude Resentment and Appreciation Test' (GRAT) (Watkins et al., 2003), 'Transpersonal Gratitude Scale' (TGS) (Hlava et al., 2014), 'Multi-Component Gratitude Measure' (MCGM) (Morgan et al., 2017) and 'Existential Gratitude Scale' (EGS) (Jans-Beken & Wong, 2019). Subsequently, these scales have been explored and validated in different cultural contexts (please see, Oğuz-Duran, 2017; Garg et al., 2021; Garg & Mehak, 2021; Hammer & Brenner, 2017; Langer et al., 2016). The third cluster also highlights that gratitude is positively associated with various psychological constructs. The direct positive relationship between gratitude and happiness has been reported by numerous researchers including, Jackowska et al. (2016), Morgan et al. (2017), McCullough et al. (2002), Watkins et al. (2003), Seligman et al. (2005), Toepfer et al. (2011) and Garland et al. (2010).

3.7.4 | Cluster-4: Gratitude among children and adolescent (in yellow)

It includes keywords like counting blessings, gratefulness, psychology, children, adolescence and intervention. This cluster seems to have special focus on gratitude studies conducted on children and adolescents. While assessing gratitude among children and adolescents, researchers firstly tried to explore whether presently available measures (developed and tested on adults) suit children and adolescents. Froh et al. (2011) concluded that the GQ-6, GRAT and GAC could be used for 14–19 year olds and the GQ-6 and GAC better suits 10–13 year olds. And, parental reports or new scales are required to measure gratitude among children below the age of 10 years. Recently, Hussong et al. (2018), in their raise grateful children project, proposed three new measures of children's gratitude: (a) parent-report daily diary; (b) a child-report scenario; (c) a behavioural task. Other studies concluded significant cultural variations in expression of gratitude among children. American and Russian children express concrete and connective gratitude, respectively (Tudge et al., 2015). Studies have reported that parents, teachers and friends' gratuitous behaviour has a significant effect on the gratitude level of children. Several studies have focused on parental role in development of gratitude among children (please see Hoy et al., 2013; Hussong et al., 2018; Rothenberg et al., 2017). There are several studies in the context of students too. For instance, Garg, Kumari, et al. (2022), Garg, Mahipalan, et al. (2022), Garg, Verma, et al. (2022) reported significant moderating effect of gratitude in the relationship of positive reframing and technostress among students. Students high on gratitude are less likely to develop cyberbullying attitude

1467839x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

(Bansal et al., 2023). Gratitude may lead to better academic performance among students (Carmona-Halty et al., 2019; Clarkson, 2020; Slåtten et al., 2021). It is reported that sense of gratefulness helps students overcome challenges in their educational journey in schools and colleges (King et al., 2023). Also, gratitude has been reported to have positive linkages with academic engagement (Cui et al., 2023; Zhen et al., 2019), learning engagement (Jin & Wang, 2019), academic self-efficacy (Yudiani et al., 2023), effective learning (Valdez et al., 2022) and academic motivation (Nawa & Yamagishi, 2021).

3.7.5 | Cluster-5: Mediating and moderating studies (in purple)

It comprises keywords like mediating roles, moderators and students. The fifth cluster includes two types of studies: (a) Gratitude studies with some other variable acting as mediating and moderating variables; (b) Gratitude studies with gratitude acting as mediator or moderator. A few mediators explored in gratitude studies are psychological needs satisfaction (Tian et al., 2016), life satisfaction (Lambert et al., 2009), perceived social support (Gabana et al., 2018), resilience (Garg, Kumari, et al., 2022), positive reframing (Garg, Verma, et al., 2022) and constructive deviance (Garg, Kumari, et al., 2022). Similarly, a few previously examined moderators are positive affect (Froh et al., 2009), personality type (Senf & Liau, 2012), spiritual climate (Garg, Mahipalan, et al., 2022) and social intelligence (Garg et al., 2020). In another set of studies, the mediating and moderating effects of gratitude are examined. Lantz et al. (2021) examined the mediating effect of gratitude amid the association of religiosity and well-being. The 10 keywords, that co-occurred most frequently with gratitude are positive psychology (total link strength = 144), depression (total link strength = 138), happiness (total link strength = 123), validation (total link strength = 120), social support (total link strength = 114), count blessing (total link strength = 113), emotions (total link strength = 107), life (total link strength = 104), stress (total link strength = 94) and health (total link strength = 92).

3.8 | Recent trends

One thousand five hundred fifty-four papers published during 2017–2023 were used to visualise the recent trends in gratitude literature using VOSviewer. Figure 5 reports four keyword clusters. The most recent cluster (in yellow) shows that current research in gratitude focuses primarily on resilience, covid-19, burnout, posttraumatic growth, symptoms and mediating role. The second cluster (in yellowish-green) reveals a trend towards examining gratitude's role in well-being. The cluster includes happiness, optimism, self-esteem, and positive affect. The third cluster (in blueish-green) reflects a growing trend towards examining the psychometric properties of gratitude. It includes validation studies, scale, reliability, fit indices, etc. And the last cluster (in blueish) indicates a recent trend of exploring gratitude in different streams and populations. It includes keywords like religion, spirituality, social support, adolescence, children, etc.

4 | DISCUSSION

The present study analysed 2982 publications on gratitude from 2001 to September 2023 to examine the current and future trends in gratitude research. The findings reveal that the last 5–6years (2018–2023)

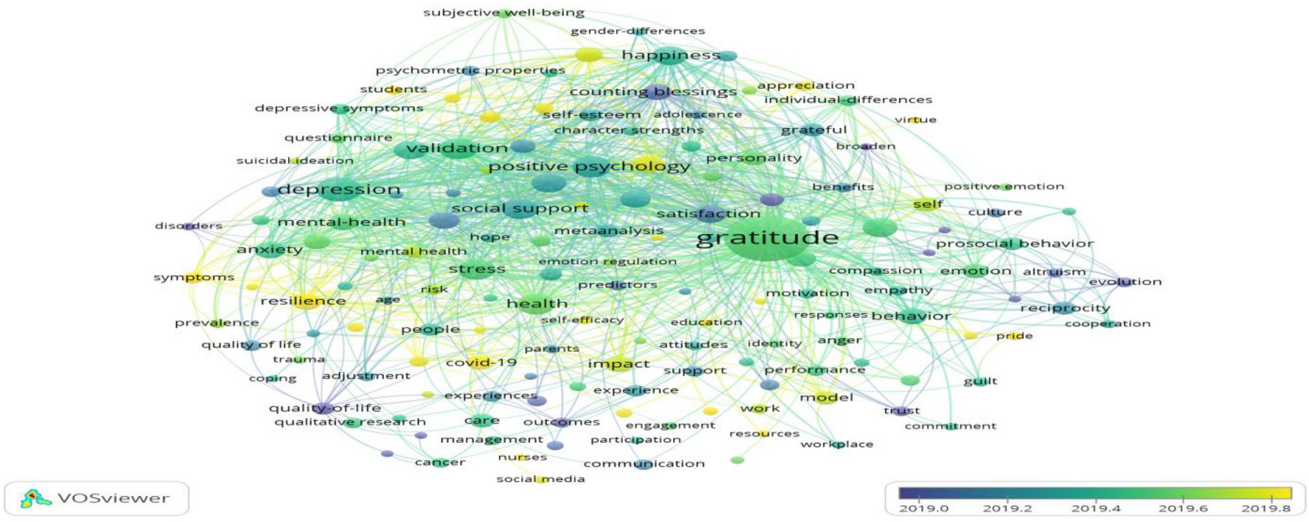


FIGURE 5 Recent trends in gratitude research. *Source*: Delineated from VOSviewer.

1467839x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

witnessed sharp jumps in both publication and citation. It is also observed that gratitude has been widely researched in the areas of psychology, social sciences, business economics, psychiatry and the public environment. However, research area-wise trends suggest that research in traditional domains of gratitude like religion and philosophy is declining. However, there has been a spurt in scientific explorations in the research areas of psychology, psychiatry, health care services and business economics in the last 5 years. It suggests two vital recent trends in gratitude research. First, there has been a clear trend of logical and scientific explorations of gratitude in recent years. Earlier gratitude studies were more inclined towards exploring the philosophical and religious connotations of gratitude. Contemporary studies on gratitude employ scientific methods like experimental research design, quasi-experimental prototypes or sophisticated statistical tools like structural equation modelling (SEM) or partial least squares regressions (PLS).

Second, recent publications are more on exploring the utilities and benefits of gratitude in the day-to-day lives of people. Researchers in business and management have started investigating the organisational utilities of gratitude. They suggest an association between gratitude and work-family enrichment (Garg, 2022), team innovation (Li et al., 2021), workplace mistreatment (Locklear et al., 2021), customer loyalty (Audrain-Pontevia & Garnier, 2021) and employees' helping behaviour (Sawyer et al., 2022). Similarly, contemporary public health scholars are examining the role of gratitude in modern-day lifestyle disease treatment. These studies suggest a significant association of gratitude with coping during COVID-19 (Beames et al., 2021), daily hassles (Tachon et al., 2021), stress among nurses (Lee et al., 2021), mental well-being among people with substance use disorders (Leung & Tong, 2017) and the quality of life of patients with chronic schizophrenic (Kim & Jun, 2021). Similarly, there has been a shift in the subject matter of psychological studies on gratitude. Earlier studies mainly emphasised the conceptualisation of gratitude. However, current studies investigate the linkage between gratitude and other psychological constructs like resilience, a sense of coherence (Fujitani et al., 2017), forgiveness and sense of humour (Chamorro-Garrido et al., 2021), and bullying and prosocial behaviour (Oriol et al., 2021).

The recent trend analysis suggests that the most recent cluster includes keywords like resilience, covid-19, burnout and posttraumatic stress disorders. Quite plausible, positive psychology has been dominated by corona-related researchers like Jans-Beken (2021), Oliveira et al. (2021), Fishman (2020), Datu and Fincham (2022), Belkin and Kong (2021) and Day et al. (2020). A search of the word 'gratitude and covid', when searched on WoS in the title, abstract or keywords of the publications for the year from 2019 to 2021, reflects 107 articles. A few major themes of these studies are 'gratitude as a coping mechanism' (Beames et al., 2021; Gaeta et al., 2021; Jans-Beken, 2021), 'gratitude and prosocial behavior' (Mesurado et al., 2021), 'gratitude and mental well-being' (Geier & Morris, 2022), 'gratitude and stress' (Liljestrand & Martin, 2021), and 'gratitude and psychological capital' (Beames et al., 2021).

Furthermore, trend analysis also highlights that the second most recent cluster is about gratitude and well-being. A search of the word 'gratitude and well-being' when searched on WoS in the title, abstract, or keywords of the publications for the year from 2019 to 2021 reflects 336 articles. It reflects a clear inclination of contemporary researchers towards examining the role of gratitude in ensuring the well-being of practitioners. A close examination suggests that there are experimental and non-experimental studies that explore both direct and indirect associations between gratitude and well-being. The experimental studies examine the difference in the well-being of experimental and control groups (please see Chamorro-Garrido et al., 2021; Southwell & Gould, 2016). The indirect studies investigate the role of various mediating variables like interpersonal helping behaviour (Kersten et al., 2021), spirituality (Perez et al., 2021), resilience and social support (Kong et al., 2020) and cognitive and affective variables (Oriol et al., 2021).

In addition, trend analysis also suggests that the third recent cluster includes studies exploring the psychometric analysis of the gratitude scale. The development of a new scale or psychometric analysis of the existing scale is the first step towards extending the knowledge horizons of a theme in different cultures or populations (Sharma et al., 2022). Recent studies have examined the psychometric properties of existing gratitude scales like GQ-6, GRAT-16, etc., in different cultural contexts like India (Dixit & Sinha, 2021; Garg et al., 2021), France (Tachon et al., 2021), Vietnam (Tran et al., 2021), Philippines (Llenares & Almeda, 2021) and Japan (Komase et al., 2020). These studies indicate that gratitude studies are poised to flourish in new cultural contexts. Also, recent studies seek to develop new scales of gratitude in different contexts. For instance, gratitude at work scale (Cain et al., 2018; Youssef-Morgan et al., 2022), existential gratitude scale (Jans-Beken & Wong, 2019) and transpersonal gratitude scale (Hlava et al., 2014). Future researchers may explore the possibilities of developing new scales in different contexts, like the group gratitude scale, which may inquire into one's gratitude for a group rather than the individual. Similarly, new gratitude scales are being developed in different religious contexts like the Christian gratitude scale (Knabb et al., 2021) and the Islamic gratitude scale (Rusdi et al., 2021). These findings suggest that future researchers may try to develop new

1467839x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

gratitude scales for different religious backgrounds like Hinduism, Sikhism, Persian, etc.

Further, the fourth recent cluster derived during trend analysis highlights the growth of gratitude research in specific streams and populations like adolescence, children, religion, students, health workers, working professionals, etc. For instance, there is a growing trend towards analysing gratitude in the workplace. Contemporary researchers have suggested that gratitude is positively related to lower levels of workplace materialism (Unanue et al., 2021), job satisfaction and job performance (Cortini et al., 2019), work-family enrichment (Garg, 2022) and OCBs and pro-social behaviour (Di Fabio et al., 2017; Sarkar & Garg, 2020; Saxena et al., 2020). Similarly, there is a significant rise in gratitude research among health care workers.

This study offers insights into historical and current trends in gratitude research. This bibliometric analysis of gratitude research over the last 20 years would benefit students, new researchers and academicians. Similarly, meaningful information on the most cited publications, most prolific journals, authors and countries would be helpful to one interested in any scientific investigation of gratitude. The identification of recent research topics would definitely encourage contemporary and future researchers to examine new areas of gratitude research. For instance, one of the most recent trends is the rise in studies on the development of new scales and the examination of the psychometric properties of existing scales. It is quite understandable given the culture-sensitive nature of gratitude. It would entice future researchers to develop new scales and examine the psychometric properties of existing scales in different cultural contexts. The findings of the study may also help in the development of specific gratitude interventions. For instance, the results highlight two major research clusters: gratitude and health, and gratitude among children and adolescents. Researchers and practitioners might be interested in developing gratitude interventions specifically for children and adolescents, like video games that rewards gratitude and visits to orphanages or old-age homes. Academicians are suggested to develop specific courses on gratitude. Some universities, like Delhi Technological University, India, have already started specific courses on the 'art and science of gratitude'. This study would surely help in designing better curriculum and learning outcomes. Also, educational institutions across the globe are witnessing a sharp rise in centres and courses on human values. Any such centre or course on happiness, mindfulness, compassion, or love is incomplete without an adequate understanding of the concept of gratitude. Literature and practitioners regard gratitude as a potent mechanism of happiness, mindfulness, and compassion. Furthermore, based on recent directions, the present study offers some foresight into future trends in gratitude research, which may guide future researchers in their journey.

5 | LIMITATIONS AND CONCLUSION

Despite a thorough investigation of the existing research on gratitude, this bibliometric analysis has two major limitations. Firstly, this study is based on data retrieved from the WoS database. Although search, retrieval and analysis using the WoS database have been sufficiently justified and used in previous bibliometric studies (Donthu et al., 2021; Ohlan et al., 2022), there are significant possibilities of missing out on articles on gratitude that are not indexed in the WoS database. For instance, there could be journal articles indexed in Scopus but not in the WoS database. Future research may use some other software that allows researchers to integrate data from WoS and Scopus (after avoiding duplication of articles). This consolidated data could further be subjected to bibliometric analysis. Alternatively, future research could utilise the Scopus database to refute or support the generalisability of the present study's findings. Secondly, given a very large database of 2982 articles, the present bibliometric analysis does not analyse the underpinning theories, dimensions, constructs, interventions and methodologies of the articles reviewed. A detailed analysis of each area would require a new thematic review based on the area of interest. Future researchers are encouraged to separately focus on reviewing gratitude theories, dimensions or methodologies to deep dive into a unique analysis theme. Such studies may be based on a small-scale framework- or theory-based review. For instance, a small-scale review of gratitude interventions may be beneficial.

The present study provides a comprehensive encapsulation of gratitude research using a bibliometric analysis of articles published between 2001 and September 2023. It offers a wide range of information on publications and citations over the years, the most cited articles, the most prolific authors, journals and countries, and important research areas. Most importantly, trend analysis suggests four clusters of research works, that shed light on not only the current status of gratitude research but also foresight into future directions too.

AUTHOR CONTRIBUTIONS

Naval Garg: Conceptualization; data curation; formal analysis; investigation; methodology; visualization; writing – original draft; writing – review and editing.

FUNDING INFORMATION

No funding from any institution.

CONFLICT OF INTEREST STATEMENT

No conflict of interest.

DATA AVAILABILITY STATEMENT

Data was procured from Web of Science database. It could be availed on reasonable request.

1467839x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

ETHICS STATEMENT

All procedures performed in the study are carried out in accordance with the ethical standards of the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

PREREGISTRATION STATEMENT

The study was not preregistered.

RESEARCH MATERIALS AVAILABILITY STATEMENT

Data was procured from Web of Science database.

ORCID

Naval Garg https://orcid.org/0000-0001-6694-9663

REFERENCES

Adler, M. G., & Fagley, N. S. (2004). Appreciation: Individual differences in finding value and meaning as a unique predictor of subjective well-being. *Journal of Personality*, *73*(1), 79–114. https://doi.org/10.1111/j.1467-6494.2004.00305.x

Aguinis, H., Dalton, D. R., Bosco, F. A., Pierce, C. A., & Dalton, C. M. (2010). Meta-analytic choices and judgment calls: Implications for theory building and testing, obtained effect sizes, and scholarly impact. *Journal of Management*, *37*(1), 5–38. https://doi.org/10.1177/0149206310377113

Algoe, S. B., & Haidt, J. (2009). Witnessing excellence in action: The "other-praising" emotions of elevation, gratitude, and admiration. *The Journal of Positive Psychology*, *4*(2), 105–127. https://doi.org/10.1080/17439760802650519

Ali, S. A., Ahmed, M., Bhatti, O. K., & Farooq, W. (2019). Gratitude and its conceptualization: An Islamic perspective. *Journal of Religion & Health*, *59*(4), 1740–1753. https://doi.org/10.1007/s10943-019-00850-6

Andersson, L. M., Giacalone, R. A., & Jurkiewicz, C. L. (2007). On the relationship of hope and gratitude to corporate social responsibility. *Journal of Business Ethics*, *70*(4), 401–409. https://doi.org/10.1007/s10551-006-9118-1

Audrain-Pontevia, A., & Garnier, I. (2021). Are your customers grateful? How customer gratitude impacts loyalty programme effectiveness. *International Journal of Retail & Distribution Management*, *49*(12), 1660–1679. https://doi.org/10.1108/ijrdm-10-2020-0426

Bansal, S., Garg, N., & Singh, J. (2023). Weight-based teasing and depressive symptoms among Indian college students: Exploring the moderating effect of gratitude. *Kybernetes*. https://doi.org/10.1108/k-10-2022-1486

Bartlett, M. Y., & DeSteno, D. (2006). Gratitude and prosocial behavior: Helping when it costs you. *Psychological Science*, *17*(4), 319–325. https://doi.org/10.1111/j.1467-9280.2006.01705.x

Beames, J. R., Li, S. H., Newby, J. M., Maston, K., Christensen, H., & Werner-Seidler, A. (2021). The upside: Coping and psychological resilience in Australian adolescents during the COVID-19 pandemic. *Child and Adolescent Psychiatry and Mental Health*, *15*(1), 77. https://doi.org/10.1186/s13034-021-00432-z

Belkin, L. Y., & Kong, D. T. (2021). Supervisor companionate love expression and elicited subordinate gratitude as moral-emotional facilitators of voice amid COVID-19. *The Journal of Positive Psychology*, *17*, 1–15. https://doi.org/10.1080/17439760.2021.1975157

Boggiss, A., Consedine, N. S., Brenton-Peters, J., Hofman, P. L., & Serlachius, A. (2020). A systematic review of gratitude interventions: Effects on physical health and health behaviors. *Journal of Psychosomatic Research*, *135*, 110165. https://doi.org/10.1016/j.jpsychores.2020.110165

Cain, I. H., Cairo, A., Duffy, M., Meli, L., Rye, M. S., & Worthington, E. L. (2018). Measuring gratitude at work. *The Journal of Positive Psychology*, *14*(4), 440–451. https://doi.org/10.1080/17439760.2018.1484936

Carmona-Halty, M., Schaufeli, W. B., Llorens, S., & Salanova, M. (2019). Satisfaction of basic psychological needs leads to better academic performance via increased psychological capital: A three-wave longitudinal study among high school students. *Frontiers in Psychology*, *10*, 2113. https://doi.org/10.3389/fpsyg.2019.02113

Chamorro-Garrido, A., Ramírez-Fernández, E., & Ortega-Martínez, A. R. (2021). Autobiographical memory, gratitude, forgiveness and sense of humor: An intervention in older adults. *Frontiers in Psychology*, *12*, 731319. https://doi.org/10.3389/fpsyg.2021.731319

Chan, D. W. (2011). Burnout and life satisfaction: Does gratitude intervention make a difference among Chinese school teachers in Hong Kong? *Educational Psychology*, *31*(7), 809–823. https://doi.org/10.1080/01443410.2011.608525

Clarkson, R. K. (2020). Effect of gratitude on life satisfaction and perceived academic performance of psychology students. [Doctoral dissertation]. The University of Waikato, Hamilton.

Cortini, M., Converso, D., Galanti, T., Di Fiore, T., Di Domenico, A., & Fantinelli, S. (2019). Gratitude at work works! A mix-method study on different dimensions of gratitude, job satisfaction, and job performance. *Sustainability*, *11*(14), 3902. https://doi.org/10.3390/su11143902

Cregg, D. R., & Cheavens, J. S. (2020). Gratitude interventions: Effective self-help? A meta-analysis of the impact on symptoms of depression and anxiety. *Journal of Happiness Studies*, *22*(1), 413–445. https://doi.org/10.1007/s10902-020-00236-6

Cui, H., Bi, X., Chen, W., Gao, T., Qing, Z., Shi, K., & Ma, Y. (2023). Gratitude and academic engagement: Exploring the mediating effects of internal locus of control and subjective well-being. *Frontiers in Psychology*, *14*, 1287702. https://doi.org/10.3389/fpsyg.2023.1287702

Dabic, M., González-Loureiro, M., & Harvey, M. (2013). Evolving research on expatriates: What is 'known' after four decades (1970–2012). *The International Journal of Human Resource Management*, *26*(3), 316–337. https://doi.org/10.1080/09585192.2013.845238

Damasio, A., & Carvalho, G. (2013). The nature of feelings: Evolutionary and neurobiological origins. *Nature Reviews. Neuroscience*, *14*(1), 143–152. https://doi.org/10.1038/nrn3403

Datu, J. A. D., & Fincham, F. D. (2022). Gratitude, relatedness needs satisfaction, and negative psychological outcomes during the COVID-19 pandemic: A short-term longitudinal study. *Journal of Clinical Psychology*, *78*, 2525–2537. https://doi.org/10.1002/jclp.23364

Davis, D. E., Choe, E., Meyers, J., Wade, N. G., Varjas, K., Gifford, A., Quinn, A., Hook, J. N., Van Tongeren, D. R., Griffin, B. J., & Worthington, E. L. (2016). Thankful for the little things: A meta-analysis of gratitude interventions. *Journal of Counseling Psychology*, *63*(1), 20–31. https://doi.org/10.1037/cou0000107

Day, G., Robert, G., & Rafferty, A. M. (2020). Gratitude in health care: A meta-narrative review. *Qualitative Health Research*, *30*(14), 2303–2315. https://doi.org/10.1177/1049732320951145

Di Fabio, A., Palazzeschi, L., & Bucci, O. (2017). Gratitude in organizations: A contribution for healthy organizational contexts. *Frontiers in Psychology*, *8*, 2025. https://doi.org/10.3389/fpsyg.2017.02025

Dickens, L. (2017). Using gratitude to promote positive change: A series of meta-analyses investigating the effectiveness of gratitude interventions. *Basic and Applied Social Psychology*, *39*(4), 193–208. https://doi.org/10.1080/01973533.2017.1323638

Disabato, D. J., Kashdan, T. B., Short, J. L., & Jarden, A. (2016). What predicts positive life events that influence the course of

1467839x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

depression? A longitudinal examination of gratitude and meaning in life. *Cognitive Therapy and Research*, *41*(3), 444–458. https://doi.org/10.1007/s10608-016-9785-x

Dixit, S. K., & Sinha, J. (2021). Adaptation and validation of the Gratitude Questionnaire (GQ-6) for the Indian context. *Current Psychology*, *42*, 8722–8732. https://doi.org/10.1007/s12144-021-02143-2

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology*, *88*(5), 736–748. https://doi.org/10.1037/0022-3514.88.5.736

Emmons, R. A., & Kneezel, T. T. (2005). Giving thanks: Spiritual and religious correlates of gratitude. *Journal of Psychology and Christianity*, *24*(2), 140–148.

Emmons, R. A., & McCullough, M. E. (2003). Counting blessings versus burdens: An experimental investigation of gratitude and subjective well-being in daily life. *Journal of Personality and Social Psychology*, *84*(2), 377–389. https://doi.org/10.1037/0022-3514.84.2.377

Fishman, M. D. (2020). The silver linings journal: Gratitude during a pandemic. *Journal of Radiology Nursing*, *39*(3), 149–150. https://doi.org/10.1016/j.jradnu.2020.05.005

Fredrickson, B. L. (2004). Gratitude, like other positive emotions, broadens and builds. In R. A. Emmons & M. E. McCullough (Eds.), *The psychology of gratitude* (pp. 145–166). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195150100.003.0008

Fredrickson, B. L., Tugade, M. M., Waugh, C. E., & Larkin, G. R. (2003). What good are positive emotions in crisis? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001. *Journal of Personality and Social Psychology*, *84*(2), 365–376. https://doi.org/10.1037/0022-3514.84.2.365

Froh, J. J., Fan, J., Emmons, R. A., Bono, G., Huebner, E. S., & Watkins, P. C. (2011). Measuring gratitude in youth: Assessing the psychometric properties of adult gratitude scales in children and adolescents. *Psychological Assessment*, *23*(2), 311–324. https://doi.org/10.1037/a0021590

Froh, J. J., Kashdan, T. B., Ozimkowski, K. M., & Miller, N. (2009). Who benefits the most from a gratitude intervention in children and adolescents? Examining positive affect as a moderator. *The Journal of Positive Psychology*, *4*(5), 408–422. https://doi.org/10.1080/17439760902992464

Fujitani, T., Ohara, K., Kouda, K., Mase, T., Miyawaki, C., Momoi, K., Okita, Y., Furutani, M., & Nakamura, H. (2017). Association of social support with gratitude and sense of coherence in Japanese young women: a cross-sectional study. *Psychology Research and Behavior Management*, *10*, 195–200. https://doi.org/10.2147/prbm.s137374

Gabana, N. T., Steinfeldt, J. A., Wong, Y. J., Chung, Y. D., & Svetina, D. (2018). Attitude of gratitude: Exploring the implementation of a gratitude intervention with college athletes. *Journal of Applied Sport Psychology*, *31*(3), 273–284. https://doi.org/10.1080/10413200.2018.1498956

Gaeta, M. L., Gaeta, L., & Rodriguez, M. D. S. (2021). The impact of COVID-19 home confinement on Mexican university students: Emotions, coping strategies, and self-regulated learning. *Frontiers in Psychology*, *12*, 642823. https://doi.org/10.3389/fpsyg.2021.642823

Garg, N. (2017). Workplace spirituality and organizational performance in Indian context: Mediating effect of organizational commitment, work motivation and employee engagement. *South Asian Journal of Human Resources Management*, *4*(2), 191–211. https://doi.org/10.1177/2322093717736134

Garg, N. (2022). Gratitude and work-family enrichment among Indian female workforce: Exploring the mediating role of psychological capital. *International Journal of Work Organisation and Emotion*, *13*(1), 1. https://doi.org/10.1504/ijwoe.2022.10043234

Garg, N. (2023). Development and validation of Hindu Gratitude Scale (HGS-15): A RNAS perspective. *Journal of Religion & Health*, *62*(5), 3622–3639. https://doi.org/10.1007/s10943-023-01831-6

Garg, N., Jain, A., & Punia, B. (2020). Gratitude, social intelligence, and leadership among university teachers: Mediation and moderation analysis. *International Journal of Organizational Analysis*, *29*(2), 368–388. https://doi.org/10.1108/ijoa-01-2020-2016

Garg, N., Katiyar, N., & Mehak. (2021). Gratitude Questionnaire (GQ-6)—Exploring psychometric properties in India. *Journal of Religion and Health*, *60*(5), 3716–3731. https://doi.org/10.1007/s10943-021-01419-y

Garg, N., Kumari, S., & Punia, B. K. (2022). Interrelation between gratitude and vitality among students: Exploring mediating role of resilience. *Psychological Studies*, *67*(2), 200–207. https://doi.org/10.1007/s12646-022-00654-x

Garg, N., Mahipalan, M., Poulose, S., & Burgess, J. (2022). Does gratitude ensure workplace happiness among university teachers? Examining the role of social and psychological capital and spiritual climate. *Frontiers in Psychology*, *13*, 849412. https://doi.org/10.3389/fpsyg.2022.849412

Garg, N., & Mehak. (2021). Gratitude resentment and appreciation scale (GRAT-16): Analyzing psychometrics properties in the Indian context. *Current Psychology*, *42*, 8771–8780. https://doi.org/10.1007/s12144-021-02169-6

Garg, N., Verma, S., & Palframan, J. T. (2022). Positive reframing as a mediator between gratitude and technostress amongst Indian students during the COVID-19 pandemic. *Kybernetes*, *52*(8), 2810–2825. https://doi.org/10.1108/k-12-2021-1250

Garland, E. L., Fredrickson, B. L., Kring, A. M., Johnson, D., Meyer, P. S., & Penn, D. L. (2010). Upward spirals of positive emotions counter downward spirals of negativity: Insights from the broaden-and-build theory and affective neuroscience on the treatment of emotion dysfunctions and deficits in psychopathology. *Clinical Psychology Review*, *30*(7), 849–864. https://doi.org/10.1016/j.cpr.2010.03.002

Geier, M. T., & Morris, J. (2022). The impact of a gratitude intervention on mental well-being during COVID-19: A quasi-experimental study of university students. *Applied Psychology: Health and Well-Being*, *14*, 937–948. https://doi.org/10.1111/aphw.12359

Grant, A. M., & Wrzesniewski, A. (2010). I won't let you down… Or will I? Core self-evaluations, other-orientation, anticipated guilt and gratitude, and job performance. *Journal of Applied Psychology*, *95*(1), 108–121. https://doi.org/10.1037/a0017974

Hammer, J. H., & Brenner, R. E. (2017). Disentangling gratitude: A theoretical and psychometric examination of the Gratitude Resentment and Appreciation Test–Revised Short (GRAT–RS). *Journal of Personality Assessment*, *101*(1), 96–105. https://doi.org/10.1080/00223891.2017.1344986

Heider, F. (1958). *The psychology of interpersonal relations*. Wiley.

Hill, P. L., Allemand, M., & Roberts, B. W. (2013). Examining the pathways between gratitude and self-rated physical health across adulthood. *Personality and Individual Differences*, *54*(1), 92–96. https://doi.org/10.1016/j.paid.2012.08.011

Hlava, P., Elfers, J., & Offringa, R. (2014). A transcendent view of gratitude: The transpersonal gratitude scale. *International Journal of Transpersonal Studies*, *33*(1), 1–14.

Hoy, B. D., Suldo, S. M., & Mendez, L. R. (2013). Links between parents' and children's levels of gratitude, life satisfaction, and Hope. *Journal of Happiness Studies*, *14*(4), 1343–1361. https://doi.org/10.1007/s10902-012-9386-7

Huang, Y., Xu, C., Zhang, X., & Li, L. (2022). Bibliometric analysis of landslide research based on the WOS database. *Natural*

1467839x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

Hazards Research, *2*(2), 49–61. https://doi.org/10.1016/j.nhres.2022.02.001

Hussong, A. M., Langley, H. A., Rothenberg, W. A., Coffman, J. L., Halberstadt, A. G., Costanzo, P. R., & Mokrova, I. L. (2018). Raising grateful children one day at a time. *Applied Developmental Science*, *23*(4), 371–384. https://doi.org/10.1080/10888691.2018.1441713

Jackowska, M., Brown, J. M., Ronaldson, A., & Steptoe, A. (2016). The impact of a brief gratitude intervention on subjective well-being, biology and sleep. *Journal of Health Psychology*, *21*(10), 2207–2217. https://doi.org/10.1177/1359105315572455

Jans-Beken, L. (2021). A perspective on mature gratitude as a way of coping with COVID-19. *Frontiers in Psychology*, *12*, 632911. https://doi.org/10.3389/fpsyg.2021.632911

Jans-Beken, L., & Wong, P. T. (2019). Development and preliminary validation of the Existential Gratitude Scale (EGS). *Counselling Psychology Quarterly*, *34*(1), 72–86. https://doi.org/10.1080/09515070.2019.1656054

Jin, G., & Wang, Y.-Z. (2019). The influence of gratitude on learning engagement among adolescents: The multiple mediating effects of teachers' emotional support and students' basic psychological needs. *Journal of Adolescence*, *77*(1), 21–31. https://doi.org/10.1016/j.adolescence.2019.09.006

Kern, M. L., Waters, L. E., Adler, A., & White, M. A. (2014). A multidimensional approach to measuring well-being in students: Application of the PERMA framework. *The Journal of Positive Psychology*, *10*(3), 262–271. https://doi.org/10.1080/17439760.2014.936962

Kerr, S., O'Donovan, A., & Pepping, C. A. (2014). Can gratitude and kindness interventions enhance well-being in a clinical sample? *Journal of Happiness Studies*, *16*(1), 17–36. https://doi.org/10.1007/s10902-013-9492-1

Kersten, A., Van Woerkom, M., Kooij, D. T. A. M., & Bauwens, R. (2021). Paying gratitude forward at work. *Journal of Personnel Psychology*, *21*, 137–148. https://doi.org/10.1027/1866-5888/a000296

Kim, J. Y., & Jun, W. H. (2021). The effect of gratitude and self-encouragement on depression in psychiatric inpatients in Korea. *Archives of Psychiatric Nursing*, *35*(3), 323–328. https://doi.org/10.1016/j.apnu.2020.12.004

King, R. B., Zhang, R., Jian-Ping, W., Xie, F., & Cai, Y. (2023). Gratitude is its own reward: How grateful students have better motivation and engagement. *Educational Psychology*, *43*(4), 284–303. https://doi.org/10.1080/01443410.2023.2205069

Knabb, J. J., Vazquez, V. E., Wang, K. T., & Pate, R. A. (2021). The Christian Gratitude Scale: An emic approach to measuring thankfulness in every season of life. *Spirituality in Clinical Practice*, *10*, 304–315. https://doi.org/10.1037/scp0000278

Komase, Y., Watanabe, K., Sasaki, N., & Kawakami, N. (2020). Reliability and validity of the Japanese version of the Gratitude at Work Scale (GAWS). *Journal of Occupational Health*, *62*(1), e12185. https://doi.org/10.1002/1348-9585.12185

Kong, F., Yang, K., Yan, W., & Li, X. (2020). How does trait gratitude relate to subjective well-being in Chinese adolescents? The mediating role of resilience and social support. *Journal of Happiness Studies*, *22*(4), 1611–1622. https://doi.org/10.1007/s10902-020-00286-w

Krause, N., Bruce, D., Hayward, R. D., & Woolever, C. (2014). Gratitude to god, self-rated health, and depressive symptoms. *Journal for the Scientific Study of Religion*, *53*(2), 341–355. https://doi.org/10.1111/jssr.12110

Lambert, N. M., Fincham, F. D., Braithwaite, S. R., Graham, S. M., & Beach, S. R. H. (2009). Can prayer increase gratitude? *Psychology of Religion and Spirituality*, *1*(3), 139–149. https://doi.org/10.1037/a001673

Langer, Á. I., Ulloa, V., Aguilar-Parra, J. M., Araya, C. E., & Brito, G. (2016). Validation of a Spanish translation of the Gratitude Questionnaire (GQ-6) with a Chilean sample of adults and high schoolers. *Health and Quality of Life Outcomes*, *14*(1), 53. https://doi.org/10.1186/s12955-016-0450-6

Lantz, E. D., Stearns, M., McKay, I. T., & Nadorff, D. (2021). The mediating effect of gratitude on the relation between religiosity and well-being in samples of undergraduate students and adults formerly raised by grandparent caregivers. *Journal of Happiness Studies*, *23*(3), 969–984. https://doi.org/10.1007/s10902-021-00435-9

Lee, J. Y., Kim, M., Jhon, M., Kim, H., Kang, H. J., Ryu, S., Kim, J. M., & Kim, S. W. (2021). The association of gratitude with perceived stress among nurses in Korea during COVID-19 outbreak. *Archives of Psychiatric Nursing*, *35*(6), 647–652. https://doi.org/10.1016/j.apnu.2021.10.002

Leung, C., & Tong, E. M. W. (2017). Gratitude and drug misuse: Role of coping as mediator. *Substance Use & Misuse*, *52*(14), 1832–1839. https://doi.org/10.1080/10826084.2017.1312449

Li, C., Dong, Y., Wu, C., Brown, M. E., & Sun, L. (2021). Appreciation that inspires: The impact of leader trait gratitude on team innovation. *Journal of Organizational Behavior*, *43*(4), 693–708. https://doi.org/10.1002/job.2577

Liljestrand, R., & Martin, S. (2021). Stress and resilience among healthcare workers during the COVID-19 pandemic. *Rehabilitation Nursing*, *46*(6), 300–304. https://doi.org/10.1097/rnj.0000000000000344

Llenares, I., & Almeda, C. (2021). Psychometric quality of the Gratitude Questionnaire in Filipino college students. *Behavioral Psychology/Psicología Conductual*, *29*(3), 667–680. https://doi.org/10.51668/bp.8321309n

Locklear, L. R., Taylor, S. G., & Ambrose, M. L. (2021). How a gratitude intervention influences workplace mistreatment: A multiple mediation model. *Journal of Applied Psychology*, *106*(9), 1314–1331. https://doi.org/10.1037/apl0000825

Lyubomirsky, S., King, L. A., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin*, *131*(6), 803–855. https://doi.org/10.1037/0033-2909.131.6.803

McCullough, M. E., Emmons, R. A., & Tsang, J.-A. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology*, *82*(1), 112–127. https://doi.org/10.1037//0022-3514.82.1.112

McCullough, M. E., Kilpatrick, S. D., Emmons, R. A., & Larson, D. B. (2001). Is gratitude a moral affect? *Psychological Bulletin*, *127*(2), 249–266. https://doi.org/10.1037/0033-2909.127.2.249

Mesurado, B., Resett, S., Tezón, M., & Vanney, C. E. (2021). Do positive emotions make you more prosocial? Direct and indirect effects of an intervention program on prosociality in Colombian adolescents during social isolation due to COVID-19. *Frontiers in Psychology*, *12*, 710037. https://doi.org/10.3389/fpsyg.2021.710037

Morgan, B., Gulliford, L., & Kristjánsson, K. (2017). A new approach to measuring moral virtues: The multi-component gratitude measure. *Personality and Individual Differences*, *107*, 179–189. https://doi.org/10.1016/j.paid.2016.11.044

Nawa, N. E., & Yamagishi, N. (2021). Enhanced academic motivation in university students following a 2-week online gratitude journal intervention. *BMC Psychology*, *9*(1), 71. https://doi.org/10.1186/s40359-021-00559-w

Oğuz-Duran, N. (2017). The revised short gratitude, resentment, and appreciation test (S-GRAT): Adaptation for Turkish college students. *Journal of Happiness & Well-Being*, *5*(1), 23–37.

Ohlan, R., Singh, R., Kaur, S., & Ohlan, A. (2022). A bibliometric analysis of first 45 years of *Journal of Management. Serials Review*, *48*, 1–22. https://doi.org/10.1080/00987913.2022.2066964

Oliveira, R., Baldé, A., Madeira, M., Ribeiro, T., & Arriaga, P. (2021). The impact of writing about gratitude on the intention to engage in prosocial behaviors during the COVID-19 outbreak. *Frontiers in Psychology*, *12*, 588691. https://doi.org/10.3389/fpsyg.2021.588691

1467839x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

Oriol, X., Varela, J. J., & Miranda, R. (2021). Gratitude as a protective factor for cyberbullying victims: Conditional effects on school and life satisfaction. *International Journal of Environmental Research and Public Health*, *18*(5), 2666. https://doi.org/10.3390/ijerph18052666

Park, N., Peterson, C., & Seligman, M. E. (2004). Feeling and believing: The influence of emotion on trust. *Journal of Social and Clinical Psychology*, *23*(5), 603–619. https://doi.org/10.1521/jscp.23.5.603.50748

Perez, J. A., Peralta, C. O., & Besa, F. B. (2021). Gratitude and life satisfaction: The mediating role of spirituality among Filipinos. *Journal of Beliefs & Values*, *42*(4), 511–522. https://doi.org/10.1080/13617672.2021.1877031

Petrocchi, N., & Couyoumdjian, A. (2015). The impact of gratitude on depression and anxiety: The mediating role of criticizing, attacking, and reassuring the self. *Self and Identity*, *15*(2), 191–205. https://doi.org/10.1080/15298868.2015.1095794

Polak, E. L., & McCullough, M. E. (2006). Is gratitude an alternative to materialism? *Journal of Happiness Studies*, *7*(3), 343–360. https://doi.org/10.1007/s10902-005-3649-5

Renshaw, T. L., & Steeves, R. M. O. (2016). What good is gratitude in youth and schools? A systematic review and meta-analysis of correlates and intervention outcomes. *Psychology in the Schools*, *53*(3), 286–305. https://doi.org/10.1002/pits.21903

Rosenberg, E. L. (1998). Levels of analysis and the organization of affect. *Review of General Psychology*, *2*(3), 247–270. https://doi.org/10.1037/1089-2680.2.3.247

Rothenberg, W. A., Hussong, A. M., Langley, H. A., Egerton, G. A., Halberstadt, A. G., Coffman, J. L., Mokrova, I., & Costanzo, P. R. (2017). Grateful parents raising grateful children: Niche selection and the socialization of child gratitude. *Applied Developmental Science*, *21*(2), 106–120. https://doi.org/10.1080/10888691.2016.1175945

Rusdi, A., Sakinah, S., Bachry, P. N., Anindhita, N., & Hasibuan, M. A. I. (2021). The development and validation of the Islamic Gratitude Scale (IGS-10). *Psikis: Jurnal Psikologi Islami*, *7*(2), 120–142. https://doi.org/10.19109/psikis.v7i2.7872

Sarkar, A., & Garg, N. (2020). "Peaceful workplace" only a myth? *International Journal of Conflict Management*, *31*(5), 709–728. https://doi.org/10.1108/ijcma-11-2019-0217

Sawyer, K., Thoroughgood, C., Stillwell, E. E., Duffy, M. K., Scott, K. L., & Adair, E. A. (2022). Being present and thankful: A multi-study investigation of mindfulness, gratitude, and employee helping behavior. *Journal of Applied Psychology*, *107*(2), 240–262. https://doi.org/10.1037/apl0000903

Saxena, A., Garg, N., Punia, B. K., & Prasad, A. (2020). Exploring role of Indian workplace spirituality in stress management: a study of oil and gas industry. *Journal of Organizational Change Management*, *33*(5), 779–803. https://doi.org/10.1108/jocm-11-2019-0327

Seligman, M. E. P., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress: Empirical validation of interventions. *American Psychologist*, *60*(5), 410–421. https://doi.org/10.1037/0003-066x.60.5.410

Senf, K., & Liau, A. K. (2012). The effects of positive interventions on happiness and depressive symptoms, with an examination of personality as a moderator. *Journal of Happiness Studies*, *14*(2), 591–612. https://doi.org/10.1007/s10902-012-9344-4

Sharma, N., Pathak, A., Lavanya, B. L., Garg, N., & Lata, K. (2022). Exploring the psychometric properties of personal optimism and self-efficacy optimism-extended (POSO-E) scale among Indian teachers. *Benchmarking: An International Journal*, *30*, 2234–2247. https://doi.org/10.1108/bij-01-2022-0054

Simmel, G. (1950). *The sociology of Georg Simmel*. Free Press.

Singh, L. B., Kumar, A., & Srivastava, S. (2021). Academic burnout and student engagement: A moderated mediation model of internal locus of control and loneliness. *Journal of International Education in Business*, *14*(2), 219–239. https://doi.org/10.1108/JIEB-03-2020-0020

Singh, S., Dhir, S., Das, V. M., & Sharma, A. (2020). Bibliometric overview of the technological forecasting and social change journal: Analysis from 1970 to 2018. *Technological Forecasting and Social Change*, *154*, 119963. https://doi.org/10.1016/j.techfore.2020.119963

Slåtten, T., Lien, G., Evenstad, S. B. N., & Onshus, T. (2021). Supportive study climate and academic performance among university students: The role of psychological capital, positive emotions and study engagement. *International Journal of Quality and Service Sciences*, *13*(4), 585–600. https://doi.org/10.1108/ijqss-03-2020-0045

Southwell, S., & Gould, E. (2016). A randomised wait list-controlled pre–post–follow-up trial of a gratitude diary with a distressed sample. *The Journal of Positive Psychology*, *12*(6), 579–593. https://doi.org/10.1080/17439760.2016.1221127

Tachon, G., Shankland, R., Marteau-Chasserieau, F., Morgan, B., Leys, C., & Kotsou, I. (2021). Gratitude moderates the relation between daily hassles and satisfaction with life in university students. *International Journal of Environmental Research and Public Health*, *18*(24), 13005. https://doi.org/10.3390/ijerph182413005

Tangney, J. P., Stuewig, J., & Mashek, D. (2007). Moral emotions and moral behavior. *Annual Review of Psychology*, *58*(1), 345–372. https://doi.org/10.1146/annurev.psych.56.091103.070145

Tian, L., Pi, L., Huebner, E. S., & Du, M. (2016). Gratitude and adolescents' subjective well-being in school: The multiple mediating roles of basic psychological needs satisfaction at school. *Frontiers in Psychology*, *7*, 1409. https://doi.org/10.3389/fpsyg.2016.01409

Toepfer, S. M., Cichy, K. E., & Peters, P. (2011). Letters of gratitude: Further evidence for author benefits. *Journal of Happiness Studies*, *13*(1), 187–201. https://doi.org/10.1007/s10902-011-9257-7

Tran, T. A. T., Nguyen Phuoc, C. T., Dinh, H. V. T., & Nguyen, V. T. (2021). The Vietnamese Version of the Gratitude Questionnaire (GQ) and the Revised Short Gratitude, Resentment, and Appreciation Test (S-GRAT): Psychometric properties among adolescents. *Japanese Psychological Research*, *64*(3), 295–307. https://doi.org/10.1111/jpr.12321

Tudge, J., De Lucca Freitas, L. B., & O'Brien, L. (2015). The virtue of gratitude: A developmental and cultural approach. *Human Development*, *58*(4–5), 281–300. https://doi.org/10.1159/000444308

Unanue, J., Oriol, X., Oyanedel, J. C., Rubio, A., & Unanue, W. (2021). Gratitude at work prospectively predicts lower workplace materialism: A three-wave longitudinal study in Chile. *International Journal of Environmental Research and Public Health*, *18*(7), 3787. https://doi.org/10.3390/ijerph18073787

Valdez, J. P. M., Datu, J. A. D., & Chu, S. K. W. (2022). Gratitude intervention optimizes effective learning outcomes in Filipino high school students: A mixed-methods study. *Computers & Education*, *176*, 104268. https://doi.org/10.1016/j.compedu.2021.104268

Watkins, P. (2014). *Gratitude and the good life: Toward a psychology of appreciation*. Springer Publisher.

Watkins, P. C., Woodward, K., Stone, T., & Kolts, R. L. (2003). Gratitude and happiness: Development of a measure of gratitude, and relationships with subjective well-being. *Social Behavior and Personality: An International Journal*, *31*(5), 431–451. https://doi.org/10.2224/sbp.2003.31.5.431

Wood, A. M., Froh, J. J., & Geraghty, A. W. (2010). Gratitude and well-being: A review and theoretical integration. *Clinical Psychology Review*, *30*(7), 890–905. https://doi.org/10.1016/j.cpr.2010.03.005

Wood, A. M., Joseph, S., Lloyd, J., & Atkins, S. (2009). Gratitude influences sleep through the mechanism of pre-sleep cognitions. *Journal of Psychosomatic Research*, *66*(1), 43–48. https://doi.org/10.1016/j.jpsychores.2008.09.002

1467839x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

Wood, A. M., Joseph, S., & Maltby, J. (2009). Gratitude predicts psychological well-being above the Big Five facets. *Personality and Individual Differences*, *46*(4), 443–447. https://doi.org/10.1016/j.paid.2008.11.012

Wood, A. M., Maltby, J., Gillett, R., Linley, P. A., & Joseph, S. (2008). The role of gratitude in the development of social support, stress, and depression: Two longitudinal studies. *Journal of Research in Personality*, *42*(4), 854–871. https://doi.org/10.1016/j.jrp.2007.11.003

Youssef-Morgan, C. M., Van Zyl, L. E., & Ahrens, B. L. (2022). The Work Gratitude Scale: Development and evaluation of a multidimensional measure. *Frontiers in Psychology*, *12*, 795328. https://doi.org/10.3389/fpsyg.2021.795328

Yudiani, E., Khosiyah, S., & Umer, A. (2023). The effect of gratitude and academic self efficacy on academic engagement in students. *Psikis: Jurnal Psikologi Islami*, *9*(1), 154–160. https://doi.org/10.19109/psikis.v9i1.16731

Zhen, R., Liu, R., Ding, Y., Jiang, R., Jiang, S., & Hong, W. (2019). Gratitude and academic engagement among primary students: Examining a multiple mediating model. *Current Psychology*, *40*(5), 2543–2551. https://doi.org/10.1007/s12144-019-00202-3

Zupic, I., & Čater, T. (2014). Bibliometric methods in management and organization. *Organizational Research Methods*, *18*(3), 429–472. https://doi.org/10.1177/1094428114562629

How to cite this article: Garg, N. (2024). Gratitude research: Review and future agenda using bibliometric analysis of the studies published in the last 20 years. *Asian Journal of Social Psychology*, *00*, 1–18. https://doi.org/10.1111/ajsp.12621

1467839x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsp.12621 by Delhi Technological University, Wiley Online Library on [09/05/2024]. 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

Green Emission in Thermally Stable Er^{3+} Doped Lead-Free Perovskite Phosphor for Solid-State Lighting and Optical Thermometry Applications

Ishant Kumar, Avinash Kumar, Sandeep Kumar, Vikas Sangwan, and Arvind K. Gathania

***Abstract*—$CaTiO_3$:xEr^{3+} (0.1≤x≤7 mole%) phosphor materials were meticulously prepared through the solution combustion method. Comprehensive analyses employing Powder X-ray Diffraction (PXRD), Field Emission Scanning Electron Microscopy (FESEM), and Fourier Transform Infrared Spectroscopy (FTIR) were conducted to investigate the phase, morphology, and vibrational characteristics of the synthesized phosphors. A detailed luminescence study was undertaken using photoluminescence spectroscopy, revealing distinctive 4f-4f transitions associated with Er^{3+} in both excitation and emission spectra. The application of Dexter's theory provided insights into the quenching mechanism inherent in $CaTiO_3$:Er^{3+} phosphors. Photometric studies were carried out to evaluate the suitability of the synthesized material for solid-state lighting applications. Furthermore, to assess the thermal stability of the phosphor material, the activation energy was computed. The fluorescence intensity ratio of thermally coupled energy levels of Er^{3+} ion was utilized to assess the temperature-sensing capabilities of the synthesized material. This material may be suitable for solid-state lighting and optical thermometry applications.**

***Index Terms*—Perovskite, phosphor, solid-state lighting, and optical thermometry.**

I. Introduction

LUMINESCENT materials, commonly denoted as phosphor materials, constitute inorganic compounds deliberately infused with impurities such as rare earth and transition metal ions. Solid-state lasers, optical temperature sensors, field emission displays, plasma display panels, and solid-state lighting exemplify a diverse array of applications associated with phosphors [1], [2], [3], [4], [5]. The versatility and efficacy of these materials make them integral components in a spectrum of cutting-edge technologies. From precise laser systems to highly sensitive temperature monitoring devices, and from state-of-the-art display technologies to innovative lighting solutions, the multifaceted utility of these materials underscores their significance in various industries [6], [7], [8], [9]. Conventional lighting sources, such as incandescent and fluorescent lights, constitute a significant portion of global electricity consumption, exceeding one-fourth of the total energy used worldwide. In response to this considerable challenge, material scientists have dedicated substantial efforts to replace these traditional sources with energy-efficient solid-state lighting alternatives. White light-emitting diodes (WLEDs) incorporating phosphor conversion present a particularly advantageous solution for energy conservation [10], [11]. Commercially, WLEDs are typically fabricated using an InGaN blue chip and Ce^{3+} yellow phosphor [12]. However, this production method falls short of meeting the criteria for high-quality interior illumination due to the absence of a red-light component in the overall emission. This results in an elevated correlated color temperature (CCT) of WLEDs and a correspondingly diminished color rendering index. To meet the standards of superior lighting quality, WLEDs must exhibit a lower correlated color temperature for indoor lighting applications and demonstrate an enhanced color rendering index [13], [14]. Achieving these characteristics is paramount to ensuring optimal visual comfort, color fidelity, and overall lighting performance in diverse applications ranging from commercial and residential settings to industrial environments. Consequently, researchers adopted an alternative strategy to enhance the quality of WLEDs by encapsulating a Near-Ultraviolet (NUV) blue chip with a composite of red, green, and blue phosphors [15], [16]. The identification of green, red, and blue phosphors responsive to NUV light becomes paramount in this context. Thus, our focus centres on the development of a novel green phosphor to serve as a color-mixing element for phosphor-converted WLEDs utilizing NUV chips.

Temperature constitutes a fundamental physical parameter across various domains of natural science, encompassing biological systems, industrial processes, and scientific investigations. Consequently, precise temperature monitoring assumes paramount importance in numerous aspects of life. Contact thermometers, relying on the thermal expansion of liquids and metals, serve as conventional instruments for direct temperature measurement through physical contact with the target object. Nonetheless, a limitation of these sensors lies in their inability to gauge the temperature of inaccessible entities, such as

Manuscript received 16 March 2024; revised 15 April 2024; accepted 17 April 2024. Date of publication 23 April 2024; date of current version 2 May 2024. *(Corresponding author: Arvind K. Gathania.)*

Ishant Kumar is with the Department of Physics and Photonics Science, National Institute of Technology, Hamirpur 177005, India, and also with the Department of Physics, Rabindranath Tagore Government College, Sarkaghat District, Mandi 175024, India.

Avinash Kumar, Sandeep Kumar, and Arvind K. Gathania are with the Department of Physics and Photonics Science, National Institute of Technology, Hamirpur 177005, India (e-mail: akgathania@nith.ac.in).

Vikas Sangwan is with the Department of Applied Physics, Delhi Technological University, Delhi 110 042, India.

Digital Object Identifier 10.1109/JPHOT.2024.3392566

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

those found in high-voltage power plants, oil refineries, coal mines, high-pressure power plants, and corrosive environments. Consequently, achieving precise temperature measurements remains a focal point for scientific researchers. To surmount the shortcomings of traditional contact thermometers, non-contact optical thermometry has garnered considerable interest among the scientific community. This attention is attributable to the fact that temperature profoundly influences fluorescence intensity, a pivotal characteristic of luminescent materials [17]. Due to its advantageous attributes, encompassing elevated sensitivity, precise spatial resolution, rapid response times, and the capability for real-time monitoring, the Fluorescence Intensity Ratio (FIR) method has garnered substantial research attention. This method is based on the examination of emissions from rare earth ion's thermally related energy levels, with a particular emphasis on the intensity ratio that varies with temperature [18], [19], [20], [21]. This feature endows it with suitability for the accurate temperature detection of typically inaccessible objects. Specifically, certain rare earth ions namely Er^{3+}, Ho^{3+}, Tm^{3+}, and Nd^{3+}- manifest pairs of thermally coupled energy levels, rendering them particularly well-suited for application in FIR based optical thermometers [22], [23], [24], [25], [26], [27]. The design and development of a phosphor material for an optical temperature sensor necessitate careful consideration of two primary factors. Firstly, the selection of an optically active ion capable of modulating its emission in response to temperature is crucial. Secondly, the choice of a host matrix assumes significance, as it profoundly influences the temperature-sensing capabilities and emission intensity of the phosphor material. In creating phosphor materials for optical temperature sensors, it's best to choose host matrices that are stable physically and chemically, have low phonon energy, and wide band gap [28]. Numerous phosphors, like $PbLa_{2-2x}Eu_{2x}W_4O_{16}$ [29] $CsPbCl_3$:Bi^{3+}, Eu^{3+} [30], $PbZrTiO_3$:Er^{3+},Yb^{3+} [31] etc. incorporate lead, a substance known for its deleterious impact on both human health and the environment due to its toxic radiation effects [32]. Considering these considerations, a collaborative endeavour has been undertaken to investigate alternative materials with environmentally sustainable attributes, serving as replacements for lead-based counterparts. The intricate oxide $CaTiO_3$ has been chosen as the host matrix in our study due to its non-toxic nature, commendable chemical and thermal stability, and relatively low phonon energy of approximately 470 cm^{-1} [33]. The orthorhombic structure of $CaTiO_3$ classifies it within the perovskite family, characterized by the generic chemical formula ABO_3. Here, B represents a cation featuring a relatively small ionic radius, exemplified by Ti^{4+}, while A denotes a cation with a larger ionic radius, exemplified by Ca^{2+}. The crystal structure of $CaTiO_3$ incorporates CaO_8 polyhedra and TiO_6 octahedra [34]. Possessing a notable band gap of 3.5 eV, $CaTiO_3$ stands out as one of the preeminent wide band gap oxides [35], [36]. Researchers have widely adopted $CaTiO_3$ as a host matrix for the formulation and advancement of rare earth-doped nano phosphors due to their distinctive attributes. These include elevated luminous efficiency, minimal phonon energy, a broadband gap, facile synthetic conditions, environmental compatibility, and robust thermal and chemical stability [37]. There has been considerable exploration of rare earth doped $CaTiO_3$ to enhance its luminescent properties, showcasing its applicability across various industrial domains [38], [39]. A key rationale for selecting Er^{3+} ions as dopants resides in their distinct energy levels, specifically $^2H_{11/2}$ and $^4S_{3/2}$, which exhibit thermal coupling and yield green light emission. This characteristic renders Er^{3+} ions particularly advantageous for applications spanning optical thermometry and the development of nano phosphors emitting green light [7].

While up conversion phosphor-based optical thermometers have been extensively scrutinized and find application in diverse fields, the competitive advantages offered by down-converting nano phosphors warrant a more comprehensive investigation for optical thermometer design. In the current study, we delve into the bifunctional attributes of $CaTiO_3$ doped with Er^{3+} ions, specifically focusing on its capabilities in temperature sensing and solid-state lighting. To the best of our knowledge, the temperature-dependent down-conversion behaviour of $CaTiO_3$:Er^{3+} has not been systematically explored. Our research is dedicated to elucidating the down-conversion characteristics of the $CaTiO_3$:Er^{3+} phosphor in relation to temperature variations, to establish its suitability as an optimal candidate for optical thermometry applications.

This study explores the dual applications of Er^{3+} doped phosphor, specifically focusing on solid-state illumination and optical thermometry. The solution combustion method was utilized to synthesize Er^{3+} doped phosphors due to its numerous benefits, including cost-effectiveness, time efficiency, and environmentally friendly nature [40]. The resulting material demonstrates commendable color purity, good thermal stability, and high sensitivity, indicating promising prospects for applications in solid-state lighting and optical thermometry.

II. Experimental

A. Phosphor Preparation

The preparation of a series of $CaTiO_3$:x Er^{3+} ($0.1 \leq x \leq 7$ mole%) phosphors using the solution combustion method involved several steps. Analytical reagent grade $Ti(OCH_2CH_2CH_2CH_3)_4$, HNO_3 (70%), $Ca(NO_3)_2 \cdot 4H_2O$, $Er(NO_3)_3 \cdot 5H_2O$, and NH_2CONH_2 were the starting materials. The synthesis process initiated with the preparation of a titanium nitrate solution. In a 100 ml beaker, 10 ml of double-distilled water and 1-2 ml of HNO_3 were added to $Ti(OCH_2CH_2CH_2CH_3)_4$. The mixture was agitated on a magnetic stirrer until a clear solution of titanium nitrate was obtained. Subsequently, stoichiometric quantities of nitrates of calcium, erbium, and urea were incorporated into the titanium nitrate solution. The resulting mixture was continuously stirred on a magnetic stirrer at an elevated temperature of approximately 80–90 °C until it transformed into a dense slurry. The obtained slurry was then subjected to a furnace heated to 500 °C to produce a fluffy powder. To enhance the crystallinity of the phosphor powder, it underwent annealing at 800 °C for a duration of 2 h. This multi-step process aimed at the preparation of high-quality phosphors for further applications.

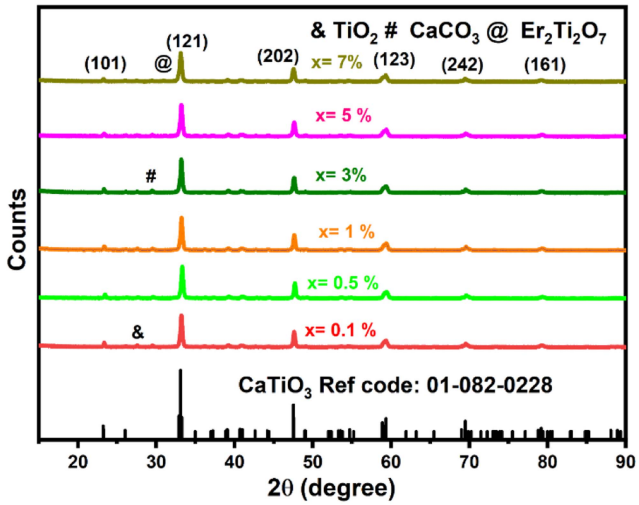


Fig. 1. PXRD patterns of $CaTiO_3$:x Er^{3+} (0.1≤x≤7 mole%).

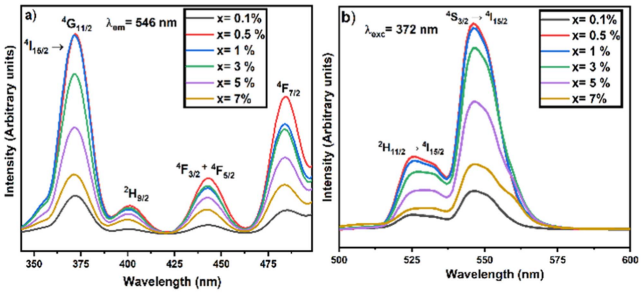


Fig. 2. (a) Excitation and (b) emission spectra of $CaTiO_3$:x Er^{3+} (0.1≤x≤7 mole%).

B. Characterization

The characterization performed in the present manuscript is same as reported in our earlier work [7], [28] and their details can be found from there.

III. Results and Discussions

A. PXRD Study

Fig. 1 presents the powder X-ray Diffraction (PXRD) pattern of $CaTiO_3$:xEr^{3+} (0.1≤x≤7 mole%) phosphor. The PXRD patterns of all the samples matches closely with the diffraction patterns of $CaTiO_3$ with ICSD (Inorganic crystal structure database) Reference Code 01-082-0228. $CaTiO_3$:xEr^{3+} (0.1≤x≤7 mole%) crystallizes into orthorhombic phase with space group Pbnm [34]. Most intense diffraction peaks of the phosphor are effectively indexed to (101), (121), (202), and (123) (242) (161). The average crystallite size was determined using Scherer's formula [41]:

$$D = \frac{K\lambda}{\beta \mathrm{Cos}\,\theta} \quad (1)$$

where 'D' represents the average crystalline size, λ is the X-ray wavelength (1.54056 Å), K is a constant, θ is the Bragg's angle, and β is the full width at half maximum (FWHM). The calculated crystallite size falls within the range of 19–22 nm. Small impurity reflections marked by ampersand (&), hash (#), and at the rate of (@) were also observed due to presence of TiO_2, $CaCO_3$ and $Er_2Ti_2O_7$ phases.

The Rietveld refinement analysis of $CaTiO_3$:xEr^{3+} (0.1≤x≤7 mole%) was performed on major phase ($CaTiO_3$) using the FULLPROF suite program. The Pseudo-Voigt function was employed to fit various parameters of the data points. Fig. S1 illustrates the Rietveld refinement of the XRD patterns of $CaTiO_3$:xEr^{3+} (0.1≤x≤7 mole%). The detailed information about the various structural parameters evaluated from the Rietveld refinement analysis of various samples is presented in Table S1.

B. FESEM Study

Fig. S2 present high-magnification Field Emission Scanning Electron Microscopy (FESEM) images, detailing the morphological characteristics of $CaTiO_3$:0.5 mole% Er^{3+}. The FESEM images depict irregularly agglomerated distorted spherical-like particles. The particle distribution, as illustrated in the histogram within Fig. S2(a), reveals a predominant particle size clustering within the range of 20-100 nm. The average particle size is 55 nm. Further insight into the elemental composition was provided by the Energy Dispersive X-ray Spectroscopy (EDS) spectra, displayed in Fig. S2(b), affirming the presence of Ca, Ti, O, and Er within the $CaTiO_3$:0.5 mole% Er^{3+} phosphors.

C. FTIR Study

The investigation into the vibrational bands of $CaTiO_3$:x Er^{3+} at varying concentrations (x = 0.1–7 mol%), was conducted through Fourier Transform Infrared (FT-IR) studies, as depicted in Fig. S3. The peaks observed at 434 cm^{-1} and broadband from 525–745 cm^{-1} correspond to the stretching vibrations of Ti–O and the bridge stretching modes of Ti–O–Ti, respectively. These spectral features reflect the structural characteristics associated with the environment surrounding the TiO_6 octahedra in the ABO_3 perovskite [42]. The shoulder peak at 875 cm^{-1} appears due to the presence of out of plane bending of CO_3^{2-} ion present in $CaCO_3$ [43], [44]. Furthermore, the band at 1442 cm^{-1} was ascribed to the asymmetric stretching vibration of the CO_3^{2-} group present in $CaCO_3$ [45], [46].

D. Room Temperature Photoluminescence Study

Photoluminescence excitation and emission spectra at room temperature were systematically recorded to analyze the impact of doping on the luminescence characteristics of $CaTiO_3$:Er^{3+}. The Photoluminescence Excitation (PLE) and Photoluminescence (PL) spectra of $CaTiO_3$:Er^{3+} are illustrated in Fig. 2(a) and (b), respectively. The PLE spectra were acquired within the range of 300–500 nm, with the emission wavelength fixed at 546 nm. In line with our previous investigations, the PLE of Er^{3+} doped $CaTiO_3$ manifested multiple narrow bands between 350 and 500 nm, corresponding to distinct 4f electronic transitions of Er^{3+} ions. Notably, sharp bands at 372, 401 nm, 442, and 484

nm denote $^4I_{15/2} \rightarrow {}^4G_{11/2}$, $^4I_{15/2} \rightarrow {}^2H_{9/2}$, $^4I_{15/2} \rightarrow {}^4F_{3/2} + {}^4F_{5/2}$, and $^4I_{15/2} \rightarrow {}^4F_{7/2}$ transitions of Er^{3+} ions, respectively [7], [47]. For the analysis of PL properties, emission spectra were recorded at 372 nm excitation for all synthesized samples. The emission spectra exhibits characteristic green emission with peaks centred at 526 nm and 546 nm, corresponding to $^2H_{11/2} \rightarrow {}^4I_{15/2}$ and $^4S_{3/2} \rightarrow {}^4I_{15/2}$ transitions [22], [48]. Notably, there was no change in peak profile with variation in doping concentration. The partial energy diagram illustrating various excitation and emission profiles are presented in Fig. S4 (a). With an increase in doping amount, there was a rise in emission intensity up to 0.5 mole percentage, attributed to the increased luminescence centres. However, beyond 0.5 mole percentage, a decline in emission intensity occurred due to concentration quenching.

When the spatial separation between activators exceeds a specific threshold, an escalation in non-radiative transitions occurs. This phenomenon is attributed to the presence of cross-relaxation channels, which act as inhibitory sites. The activation of cross-relaxation channels occurs in energy levels with comparable energy differences between them. In the context of an energy transfer process, such as A + B → C + D, cross-relaxation channels become feasible if the energy difference between A and C closely matches the energy difference between B and D. Various cross-relaxation channels may be accountable for the heightened non-radiative transitions in the synthesized phosphors, and these are detailed below [7]:

$$\text{CRC1: } {}^2H_{11/2}(\text{Er1}) + {}^4I_{15/2}(\text{Er2}) \rightarrow {}^4I_{13/2}(\text{Er1}) + {}^4I_{9/2}(\text{Er2}) \quad (2)$$

$$\text{CRC2: } {}^4S_{3/2}(\text{Er1}) + {}^4I_{15/2}(\text{Er2}) \rightarrow {}^4I_{9/2}(\text{Er1}) + {}^4I_{13/2}(\text{Er2}). \quad (3)$$

Within CRC1, the Erbium ion in the excited state $^2H_{11/2}$ undergoes relaxation to the energy state $^4I_{13/2}$, subsequently transferring the remaining energy to an adjacent Erbium ion. As depicted in Fig. S4 (b), this energy transfer through CRC1 induces stimulation of a neighbouring Erbium ion in the ground state $^4I_{15/2}$ to the excited state $^4I_{9/2}$ level. In CRC2, the Erbium ion initially in the excited state $^4S_{3/2}$ undergoes relaxation to the $^4I_{9/2}$ state and transfers the remaining energy to another Erbium ion upon being excited from the ground state $^4I_{15/2}$ to the higher state $^4I_{13/2}$. These cross-relaxation channels contribute to the increase in non-radiative transitions, and similar channels have been observed in our previous work [7], confirming their role in the concentration quenching. The influence of exchange interaction is prominent when the critical distance (R_c) between adjacent activator ions is less than 5 Å, whereas distance exceeding 5 Å lead to multipole interactions. According to Blasses, the distance between adjacent Er^{3+} ions can be determined using the equation [49]:

$$R_c = 2\left[\frac{3V}{4\pi Z X_c}\right]^{\frac{1}{3}}. \quad (4)$$

Here, R_c represents the critical distance, V denotes the unit cell volume, X_c represents the critical concentration, and Z represents the number of sites available for dopant ion occupancy per unit cell in the host lattice. For the $CaTiO_3$ host with Z equal to 4, V equal to 224.241 $Å^3$, and X_c equal to 0.005, the calculated critical distance for energy transfer is 27.78 Å. Given that this value exceeds 5 Å, it indicates the prevalence of multipole-multipole interactions in $CaTiO_3$:Er^{3+}. The determination of multipole interactions can be carried out using the Dexter formula for energy transfer [50]:

$$\frac{I}{X} = k\left[1 + l(X)^{\frac{m}{3}}\right]^{-1}. \quad (5)$$

Here, I denotes PL intensity, and k and l are arbitrary constants, while X denotes the critical concentration. The value of m signifies the type of multipolar interaction, with various constant values corresponding to different types of multipolar interactions. Specifically, for values of m = 3, 6, 8, and 10, exchange interactions, dipole-dipole, dipole-quadrupole, and quadrupole-quadrupole interactions occur, respectively. The determination of the m value involves calculating it from the slope (m/3) of the graph log(I/X) vs log(X), as illustrated in Fig. S5. The computed value of m is 5.28, closely aligning with 6. This proximity implies that dipole-dipole interactions predominate in the synthesized phosphor.

E. Photometric Analysis

The determination of CIE (Commission Internationale de L'Eclairage) coordinates for the $CaTiO_3$:x Er^{3+} (0.1≤x≤7 mole%) phosphors followed the 1931 CIE standard. The CIE diagram, corresponding to an excitation wavelength of 372 nm, is depicted in Fig. S6, with the corresponding estimated values detailed in Table S2 (The CIE coordinates for different samples are indicated numerically, corresponding to the serial numbers listed in Table S2). The CIE coordinates of the optimized phosphor, $CaTiO_3$:0.5%Er^{3+}, were found to be (0.28195, 0.70117). It has been documented that a correlated color temperature (CCT) below 5000 K is suitable for warm white light used for inside home application whereas CCT value greater than 5000 K value is suitable for outside lighting applications [51]. Furthermore, color purity, an essential parameter WLED application. The CCT and color purity values of all the samples have been evaluated using the formula described in our previous work [7] and enlisted in Table S2. The high color purity and CCT (>5000 K) value indicates that synthesized phosphor can be used as cool green emitting phosphor for outside lighting applications.

F. Decay Curve

The decay characteristics of $CaTiO_3$:0.5% Er^{3+} phosphor was analysed, as depicted in Fig. S7. Remarkably, the decay behavior of $CaTiO_3$: 0.5% Er^{3+} phosphor is effectively described by a third-order exponential decay model (Exp Dec 3). The governing equation for Exp Dec 3 is denoted as equation (6) [52]:

$$I(t) = I_0 + A_1 e^{\frac{-t}{\tau_1}} + A_2 e^{\frac{-t}{\tau_2}} + A_3 e^{\frac{-t}{\tau_3}} \quad (6)$$

Here, A_1, A_2, and A_3 are constants, while I and I_0 are luminescent intensity at time t and initial time respectively. τ_1, τ_2, and τ_3 signify decay times. For $CaTiO_3$:Er^{3+}, specific decay

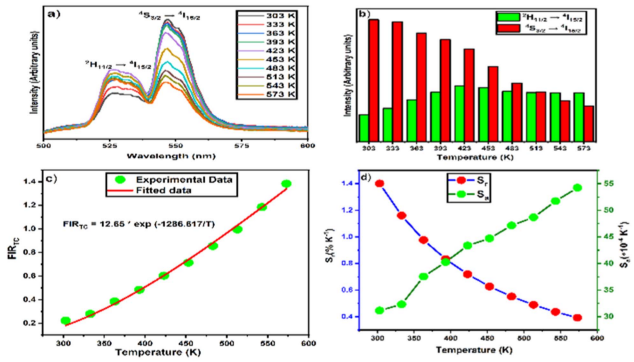


Fig. 3. Thermal evolution of (a) PL emission spectra of $CaTiO_3$: 0.5% Er^{3+} phosphor under 372 nm excitation (b) thermally coupled levels of Er^{3+} ion (c) FIR of thermally coupled levels of Er^{3+} ion and their fitting using Boltzmann's relation (d) relative and absolute sensitivity.

parameters were determined as follows: $A_1 = 0.72824$, $A_2 = 0.26383$, $A_3 = 0.33172$, $\tau_1 = 2.00036\ \mu s$, $\tau_2 = 15.39955\ \mu s$, and $\tau_2 = 76.42103\ \mu s$. The average lifetime value for the $CaTiO_3$:Er^{3+} phosphor was computed utilizing equation (7):

$$\tau_{av} = \frac{\sum A_i \tau_i^2}{\sum A_i \tau_i} \tag{7}$$

where, τ_{av}, τ_i, and A_i represent average lifetime, decay times, and constants. Through these calculations, the average lifetime was determined to be approximately 64.88 μs.

G. Thermal Stability

The influence of temperature on the PL characteristics of phosphor-based materials is a significant consideration for phosphor usage in numerous kinds of solid-state lighting applications. Usually, phosphors employed in illuminating devices necessitate excellent thermal stability as blue/UV LED chips generate temperatures of 150 °C during functioning time. Furthermore, the thermal quenching properties of the versatile phosphor are also beneficial for optical thermal sensing applications. By analyzing the emission profiles at varying temperatures, the thermal quenching features of the phosphor have been explored. The emission spectral profiles of the synthesized phosphor at 372 nm excitation as the temperature increased from 303 to 573 K, are illustrated in Fig. 3(a). The peak intensities of emission were dependent on temperature and the peak intensities associated with the transitions $^2H_{11/2} \rightarrow {}^4I_{15/2}$ and $^4S_{3/2} \rightarrow {}^4I_{15/2}$ vary with temperature. The emission peak related to the $^2H_{11/2} \rightarrow {}^4I_{15/2}$ transition at 526 nm increases with an upsurge in temperatures. Simultaneously, the emission peak at 546 nm is linked to the $^4S_{3/2} \rightarrow {}^4I_{15/2}$ transition, which diminishes with temperature. The $^2H_{11/2}$ energy level was thermally populated from the $^4S_{3/2}$ energy level via external temperature, which resulted in an intensity improvement in the $^2H_{11/2} \rightarrow {}^4I_{15/2}$ transition. This kind of variation in the PL spectral profiles occurs because of the thermally coupled and very little energy difference among the $^2H_{11/2}$ and $^4S_{3/2}$ energy levels of the dopant (Er^{3+}) ions. Furthermore, one of the most essential criteria of thermal stability for the synthesized phosphor is the activation energy, which was determined using the Arrhenius equation as shown below [53]:

$$I_T = \frac{I_0}{1 + Cexp\left(-\frac{\Delta E}{K_B T}\right)}. \tag{8}$$

The emission intensities at specific temperature T (K) and initial temperature are denoted as I_T and I_0. The Boltzmann and arbitrary constant parameters are represented as K_B and C. The activation energy of the phosphor is noted ΔE. The Arrhenius equation has been applied to assess the ΔE based on the slope of a straight line between $\ln[(I_0/I_T) - 1]$ and $1/K_BT$. The slope value was determined to be 0.279, as shown in Fig. S8. For the synthesized phosphor, ΔE is found to be 0.279 eV. Henceforth, the influence of temperature on the emission profiles of the synthesized phosphor indicates that the phosphor has excellent thermal stability.

H. Thermal Sensing

Fig. 3(a) present the thermal evolution of the emission profile of phosphor excited by 372 nm excitation within the temperature range of 303–573 K. The variation in CIE coordinates with T is negligible as depicted in Fig. S9. The emission intensity at 526 nm is observed to escalate relative to the emission at 546 nm with increasing temperature, as depicted in Fig. 3(b). This phenomenon emerges due to the thermal interconnection between two emitting levels. The relative population of these thermally coupled energy levels, separated by thermal energy (of the order of K_BT), conforms to a Boltzmann-type distribution. This distribution significantly influences the transitions between the two levels, particularly at elevated temperatures. The FIR associated with the thermally coupled levels can be described by the following expression [54]:

$$FIR_{TC} = \frac{I_{526}}{I_{546}} = A\, e^{\left(\frac{-\Delta E}{K_B T}\right)} \tag{9}$$

where I_{526} and I_{546} represent the integrated emission intensity of two thermally coupled (TC) green emitting bands, ΔE represents the energy difference between two TC levels. K_B, T, and A represent Boltzmann's constant, absolute temperature, and constant in relation to degeneracy, spontaneous radiative transition rate, and angular frequency of the respected transition. In Fig. 3(c), the experimentally determined FIR values are presented against temperature (T). The FIR data is fitted using the Boltzmann formula, and the value of $\Delta E/K_B$ is derived from the gradient of the fitted formula. Additionally, ΔE is determined to be approximately 894.25 cm^{-1}. By analysing emission spectra, the value of ΔE is found to be 696.39 cm^{-1}. The presence of non-radiative cross-relaxation channels and the overlap of two thermally coupled energy levels account for the observed subtle energy differential of 197.86 cm^{-1} [54]. In optical temperature-sensing, certain performance parameters like relative and absolute sensitivity are crucial for comparing

various optical temperatures. The absolute and relative sensitivities are defined as [55], [56]:

$$S_a = \left|\frac{d\left(FIR_{TC}\right)}{dT}\right| = \left|\frac{\Delta E}{K_B T^2} \times FIR_{TC}\right| \quad (10)$$

$$S_r = \left|\frac{1}{FIR_{TC}} \times \frac{d\left(FIR_{TC}\right)}{dT}\right| = \left|\frac{\Delta E}{K_B T^2}\right|. \quad (11)$$

The temperature-dependent behaviour of S_a and S_r is evident in Fig. 3(d), showing a steady increase with temperature. At 573 K, S_a attains its peak value, reaching 54.23 $\times$ 10^{-4}% K^{-1}. Conversely, S_r displays a consistent decreasing pattern with temperature, with its maximal value of 1.40% K^{-1} recorded at 303 K. A comparative analysis of S_a and S_r with other temperature sensing phosphors is enlisted in Table S3.

IV. Conclusion

In the current work, Er^{3+} doped $CaTiO_3$ has been successfully synthesized through solution combustion method. The obtained material underwent structural confirmation and morphology analysis via Powder X-ray Diffraction (PXRD) and Field Emission Scanning Electron Microscopy (FESEM), respectively. Additionally, Fourier Transform Infrared Spectroscopy (FTIR) was employed for vibrational characterization. Photoluminescence (PL) emission spectra were recorded at an excitation wavelength of 372 nm, revealing a distinctive green emission attributable to specific transitions. Notably, our investigation indicates a notable increase in PL emission with the initial rise in Er^{3+} concentration. However, beyond a critical concentration (x = 0.5 mol%), there is a discernible decline in emission intensity. Dexter theory elucidates that dipole-dipole interactions play a pivotal role in the observed quenching phenomenon. Calorimetric analysis of the material yields CIE coordinates, CCT, and color purity values, denoting its suitability as a green component in solid-state lighting applications. Given the thermal coupling of transitions, our study delves into potential applications, revealing the synthesized material's efficacy in optical thermometry. Maximum relative sensitivity of 1.40% K^{-1} and absolute sensitivity of 54.23 $\times$ 10^{-4}% K^{-1} were observed at temperatures of 303 K and 573 K, respectively. In summary, our findings underscore the multifunctional potential of the synthesized material. It exhibits promise for applications in both solid-state lighting and non-contact optical thermometry, establishing its viability as a bifunctional material with diverse practical implications.

References

[1] H. Scheife, G. Huber, E. Heumann, S. Bär, and E. Osiac, "Advances in up-conversion lasers based on Er^{3+} and Pr $^{3+}$," *Opt. Mater.*, vol. 26, pp. 365–374, 2004, doi: 10.1016/j.optmat.2003.10.010.

[2] T. V. Gavrilović, D. J. Jovanović, V. Lojpur, and M. D. Dramićanin, "Multifunctional Eu^{3+}- and Er^{3+}/Yb^{3+}-doped $GdVO_4$ nanoparticles synthesized by reverse micelle method," *Sci. Rep.*, vol. 4, 2014, Art. no. 4209, doi: 10.1038/srep04209.

[3] T. Wang, X. Xu, D. Zhou, J. Qiu, and X. Yu, "Red phosphor $Ca_2Ge_7O_{16}$:Eu^{3+} for potential application in field emission displays and white light-emitting diodes," *Mater. Res. Bull.*, vol. 60, pp. 876–881, 2014, doi: 10.1016/j.materresbull.2014.09.091.

[4] L. Y. Zhou, F. Z. Gong, J. X. Shi, M. L. Gong, and H. B. Liang, "A novel red phosphor $Na_2Ca_4Mg_2Si_4O_{15}$:Eu^{3+} for plasma display panels," *Mater. Res. Bull.*, vol. 43, pp. 2295–2299, 2008, doi: 10.1016/j.materresbull.2007.08.014.

[5] Y. Zhang, J. Xu, Q. Cui, and 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.*, vol. 7, 2017, Art. no. 42464, doi: 10.1038/srep42464.

[6] J. Rajagukguk, R. Situmorang, M. Djamal, and R. Rajaramakrishna, "Structural, spectroscopic and optical gain of Nd^{3+} doped fluorophosphate glasses for solid state laser application," *J. Lumin.*, vol. 216, 2019, Art. no. 116738, doi: 10.1016/j.jlumin.2019.116738.

[7] I. Kumar, A. Kumar, S. Kumar, H. Thakur, N. Dhiman, and A. K. Gathania, "n-UV triggered green emitting Er^{3+} doped Zirconia: A bifunctional material for solid-state lighting and optical thermometry," *Ceramics Int.*, vol. 49, pp. 36979–36987, 2023, doi: 10.1016/j.ceramint.2023.09.030.

[8] C. Kim et al., "Phosphors for plasma display panels," *J. Alloys Compounds*, vol. 311, pp. 33–39, 2000, doi: 10.1016/S0925-8388(00)00856-2.

[9] J. Xue, H. Mi, B. Chun, S. Heum, and J. Hwan, "Dual-functional of non-contact thermometry and field emission displays via efficient Bi^{3+} $\rightarrow$ Eu^{3+} energy transfer in emitting-color tunable $GdNbO_4$ phosphors," *Chem. Eng. J.*, vol. 382, 2020, Art. no. 122861, doi: 10.1016/j.cej.2019.122861.

[10] Z. Xia and A. Meijerink, "Ce^{3+} doped garnet phosphors : Composition modification, luminescence properties and applications," *Chem. Soc. Rev.*, vol. 46, pp. 275–299, 2016, doi: 10.1039/b000000x.

[11] V. Singh et al., "Optik luminescence features of Mn^{2+} -doped Zn_2 SiO_4: A green color emitting phosphor for solid-state lighting," *Optik (Stuttg)*, vol. 225, 2021, Art. no. 165715, doi: 10.1016/j.ijleo.2020.165715.

[12] B. Li et al., "Cyan-emitting $Ba_3Y_2B_6O_{15}$: Ce^{3+}, Tb^{3+} phosphor : A potential color converter for near-UV-excited white LEDs," *J. Lumin.*, vol. 211, pp. 388–393, 2019, doi: 10.1016/j.jlumin.2019.04.001.

[13] Z. J. Jiasong Zhong, D. Chen, W. Zhao, Y. Zhou, H. Yu, and L. Chen, "Garnet-based $Li_6CaLa_2Sb_2O_{12}$:Eu^{3+} red phosphors: A potential color-converting material for warm white light-emitting," *J. Mater. Chem. C.*, vol. 4, pp. 5559–5563, 2016, doi: 10.1039/C6TC01484D.

[14] H. Thakur, A. K. Gathania, I. Kumar, and R. K. Singh, "Tunable white light photoluminescence of a single phase Tm^{3+}/Tb^{3+}/Eu^{3+} codoped $GdPO_4$ phosphor," *Mater. Today Commun.*, vol. 36, 2023, Art. no. 106458, doi: 10.1016/j.mtcomm.2023.106458.

[15] P. Du and J. S. Yu, "Near-ultraviolet light induced visible emissions in Er^{3+}-activated La_2MoO_6 nanoparticles for solid-state lighting and non-contact thermometry," *Chem. Eng. J.*, vol. 327, pp. 109–119, 2017, doi: 10.1016/j.cej.2017.06.069.

[16] I. Kumar, A. Kumar, S. Kumar, H. Thakur, and A. K. Gathania, "Synthesis, structure, and photoluminescence analysis of thermally stable Sm^{3+} doped YPO_4 phosphor," *J. Phys. Conf. Ser.*, vol. 2663, 2023, Art. no. 012002, doi: 10.1088/1742-6596/2663/1/012002.

[17] Y. Zhao, X. Wang, Y. Zhang, Y. Li, and X. Yao, "Optical temperature sensing of up-conversion luminescent materials: Fundamentals and progress," *J. Alloys Compounds*, vol. 817, 2020, Art. no. 152691, doi: 10.1016/j.jallcom.2019.152691.

[18] X. Wang, Q. Liu, Y. Bu, C. S. Liu, and T. Yan, "Optical temperature sensing of rare-earth ion doped phosphors," *RSC Adv.*, vol. 5, pp. 86219–86236, 2015, doi: https://doi.org/10.1039/C5RA16986K.

[19] R. Monika, R. S. Yadav, A. Rai, and S. B. Rai, "NIR light guided enhanced photoluminescence and temperature sensing in Ho^{3+}/Yb^{3+}/Bi^{3+} co-doped $ZnGa_2O_4$ phosphor," *Sci. Rep.*, vol. 11, 2021, Art. no. 4148, doi: 10.1038/s41598-021-83644-9.

[20] F. Men et al., "Thermal-enhanced near-infrared upconversion luminescence of Er^{3+} for high-sensitive optical temperature sensing," *J. Lumin.*, vol. 236, 2021, Art. no. 118153, doi: 10.1016/j.jlumin.2021.118153.

[21] K. Pavani et al., "Analogy of different optical temperature sensing techniques in $LaNbO_4$: Er^{3+} /Yb $^{3+}$ phosphor," *J. Lumin.*, vol. 235, 2021, Art. no. 117992, doi: 10.1016/j.jlumin.2021.117992.

[22] I. Kumar, A. Kumar, S. Kumar, G. B. Nair, H. C. Swart, and A. K. Gathania, "Simultaneous realization of FIR-based multimode optical thermometry and photonic molecular logic gates in Er^{3+} and Yb^{3+} co-doped $SrTiO_3$ phosphor," *Physica Scripta*, vol. 98, 2023, Art. no. 105532, doi: 10.1088/1402-4896/acfa2b.

[23] H. Zhou et al., "Optical temperature sensing properties of Tm^{3+} /Yb^{3+} co-doped LuAG polycrystalline phosphor based on up-conversion luminescence," *J. Lumin.*, vol. 229, 2021, Art. no. 117656, doi: 10.1016/j.jlumin.2020.117656.

[24] D. Wawrzynczyk, A. Bednarkiewicz, M. Nyk, and M. Samoc, "Nanoscale neodymium (III) doped fluoride nanoparticles as non-contact optical," *Nanoscale*, vol. 4, pp. 6959–6961, 2012, doi: 10.1039/c2nr32203j.

[25] H. Thakur, A. K. Gathania, S. Kachhap, S. K. Singh, and R. K. Singh, "Coprecipitation synthesis, structural, optical properties, and thermometry application of Tm^{3+}/Yb^{3+} co-doped YPO_4 phosphor," *J. Lumin.*, vol. 254, 2023, Art. no. 119513, doi: 10.1016/j.jlumin.2022.119513.

[26] K. Li, D. Zhu, and H. Lian, "Up-conversion luminescence and optical temperature sensing properties in novel $KBaY(MoO_4)_3:Yb^{3+}, Er^{3+}$ materials for temperature sensors," *J. Alloys Compounds*, vol. 816, 2020, Art. no. 152554, doi: 10.1016/j.jallcom.2019.152554.

[27] K. Li, D. Zhu, and C. Yue, "Exceptional low-temperature fluorescence sensing properties in novel $KBaY(MoO_4)_3:Yb^{3+},Ho^{3+}$ materials based on FIR of Ho^{3+} transitions $^5F_{5(1)} \rightarrow {}^5I_8/{}^5S_2 \rightarrow {}^5I_8$," *J. Mater. Chem. C.*, vol. 10, pp. 6603–66010, 2022.

[28] I. Kumar and A. K. Gathania, "Photoluminescence and quenching study of the Sm^{3+} doped $LiBaPO_4$ phosphor," *J. Mater. Sci. Mater. Electron.*, vol. 33, pp. 328–341, 2021, doi: 10.1007/s10854-021-07301-7.

[29] A. S. Altowyan et al., "Synthesis and properties of novel $Pb/La/WO_4:Eu^{3+}$ red phosphors for white light emitting diodes (WLEDs)," *Optik (Stuttg)*, vol. 268, 2022, Art. no. 169805, doi: 10.1016/j.ijleo.2022.169805.

[30] W. Wang, S. Song, B. Cao, and J. Li, "Bi^{3+} and Eu^{3+} co-doped $CsPbCl_3$ perovskite quantum dots with efficient controllable blue emission via energy transfer," *J. Lumin.*, vol. 247, 2022, Art. no. 118901, doi: 10.1016/j.jlumin.2022.118901.

[31] P. P. Sukul and H. C. Swart, "Synergistic red dominancy over green up-conversion studies in $PbZrTiO_3:Er^{3+}/Yb^{3+}$ phosphor synthesized via two different modified technique and flexible thin-film thermometer demonstration on C1 : $PbZrTiO_3$ @ P," *J. Alloys Compounds*, vol. 966, 2023, Art. no. 171656, doi: 10.1016/j.jallcom.2023.171656.

[32] E. V. Ushakova, S. A. Cherevkov, V. A. Kuznetsova, and A. V. Baranov, "Lead-free perovskites for lighting and lasing applications: A minireview," *Mater. (Basel)*, vol. 12, pp. 1–27, 2019, doi: 10.3390/ma122333845.

[33] E. Tm et al., "Red, green, blue and bright white upconversion luminescence of $CaTiO_3:Er^{3+}/Tm^{3+}/Yb^{3+}$ nanocrystals," *J. Alloys Compounds*, vol. 541, pp. 505–509, 2012, doi: 10.1016/j.jallcom.2012.07.008.

[34] S. Sasidharan, G. Jyothi, and K. G. Gopchandran, "Solution combustion synthesis and luminescence dynamics of $CaTiO_3:Eu^{3+}, Y^{3+}$," *J. Lumin.*, vol. 235, 2021, Art. no. 118048, doi: 10.1016/j.jlumin.2021.118048.

[35] A. Nag Bhargavi and G. Khare, "Luminescence studies of perovskite structured titanates: A review," *Opt. Spectrosc.*, vol. 118, pp. 902–917, 2015, doi: 10.1134/S0030400X15060156.

[36] H. Mizoguchi, K. Ueda, and M. Orita, "Decomposition of water by a $CaTiO_3$ photocatalyst under UV light irradiation," *Mater. Res. Bull.*, vol. 37, pp. 2401–2406, 2002, doi: 10.1016/S0025-5408(02)00974-1.

[37] X. Xu, Y. Tang, F. Mo, L. Zhou, and B. Li, "Synthesis and luminescent properties of $CaTiO_3:Eu^{3+}, Al^{3+}$ phosphors," *Ceram. Int.*, vol. 40, pp. 10887–10892, 2014, doi: 10.1016/j.ceramint.2014.03.083.

[38] J. Mu, J. Liu, and L. Gao, "Upconversion fluorescence modulation of $CaTiO_3:Yb^{3+}/Er^{3+}$ nanocubes via Zn^{2+} introduction," *Optoelctronics Lett.*, vol. 18, pp. 0129–0134, 2022, doi: 10.1007/s11801-022-1125-7.

[39] Y. Pan et al., "Synthesis and red luminescence of Pr^{3+}-doped $CaTiO_3$ nano- phosphor from polymer precursor," *J. Solid State Chem.*, vol. 174, pp. 69–73, 2003, doi: 10.1016/S0022-4596(03)00175-0.

[40] Y. R. Parauha, V. Sahu, and S. J. Dhoble, "Prospective of combustion method for preparation of nanomaterials: A challenge," *Mater. Sci. Eng. B.*, vol. 267, 2021, Art. no. 115054, doi: 10.1016/j.mseb.2021.115054.

[41] S. Thakur and A. K. Gathania, "Investigation of optical properties of YVO_4–Er^{3+} nano-phosphors at different Er^{3+} concentrations and calcination temperatures," *J. Mater. Sci. Mater. Electron.*, vol. 27, pp. 1988–1993, 2016, doi: 10.1007/s10854-015-3982-3.

[42] Y. Wang, C. G. Niu, L. Wang, Y. Wang, X. G. Zhang, and G. M. Zeng, "Synthesis of fern-like $Ag/AgCl/CaTiO_3$ plasmonic photocatalysts and their enhanced visible-light photocatalytic properties," *RSC Adv.*, vol. 6, pp. 47873–47882, 2016, doi: 10.1039/c6ra06435c.

[43] D. Chakrabarty and S. Mahapatra, "Aragonite crystals with unconventional morphologies," *J. Mater. Chem.*, vol. 9, pp. 2953–2957, 1999, doi: 10.1039/A905407C.

[44] D. Meroni, L. Porati, F. Demartin, and D. Poelman, "Sol-gel synthesis of $CaTiO_3:Pr^{3+}$ red phosphors: Tailoring the synthetic parameters for luminescent and afterglow applications," *ACS Omega*, vol. 2, pp. 4972–4981, 2017, doi: 10.1021/acsomega.7b00761.

[45] N. P. Bhagya, P. A. Prashanth, R. Hari, B. M. Nagabhushana, and R. S. Raveendra, "Photoluminescence studies of Eu^{3+} activated $SrTiO_3$ nanophosphor prepared by solution combustion approach," *Opt.-Int. J. Light Electron Opt.*, vol. 145, pp. 678–687, 2017, doi: 10.1016/j.ijleo.2017.07.003.

[46] H. K. Yang et al., "Luminescent characteristics of $CaTiO_3:Pr^{3+}$ thin films prepared by pulsed laser deposition method with various substrates," *Appl. Surf. Sci.*, vol. 255, pp. 5062–5066, 2009, doi: 10.1016/j.apsusc.2008.12.068.

[47] E. Greenberg and R. D. E. Pape, "Energy transfer between fluoride manganese (III) and erbium (III) in various glasses," *J. Solid State Chem.*, vol. 53, pp. 236–245, 1984, doi: 10.1016/0022-4596(84)90097-5.

[48] H. R. Girisha, B. R. Radha krushna, D. R. Lavanya, P. B. Daruka, S. C. Sharma, and H. Nagabhushana, "Anti-counterfeiting, latent fingerprint detection and optical thermometry using a multi-stimulus down-converting $La_2CaZnO_5:Er^{3+}$ phosphor," *Opt. Mater.*, vol. 134, 2022, Art. no. 113053, doi: 10.1016/j.optmat.2022.113053.

[49] I. Kumar, A. Kumar, and A. K. Gathania, "Activation of zirconia nanophosphors with Eu^{3+} to demonstrate multifunctional optical applications," *Mater. Chem. Phys.*, vol. 292, 2022, Art. no. 126846, doi: 10.1016/j.matchemphys.2022.126846.

[50] I. Charak, M. Manhas, A. K. Bedyal, A. Vij, H. C. Swart, and V. Kumar, "Synthesis, luminescence and photometric investigation of $Sr_2B_2O_5:Dy^{3+}$ phosphor for UV-based white LEDs," *Appl. Phys. A.*, vol. 129, 2023, Art. no. 222, doi: 10.1007/s00339-023-06488-2.

[51] P. Sukul, K. Kumar, and H. Swart, "Erbium energy bridging upconversion mechanism studies on $BAKL:Er^{3+}/Yb^{3+}$ glass-ceramics and simultaneous enhancement of color purity of the green luminescence," *Dalton Trans.*, vol. 51, pp. 2827–2839, 2022, doi: 10.1039/D1DT03918K.

[52] Y. Cui et al., "Improving the temperature-sensing performance of the $SrZn_{0.33}Nb_{0.67}O_3:Pr^{3+}$ phosphor via Ga^{3+} doping," *Mater. Adv.*, vol. 3, pp. 3267–3277, 2022, doi: 10.1039/d1ma01247a.

[53] V. Sangwan, M. Jayasimhadri, and D. Haranath, "Colour-tunable and warm white light emitting thermally stable Dy^{3+}/Sm^{3+} co-activated tungstate-tellurite glasses for photonic applications," *J. Lumin.*, vol. 266, 2024, Art. no. 120276, doi: 10.1016/j.jlumin.2023.120276.

[54] P. Du, L. Luo, Q. Yue, and W. Li, "The simultaneous realization of high- and low-temperature thermometry in Er^{3+}/Yb^{3+}-codoped Y_2O_3 nanoparticles," *Mater. Lett.*, vol. 143, pp. 209–211, 2015, doi: 10.1016/j.matlet.2014.12.123.

[55] J. Zhang, Y. Zhang, and X. Jiang, "Investigations on upconversion luminescence of $K_3Y(PO_4)_2:Yb^{3+}$-$Er^{3+}/Ho^{3+}/Tm^{3+}$ phosphors for optical temperature sensing," *J. Alloys Compd.*, vol. 748, pp. 438–445, 2018, doi: 10.1016/j.jallcom.2018.03.127.

[56] W. Ran et al., "Application of thermally coupled energy levels in Er^{3+} doped $CdMoO_4$ phosphors: Enhanced solid-state lighting and non-contact thermometry," *Mater. Res. Bull.*, vol. 117, pp. 63–71, 2019, doi: 10.1016/j.materresbull.2019.04.035.

Check for updates

HumourHindiNet: Humour detection in Hindi web series using word embedding and convolutional neural network

AKSHI KUMAR, Goldsmiths University of London, London, United Kingdom of Great Britain and Northern Ireland
ABHISHEK MALLIK, Delhi Technological University, New Delhi, India
SANJAY KUMAR*, Delhi Technological University, New Delhi, India

Humour is a crucial aspect of human speech, and it is, therefore, imperative to create a system that can offer such detection. While data regarding humour in English speech is plentiful, the same cannot be said for a low-resource language like Hindi. Through this paper, we introduce two multimodal datasets for humour detection in the Hindi web series. The dataset was collected from over 500 minutes of conversations amongst the characters of the Hindi web series *Kota–Factory* and *Panchayat*. Each dialogue is manually annotated as Humour or Non-Humour. Along with presenting a new Hindi language-based Humour detection dataset, we propose an improved framework for detecting humour in Hindi conversations. We start by preprocessing both datasets to obtain uniformity across the dialogues and datasets. The processed dialogues are then passed through the Skip-gram model for generating Hindi word embedding. The generated Hindi word embedding is then passed onto three convolutional neural network (CNN) architectures simultaneously, each having a different filter size for feature extraction. The extracted features are then passed through stacked Long Short-Term Memory (LSTM) layers for further processing and finally classifying the dialogues as Humour or Non-Humour. We conduct intensive experiments on both proposed Hindi datasets and evaluate several standard performance metrics. The performance of our proposed framework was also compared with several baselines and contemporary algorithms for Humour detection. The results demonstrate the effectiveness of our dataset to be used as a standard dataset for Humour detection in the Hindi web series. The proposed model yields an accuracy of 91.79 and 87.32 while an F1 score of 91.64 and 87.04 in percentage for the *Kota – Factory* and *Panchayat* datasets, respectively.

CCS Concepts: • **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Paper; Generate the Correct Terms for Your Paper.

Additional Key Words and Phrases: Convolutional Neural Network (CNN), Hindi Web Series, Humour Detection, Long Short-Term Memory (LSTM), Low-Resource Languages, Social networks, Skip-gram Hindi Word Embedding

1 INTRODUCTION

In recent years, low-resource languages have been the epicenter of numerous research problems for Natural Language Processing (NLP) tasks such as the detection of fake news, hate speech, offensive language, sentiment analysis, emotion classification, and many others [9, 13, 15, 16, 27]. Moreover, the recent developments in language

*corresponding Author

Authors' addresses: Akshi Kumar, akshi.kumar@gold.ac.uk, Department of Computing, Goldsmiths University of London, London, United Kingdom of Great Britain and Northern Ireland; Abhishek Mallik, abhishekmallik265@gmail.com, Computer Science & Engineering, Delhi Technological University, New Delhi, Delhi, India; Sanjay Kumar, sanjay.kumar@dtu.ac.in, Computer Science & Engineering, Delhi Technological University, New Delhi, Delhi, India.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM 2375-4699/2024/4-ART
https://doi.org/10.1145/3661306

Just Accepted

analysis have helped detection and mitigation of humour in conversations [8, 29]. Humour can be characterized as the attribute of being entertaining or comical, particularly as manifested in writing or speech. We frequently integrate humor into our discussions to elevate the conversation's quality or convey points in a more lighthearted manner. Since humor is typically conveyed through speech, it becomes imperative to devise programs capable of identifying humor in spoken language. Spoken humour is dependent on a multitude of factors, including but not limited to context, timing, and delivery [26]. Humour helps in effectively establishing communication channels and social relationships [7]. Humour detection refers to the task of detecting humour in various forms of conversation. The task of recognizing humour in conversations or day-to-day exchanges is called humour detection. Over the years, humour detection has gained great popularity in industry and academia. However, detecting humour is not a simple task due to several reasons the idiosyncrasies of humour are opposed to the dynamics of regular language. Moreover, humour is often expressed through subtle facial expressions body movements, or hand gestures that aren't easily captured. The use of sarcasm in between lines also adds to the difficulty of detecting humour in conversations. humour detection also varies as per age, gender, or ethnic background [32]. Humour detection frameworks can be used in sentiment analysis and inference [12]. With thorough analysis, they're able to interpret the conversations as humorous or not [6].

Humour detection in low-resource languages like Hindi, Urdu, Bengali, Telugu, and others is challenging due to limited data, cultural specificity, and the lack of pre-trained models. The scarcity of diverse and annotated data, coupled with the difficulty in capturing cultural nuances, poses obstacles to developing accurate and robust humour detection models for languages with fewer resources. Even though extensive work has been done in the field of humour detection. However, its application to the Indian subcontinent has still been very restricted. Amongst the various languages being spoken in India, Hindi is the most popular one, with about half a billion Hindi speakers in India alone. While there is sufficient data in the field of Humour detection for English, support for Indian languages is still required. The conventional methods used for the English language can't be used for Hindi as they differ significantly from each other across the dimensions. The methodology needed for the Hindi language is very different from the English language, from preprocessing to feature extraction to classification owing to the grammatical, syntactical, and semantical features of both languages.

The lack of availability of humour detection datasets in the Hindi language has motivated us to create a multi-modal dataset for humour detection in the Hindi web series. The datasets were collected from two popular Hindi web series. The entire scene was split into various dialogues, and then each dialogue was manually annotated as Humour or Non-Humour. The entire dataset spanned through the 1st season of both the web series and over 500 minutes of run-time. The exhaustive Hindi dataset proposed by us can be used as a benchmark dataset for Humour detection in the Hindi language or for other text classification or NLP based tasks related to the Hindi language.

In our pursuit of humour detection in the Hindi web series, we confront not only the scarcity of dedicated datasets but also the paucity of tailored word embedding techniques for the Hindi language. This research endeavors to bridge these critical gaps by proposing an innovative architecture that seamlessly integrates word embedding and convolutional neural networks to decode humour in the context of Hindi entertainment. The proposed work starts with data preprocessing, where we meticulously cleanse the raw dialogue excerpts harvested from web series transcripts. This preparatory phase meticulously rids the data of superfluous noise and extraneous special characters, establishing a foundation of data uniformity. Next, we introduce a pioneering Hindi-centric word embedding framework, cultivated through the training of a Skip-gram model on a corpus of Hindi text. This novel word embedding schema discerningly captures the nuances of Hindi language semantics and the intricate contextual intricacies embedded within the dialogues. Subsequently, we employ a trio of distinct convolutional neural network architectures, each wielding a unique filter size, to meticulously sift through the word embeddings. These parallel networks collectively unearth a rich tapestry of features, each tailored to a specific facet of the data, be it structural, semantic, or context-driven. These extracted features then embark on a transformative

Just Accepted

journey through stacked LSTM layers, where they undergo deep processing to uncover latent interdependencies among the words, a crucial step in deciphering humour's intricate timing and subtleties. These LSTM layers, serving as the final arbiter, adeptly categorize the dialogues into their respective humour or non-humour domains, encapsulating the essence of our pioneering HumourHindiNet architecture. We run intensive simulations on both datasets and evaluate several performance metrics. We compare the performance of our proposed framework with various baseline algorithms and several recently proposed Humour detection techniques. The experimental analysis reveals the utility of our proposed dataset to be used as a benchmark dataset for Humour detection in the Hindi web series. The experimental results also reveal the superiority of our proposed framework. The major contributions of this work can be summarised below.

(i) Two datasets for humour detection in the Hindi language are introduced. The dataset was collected from two popular Hindi web series, and each dialogue was manually annotated.
(ii) We perform meticulous data preprocessing techniques, involving noise reduction and special character removal, laying a strong foundation for robust analysis. These preprocessing methods are valuable not only for humour detection but also for enhancing the overall quality of other related tasks involving Hindi text.
(iii) We introduce a novel Hindi word embedding model, a custom Skip-gram model trained on a Hindi text corpus. This equips the model to capture the subtle nuances and contextual intricacies specific to the Hindi language.
(iv) The concurrent use of multiple CNNs, each with distinct filter sizes, exemplifies an innovative approach to feature extraction. This technique allows the model to extract a diverse range of features from the input data, enhancing its ability to discern humour through structural, semantic, and contextual cues.
(v) We leverage stacked LSTM layers for sequential dependency analysis, which enables our model to delve deeper into the linguistic intricacies of humour, capturing nuances such as timing and word interdependencies.

The rest of the paper is organized as follows. Section 2 discusses some recent work done in the field of Humour detection. Section 3 describes the various details of the Hindi Humour detection datasets created by us. Section 4 illustrates the various phases of our proposed framework. The experimental results obtained by us are discussed in Section 6. Section 7 presents the concluding remarks and the scope for future improvement for this work.

2 RELATED WORK

In this section, we discuss some of the recent work done in the field of Humour detection. Modeling humour is fairly difficult as compared to other text classification tasks due to several factors like idiosyncrasy and contextual dependencies [10]. Idiosyncrasy refers to the use of twisted words, hand or body gestures, or misalignment of words to get across the humour. On the other hand, contextual dependencies refer to the fact that most of the time, a punchline is based on the conversations leading up to that point. Some of the English datasets that have been proposed in the recent study are: Big Bang Theory [1], Ted Laughter [4], PTT Jokes [5], Pun of the Day [30], and 16000 One-Liners [21]. Recently, the field of humour detection has drawn a lot of attention from both the industry and the research community. Consequently, a lot of work has been done to improve the performance of the humour detection frameworks. The existing works can be broadly classified into three main categories, namely, Language analysis, Feature engineering, and machine learning-based approaches.

Language analysis-based approaches refer to the analysis of the multi-modalities of the text, audio, and video to categorize the text as humourous or non-humourous [31]. Wang et al. [28] considered not only the literal meanings of the words but also the context in which they were spoken. They proposed a framework to model the visual and acoustic patterns that appear in the spoken word segments. Moreover, they also capture the dynamic nature of humour by modifying word representations based on shifting non-verbal behaviors. Pham et al. [24] propose a method to learn joint representations for source and target modalities by using the source modality as

Just Accepted

input and injecting the target modality at runtime. This reduces disturbance when making transitions and makes the model robust. A conversational memory network to detect emotions in video conversations was proposed by Hazarika et al. [11]. They leverage historical contextual information using a deep neural network architecture. They present a multimodal approach by combining textual, audio, and video features with the vanilla-gated recurrent neural network. It helped them in modeling the past utterances of each speaker into memories. The created memories are then merged using an attention framework to capture the inter-speaker dependencies. Liang et al. [18], proposed a Recurrent Multistage Fusion Network (RMFN) model. It solves the problem of fusing the various modalities by decomposing them into various stages, with each of them focused on a subset of multimodal signals. This helps in the particular and efficient fusion of the modalities. The multi-stage fusion approach builds upon the intermediate representations of the previous stages to model the cross-modal interactions. They finally used a recurrent neural network along with the fusion approach to integrate the temporal and cross-modal interactions. Chauhan et al. [3] proposed a Multimodal Multiparty Hindi Dataset For humour recognition in conversations. It has 6,191 utterances from 13 episodes of a top-rated TV series "Shrimaan Shrimati Phir Se". The various utterances of the episodes are labeled as humour and non-humour, containing acoustic, visual, and textual modalities. They also propose several multimodal baseline algorithms to show contextual and multimodal information for humour recognition in conversations.

Feature engineering-based approaches refer to explicitly modeling features based on the task and data at hand. It usually requires domain expertise and proficiency. Various works have been done to explore the feature engineering-based approach by developing features along the lines of effective dimensions [20], distribution of part of speech [14], stylistic [23], etc. Mihalcea and Strapparava [21] defined stylistic features by working on adult slang, antonyms, and alliteration. Yang et al. [30], engineered four structural features for humour and created characteristic sets for each of the structural features. Based on experiments they concluded that incongruity and ambiguity are the best performers as compared to the other latent semantic structural features. Even though feature engineering-based approaches give good results they have a major limitation in terms of generalization. This is due to the fact that feature engineering has to be done individually for each dataset instead of the task in general.

The popularity of deep learning methods for NLP based tasks has paved the way for exploring the applications of deep learning methods in the field of humour detection. Oliveira and Rodrigo [6] presented a study for detecting humour in Yelp reviews. They used bag-of-words and word vectors as the features for the dialogues. They started the classification study with shallow methods like Random forest and linear discriminant. Then they moved on to more complex methods like deep feed-forward neural networks, recurrent neural networks, and convolutional neural networks. Bertero and Fung [2] presented the first-ever application of using long short-term memory for humour detection in dialogues of popular sitcoms. They used the canned laughter in the audio files as annotations for categorizing the dialogues as humourous or non-humourous. They use long short-term memory for modeling the punchline relation of the conversational humour while the dialogue encodings were obtained from convolutional neural networks. Bertero and Fung [1] proposed a deep learning-based approach by combining the word-level and audio-level features using a linear-chain Conditional Random Field over the convolutional neural network and recurrent neural network. They also generate a new humour detection dataset using a very popular sitcom "*The Big Bang Theory*". Chen and Soo [5], implemented a CNN based approach with extensive filter sizes, numbers, and highway networks to enhance the depth of the deep learning architecture. They also collected and constructed four datasets for conducting experiments on humour detection, with distinct joke types in both English and Chinese. Kumar et al. [17], proposed DeepHumor, an automatic humour detection model. It is an amalgamation of CNN, LSTM, and highway networks. They used CNN layers for feature extraction, and the LSTM layers were used for sequence learning. The performance of the model is enhanced using the highway network. To overcome the overfitting problem, they also add dropout layers to their architectures. They

Just Accepted

compare the performance of their proposed model with several recently developed techniques over the Yelp user review dataset.

Table 1. Statistical details about the Hindi dataset for Humour detection created by us

Description	$Kota - Factory$	$Panchayat$
Number of Episodes	5	8
Total number of instances	2986	4145
Total number of "$Humour$" instances	1344	1713
Total number of "$Non - Humour$" instances	1642	2432

3 CREATED DATASET

In this section, we illustrate the Hindi dataset for humour detection created by us. For this purpose, we have gathered data from two popular Hindi web series, namely, $Kota - Factory$ and $Panchayat$. The data is collected from the first season of both web series. $Kota - Factory$ has 5 episodes while $Panchayat$ has 8 episodes. We have extracted utterances from different scenes from each episode and manually annotated them as "$Humour$" or "$Non - Humour$"—the various differentiators and context help in classifying the dialogues or utterances into their respective categories. The statistical details of both datasets are presented in Table 1. It gives us a clearer depiction of the empirical view of the data. Table 2 and Table 3 show the sample dialogues along with their respective categories for $Kota - Factory$ and $Panchayat$ datasets, respectively. For instance, "He founded this Jurassic park" dialogue in the "$Kota - Factory$" example refers to the coaching institute in the context of being referred to as a Jurassic park, hence it is a humour instance. While the "I'll let you know when we get there. We will not kidnap you" example from the "$Panchayat$" dataset, refers to the bus conductor saying that he'll drop the passenger at his destination and won't kidnap him. Hence, this can be considered as humour too. The exhaustive Hindi dataset proposed by us can be used as a benchmark dataset for Humour detection in the Hindi language or for other text classification or NLP-based tasks related to the Hindi language. The dataset can be accessed from here when the paper gets published HumourHindiNet dataset. The proposed datasets might be subject to some biases such as selection bias, as the choice of specific web series may limit the diversity of humour styles represented; subjectivity bias, as manual annotation of utterances, may lead to varied interpretations of comedic elements; and sampling bias, as extracting utterances from specific scenes may overlook nuanced comedic moments elsewhere in the series. By acknowledging and addressing these biases, researchers can enhance the validity and generalizability of their findings in humour detection research.

4 PROPOSED WORK

In this section, we illustrate our approach to the detection of humour in the Hindi web series by leveraging the power of word embedding and convolutional neural networks. The proposed model utilizes word embedding techniques that transform individual Hindi words into dense vectors, preserving semantic relationships and contextual nuances crucial for understanding humour. These embeddings serve as the initial input to a CNN, specifically tailored to capture complex linguistic structures inherent in comedic content. The CNN acts as a feature extractor, identifying subtle cues such as wordplay, sarcasm, and comedic timing by analyzing the input text at different levels of abstraction. Importantly, the incorporation of stacked LSTM layers further enhances the model's capability by capturing sequential dependencies and temporal context within the text. The stacked LSTM layers further enable the model to capture sequential dependencies, vital in discerning comedic timing and punchlines. This allows HumourHindiNet to recognize the sequential flow of jokes, discern comedic timing, and accurately identify punchlines, contributing to a more nuanced understanding of comedic content. By amalgamating these

Just Accepted

Table 2. Sample dialogues along with their categories for *Kota – Factory* Dataset.

Dialogue in Hindi	Dialogue in English	Class
इस जुरासिक पार्क की स्थापना उन्होंने की थी।	He founded this Jurassic Park.	Humour
महिला होगी तेरी माँ, लड़की कहते हैं	woman will be your mother, people say girl	Humour
फ्रिज सार्वजनिक है, मैं नहीं।	The fridge is public, I'm not.	Humour
क्या तुम पीयूष से बात कर सकते हो?	Can you talk to Piyush?	Non-Humour
हमें तब ही एडमिशन लेना चाहिए था।	We should have taken admission only then.	Non-Humour
और आईआईटी में सीट कितनी हैं ?	And how many seats are there in IIT?	Non-Humour

Table 3. Sample dialogues along with their categories for *Panchayat* Dataset.

Dialogue in Hindi	Dialogue in English	Class
जब हम वहां पहुचेंगे तो आपको बता दूंगा। हम तुम्हारा अपहरण नहीं करेंगे।	I'll let you know when we get there. We will not kidnap you.	Humour
और आप मिट्टी के पुत्र होने का मौका छोड़ना चाहते हैं	And you wanna miss the chance to be the son of the soil	Humour
अगर आपको मेरा काम इतना पसंद है तो आप भी अप्लाई क्यों नहीं करते।	If you like my work so much why don't you apply too.	Humour
झे फुलेरा नामक गांव में तैनात किया गया है।	I have been posted in a village called Phulera.	Non-Humour
मुझे कॉलिज में बहुत मेहनत करनी चाहिए थी।	I should have worked very hard in college.	Non-Humour
आप केवल नकारात्मक को उजागर कर रहे हैं। सकारात्मक पर भी ध्यान दें।	You are only exposing the negative. Focus on the positive too.	Non-Humour

techniques, HumourHindiNet offers a sophisticated and effective solution to the challenging task of humour detection in Hindi web series. Furthermore, its contributions extend beyond natural language processing, as it facilitates deeper insights into entertainment analytics, enabling better analysis and understanding of comedic elements within digital content produced in the Hindi language.

The various stages of the proposed framework are described below.

4.1 Data Preprocessing

In the data preprocessing phase, we diligently address the challenges presented by the raw transcripts extracted from the Hindi web series. These transcripts often contain a myriad of special characters, URLs, and other extraneous elements that introduce non-pertinent noise into the dialogues, potentially obstructing accurate humour detection. To mitigate this, we employ a range of text-cleaning techniques, a standard practice in Natural Language Processing, to systematically rid the data of such distractions. Given the linguistic complexities of the Hindi language, we harness the capabilities of the iNLTK library, a publicly available resource equipped with fundamental functions tailored for Natural Language Processing in Indian languages. This library aids in refining our approach to processing the obtained Hindi dialogues, ensuring linguistic subtleties are preserved. Furthermore, recognizing the pivotal role of class labels in our model's performance, we encode the dialogue labels

Just Accepted

as 2-dimensional binary arrays. In this encoding scheme, dialogues categorized as humour are represented as [1, 0], while those classified as non-humour are denoted as [0, 1]. This encoding enhances our model's processing capabilities, facilitating more efficient and precise classification of humour and non-humour dialogues in the subsequent stages of our analysis.

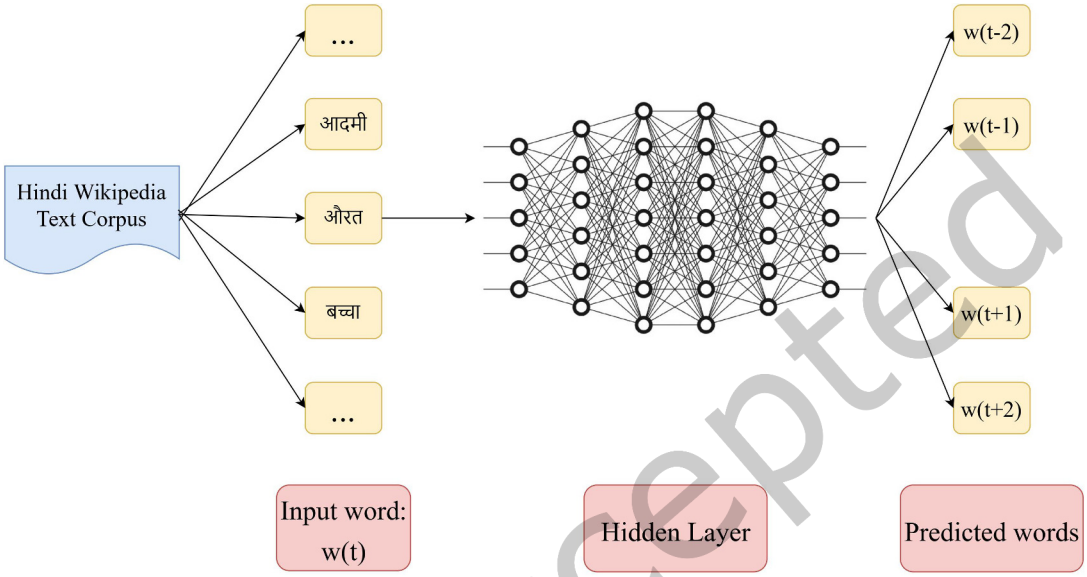


Fig. 1. Process of generating Hindi Word Embeddings using Skip-gram architecture.

Just Accepted

4.2 Skip-gram model based Hindi Word Embeddings generation

In this section, we address the need to transform the processed Hindi dialogues, obtained in the previous stage, into numerical vectors suitable for machine learning and deep learning models. Since these dialogues remain in their original Hindi form, they can't be directly utilized with such models. To bridge this gap, we employ a method commonly employed in NLP: word embeddings. While Word2Vec and Global Vectors (GloVe) are popular choices for generating word embeddings, they are primarily tailored for the English language, rendering them less effective for Hindi text.

To overcome this limitation, we embark on the creation of our Hindi-specific word embeddings. To achieve this, we employ a Skip-gram model, a technique pioneered by Mikolov et al. [22]. The Skip-gram model stands out for its ability to predict context based on a target word, a departure from the Continuous Bag of Words (CBOW) model, which predicts the target word based on context [19]. This predictive capability aligns seamlessly with our goal of humour detection, as it aids our framework in better comprehending the contextual nuances of humour or non-humour in dialogues. Our Skip-gram model is meticulously trained on the Hindi Wikipedia Corpus, a publicly accessible resource here. During training, we iteratively update the feature vectors to ensure that target words and their corresponding contexts are positioned closer to each other in the feature space. As a result, we generate feature vectors of dimension 300 for each word within the dialogues. This dimensionality strikes a balance between capturing a rich information set from the dialogue corpus while maintaining computational scalability, ultimately bolstering the effectiveness of our humour detection framework. The process of generating the Hindi word embedding using the Skip-gram architecture is shown in Figure 1

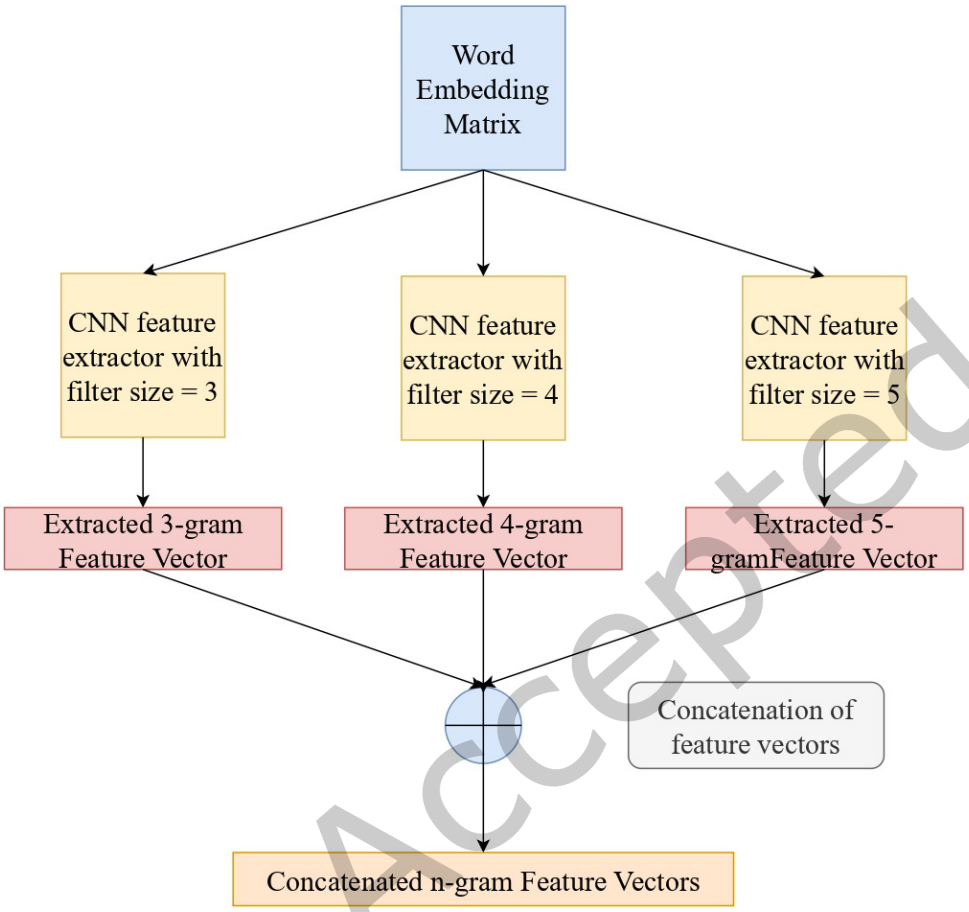


Fig. 2. Process of extracting n-gram features using CNN architecture.

Just Accepted

4.3 N-gram feature extraction using CNN

In this section of our model, we delve into the process of extracting vital features from the feature vectors acquired in the preceding stages. These feature vectors are transformed into a 2-dimensional matrix, a pivotal step in preparing them for efficient processing by a CNN. This matrix representation mimics the format of an image, aligning well with the CNN architecture, which inherently operates on 2-dimensional data.

Within our framework, we concurrently feed these 2-dimensional dialogue matrices into three distinct CNN architectures, each employing filter sizes of 3, 4, and 5. This simultaneous processing enables the extraction of n-gram features, encompassing 3-grams, 4-grams, and 5-grams, dispersed throughout the dialogues. This approach synergizes with the context-capturing prowess of the Skip-gram-based word embeddings, as the CNN-based n-gram feature extraction probes various word prefixes and suffixes to unearth intricate contextual nuances. To further harness the information gleaned from these n-gram features, we consolidate them through a concatenation layer. This amalgamation enhances feature utilization and preserves the sentence's semantic structure, judiciously situating context and target words to facilitate more precise humour detection. The process of extracting the n-gram features using parallel CNN architectures is shown in Figure 2

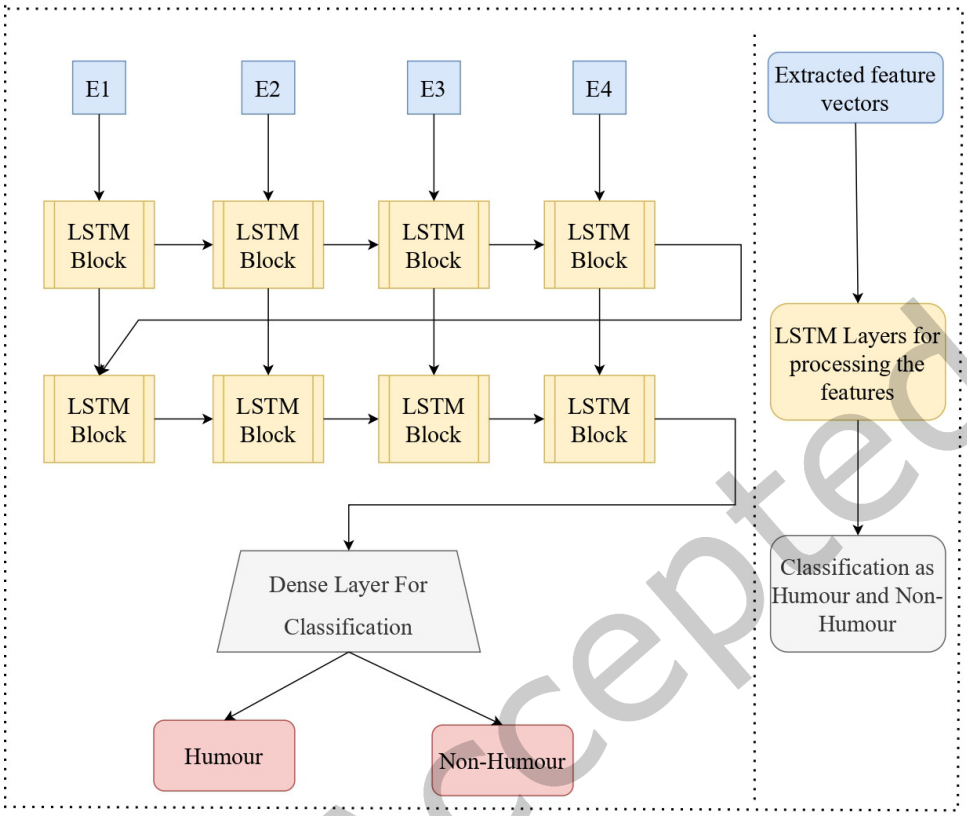


Fig. 3. Process of classification of dialogues using stacked LSTM layers

Just Accepted

4.4 Humour Classification using Stacked LSTM Layers

In this section of our model, we take the extracted features from the previous stages and subject them to the scrutiny of stacked LSTM layers. This critical phase serves as the decision-making core of our humour detection framework, where dialogues are definitively classified as either Humour or Non-Humour. LSTMs were chosen due to their well-established effectiveness in handling data holistically and their exceptional aptitude for processing sequential data while retaining contextual nuances. The recurrent nature of LSTMs, coupled with their memory cells, equips them to capture long-term dependencies without falling prey to the vanishing gradient problem. Consequently, the stacked LSTM layers represent an optimal choice for processing the extracted features. They adeptly explore these features, ferreting out hidden long-term contextual dependencies that are often pivotal in humour comprehension The processed features are then passed into a deep neural network to ultimately assign the appropriate labels of Humour or Non-Humour to the dialogues, constituting the final classification step in our approach. The process of classifying the dialogues as Humour or Non-Humour is shown in Figure 3.

5 EVALUATION METRICS

In this section, we explain the various evaluation metrics used for estimating the performance measure of our proposed algorithm are described as follows.

(i) Accuracy: Accuracy is an indicator of the share of correctly made classifications to the total number of classifications. Mathematically, it can be expressed as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

(ii) Precision: Precision is an indicator of the share of positive classifications made that were actually true. Mathematically it can be expressed as follows.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

(iii) Recall: It is a performance metric, which attempts to implicate what share of actual positives were identified correctly. For a binary classification task, recall can be expressed mathematically as follows.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

(iv) F1 score: F1 score indicates the balance between precision and recall. Formally speaking, it is the harmonic mean of both precision and recall. Mathematically it is expressed as follows.

$$F1\ score = \frac{2 * (Precision) * (Recall)}{Precision + Recall} \tag{4}$$

Here, True positive (TP) - Correctly identified, True Negative (TN) - Correctly rejected, False positive (FP) - Incorrectly identified, and False negative (FN) - Incorrectly rejected.

6 EXPERIMENTAL RESULTS

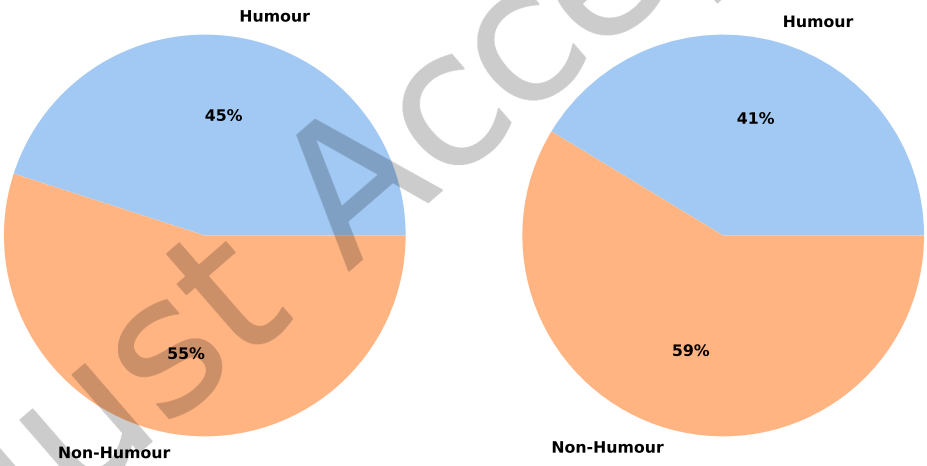
In this section, we discuss the experimental analysis of the proposed work. All the experiments are conducted on both the datasets proposed in Section 3, and several standard performance metrics like Accuracy, Precision, Recall, and F1 Score are evaluated for the same as mentioned in Section 5. For conducting the experiments, we split the entire dataset into 80:20 ratios by keeping 80% of the dataset for training and 20% of the dataset for testing the performance. We evaluate the performance of several baseline machine learning algorithms on both the proposed datasets and compare it with our proposed framework. Moreover, we also compare the performance of our proposed framework with several recent humour detection frameworks.

We also evaluate the AUC values of our algorithm to understand the bias of our model. Time and space complexity analysis are also presented for our algorithm. Figure 4, shows the percentage of humour and non-humour dialogues or utterances in both the proposed datasets. From Figure 4, we see that 45% of the dialogues of the **Kota-Factory** dataset are humour while 41% of the dialogues of the *Panchayat* dataset are humour. This shows that both datasets are well-balanced in terms of the number of samples in both classes. Table 4 lists the various hyperparameters for our framework. We employ two LSTM layers and two Dense layers. We use three CNN layers having filter sizes of 3, 4, and 5. The dropouts and recurrent dropouts values used for the two LSTM layers are 0.25 and 0.20 respectively. The optimizer used to compile the model is Adam. The activation function used in the first dense layer is ReLu while that used for the second layer is Sigmoid. The loss function used to monitor the loss is Binary-crossentropy. We used 50 epochs as around 50 epochs the training loss stabilised. The batch size used is 64. We also used callbacks to train our model. The callback method used is ReduceLROnPlateau, which reduces the learning rate when the monitored metric, which in this case is validation accuracy, has stopped improving. We also use 5-fold cross-validation and L1 regularization to improve the model performance. The obtained experimental results are as follows.

Just Accepted

Table 4. Hyperparameters for SENet

Hyperparameter	Description or Value
Number of LSTM layers	2
Number of CNN layers	3
Filter Sizes	3, 4, 5
Number of Dense layers	2
Dropout rate	0.25
Recurrent dropout rate	0.20
Optimizer	Adam
Activation function	ReLu and Sigmoid
Loss function	Binary-crossentropy
Regularization	L1 Regularization
Resampling method	Cross-Validation
Type of Cross-Validation	5-fold
Number of epochs	50
Batch size	64
Callbacks	ReduceLROnPlateau



(a) *Kota – Factory* (b) *Panchayat*

Fig. 4. Representation of the percentage of humour and non-humour samples in both the proposed datasets

Just Accepted

6.1 Comparison with Baseline algorithms

In this section, we compare the performance of our proposed HumourHindiNet framework with several baseline machine learning and deep learning-based approaches for humour detection on both of our proposed datasets. We also use a good number of supervised and unsupervised algorithms, along with several language-based models.

Table 5. Performance comparison of our proposed HumourHindiNet framework with several baseline machine learning and deep learning algorithms on *Kota – Factory* dataset.

Methods	Accuracy	Precision	Recall	F1 Score
K-Means Clustering	67.61	64.12	76.68	69.84
KNN	75.91	78.32	72.38	75.23
Logistic Regression	78.92	75.92	85.38	80.37
Gaussian Naive Bayes	77.28	75.8	79.03	77.38
Decision Trees	80.47	78.58	83.49	80.96
Random Forest	85.22	83.3	86.77	85.0
ANN	88.59	87.34	**90.07**	88.69
CNN	73.08	73.9	71.06	72.45
RNN	83.94	87.4	79.86	83.46
GloVe	89.69	95.37	83.27	88.91
BERT	89.87	92.86	86.19	89.4
HumourHindiNet	**91.79**	**96.29**	87.41	**91.64**

The various machine learning algorithms used are K-Means Clustering, K Nearest Neighbors, Logistic Regression, Gaussian Naive Bayes, Decision Trees, and Random Forests. We use Artificial Neural Networks (ANNs), CNNs, and Recurrent Neural Networks (RNNs) as the deep learning methods. The GloVe embedding method and the Bidirectional Encoder Representations from Transformers (BERT) as the language-based models. The obtained results are presented in Table 5 for the *Kota – Factory* dataset and Table 6 for the *Panchayat* dataset. For the *Kota – Factory* dataset, we see that the unsupervised clustering algorithms are the weakest performers. However, the performance improves as we go to the supervised algorithms as can be seen by the good results achieved by Random Forest, Decision Trees, etc. Deep learning-based models like ANN and RNN are able to appropriately model the non-linear multi-dimensional relationships of the dialogues to generate even better results. Amongst all the baseline models used for the *Kota – Factory* dataset, the language-based models are the best performers. This can be attributed to the optimal language dependency modeling done by the language-based models. The best performer for the *Kota – Factory* dataset is our proposed HumourHindiNet model. For the *Panchayat* dataset, all the unsupervised and supervised machine learning models along with the deep learning models give sub-optimal results which are in close agreement with each other. The language-based models are the only models amongst the baseline algorithms to cross the 80% accuracy mark. However, they're still second to our proposed HumourHindiNet. The optimal performance of our proposed HumourHindiNet model can be attributed to the ability of our model to appropriately model the latent language dependencies along with using proper deep learning architectures for feature extraction.

6.2 Comparison with contemporary models

In this section, we compare the performance of our proposed framework with several recently proposed humour detection models. Most of the recently proposed models can be broadly categorized as follow: Language-model based, Feature engineering-based, and Deep learning based. To obtain a good sense of comparative study, we have selected a balance of algorithms from all these categories as follows. We use the deep-learning architectures of the various contemporary as it is. However, we convert the Hindi dialogues from our datasets to be compatible with the input of these models. We use our skip-gram-based model to tokenize the Hindi dialogues. For feature-engineering-based models, we generate the humour anchors for the Hindi language similar to the original work.

Just Accepted

Table 6. Performance comparison of the proposed HumourHindiNet framework with several baseline machine learning and deep learning algorithms on *Panchayat* dataset.

Methods	Accuracy	Precision	Recall	F1 Score
K-Means Clustering	76.37	73.45	81.23	77.14
KNN	75.18	80.0	69.26	74.24
Logistic Regression	75.91	77.6	71.85	74.61
Gaussian Naive Bayes	75.55	81.48	68.99	74.72
Decision Trees	78.83	81.92	75.53	78.6
Random Forest	78.01	82.4	70.72	76.11
ANN	78.56	84.03	70.33	76.57
CNN	74.09	70.95	79.44	74.95
RNN	78.92	79.52	78.22	78.86
GloVe	84.76	83.0	86.28	84.61
BERT	85.77	84.11	**87.55**	85.79
HumourHindiNet	**87.32**	**90.15**	84.14	**87.04**

Table 7. Performance comparison of our proposed HumourHindiNet framework with several recently proposed humour detection algorithms on $Kota - Factory$ dataset.

Methods	Accuracy	Precision	Recall	F1 Score
Bertero and Fung [2]	72.84	68.36	85.5	75.97
Chen and Lee [4]	78.91	76.37	82.3	79.22
Chen and Soo [5]	84.22	84.34	84.34	84.34
Yang et al. [30]	88.62	88.43	**89.22**	88.82
Oliveira and Rodrigo [6]	84.98	87.19	82.79	84.93
Kumar et al. [17]	85.76	84.10	87.54	85.79
HumourHindiNet	**91.79**	**96.29**	87.41	**91.64**

The obtained results are presented in Table 7 for the $Kota - Factory$ dataset and Table 8 for the *Panchayat* dataset. For both the datasets, we observe that the language-model based Chen and Lee [4] model is the second worst performer after the deep learning-based Bertero and Fung [2]. On the other hand, the deep learning-based models, namely, Chen and Soo [5], and Rodrigo and Rodrigo [6], are the best performers in terms of the recently proposed algorithms due to their ability to uncover and train on the various intricacies of the language and working to explore a deep understanding of the problem-specific details. The feature-engineering based Yang et al. [30], gives optimal and stable performance for both the datasets, due to its appropriate task-specific feature engineering, making it suitable for humour detection.

The proposed HumourHindiNet's superior performance can be primarily attributed to its ability to capture contextual dependencies within the dialogues, facilitated by the LSTM section of our model. Unlike traditional algorithms that may struggle to grasp the sequential nature of humour, the proposed HumourHindiNet' effectively leverages stacked LSTM layers to capture the intricate temporal relationships inherent in comedic content. This enables the model to discern not only individual humorous elements but also the overarching comedic structure within the dialogues, leading to more accurate predictions. Additionally, the utilization of appropriate preprocessing techniques ensures uniformity and well-structured dialogue across the datasets, reducing noise and enhancing the model's ability to extract relevant features. Furthermore, the CNN-based n-gram feature

Just Accepted

Table 8. Performance comparison of our proposed HumourHindiNet framework with several recently proposed humour detection algorithms on *Panchayat* dataset.

Methods	Accuracy	Precision	Recall	F1 Score
Bertero and Fung [2]	72.38	66.43	87.89	75.67
Chen and Lee [4]	84.98	82.96	88.66	85.71
Chen and Soo [5]	86.8	87.87	86.59	87.22
Yang et al. [30]	84.37	84.01	85.76	84.88
Oliveira and Rodrigo [6]	85.73	84.68	**87.72**	86.18
Kumar et al. [17]	84.76	83.01	86.27	84.68
HumourHindiNet	**87.32**	**90.15**	84.14	**87.04**

extraction mechanism employed by the proposed HumourHindiNet enables a deeper understanding of latent language dependencies, allowing the model to uncover subtle linguistic nuances that contribute to the comedic effect. By combining these advanced techniques, the proposed HumourHindiNet achieves superior performance across all performance metrics, establishing it as the top performer in humour detection for Hindi web series.

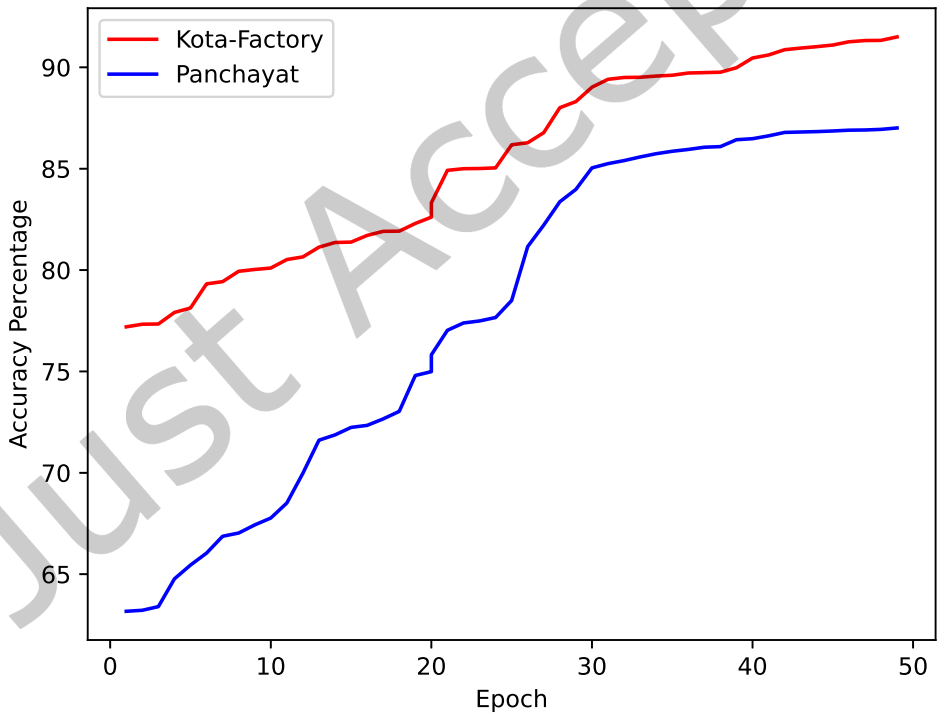


Fig. 5. The Accuracy values obtained by the proposed model for the two datasets across all the epochs

Just Accepted

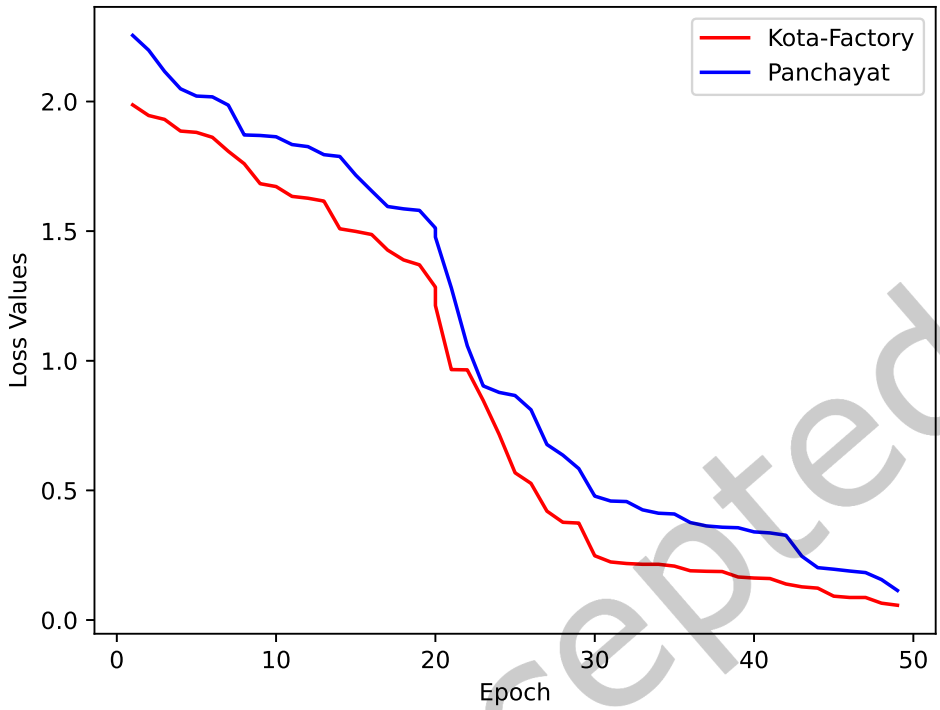


Fig. 6. The Loss values obtained by the proposed model for the two datasets across all the epochs.

Table 9. Accuracy Comparisons with Contemporary models on Non-Hindi (English) Datasets.

Methods	**16000 One-Liners**	**Pun of the Day**	**Short Jokes**
Bertero and Fung [2]	79.6	83.6	84.3
Chen and Lee [4]	84.8	86.1	87.6
Chen and Soo [5]	89.7	89.4	90.6
Yang et al. [30]	79.7	85.4	87.6
Oliveira and Rodrigo [6]	87.5	88.7	88.1
Kumar et al. [17]	86.6	86.6	83.2
HumourHindiNet	90.2	90.7	91.1

Just Accepted

6.3 Comparison with Non-Hindi Datasets

In this section, we compare the performance of different contemporary models with our proposed model on non-Hindi specifically English datasets. We utilize several benchmark humour English datasets, namely, 16000 One-Liners [21], Pun of the Day [30], and Short Jokes Kaggle. We modify the proposed approach by replacing skip-gram based Hindi embedding with skip-gram-based English embedding using an English text corpus in Section 4.2. This helps the proposed model to generate word embedding for the English words in these datasets. The generated embedding is then fed to the "N-gram based feature extraction using CNN" module as mentioned in Section 4.3. This enables our model to adapt to non-Hindi Humour datasets. Table 9 lists the results of the

proposed model along with some contemporary models on above mentioned English datasets. The obtained results highlight HumourHindiNet's superior accuracy for these datasets, underscoring its advanced approach that combines word embeddings, CNNs, and stacked LSTM layers. This blend allows for a nuanced understanding of humor's complex linguistic and sequential nature, evident in its leading accuracy figures: 90.2% on 16000 One-Liners, 90.7% on Pun of the Day, and 91.1% on Short Jokes. Such performance not only demonstrates HumourHindiNet's adeptness at humor detection, particularly in the nuanced and culturally rich context of Hindi web series but also for English datasets.

6.4 Accuracy and Loss vs Epoch Analysis

In this section, we perform the Accuracy vs Epoch and Loss vs Epoch analysis. Figure 5 shows the Accuracy results obtained by our proposed framework for different epochs. For the $Kota - Factory$ dataset, the accuracy significantly improved from 77.2% to a remarkable peak of 91.79% throughout 50 epochs. This ascent demonstrates the model's ability to learn and capture humour patterns within the Hindi web series content. While the accuracy exhibits impressive growth early in training, it begins to plateau towards the latter epochs, suggesting that the model has captured a substantial portion of the humour-related features present in the dataset. In the case of the *Panchayat* dataset, a similar trend is observed. The accuracy starts at 63.17% and gradually climbs to a peak of 87.32% after 50 epochs. This growth indicates the model's effectiveness in discerning humour in this specific dataset. Like the $Kota - Factory$ dataset, the accuracy curve levels off in the later epochs, signaling a diminishing rate of improvement as the model has likely captured the predominant humour-related patterns within the data.

Figure 6 shows the accuracy results obtained by our proposed framework for different epochs. In terms of loss, the $Kota - Factory$ dataset exhibited a substantial decrease during training. The loss started at 1.987 and dropped consistently as the model learned. This decline reflects the model's ability to reduce the errors in its predictions. Similar to the accuracy trend, the loss curve begins to plateau towards the later epochs, indicating that the model's learning rate has slowed, and is approaching convergence. The *Panchayat* dataset showcases a comparable loss reduction over the 50 epochs. Starting at 2.255, the loss steadily diminishes as the model refines its understanding of humour within this dataset. However, similar to the $Kota - Factory$ dataset, the loss reduction rate decelerates in the later epochs, signifying that the model has captured the primary humour-related patterns in the data, and further significant loss reduction becomes challenging. In summary, both datasets exhibit notable improvements in accuracy and reductions in loss over the training period, indicating that the model is effective at humour detection. The plateauing of these trends in the later epochs suggests that the model's learning rate slows as it approaches convergence, having already captured a substantial portion of the underlying humour patterns within the respective datasets.

6.5 AUC Analysis

In this section, we evaluate the AUC values for our algorithm. Since the number of data points in our datasets is extremely less as compared to other pre-trained models, hence, the high performance of our model could be due to the inherent bias of our model. To evaluate the bias of our model, we calculate the AUC values for our model. An AUC value near 0 shows that the model has the worst distinguishing capability and it is classifying positive samples as negative and vice-versa. An AUC value of 0.5 means that the model has no distinguishing capability while a value close to 1 represents that the model can appropriately distinguish amongst the classes. Figure 7 shows the AUC values obtained by our model for the two datasets. From the figure, we can see that the obtained AUC values are well over 0.5, hence it shows that our model has great distinguishing capability and doesn't suffer from bias.

Just Accepted

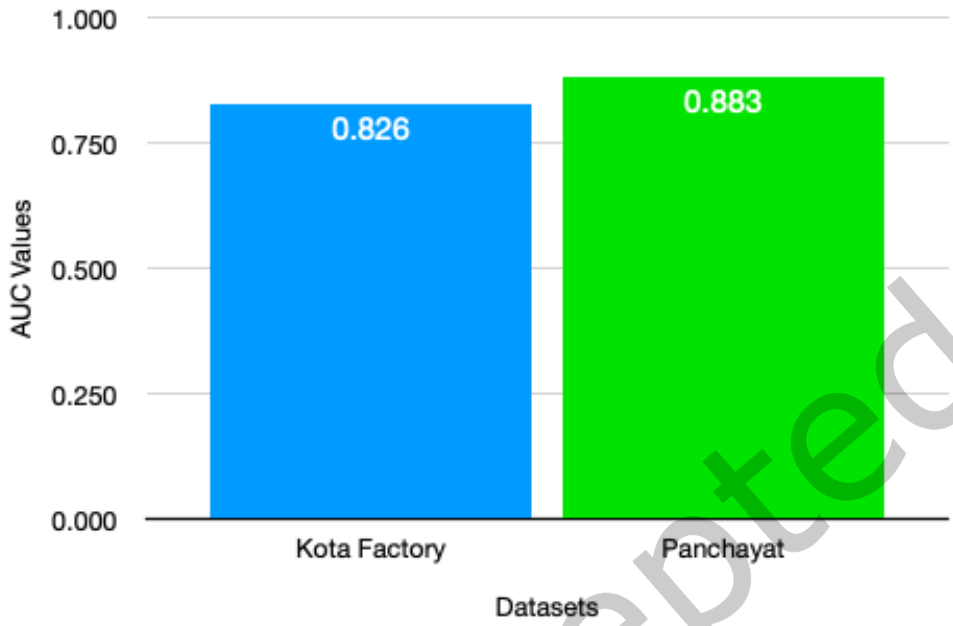


Fig. 7. The AUC values obtained by our model for the two datasets.

Just Accepted

6.6 Time Complexity

In this section, we study the time complexity of the various components of our proposed model. Our model involves various steps, namely, (i) Skip-gram model based word embedding generation, (ii) feature extraction using CNN layers, (iii) stacked LSTM layers, and (iv) a deep neural network classifier. The time complexity of the various steps is given below.

(i) Skip-gram model-based word embedding generation: Let N be the size of the entire text corpus, D be the size of the desired embeddings, and the V be the vocabulary of the unique words in the text corpus. We perform a binary search on the unique words vocabulary and we generate a D dimensional embedding for all the N words in the corpus, hence the time complexity for the skip-gram model would be $O(N \times D \times log(V))$.

(ii) Feature extraction using CNN layers: Let D be the size of the word embedding extracted in the previous step. Let D be expressed as a 2D matrix of size $n \times d$ for convolutions. Let k be the kernel size and f be the filter size. Since, we evaluate convolutions for k kernels across the entire $n \times d$, hence the time complexity becomes $O(k \times n \times d)$. Since this is repeated for the entire filter size f, hence the final time complexity comes out to be $O(f \times k \times n \times d)$.

(iii) As per the works of Raja [25], let I be the number of inputs, K be the number of outputs and H be the number of cells in the hidden layers. The time complexity for the LSTM model can then be defined as $O(4IH + 4H2 + 3H + HK)$.

(iv) Deep neural network classifier: For a deep neural network classifier, the output of every node in the input layer is multiplied by the input of every node in the first hidden layer, and the output of every node of the last hidden layer is multiplied with the input of every node of the output layer. Moreover, a similar multiplication happens across the nodes of all the layers of the hidden layers. Hence, let P be the product

of the number of layer-wise nodes in the input layer, all the hidden layers, and the output layer. Let E be the number of epochs and S be the total number of training samples. Hence, the time complexity of a deep neural network classifier is $O(E \times S \times P)$.

(v) Overall time-complexity: Combining the above four phases of our model, we arrive at an overall time complexity of $O(N \times D \times log(V) + f \times k \times n \times d + 4IH + 4H2 + 3H + HK + E \times S \times P)$

6.7 Space Complexity

In this section, we study the space complexity of the various components of our proposed model. Our model involves various steps, namely, (i) Skip-gram model based word embedding generation, (ii) feature extraction using CNN layers, (iii) stacked LSTM layers, and (iv) a deep neural network classifier. The time complexity of the various steps is given below.

(i) Skip-gram model-based word embedding generation: Let N be the size of the entire text corpus, D be the size of the desired embeddings. Since we calculate and store the embedding for every word in the text corpus, hence the space complexity for the skip-gram model would be $O(N \times D)$.

(ii) Feature extraction using CNN layers: For CNN layers, the output dimensions decrease after every layer. Hence, the largest space that we need is the output of the first layer. Let D be the size of the word embedding extracted in the previous step and let f be the filter size. The dimension of the output of the CNN layer is $((D - f)/2) + 1$. Hence the space complexity becomes $O(((D - f)/2) + 1)$. Since $f << D$, this can also be written as $O(D)$.

(iii) As per the works of Raja [25], let I be the number of inputs, and H be the number of cells in the hidden layers. The space complexity for the LSTM model can then be defined as $O(4 \times (I + H) \times H + K)$.

(iv) Deep neural network classifier: For a deep neural network classifier, the weights are stored for every neuron in every layer of the network. Let P be the product of the number of layer-wise nodes in the input layer, all the hidden layers, and the output layer. Hence, the space complexity of a deep neural network classifier is $O(P)$.

(v) Overall space-complexity: Combining the above four phases of our model, we arrive at an overall time complexity of $O(N \times D + D + 4 \times (I + H) \times H + K + P)$. Since, $D << N \times D$ and $K << 4 \times (I + H) \times H$, hence the overall space complexity can be written as $O(N \times D + 4 \times (I + H) \times H + P)$

6.8 Advantages and Limitations

The proposed approach offers several notable advantages. Firstly, it leverages word embedding techniques to enhance the contextual understanding of Hindi dialogues. This is particularly crucial for humour detection, as humour often relies on subtle contextual cues. By representing words in a dense vector space, semantic relationships are preserved, significantly improving the model's grasp of contextual humour. Secondly, the integration of convolutional neural networks, such as proposed HumourHindiNet, provides an automatic feature learning mechanism. This reduces the need for labor-intensive handcrafted feature engineering and potentially results in more robust and adaptable models. Moreover, the inclusion of stacked LSTM layers allows the model to capture sequential dependencies within dialogues, an essential aspect of humour recognition. Lastly, the research addresses a timely need in the digital age, where web series and digital content are gaining immense popularity. It contributes to the field of entertainment analytics, offering practical applications and insights into humour detection.

Despite its merits, the proposed approach does come with certain challenges. Firstly, its success hinges on the availability of suitable datasets for training and evaluation. If there's a scarcity of diverse and high-quality datasets specific to humour in Hindi web series, the model's performance could be constrained. Secondly, the

Just Accepted

computational complexity associated with deep learning models must be considered. These models demand substantial computational resources for training and inference, potentially limiting their accessibility to researchers with constrained computing power. Additionally, there's a risk of overfitting, especially if the dataset is limited or not well-regularized, which could affect the model's ability to generalize to new data. Lastly, deep learning models are often criticized for their lack of interpretability, making it challenging to understand the rationale behind specific humour detection decisions. This limitation might hinder their real-world applicability in certain contexts.

In summary, while the proposed approach holds promise for advancing humour detection in Hindi web series, it's essential to carefully address challenges related to data availability, computational resources, model complexity, and interpretability to maximize its potential impact in the field of natural language processing and entertainment analytics.

7 CONCLUSION

Humour detection is an important task in the field of NLP. However, relatively it is a difficult task as humour is often contorted in between conversations or expressed through feeble expressions or gestures that it isn't easily detectable. Moreover, it also requires good language-specific knowledge to detect humour appropriately. There is also a huge shortage of humour detection datasets in the Hindi language making it increasingly difficult to obtain good performance for detecting humour in the Hindi language. Following along these lines, in this paper, we proposed two Hindi datasets that are extracted from two recent popular web series, namely, $Kota - Factory$ and *Panchayat*. We also proposed Hindi word embedding and convolutional network-based framework, named HumourHindiNet, for humour detection in Hindi web series. We started by preprocessing the dataset to remove irrelevant data and obtain uniformity across datasets. The processed dataset is then used by our proposed Hindi language-based word embedding generation model to create word embeddings for all the dialogues in the introduced datasets. The generated word embeddings are then passed through parallelly stacked CNN layers to extract n-gram features based on varying filter sizes which range from 3 to 5. The extracted features are then passed through stacked LSTM for further processing and exploring the latent language and contextual dependencies. The final extracted features are then passed to a deep neural network architecture to make the final classifications as humour or non-humour. We conducted intensive experiments on both the proposed datasets and evaluate several standard performance metrics. We compared the performance of our proposed HumourHindiNet model with several baseline unsupervised and supervised machine learning algorithms, deep learning algorithms, and language-based models. We also compared our performance with several recently proposed language-based algorithms, feature-engineering based algorithms, and deep learning-based algorithms for humour detection. The obtained results demonstrated that our proposed HumourHindiNet model gives the best performance and proves its efficiency and efficacy for humour detection. This work can further be extended by using multi-language or multi-modal datasets. We can also work on creating humour detection datasets for other low-resource languages.

Just Accepted

REFERENCES

[1] Dario Bertero and Pascale Fung. 2016. Deep learning of audio and language features for humor prediction. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. 496–501.

[2] Dario Bertero and Pascale Fung. 2016. A long short-term memory framework for predicting humor in dialogues. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 130–135.

[3] Dushyant Singh Chauhan, Gopendra Vikram Singh, Navonil Majumder, Amir Zadeh, Asif Ekbal, Pushpak Bhattacharyya, Louis-philippe Morency, and Soujanya Poria. 2021. M2H2: A Multimodal Multiparty Hindi Dataset For Humor Recognition in Conversations. In *Proceedings of the 2021 International Conference on Multimodal Interaction*. 773–777.

[4] Lei Chen and Chungmin Lee. 2017. Predicting audience's laughter during presentations using convolutional neural network. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*. 86–90.

[5] Peng-Yu Chen and Von-Wun Soo. 2018. Humor recognition using deep learning. In *Proceedings of the 2018 conference of the north american chapter of the association for computational linguistics: Human language technologies, volume 2 (short papers)*. 113–117.

[6] Luke De Oliveira and Alfredo L Rodrigo. 2015. Humor detection in yelp reviews. *Retrieved on December* 15 (2015), 2019.

[7] J Devlin, MW Chang, K Lee, and KB Toutanova. 2019. Pre-training of deep bidirectional transformers for language understanding In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). *Minneapolis, MN: Association for Computational Linguistics* (2019), 4171–86.

[8] Xiaochao Fan, Hongfei Lin, Liang Yang, Yufeng Diao, Chen Shen, Yonghe Chu, and Yanbo Zou. 2020. Humor detection via an internal and external neural network. *Neurocomputing* 394 (2020), 105–111.

[9] Sayani Ghosal and Amita Jain. 2023. HateCircle and Unsupervised Hate Speech Detection Incorporating Emotion and Contextual Semantics. *ACM Transactions on Asian and Low-Resource Language Information Processing* 22, 4 (2023), 1–28.

[10] Md Kamrul Hasan, Wasifur Rahman, Amir Zadeh, Jianyuan Zhong, Md Iftekhar Tanveer, Louis-Philippe Morency, et al. 2019. UR-FUNNY: A multimodal language dataset for understanding humor. *arXiv preprint arXiv:1904.06618* (2019).

[11] Devamanyu Hazarika, Soujanya Poria, Amir Zadeh, Erik Cambria, Louis-Philippe Morency, and Roger Zimmermann. 2018. Conversational memory network for emotion recognition in dyadic dialogue videos. In *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting*, Vol. 2018. NIH Public Access, 2122.

[12] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759* (2016).

[13] Sulaiman Khan, Shah Nazir, and Habib Ullah Khan. 2023. Analysis of Cursive Text Recognition Systems: A Systematic Literature Review. *ACM Transactions on Asian and Low-Resource Language Information Processing* (2023).

[14] Chloe Kiddon and Yuriy Brun. 2011. That's what she said: double entendre identification. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*. 89–94.

[15] Dhanashree S Kulkarni and Sunil S Rodd. 2021. Sentiment Analysis in Hindi—A Survey on the State-of-the-art Techniques. *Transactions on Asian and Low-Resource Language Information Processing* 21, 1 (2021), 1–46.

[16] Sanjay Kumar. [n. d.]. Negative Stances Detection from Multilingual Data Streams in Low-resource Languages on Social Media using BERT and CNN based Transfer Learning Model. *ACM Transactions on Asian and Low-Resource Language Information Processing* ([n. d.]).

[17] Vijay Kumar, Ranjeet Walia, and Shivam Sharma. 2022. DeepHumor: a novel deep learning framework for humor detection. *Multimedia Tools and Applications* 81, 12 (2022), 16797–16812.

[18] Paul Pu Liang, Ziyin Liu, Amir Zadeh, and Louis-Philippe Morency. 2018. Multimodal language analysis with recurrent multistage fusion. *arXiv preprint arXiv:1808.03920* (2018).

[19] Bing Liu. 2020. Text sentiment analysis based on CBOW model and deep learning in big data environment. *Journal of ambient intelligence and humanized computing* 11, 2 (2020), 451–458.

[20] Yong Liu, Wei Wei, Aixin Sun, and Chunyan Miao. 2014. Exploiting geographical neighborhood characteristics for location recommendation. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*. 739–748.

[21] Rada Mihalcea and Carlo Strapparava. 2005. Making computers laugh: Investigations in automatic humor recognition. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*. 531–538.

[22] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems* 26 (2013).

[23] Reynier Ortega-Bueno, Carlos E Muniz-Cuza, José E Medina Pagola, and Paolo Rosso. 2018. UO UPV: Deep linguistic humor detection in Spanish social media. In *Proceedings of the third workshop on evaluation of human language technologies for Iberian languages (IberEval 2018) co-located with 34th conference of the Spanish society for natural language processing (SEPLN 2018)*. 204–213.

[24] Hai Pham, Paul Pu Liang, Thomas Manzini, Louis-Philippe Morency, and Barnabás Póczos. 2019. Found in translation: Learning robust joint representations by cyclic translations between modalities. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 6892–6899.

[25] Sudeep Raja. 2021. FNNs, RNNs, LSTM and BLSTM.

[26] Alison Ross. 2005. *The language of humour*. Routledge.

[27] Ramsha Saeed, Hammad Afzal, Sadaf Abdul Rauf, and Naima Iltaf. 2023. Detection of Offensive Language and ITS Severity for Low Resource Language. *ACM Transactions on Asian and Low-Resource Language Information Processing* 22, 6 (2023), 1–27.

[28] Yansen Wang, Ying Shen, Zhun Liu, Paul Pu Liang, Amir Zadeh, and Louis-Philippe Morency. 2019. Words can shift: Dynamically adjusting word representations using nonverbal behaviors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 7216–7223.

[29] Haojie Xu, Weifeng Liu, Jiangwei Liu, Mingzheng Li, Yu Feng, Yasi Peng, Yunwei Shi, Xiao Sun, and Meng Wang. 2022. Hybrid Multimodal Fusion for Humor Detection. In *Proceedings of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge*. 15–21.

[30] Diyi Yang, Alon Lavie, Chris Dyer, and Eduard Hovy. 2015. Humor recognition and humor anchor extraction. In *Proceedings of the 2015 conference on empirical methods in natural language processing*. 2367–2376.

Just Accepted

[31] Amir Zadeh, Paul Pu Liang, Navonil Mazumder, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. 2018. Memory fusion network for multi-view sequential learning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.

[32] Yftah Ziser, Elad Kravi, and David Carmel. 2020. Humor detection in product question answering systems. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 519–528.

Just Accepted

2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI) | 979-8-3503-6052-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/IATMSI60426.2024.10503284

Hypervisor Based Intrusion Detection Using Enhanced Radial Basis Neural Network on Cloud Environment

Naga Seshu Kumar Anumukonda
CSE Department
Delhi Technological University
Delhi,India.
seshu1203@gmail.com

Rajesh Kumar Yadav
CSE Department
Delhi Technological University
Delhi,India.
rkyadav6711@gmail.com

Raghava N S
ECE Department
Delhi Technological University
Delhi,India.
nsraghava@gmail.com

Abstract—Cloud computing has grown to be a significant component of the IT sector in the current digital economy. Despite its advantages such as reduced capital costs and operational costs, there are some constraints that limit its use. Previous research has proposed various intrusion detection systems and authentication systems, but they fail to detect intruders. Hypervisor-based intrusion detection systems have recently been developed that monitor both virtual machines (VM) and the host simultaneously. Furthermore, it has become popular because it provides a comprehensive view of the cloud environment and allows for more accurate threat detection. Considering these, we propose novel hypervisor-based intrusion detection in this work. Initially, we collect intrusion data and pre-process it. The best features are then chosen using an adaptive artificial flora algorithm (AAF), which lowers processing time and complexity. Then, the selected features are passed to the ERBNN classifier to classify the data as normal or intrusive data. For validation of the proposed method, KDDCup-99 and NSL-KDD datasets are used. The performance of the suggested technique is validated based on different metrics and the proposed system is implemented using Python.

***Keywords*—cloud computing, intrusion detection, feature selection, ERBNN, classification.**

I. INTRODUCTION

A hypervisor is a software layer used in cloud computing that allows the operation of numerous virtual machines (VMs) on one physical system [1]. Hypervisors create and manage VMs isolated from each other and from the underlying hardware. In addition to harming a single virtual machine, an attacker can harm several virtual machines and hypervisors by installing hacking tools on them [2]. The security of a system is determined by its ability to provide confidentiality, integrity and availability to its users [3]. With the growing significance of network-based computers in modern society, they are subject to more attacks and intrusions [4]. The term intrusion refers to a series of illegal actions conducted to compromise the security of a system [5]. Detecting intrusions into computer systems and networks by both internal users and external attackers is the primary objective of intrusion detection [6].
Host-based systems and network-based systems are the two conventional categories into which intrusion detection systems can be separated [7]. There are advantages and disadvantages to both host-based and network-based intrusion detection systems [8]. There is a limited effectiveness of network intrusion detection systems when it comes to detecting insider attacks because they are designed to address outsider attacks. Due to the shared multi-tenancy nature of cloud computing, these and other perimeter security controls may be less effective [9]. Many cloud users are virtually distributed across physical machines, which opens them up to attack via virtual or internal networks [10]. Despite the success of host-based intrusion detection systems, users are typically required to monitor and manage these systems themselves. When users have multiple instances running in the cloud, this approach can be challenging. In addition, host-based intrusion detection systems can be disabled by a skilled attacker once the event has been compromised [11].

II. LITERATURE SURVEY

A network-based intrusion detection system made up of ensembles of four classifiers machine learning technique was suggested by Singh and Ranga [12]. In order to obtain a consolidated final score, a voting mechanism was incorporated into the framework. Simulation and testing of the proposed model were conducted using standard datasets and simulators, namely CICIDS 2017 and CloudSim. According to the results, this model has a low rate of false positive alerts, as well as a high rate of detection, indicating that it is superior to state-of-the-art approaches for identifying intrusions in cloud environments. The accuracy of this work needs to be improved.

A cloud intrusion detection system (IDS) that employed online multivariate statistical analysis of change to identify unusual network activity was created by Aldribi et al. [13]. Using the hypervisor's collection of events as a starting point, an event-based IDS feature model is presented as part of traditional monolithic network IDS feature models. This model improves detection efficiency by taking advantage of the diverse and connected tendencies of event. To test this strategy, an updated cloud intrusion dataset with a wide range of attack routes was utilised. A Riemann rolling feature extraction strategy was utilised to empirically test this approach, and it was shown to be better to the commonly employed Riemann chunking approach.
Jin et al. [14] developed SwiftIDS, an IDS that provides satisfactory detection performance in high-speed networks for massive traffic analysis. The following methods are used by SwiftIDS to achieve these goals. In order to handle

979-8-3503-6052-3/24/$31.00 ©2024 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:59:31 UTC from IEEE Xplore. Restrictions apply.

massive traffic data, Light Gradient Boosting Machine addition, traffic data arriving at different time windows can also be analyzed with a parallel intrusion detection mechanism. By doing this, later incoming data will not be delayed by waiting for the end of the intrusion detection cycle of the first incoming data. Using offline tests, SwiftIDS achieved excellent time performance and satisfactory detection performance. Zafar and Rehman [15]proposed an improved intrusion detection system for cloud computing environments. An approach combining SVM and FCM clustering was used to develop Hypervisor Inspector. Three phases were incorporated into this model. As part of the first phase, the FCM clustering module was introduced, which allowed large datasets to be divided into smaller clusters, which in turn enabled the SVM to learn more effectively and more quickly. After the cluster values were assigned, different SVM modules were trained. Thirdly, the results of the Hypervisor Inspector were aggregated by a module called fuzzy aggregation. Various evaluation metrics were used to compare this technique with existing techniques. It has been demonstrated that this mechanism provides a high level of detection accuracy and low levels of false alarms when detecting various attacks.

According to Mendonca et al, [16] an intrusion detection system (IDS) is based on a tree-CNN hierarchical algorithm with an activation function called soft-root-sign (SRS). The model developed to detect DDoS, intrusion, brute force and web attacks reduced the training time with this model. The complexity level of this solution was analyzed. The results demonstrate that this hierarchical model can reduce execution time while improving detection accuracy by considering all analyzed attacks.

SaisindhuTheja and Shyam [17] proposed an effective system for detecting DoS attacks that used the adversarial crow search algorithm (OCSA) that combined the raven search algorithm (CSA) with adversarial based learning (OBL). Features were chosen by OCSA and classified using a Recurrent Neural Network (RNN) classifier/classifier. A RNN classifier was used to select the essential features based on the OCSA algorithm. An RNN classifier was used to classify the given data in the testing stage. An experimental evaluation of the technique demonstrated that it outperformed other conventional methods based on a benchmark data set.

III. PROPOSED METHODOLOGY

The main objective of this study is to present an innovative intrusion detection system for cloud environment. Figure 1 shows the proposed intrusion detection architecture. Initially, intrusion data is collected from the virtual machine's data collection agent (DAC) and pre-processed. After that, to conserve time and complexity of computation, the best features are chosen using an adaptive artificial flora algorithm. Finally, the selected features are passed to an Enhanced Radial Based Neural Network classifier, which classifies the data as normal or intrusive data.

A. Hypervisor based intrusion detection

In a cloud environment, hypervisor-based intrusion detection (HypIDS) is a type of intrusion detection system (LightGBM) is used as an intrusion detection algorithm. In deployed within the hypervisor. Monitoring and analyzing network traffic and detecting intrusions are carried out by hypervisor-level data and communications.The HypIDS offers comprehensive visibility into the cloud environment by simultaneously gathering data from virtual machines (VMs) and the host. It is feasible to watch and examine network interactions involving virtual machines (VM-VM), among VMs and hypervisors, and also between the cloud context and the outer (network-based) world. HypIDS is designed to detect a wide variety of intrusions, including malware infections, unauthorized access attempts, and denial of service attacks (DoS). To detect intrusions, handwriting-based rules and/or machine learning models are used. It is possible for HypIDS to take a number of actions if an intrusion is detected, including alerting the cloud administrator, blocking malicious traffic, or isolating the infected virtual machine.

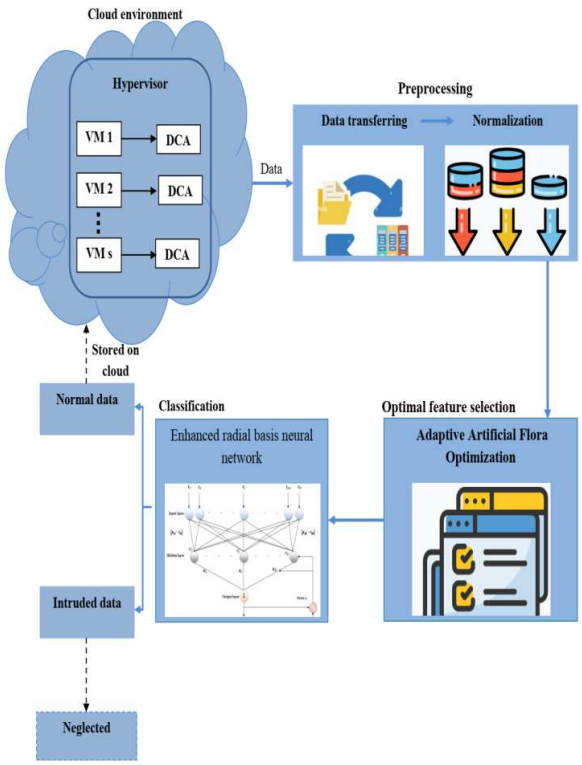


Fig.1. Overall architectural diagram for the proposed intrusion detection

B. Data pre-processing

At first, intrusion data is collected from the datasets. As part of the data collection process, the data should be pre-processed, since pre-processing is an integral part of the classification process. During the pre-processing of data, two primary stages are involved: the transfer of data and the normalization of the data. A number of non-numeric values are converted into numeric values during data transfer. There are symbolic characteristics present in all of the datasets, such as flag features and protocol features that are incompatible with classification methods (e.g., tcp, udp, etc.), as well as features related to services (e.g., telnet, ftp). As a result, each symbol must be made up of numeric values, and every class must be represented by a numeric value.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:59:31 UTC from IEEE Xplore. Restrictions apply.

The data is normalized after it has been transferred. In order to normalize, the min-max method is used. Equation (1) is used to normalize the value of each attribute.

$$A_{new} = \frac{A_{current} - A_{\min}}{A_{\max} - A_{\min}} \quad (1)$$

$A_{\min}$ → Minimum value attribute in the corresponding column

$A_{\max}$ -> Maximum value attribute in the corresponding column

After pre-processing, the pre-processed data is sent to AAF for optimal feature selection.

C. Adaptive artificial flora (AAF) algorithm

Seeds are dispersed by plants in different ways. There are two types of seed dispersal: autochory and allochory. A plant that propagates itself is called an autochory, whereas a plant that propagates via external forces is called an allochory. The process of mechanical propagation can be considered autophagy, and the process of biological propagation can be considered allochory. Because of autochory, plants can move to an appropriate habitat. Red fruits are loculicidal when they are ripe, and their shells curl to release the seeds. When impatiens fruit bursts, the seeds are spread throughout the surrounding area. A plant migration to a distant and unspecified region is portrayed in the allegory. Anemochory, as the wind speed and direction change, determines the direction and distance of seed dispersal. As a result of these breeding methods, plant research has been widened and the risk of extinction of plants has been reduced.

The range of vegetation can expand, decrease, or be displaced due to climate change, a harsh natural environment, or competition. The individual within a plant also evolves as it migrates to a new environment. As a result, plant migration may cause changes in the distribution area of plants, as well as trigger the evolution, extinction, and regeneration of plants. By dispersing seeds and reproducing, plants locate the best place to live despite the fact that they cannot move nor have intelligence.

During plant migration and reproduction, seeds are dispersed randomly within a certain distance by the original plant. Depending on the external environment, a seed has a different probability of survival.

After ripening, a plant's seeds spread when it is placed in a suitable environment. Extreme weather can cause plants to either develop and adjust to their surroundings or go perish in the area. Plants may be able to proliferate in other regions before they die out completely in one region as a result of an analogy. Seeds can be carried by plants from one area to another where they reproduce. As a result of several generations of reproduction, plants will migrate from one area to another that is more suitable for them. Evolution, extinction, and regeneration are the mechanisms by which plants locate the optimal growing environment through migration and reproduction.

Four fundamental components are taken into account in artificial floras: the parent plant, the offspring of that plant, the plant's position, and the propagation distance. An original plant is defined as one that is prepared to yield seeds. Seeds from offspring plants cannot spread seeds at that time as they are seeds from original plants. The location of a plant is its geographical location. The propagation distance refers to the extent of a seed's spread. Behavioral patterns can be categorized into three main categories: evolution, spreading, and selection. It is possible that a plant will adapt to its environment based on its evolution behavior. The spreading behavior of seeds refers to their movement, and seeds may move by autochory or allochory. Depending on the environment, flora may either survive or go extinct. In this work, we use this AAF algorithm to select features to reduce computational complexity and time consumption.

The following describes the phases involved in the AAF algorithm's feature selection process:

1) Initialization: In order for an algorithm to be optimized, the solution initialization process must be followed.Based on the AAF algorithm, the original plants were randomly initialized with N solutions, representing the number of plants in the original plant. By using this approach, the available cells in the network are considered as candidates and initialized as follows.

$$Y_N = \{y_1, y_2, ..., y_S\}_N \quad (2)$$

Where, y_S is the solution and can be expressed as

$$Y_S = [\,Selected \quad features\,]$$

2) Fitness calculation: Fitness is calculated for each initialized solution. With this method, the classification accuracy is used to calculate each solution's fitness. The best fitness is considered to be most accurate. Equation (3) represents the fitness function.

$$Fitness = Max\,(Accuracy) \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

3) Update the Solution: The proposed AAF algorithm optimization is based on three behaviours of plants, namely, evolutionary behaviour, spreading behaviour, and selective behaviour.

a) Evolutionary behaviour: Initially, seeds were spread out over a circle with a specific radius, or dispersal distance. The reproductive distance is calculated using the propagation distances of the mother plant and the grandfather plant.

$$d_q = d_{1q} \times rand\,(0,1) \times l_1 + d_{2q} \times rand\,(0,1) \times l_2 \quad (5)$$

In this equation, d_q represents the propagation distance between plants. Specifically, d_{1q} represents the new grandparent propagation distance, while d_{2q} represents the new parent propagation distance. l_1 and l_2 denote the learning coefficient, and rand(0,1) determines the uniformly distributed independent number in (0,1).

Equation (6) represents the propagation distance of a new grandparent.

$$d'_{1q} = d_{2q} \quad (6)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:59:31 UTC from IEEE Xplore. Restrictions apply.

The new parent propagation distance is determined by the standard difference among the locations of the offspring plant and the original plant.

$$d'_{2j} = \sqrt{\frac{\sum_{q=1}^{N} (S_{p,q} - S'_{p,q})^2}{N}} \tag{7}$$

Where, N represents the number of plants in the flora, $S_{p,q}$ represents the location of the original plants in the flora, and $S'_{p,q}$ represents the position of the offspring plants in the flora.

b) Spreading behaviour: In the flora algorithm, N solutions are generated at random, resulting in N plants in the flora. Using a matrix $S_{p,q}$ where p represents the dimension and q represents the number of plants in the flora, the position of the original plants is expressed.

$$S_{p,q} = rand(0,1) \times d \times 2 - d \tag{8}$$

In which d is the maximum range region and $rand(0,1)$ is an array of random numbers evenly distributed between (0,1).

Following is how the propagation function generates the position of the offspring plant:

$$S'_{p,q\times c} = D_{p,q\times c} + S_{p,q} \tag{9}$$

c is the maximum number of seeds that may be propagated by one plant, $S'_{p,q\times c}$ is the position of the offspring, $S_{p,q}$ is the position of the original plant, and $D_{p,q\times c}$ is a random number with Gaussian distribution and mean 0 and variance d_q. The original plant is generated according to Equation (8) if none of the offspring plants survive.

*c) Select behaviour:*The survival probability of the offspring plants is determined as follows:

$$S = \left| \sqrt{\frac{Fitness\ (S'_{p,q\times c})}{Fitness_{\max}}} \right| \times M_x^{(q\times c-1)} \tag{10}$$

Where M_x = selection probability. There must be a value between 0 and 1. It is possible that a progeny plant that is too distant from the original plant will be less fit. The exploratory efficiency of the algorithm is determined by M_x. It is important to have a large M_x for an easy problem in order to obtain a local optimal solution. In this generation of plants, $Fitness_{\max}$ represents the maximum fitness, and $Fitness(S'_{p,q\times c})$ represents the fitness of the qth solution.

4) Updation using Levy flight with AAF: Levy flight strategies are introduced to avoid the occurrence of local optima that fall easily, resulting in a loss of AAF diversity. Many optimization algorithms make use of the Levy plane, which is a stochastic strategy. This algorithm employs a random walk with a high step probability, which can increase its randomness. The AAF formula was enhanced by including Levy planes in step 3 to enhance its analytical capability:

$$Y_S = \left| \sqrt{\frac{Fitness\ (S'_{p,q\times c})}{Fitness_{\max}}} \right| \times M_x^{(q\times c-1)} \bullet Levy\ (D) \tag{11}$$

$$Levy\ (D) = 0.01 \times \frac{a \times \sigma}{|b|^{\frac{1}{\beta}}} \tag{12}$$

$$\sigma = \left(\frac{\Gamma(1+\beta) \times \sin(\frac{\pi\beta}{2})}{\Gamma(\frac{1+\beta}{2}) \times \beta \times 2^{(\frac{\beta-1}{2})}} \right)^{\frac{1}{\beta}} \tag{13}$$

The Levy plane function is denoted by *levy*, and the problem dimension is denoted by D. Here, β is a constant equal to 1.5 and random values between 0 and 1 are assigned to a and b.After the completion of the initial and second phases of the AAF, the iteration is complete. Each iteration of the algorithm updates the population using Eqs (5) and (11) until it reaches its last iteration. The best solution obtained during each iteration of the algorithm is given as output after completing the ROA processing.After feature selection, an improved radial-based neural network is used to classify intrusions.

D. Enhanced radial basis neural network (ERBNN)

Artificial neural networks that employ radial basis operations as activation functions are called enhanced radial basis neural networks (ERBFNNs). An input layer, a hidden layer having radial basis activation functions, and an output layer are the three layers that typically make up this model. In order to model the interaction among input and output data, the RBFNN model is built on expressing the hidden layer as a set of radial basis functions, frequently Gaussian functions.

The input, hidden, and output layers make up the three layers of the ERBFNN. An input layer and the jth input vector are connected $Y_j = [y_{j1}, y_{j2}, \cdots, y_{ji}, y_{jn}]$, with $i = 1,2,...,m$ and $j = 1,2,\cdots,n$. The total number of output nodes x_j, $j = 1,2,\cdots n$, is the same as the number of training input-output pair of data. Put differently, a pair of input nodes (represented by a matrix) and output nodes (represented by a vector) exist. In particular, the vectors of the jth hidden node are: $V_j = [V_{j1}, V_{j2}, \cdots, v_{jr}, \cdots, V_{jR}]$, $r = 1,2,\cdots R\ and\ j = 1,2,\cdots,n$. In this regard, W_{jr} describes the connection between the Kth hidden node and the jth output node. Figure 3 illustrates the structure of the ERBFNN. A number of input variables is indicated by M, a number of training sets is indicated by N, and a number of centers is indicated by R in Figure 2.

1) Input layer: This paper uses y_i as the ith variable in its analysis of expected performance. The input matrix for each pair of training data can be described as follows:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:59:31 UTC from IEEE Xplore. Restrictions apply.

$$Y = [y_{ji}]_{M\times N}, i = 1,2,3,\cdots,M,\ \ j = 1,2,3,\cdots,N. \quad (14)$$

2) Hidden layer:

T_j is defined as $[t_{j1},...,t_{jr},...,t_{jR}]$in the hidden layer. This is the jth center of the ERBFN. An Euclidean distance $\|y_{ji} - t_{jr}\|$ is calculated between input layer's i^{th} and the rth node of the hidden layer. As a result of Eq (16), the Euclidean distance can be determined. The output of the kth hidden layer can be expressed as follows by Equation (17):

$$\|y_{ji} - t_{jr}\| = \sqrt{\sum_{i=1}^{N}(y_{ji} - t_{jr})^2} \quad (15)$$

$$V_{jr} = \varphi_{jr}(\sqrt{\sum_{i=1}^{N}(y_{ji} - t_{jr})^2}) \quad (16)$$

$$\varphi(x) = E^{\frac{-x^2}{\sigma^2}} \quad (17)$$

The Gaussian distribution function $\varphi(.)$ and the smoothing parameter σ represent the Gaussian distribution function.

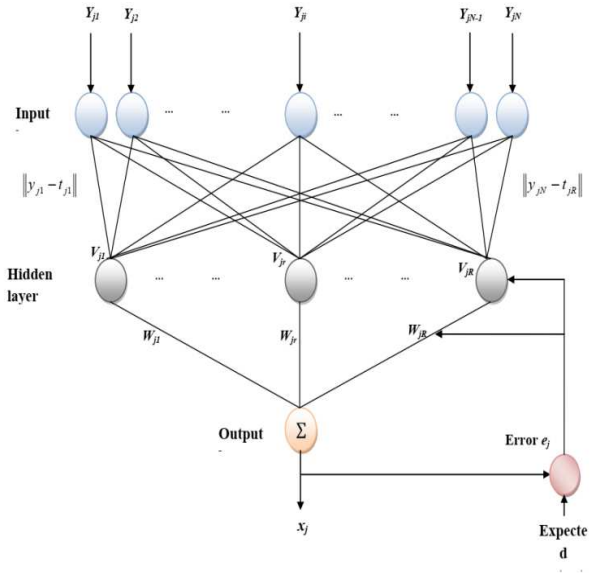


Fig. 2. An overview of the ERBFN structure based on the jth forecast.

3) Output layer: An output layer weight W_{jr} is calculated by multiplying the hidden node V_{jr} by the output node x_j.

Equation (18) describes the j^{th}output of the output layer.

$$y_j = \sum_{r=1}^{R} W_{jr} V_{jR} \quad (18)$$

IV. RESULTS AND DISCUSSION

This work proposes ERBNN-based security-aware intrusion detection. On a Windows 10 test computer with an Intel Core i5 processor and 6GB of RAM, the method is implemented in Python. The proposed work is evaluated based on accuracy, precision, recall, F-measure, specificity, detection rate, loss, error metrics, ROC curve and confusion matrix. For evaluation of the proposed method, KDDCup-99 and NSL-KDD datasets are used.

A. Metrics evaluation

An analysis of the results obtained from the presented technique is presented in this section. Here, we analyze the comparative results of the proposed and existing techniques. In order to accomplish this, we examine existing classification techniques, such as the radial basis neural network (RBNN) and the long short-term memory (LSTM).

TABLE 1. PERFORMANCE COMPARISON BETWEEN PROPOSED AND EXISTING CLASSIFICATION TECHNIQUES

Techniques	Accuracy	Precision	Recall	FMeasure
PROPOSED	97.15	97.24	97.02	97.13
RBNN	95.89	95.66	95.67	95.67
LSTM	94.06	93.91	93.82	93.86

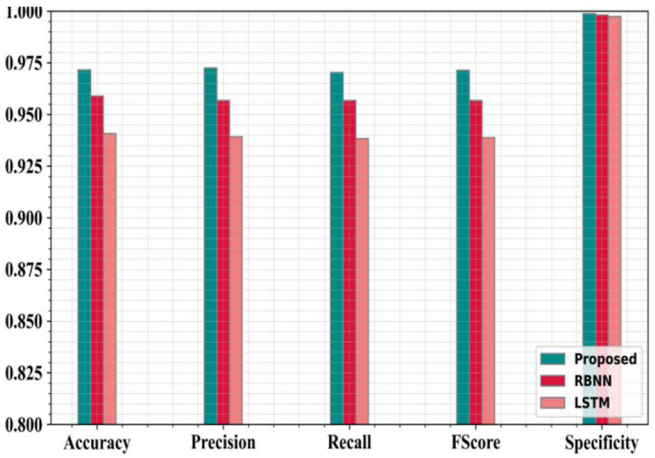


Fig. 3. Performance evaluation comparison of proposed and existing techniques

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:59:31 UTC from IEEE Xplore. Restrictions apply.

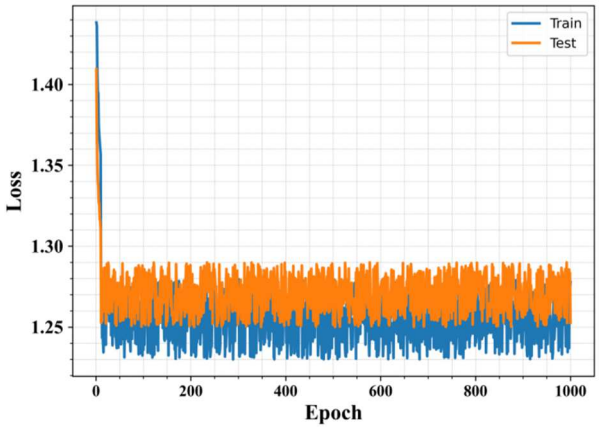


Fig. 4. Training and test validation result evaluation based on loss Vs Epoch

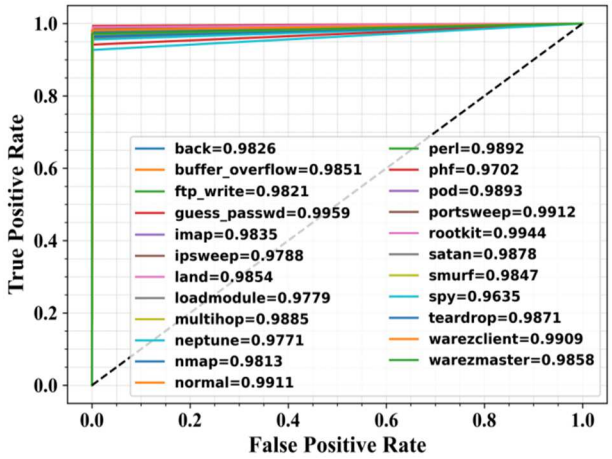


Fig. 5. The ROC curve for the proposed technique

Figure 3 shows the comparative measurement evaluation in terms of precision, accuracy, recall, F-measure and specificity. The proposed ERBNN achieves 97.15% accuracy. Our proposed ERBNN improved the accuracy by 1.26% over RBNN and 3.09% over LSTM. Our ERBNN achieves an accuracy level of 97.24%, which is 1.58% higher than RBNN and 3.33% higher than LSTM. The recall rate of the proposed technique is 97.02%, which is improved by 1.35% from RBNN and 3.2% from LSTM. The F-measure of our proposed technique is 97.13%. The proposed ERBNN improved the F-measure by 1.46% and 3.27% over the existing RBNN and LSTM techniques. Our proposed ERBNN achieves a high mark of 99.87% which is 0.06% higher than RBNN and 0.145 higher than LSTM. From Figure 4, it is demonstrated that our proposed ERBNN has the lowest test loss. Figure 5 presents the ROC curve for the proposed technique. From the above results it is proved that our proposed technique is better than the existing RBNN, LSTM.

V. CONCLUSION

The main objective of this work is to detect intrusions in a cloud-based hypervisor environment. Initially, intrusion data is collected and pre-processed. After that, an adaptive artificial flora algorithm (AAF) is used to select the optimal features to reduce computational complexity and time consumption. Then, the selected features are passed to an Enhanced Radial Based Neural Network (ERBNN) classifier to classify the data as normal or intrusive data. The performance of the proposed method is implemented in Python and the performance is analyzed based on various metrics. Our proposed task achieves 97.15%, 97.24%, 97.02%, 97.13% and 99.87% in terms of precision, accuracy, recall, F-measure and specificity, respectively.

REFERENCES

[1] A. M. Potdar, N. D G, S. Kengond, and M. M. Mulla, “Performance evaluation of docker container and virtual machine,” Procedia Computer Science, vol. 171, pp. 1419–1428, 2020.

[2] M. Mostafavi and P. Kabiri, “Detection of repetitive and irregular hypercall attacks from guest virtual machines to xen hypervisor,” Iran Journal of Computer Science, vol. 1, no. 2, pp. 89–97, Jan. 2018.

[3] A. Tchernykh, U. Schwiegelsohn, E. Talbi, and M. Babenko, “Towards understanding uncertainty in cloud computing with risks of confidentiality, integrity, and availability,” Journal of Computational Science, vol. 36, p. 100581, Sep. 2019.

[4] I. Martins et al., “Host-based ids: A review and open issues of an anomaly detection system in IOT,” Future Generation Computer Systems, vol. 133, pp. 95–113, Aug. 2022.

[5] Z. Tian et al., “A real-time correlation of host-level events in Cyber Range Service for smart campus,” IEEE Access, vol. 6, pp. 35355–35364, 2018.

[6] T. Nathiya and G. Suseendran, “An effective hybrid intrusion detection system for use in security monitoring in the virtual network layer of cloud computing technology,” Data Management, Analytics and Innovation, pp. 483–497, Sep. 2018.

[7] D. Park, S. Kim, H. Kwon, D. Shin, and D. Shin, “Host-based Intrusion Detection Model using Siamese network,” IEEE Access, vol. 9, pp. 76614–76623, 2021.

[8] M. F. Elrawy, A. I. Awad, and H. F. Hamed, “Intrusion Detection Systems for IOT-based Smart Environments: A survey,” Journal of Cloud Computing, vol. 7, no. 1, Dec. 2018.

[9] S. N. Mthunzi, E. Benkhelifa, T. Bosakowski, C. G. Guegan, and M. Barhamgi, “Cloud computing security taxonomy: From an atomistic to a holistic view,” Future Generation Computer Systems, vol. 107, pp. 620–644, Jun. 2020.

[10] B. B. Gupta and O. P. Badve, “Taxonomy of DOS and ddos attacks and desirable defense mechanism in a cloud computing environment,” Neural Computing and Applications, vol. 28, no. 12, pp. 3655–3682, Apr. 2016.

[11] Z. Inayat, A. Gani, N. B. Anuar, S. Anwar, and M. K. Khan, “Cloud-based intrusion detection and response system: Open research issues, and solutions,” Arabian Journal for Science and Engineering, vol. 42, no. 2, pp. 399–423, Jan. 2017.

[12] P. Singh and V. Ranga, “Attack and intrusion detection in cloud computing using an Ensemble Learning Approach,” International Journal of Information Technology, vol. 13, no. 2, pp. 565–571, Jan. 2021.

[13] A. Aldribi, I. Traoré, B. Moa, and O. Nwamuo, “Hypervisor-based cloud intrusion detection through online multivariate statistical change tracking,” Computers & Security, vol. 88, p. 101646, Jan. 2020.

[14] D. Jin, Y. Lu, J. Qin, Z. Cheng, and Z. Mao, “SwiftIDS: Real-time intrusion detection system based on LIGHTGBM and parallel intrusion detection mechanism,” Computers & Security, vol. 97, p. 101984, Oct. 2020.

[15] A. N. Jaber and S. U. Rehman, “FCM–SVM based Intrusion Detection System for Cloud Computing Environment,” Cluster Computing, vol. 23, no. 4, pp. 3221–3231, Mar. 2020.

[16] R. V. Mendonca et al., “Intrusion detection system based on fast hierarchical deep convolutional neural network,” IEEE Access, vol. 9, pp. 61024–61034, 2021.

[17] R. SaiSindhuTheja and G. K. Shyam, “An efficient metaheuristic algorithm based feature selection and recurrent neural network for DOS attack detection in cloud computing environment,” Applied Soft Computing, vol. 100, p. 106997, Mar. 2021.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:59:31 UTC from IEEE Xplore. Restrictions apply.

2024 3rd International Conference for Innovation in Technology (INOCON)
Karnataka, India. Mar 1-3, 2024

2024 3rd International Conference for Innovation in Technology (INOCON) | 979-8-3503-8193-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/INOCON60754.2024.10511693

Image Encryption Using Novel Chaotic System

Kriti Suneja, Bhushan Kumar, Kshitiz Jha, Lakshya Singh Chuphal
Department of Electronics and Communication
Delhi Technological University
New Delhi, Delhi
kritisuneja@dtu.ac.in, bhushankumar98178@gmail.com, jhak2001@gmail.com, lxyachuphal02@gmail.com

***Abstract* – In the modern digital world, data security has become a major concern. All kinds of data are transmitted through various channels in different forms. For safeguarding this information and maintaining privacy, encryption techniques have proved very useful. In this paper, we have designed a new 3D chaotic system to be used in image encryption algorithms. The chaotic system's random and unpredictable nature make them highly suitable for encryption techniques. We have put forward a new 3D chaotic system, calculating its Lyapunov exponents, fractal dimension and performed its bifurcation analysis for successful analysis of the proposed system. A pixel-value diffusion-based encryption algorithm, using the novel chaotic system, has been performed.**

Keywords – Chaotic system, Lyapunov exponents, Image encryption, Pseudo Random Number Generator (PRNG), Pixel diffusion.

I. INTRODUCTION

Chaotic systems are a class of dynamic system which exhibit high sensitivity to initial conditions. Even a small change in the starting conditions can cause the system to diverge drastically, resulting in totally different outcomes. Chaotic systems are deterministic and can be modelled using a set of differential equations. However, they exhibit random nature because they get highly affected by small changes in input conditions. Chaos theory is a field in mathematics that showcases the pattern of nonlinear dynamic systems, showing their highly susceptible dependence on initial conditions. Some of the applications of chaos theory are random number generator [1], unmanned aerial systems [2], DC motor [3], cryptography [4], magnetic levitation oscillation [5], memristive devices [6] and many more.

To protect data from unauthorized access, the complex and unpredictable nature of the chaotic systems is used to perform encryption. In case of digital images, the pixel value or position are altered using the chaotic system due to which it appears random. Decryption of such image without the knowledge of exact chaotic parameters becomes difficult. The sensitive dependence of chaotic system makes sure that even a minute variation in chaotic parameters result in totally different decrypted image.

We have generated a new image encryption algorithm based on pixel value diffusion using the proposed new chaotic system. Pseudo random numbers are generated using the chaotic system.

There have been numerous works on using chaotic systems for image encryption. Xing-Yuan Wang et al. [7] proposed a new image encryption algorithm based on cycle shift and chaotic system. In this technique, we generate some integers which are of the same size as is the image and use these integers to scramble the plain-text image. Xiuli Chai et al. [8] proposed a chaos-based image encryption algorithm using DNA sequence operation. In this technique, a DNA matrix is used to encode the plain image. In the second step, a wave-based permutation scheme is performed on it. Hossein Movafegh Ghadirli et al. [9] discussed various encryption algorithms for color images. Li et al. [10] proposed a new algorithm for grayscale images which utilized the concepts of bit-level scrambling and multiplication diffusion for removing the correlation between pixels.

This paper is divided into five sections. The mathematical model of the new chaotic system and its dynamic properties are described in Section 2. Section 3 presents the results of simulation of chaotic system and its bifurcation analysis. Section 4 describes the applications of image encryption process using the new chaotic system. Section 5 provides the conclusion.

II. MATHEMATICAL MODEL OF THE NEW CHAOTIC SYSTEM

The newly designed 3D chaotic system is given by the following equations:

$$\dot{x} = z + a \tag{1}$$

$$\dot{y} = -x^2 - bz \tag{2}$$

$$\dot{z} = |y| + cxy \tag{3}$$

where x, y and z are the states and a, b and c are the chaotic parameters. As can be seen from the equation (1), (2) and (3), the proposed chaotic system consists of two quadratic nonlinearities and one modulus function nonlinear term.

For the system to exhibit chaotic nature, we have taken the value of chaotic parameters as:

$$(a, b, c) = (5, 2.9, 2.6) \tag{4}$$

and initial conditions as:

$$X(0) = (x(0), y(0), z(0)) = (0.2, 0.2, 0.2) \tag{5}$$

We calculated the Lyapunov exponents using the Wolf algorithm [11]. We used the conditions mentioned in (4) and (5) and run the simulation for T = seconds. The obtained Lyapunov exponents are:

$$(L1, L2, L3) = (0.0067, 0.0009, -0.0223) \tag{6}$$

From (6), we observe that of the three Lyapunov exponents two are positive which proves that the system is indeed chaotic.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:58:55 UTC from IEEE Xplore. Restrictions apply.

The sum of the Lyapunov exponents is

$$L1 + L2 + L3 = -0.0147 \ < 0 \qquad (7)$$

The sum is less than zero indicating that the system is dissipative in nature, which again, represents its unpredictable nature.

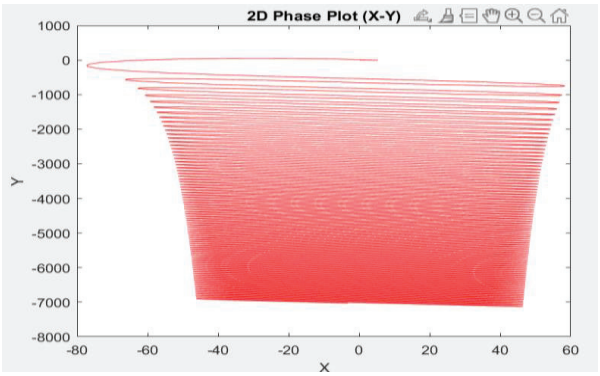
Chaotic systems exhibit fractal dimensions. The Kaplan – Yorke Fractal dimension [12] for a chaotic system is calculated as

$$D\text{xy} = 2 + \frac{L1+L2}{|L3|} \qquad (8)$$

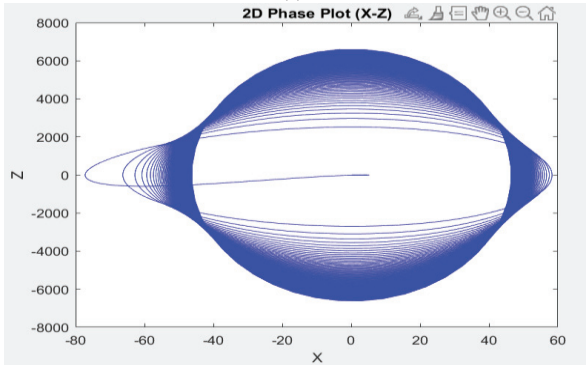
The fractal dimension of the new chaotic system comes out to be 2.3408 indicating highly chaotic nature of the system.

III. SIMULATION

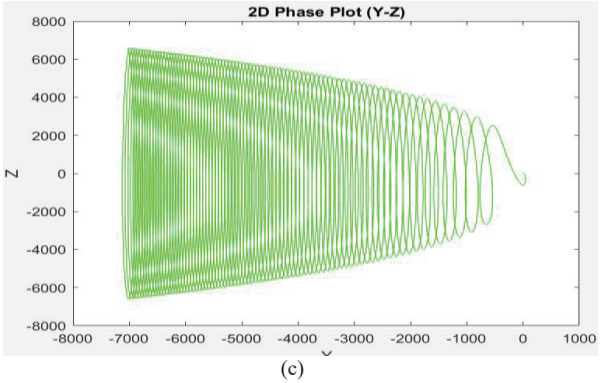
The plots obtained from the proposed chaotic system are as follows,



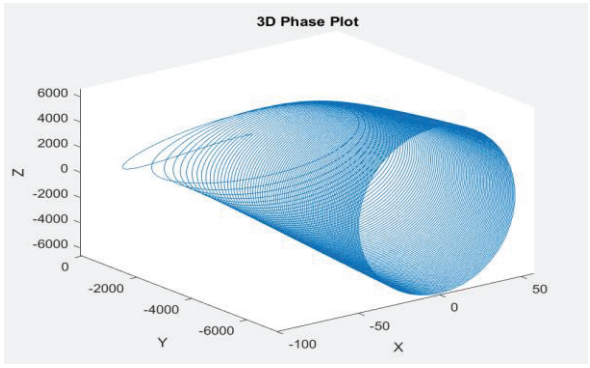
(a)



(b)



(c)



(d)

Fig. 1. (a) Phase Plot (X vs Y), (b) Phase Plot (X vs Z), (c) Phase Plot (Y vs Z), (d) Complete 3-D Phase Plot

The variation of Lyapunov exponents, can be seen in the following plot. After some initial fluctuations which was to be expected but eventually it settles down to the values mentioned previously.

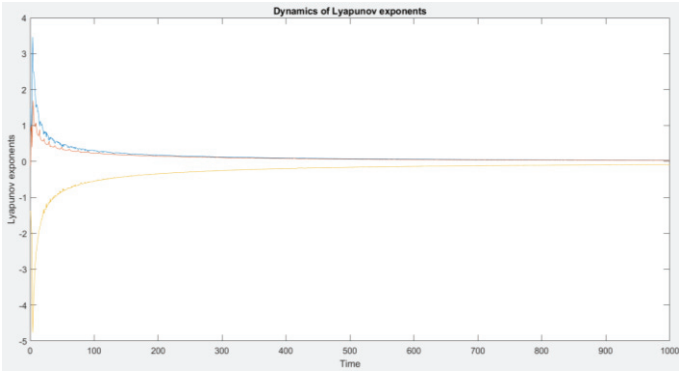
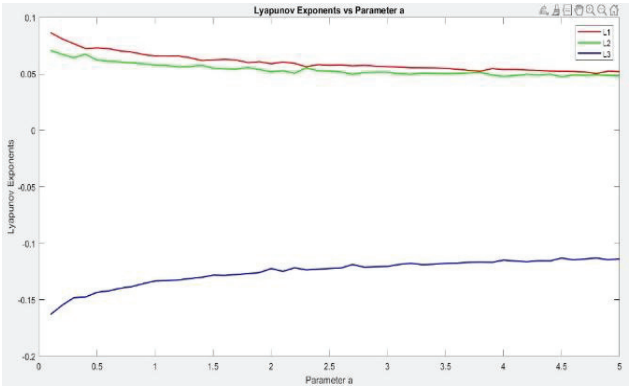


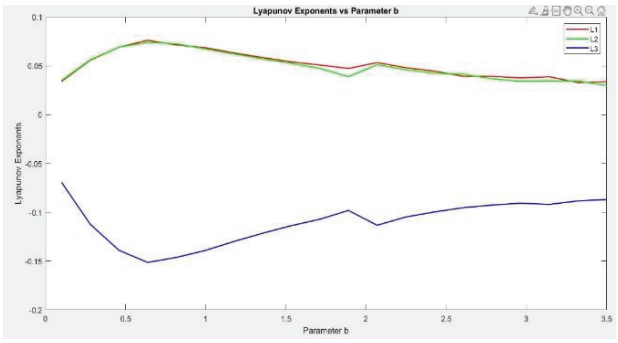
Fig. 2 Variation of Lyapunov Exponents L1, L2, L3 with time

Usually, rigid chaotic systems are not preferred. Therefore, we have used 'Bifurcation Analysis' on all the three parameters to determine the ranges in which the chaotic system preserves its chaotic and unpredictable nature.

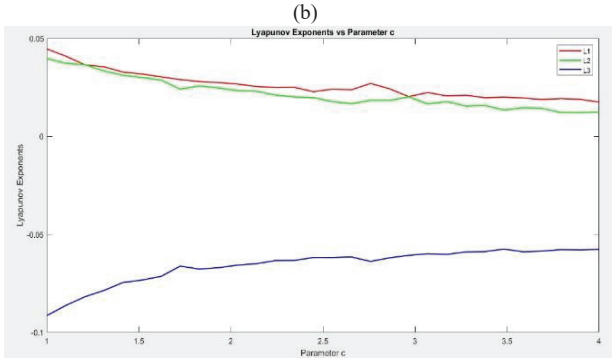


(a)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:58:55 UTC from IEEE Xplore. Restrictions apply.



(b)



(c)

Fig. 2. Bifurcation Analysis vs (a) Parameter 'a', (b) Parameter 'b', (c) Parameter 'c'

Following this, is the image encryption done using the proposed novel chaotic system and its respective decryption.

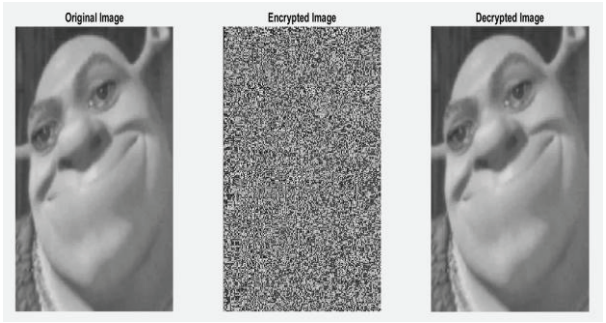


Fig. 3. Encrypted and Decrypted Image using proposed Novel Chaotic System

IV. APPLICATION

Encryption Process has been done using the Symmetric Key Block Ciphering Technique. Chaotic system's independent variable is used to make the Key for encryption. The first step includes interpretation of the image to be transmitted in the form of pixels. Following which, the data is converted into a matrix of rows and columns. In a similar manner, the Encryption Key is also reshaped to be of a matrix with same number of rows and columns as the original image matrix, so that, smooth operation of EXOR can be executed. It is to be kept in mind that this Key matrix is generated with the use of the proposed chaotic system.

EXOR operation is performed between the first column of the original image matrix taken as blocks of data and the Key matrix to get the first column of the encrypted image matrix, which in succession is then XORed with the column obtained from the EXOR operation performed between the second column of original image matrix and the second column of the Key matrix. And so on and so forth until all the columns of the encrypted matrix has been obtained according to the following equations,

$$E_1 = P_1 \oplus K_1 \tag{9}$$

$$E_2 = (P_2 \oplus K_2) \oplus R_1 \tag{10}$$

$$E_n = (P_n \oplus K_n) \oplus R_{n-1} \tag{11}$$

Decryption process can also be performed in a similar fashion with a little tuning in the equations. The same Key matrix is required to successfully decrypt the image. To get the first column of the decrypted image data, first column of the encrypted image matrix is XORed with first column of the Key matrix. Subsequently, to get the second column of the decrypted image matrix, the second column of the encrypted image matrix is XORed with its previous column and then, with the corresponding column from the key matrix. Thus, after complete decryption the original image is successfully retrieved according to the following equations,

$$D_n = E_n \oplus K_n \oplus E_{n-1} \tag{12}$$

V. CONCLUSION

Encryption performed through the proposed chaotic system showed better efficiency against unauthorized external attacks. The Lyapunov exponents L1, L2 and L3 were plotted and obtained to be 0.0067, 0.0009 and -0.0223, respectively. Also, the Fractal Dimension was evaluated to be equal to 2.3408, which was found out to be more than those of some of the more popular chaotic systems like Lorentz and Rössler Chaotic Systems. Thus, providing further unpredictability and hence, better suited for encryption techniques as a PRNG (Key Matrix).

REFERENCES

[1] F. Pareschi, G. Setti, and R. Rovatti, "Implementation and testing of highspeed CMOS true random number generators based on chaotic systems," *IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 57, no. 12, pp. 3124_3137, Dec. 2010.*

[2] H. Bi, G. Qi, J. Hu, P. Faradja, and G. Chen, "Hidden and transient chaotic attractors in the attitude system of quadrotor unmanned aerial vehicle," *Chaos, Solitons Fractals*, vol. 138, Sep. 2020*, Art. no. 109815.*

[3] Q. Wang, E. Dong, S. Du, J. Tong, H. Yu and F. Duan, "Mega-Stability and Transient Chaos Analysis of Brushless DC Motor with Winding Heat Loss," 2022 IEEE International Conference on Mechatronics and Automation (ICMA), Guilin, Guangxi, China, 2022, pp. 1801-1806, doi: 10.1109/ICMA54519.2022.9856249.

[4] A. Sambas, S. Vaidyanathan, E. Tlelo-Cuautle, B. Abd-El-Atty, A. A. A. El-Latif, O. Guille-Fernandez, Y. Hidayat, and G. Gundara, "A 3-D multi-stable system with a peanut-shaped equilibrium curve: Circuit design, FPGA realization, and an application to image encryption," *IEEE Access, vol. 8, pp. 137116137132, 2020.*

[5] M. Gao, Y. Wang, Y. Wang, and P. Wang, "Experimental investigation of non-linear multi-stable electromagnetic-induction energy harvesting mechanism by magnetic levitation oscillation," *Appl. Energy, vol. 220, pp. 856875, Jun. 2018.*

[6] C. Li, J. C. T. Wesley, H. C. I. Herbert, and T. Lu, "A memristive chaotic oscillator with increasing amplitude and frequency," *IEEE Access, vol. 6, pp. 1294512950, 2018.*

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:58:55 UTC from IEEE Xplore. Restrictions apply.

[7] Xingyuan Wang, & Gu, Sheng-Xian. (2015), "Novel image encryption algorithm based on cycle shift and chaotic system," *Optics and Lasers in Engineering. 68. 10.1016/j.optlaseng.2014.12.025.*

[8] X. Chai, Z. Gan, K. Yuan et al, "A novel image encryption scheme based on DNA sequence operations and chaotic systems," *Neural Comput & Applic 31, 219–237 (2019).*

[9] Hossein Movafegh Ghadirli, Ali Nodehi, & Rasul Enayatifar, (2019), "An overview of encryption algorithms in color images," *Signal Processing. 164. 163-185. 10.1016/j.sigpro.2019.06.010.*

[10] C.-L. Li, Y. Zhou, H.-M. Li, W. Feng, and J.-R. Du, ``Image encryption scheme with bit-level scrambling and multiplication diffusion," Multime- dia Tools Appl., vol. 80, no. 12, pp. 1847918501, May 2021.

[11] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Phys. D, Nonlinear Phenomena, vol. 16, no. 3, pp. 285317, 1985.*

[12] Z.-M. Chen, "A note on Kaplan-Yorke-type estimates on the fractal dimension of chaotic attractors", *Chaos Solitons and Fractals, vol. 3, no. 5, pp. 575–582, 1993. doi:10.1016/0960-0779(93)90007-N.*

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:58:55 UTC from IEEE Xplore. Restrictions apply.

2024 3rd International Conference for Innovation in Technology (INOCON)
Karnataka, India. Mar 1-3, 2024

Implementation of Secure Communication System using Various Chaotic Systems

S Suchit
Dept. of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
suchit981@gmail.com

Piyush Goenka
Dept. of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
piyushgoenka12@gmail.com

Deva Nand
Dept. of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
devkamboj@dce.ac.in

2024 3rd International Conference for Innovation in Technology (INOCON) | 979-8-3503-8193-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/INOCON60754.2024.10511819

***Abstract*—This paper conducts a comparative study among various chaotic systems, implementing a secure communication system. The chaotic communication system incorporates specific synchronization and masking schemes, contributing to achieving the objectives of secure communication. This comparative analysis aims to identify the optimal choice for this application, evaluating multiple crucial parameters integral to the design.The simulation for the design has been carried out in Xilinx Vivado using Verilog HDL.**

Keywords—Chaotic system, Synchronization, Masking, Xilinx Vivado, Verilog HDL.

I. INTRODUCTION

Since the advent of chaos theory, discovered by Edward Lorenz in 1963 [1], there have been a lot of discussions surrounding it and its applications. It has found its use in wide range of areas like stock markets, literary theory, music and electronics as well [2]. An example of its application in electronics is the Chua's circuit, which is a simple oscillator circuit exhibiting a variety of bifurcations and chaos [2].

Chaotic behavior, presents as irregular, aperiodic, uncorrelated, broad-band signals that defy prediction over extended periods [3]. These distinctive characteristics of chaotic systems, coupled with their sensitivity to initial conditions, makes them suitable for application in communication systems as well. This is also beneficial as the power spectrum of chaotic systems is similar to that of white noise. Several chaotic systems have been studied since the discovery of the Lorenz chaotic system.

A secure communication scheme that has been implemented, has been explored in detail in [4], and will be discussed briefly in the subsequent sections as well. Such a chaotic communication system was made possible through the self-synchronizing [5] capability of chaotic systems as well as a modulation scheme like the chaotic masking [6].

In this paper, we intend to find out the more suitable chaotic system by comparing their implementation in such a secure communication system. This comparison will be based on the synthesis results, timing information and the self-synchronizing ability of the respective chaotic systems.

To conduct this study, we have implemented the earlier mentioned design using five different chaotic systems, i.e., Lorenz, Rossler, Nose'-Hoover, Sprott-C and Rabinovich Chaotic systems [7-11].

The succeeding portions of the paper are organized as follows: Section II provides a brief exploration of the implemented design for communication and a comparison between chaotic systems. In Section III, the paper delves into the results derived from the conducted simulations, along with the conclusions drawn from these results. Lastly, Section IV explores potential avenues for future work related to the undertaken study.

II. METHODOLOGY

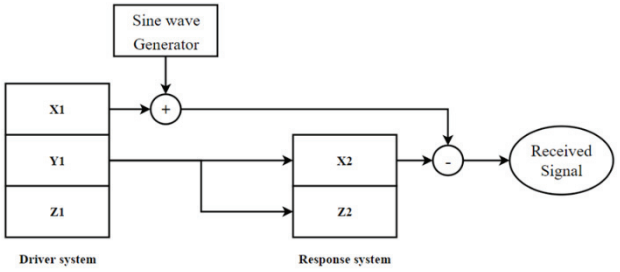


Fig. 1. Block Diagram for the Implemented Design

As shown in Fig. 1, The communication system consists of a Driver system and a Response System [4]. Both are implemented using Chaotic Systems.

The Driver system encodes the message signal, i.e., sinusoidal signal using the chaotic samples generated using the chaotic system. The Response system receives the transmitted signal and generates the same chaotic samples as the Driver to recover the original signal.

Here, the subsequent chaotic samples are generated by both Driver and Response systems by the implementation of the differential equations of the respective chaotic systems. This is done by using the Runge-Kutta 4th order method as it's an efficient method for the above purpose [12]. This approach depends on the equations that define each state within the Finite State Machine (FSM)shown in Fig 2 where, X, Y, Z represent state variables, while Fx(X, Y, Z) denotes the derivative of the state variable X. The step size, designated as h, signifies the interval between successive samples within the chaotic system. Additionally, KX1, KX2, KX3, and KX4 represent the intermediate slopes of X state variable involved in the process. Similarly, Y and Z have their intermediate slopes in the same fashion. Also, X0, Y0 and Z0 are the initial conditions.

The FSM in Fig. 2., through which the above method is being applied to generate chaotic samples in Driver and Response systems, has been implemented in Verilog HDL.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:58:01 UTC from IEEE Xplore. Restrictions apply.

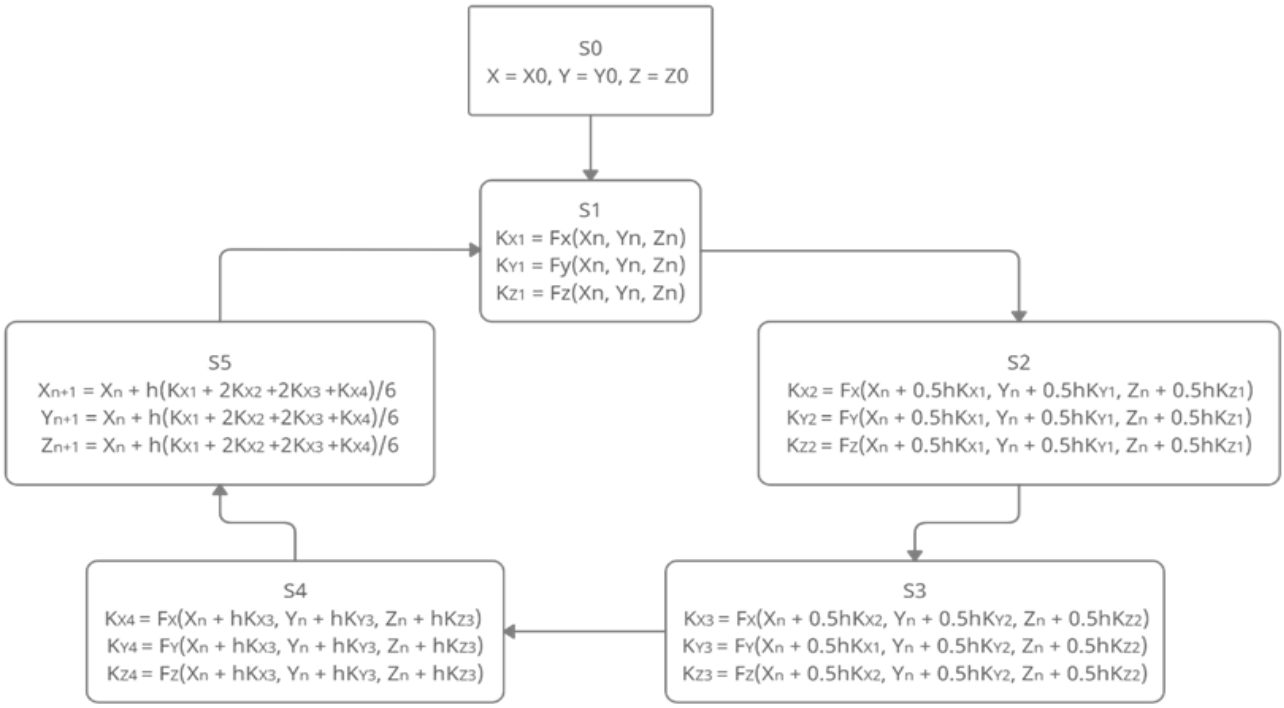


Fig. 2. Finite State Machine illustrating the implementation of a chaotic system through the RK4 method

For communication to occur, synchronization between the Driver and Response systems is essential. Both systems must be capable of generating identical chaotic samples. The synchronization method employed here is 'Complete Replacement Synchronization.' In this approach, a state variable serves as the driving signal, transmitted through the communication channel to the response system alongside the actual transmitted signal. Each sample in the driving signal replaces the corresponding state variable in the response system, ensuring it produces the same sample as the driving system.

A crucial element for ensuring a secure communication method involves implementing a modulation scheme for the encryption and decryption of a message signal. This requirement is fulfilled by employing the Chaotic Masking Scheme. In this scheme, the chaotic sample generated by the driver system is added to the sinusoidal signal intended for encryption. Subsequently, in the response system, the identical chaotic sample is subtracted from the received signal to retrieve the original sample[13]. The success of this scheme depends on the synchronization of chaotic systems at both the driver and response ends, ensuring the generated chaotic samples are identical.

III. Results and Conclusions

The design was executed by employing five distinct Chaotic Systems, each characterized by three state variables and specific variable parameters. The equations corresponding to these systems are outlined below.

1. Lorenz Chaotic System:

$$x'=a(y-x) \quad (1a)$$

$$y'=x(c-z)-z \quad (1b)$$

$$z'=xy-cz \quad (1c)$$

Where a=10, b=28, c=8/3 and initial conditions were (0,2,0).

2. Nose-Hoover Chaotic System:

$$x'=y \quad (2a)$$

$$y'=-x+yz \quad (2b)$$

$$z'=x^2-ay^2+b \quad (2c)$$

Where a=9, b=2 and initial conditions were (1,1,1).

3. Rossler system:

$$x'=-y-z \quad (3a)$$

$$y'=x+ay \quad (3b)$$

$$z'=a+z(x-c) \quad (3c)$$

Where a=0.2,b=5.7 and initial conditions were (0,0,0).

4. Spott C system:

$$x'=y \quad (4a)$$

$$y'=-x-yz \quad (4b)$$

$$x'=y2-a \quad (4c)$$

Where a=1 and initial conditions were (0,0.01,0).

5. Rabinovich system:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:58:01 UTC from IEEE Xplore. Restrictions apply.

$$x' = y(h+z) - ax \quad (5a)$$

$$y' = -x(h-z) - by \quad (5b)$$

$$z' = -cz + xy \quad (5c)$$

Where a=1.5, b=-0.3, c=1.67 and initial conditions were (5,5,5).

As noted in the preceding section, the application of the RK4 method is employed for implementing each chaotic system, with dependence on the step size 'h'. In this instance, the specific value assigned to 'h' is 0.0078125, and the design operates with a clock frequency of 20MHz.

The state variable Y serves as the driving signal to facilitate synchronization between the chaotic systems in both the drive and response systems. Additionally, samples derived from the state variable X execute chaotic masking on the message signal.

Communication between the Driver system and Response system was achieved successfully for all the mentioned chaotic systems and the simulation results of the design for the case of the Lorenz Chaotic system have been shown in Fig. 3. and Fig. 4. Here we can observe that the chaotic samples at both Driver and Response Systems have been generated successfully. As the Driver and Response chaotic samples are identical, we can conclude that both are synchronized successfully.

Further, it can be observed that the original message signal has been recovered. Hence secure communication has been achieved.

The design corresponding to each chaotic system has been synthesized on the xc7a200tffv1156-1 device using Xilinx Vivado, and the results have been compiled and compared in Table 1.

From Table 1., We can observe that the time taken to synchronize the Driver and Response system is the least in the case of the Lorenz Chaotic System but the Flip-Flops and DSPs utilized are at par. However, in the communication scheme employed, the synchronization time holds the utmost importance in establishing efficient communication. Due to such requirements of the design, as well as considering the trade-offs, the Lorenz Chaotic System is the preferred choice for such a secure communication system.

IV. Future Research Direction

The comparative analysis conducted herein, involving the deployment of a secure communication system as elucidated in preceding sections, offers valuable insights into the necessity of chaotic systems in communication., and the Lorenz Chaotic System stands out for its ability to achieve near-instantaneous synchronization, as demonstrated in previous sections.

Further work can be done to optimize the current design to efficiently utilize the hardware which may assist in lowering the delays as well.

Additional efforts can be directed towards implementing an enhanced synchronization scheme, potentially leading to improved security for the communication system. Exploring alternative masking schemes for the same purpose may also be considered.

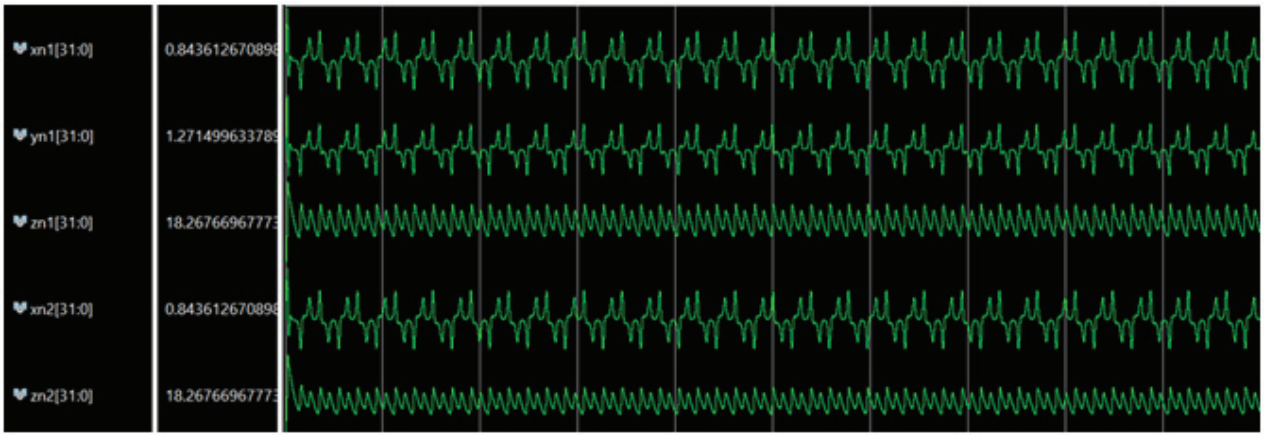


Fig. 3. Waveforms of State Variables of both Transmitter and Receiver

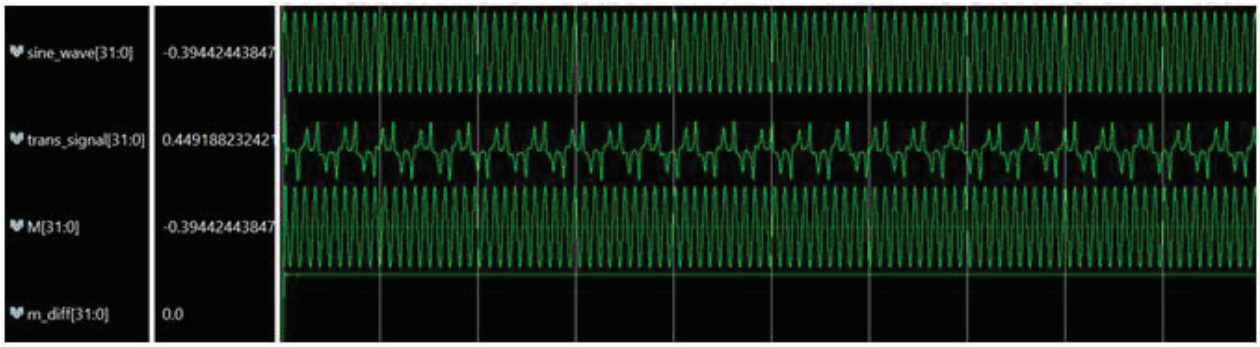


Fig. 4. Message, Transmitted, Receiver Signal and M_diff Signal

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:58:01 UTC from IEEE Xplore. Restrictions apply.

TABLE I. SIMULATION AND SYNTHESIS RESULTS

S.No.	Chaotic Systems	LUTs (in %)	FFs (in %)	DSPs (in %)	Delay (ns)	Time taken to synchronize (ms)
1.	Lorenz	9.51	0.373	23.78	50.852	0.842
2.	Nose-Hoover	7.53	0.372	26.48	55.243	123.197
3.	Rossler	8.16	0.375	20	50.497	57.850
4.	Sprott C	5.85	0.372	20.54	46.159	116.043
5.	Rabinovich	14.05	0.38	25.94	53.161	7.405

REFERENCES

[1] Lorenz, N. Edward "Deterministic Nonperiodic Flow". Journal of Atmospheric Sciences 20.2 (1963): 130-141.

[2] H.R. Biswas, M.M. Hasan and S.K. Bala. Chaos theory and its applications in our real life. *Barishal University Journal Part*, *1*(5), pp.123-140, 2018.

[3] A. Abel and W. Schwarz, "Chaos communications-principles, schemes, and system analysis," in Proceedings of the IEEE, vol. 90, no. 5, pp. 691-710, May 2002, doi: 10.1109/JPROC.2002.1015002.

[4] S. Suchit and K. Suneja, "Implementation of Secure Communication System Using Chaotic Masking," 2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT), New Delhi, India, 2022, pp. 1-5, doi: 10.1109/GlobConPT57482.2022.9938303.

[5] L. M. Pecora and T. L. Carroll, ''Synchronization in chaotic systems,'' Phys. Rev. Lett., vol. 64, pp. 821–823, Feb. 1990

[6] K. M. Cuomo and A. V. Oppenheim, ``Circuit implementation of synchronized chaos with applications to communications," *Phys. Rev. Lett.*, vol. 71, pp. 65_68, Jul. 1993. Signal masking

[7] O. E. Rossler, An equation for continuous chaos, Phys. Lett. A 57 (1976) 397-398.

[8] J. C. Sprott, Some simple chaotic flows, Phys. Rev. E 50 (1994) 647-650.

[9] İ. Pehlivan and Y. Uyaroglu, A new chaotic attractor from general Lorenz system family and its electronic experimental implementation, Turk. J. Elec. Eng. Comp. Sci. 18 (2010) 171-184

[10] A.S. Pikovski, M.I. Rabinovich, V.Y. Trakhtengerts. Onset of stochasticity in decay confinement of parametric instability, Sov. Phys. JETP, 47: 715-719, 1978.

[11] B.L. Holian,W.G. Hoover,:Numerical test of theLiouvilleequation. Phys.Rev. 34(5), 4229–4237(1986)

[12] A. Garg, B. Yadav, K. Sahu, and K. Suneja , "An FPGA based Realtime Implementation of Nosé hoover Chaotic System using different numerical Techniques," 2021 7th International Conferenceon Advanced Computing and Communication Systems (ICACCS), 2021, pp. 108-113, doi10.1109/ICACCS51430.2021.9441923.

[13] K. M. Cuomo and A. V. Oppenheim, ``Circuit implementation of synchronized chaos with applications to communications," *Phys. Rev. Lett.*, vol. 71, pp. 65_68, Jul. 1993. Signal masking

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:58:01 UTC from IEEE Xplore. Restrictions apply.

2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI) | 979-8-3503-6052-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/IATMSI60426.2024.10503150

Implicit and Explicit Sarcasm Detection via Target of Offensive Language and Cognition of Logistic Tsallis Entropy Distribution in Social Networks

Himani Pokhriyal
Department of Applied Mathematics
Delhi Technological University
Delhi, India
himpokh1698@gmail.com

Goonjan Jain
Department of Applied Mathematics
Delhi Technological University
Delhi, India
goonjan_jain@hotmail.com

***Abstract*— Social media platforms have become increasingly popular avenues for individuals to express their thoughts and emotions, especially with the advancements in computer systems and other technologies. One area of research that holds significant commercial potential is the collection and analysis of sentiments expressed in user comments, particularly in terms of their polarities. The detection, identification, and interpretation of sarcasm are of particular importance in sentiment analysis. However, automated sarcasm detection is a complex and challenging task that has not been extensively explored. This paper presents a novel approach for detecting sarcasm in social media, specifically on platforms such as Twitter and Instagram, by utilizing unsupervised mathematical optimization and probability distribution techniques. In this study, we initially employ a probabilistic distribution categorization method and a lexicon-based framework to analyze the words. Subsequently, the VADER sentiment library is utilized to ascertain the polarities of the specified sentences. By incorporating a parameter specifically designed to detect sarcasm in sentences, a sentiment score is then fed into the logistical distribution. Ultimately, we present statements that exhibit both implicit and explicit sarcasm, as well as those that contain offensive language. This proposed model serves to safeguard social media platforms against the dissemination of sarcasm, the propagation of abusive content, and instances of cyberbullying.**

Keywords—Sarcasm Detection, Logistic Distribution, Tsallis Entropy, Entropy distribution, Sentiment Analysis, Optimization.

I. INTRODUCTION

In contemporary society, individuals rely heavily on internet connectivity to access a wide range of information for personal purposes, such as obtaining relevant data, reading product reviews, acquiring knowledge, engaging in online interactions, and conducting virtual transactions. The advent of the internet has significantly enhanced the utilization of social media platforms. This is evident in the exponential growth of popular online platforms like Facebook, Twitter, LinkedIn, and others. While experts have made progress in categorizing user feedback as positive or negative, they face challenges due to the presence of sarcasm in the data, which complicates the measurement of sentiment.

Sarcasm is a rhetorical device used to mock or ridicule someone by employing a contrasting or ironic meaning [1]. Even for individuals who are adept at manipulation, identifying instances of sarcasm can be more challenging than discerning whether a situation is beneficial or detrimental [2]. The task of extracting sarcastic statements poses a complex and arduous process for computational systems. However, current techniques for sarcasm recognition have a limitation in that they primarily focus on analyzing written content responses, overlooking the valuable referential information present in comments and user data [3]. Several of the pre-defined models cannot inhibit the information contained within the text [4]. This limitation hinders the understanding of the true purpose of a sentence, as language can have a wide range of meanings. In order to address this gap in research, we have conducted a study to fill this void in knowledge. Despite the computational challenges involved, researchers are increasingly focusing on the detection of sarcasm due to its limitations in sentiment analysis [5]. This study proposes an innovative technique for detecting sarcasm in two major social media platforms, Twitter and Instagram, which are widely used for microblogging. With a large global audience, Twitter and Instagram continue to dominate the microblogging industry. This platform facilitates the sharing of ideas, perspectives, and supplementary information on a particular topic through the Twitter platform. In the present study, user attitudes and ratings expressed in comments are initially classified using the VADER sentiment analysis tool. Subsequently, the proposed approach is employed to ascertain the presence of sarcasm in statements based on the sentiment ratings and the presence of offensive language. By analyzing the tone of sarcasm and its explicit and implicit manifestations in responses, along with the occurrence of abusive language in tweets, more robust conclusions can be drawn. For the proposed study we have incorporated two datasets of online comments taken from Kaggle of Twitter and Instagram respectively.

In the course of our investigation, we have identified significant limitations in prior research endeavors and seek to address these issues through the introduction of novel methodologies. The contributions we put forth are outlined as follows:

1. The proposed algorithm, known as Logistic Distribution Learning based Sarcasm Detection using the Tsallis Entropy approach (LDL-SDTE), is a groundbreaking method that we propose. This approach leverages the lexical values and prior probabilities related to the speaker's understanding of sarcasm to effectively utilize recently acquired content. Furthermore, the incorporation of sentiment and emotion within sentences aims to enhance the classification process.
2. The suggested model aids in determining public remarks with the aim of understanding erroneous and abusive

979-8-3503-6052-3/24/$31.00 ©2024 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:00:37 UTC from IEEE Xplore. Restrictions apply.

qualities. This technique aids social media platforms in identifying harmful and abusive remarks used for digital trolling.

3. Utilizing a dynamically optimized entropy distribution strategy, we merge the sarcastic text's outcome. The result of this output combines with the algorithm's response in an appropriate and useful way to better describe implicit as well as explicit sarcasm.

This article is organized into several sections. The first section offers a concise introduction to the topic of sarcasm in sentiment analysis. The second section examines the existing literature on sentiment mining that incorporates sarcasm. The third section presents a new algorithm for detecting sarcasm. The fourth section provides a comprehensive evaluation of the proposed algorithm. Finally, the fifth section concludes the article by summarizing the key findings and offering some additional remarks.

II. Related Work

Being subjective, sarcasm recognition in text is a very difficult NLP (Natural Language Processing) topic. In a classification problem, we essentially attempt to predict whether the incoming text is sarcasm or not. Here, we discuss the primary tactics that were successfully used in order to complete this work while leveraging a range of traditional approaches with databases.

Y. Cai et al. [1] centers on the identification of sarcastic messages on Twitter. It encompasses both textual and visual elements and proposes a framework for merging hierarchical structures to serve multiple purposes. M. Bouazizi and T. O. Ohtsuki [2] introduce a classification system consisting of four categories that encompass the diverse forms of sarcasm. Additionally, they propose a pattern-based methodology (PBA) for detecting sarcasm on Twitter. H. Gregory et al. [6] aim to build upon existing research by incorporating LSTM, GRU, RoBERTa, and BERT and transformer models to investigate sarcasm in Twitter posts. Y. Liu et al. [7] aim to replicate the BERT pretraining conducted by previous studies and thoroughly assess the influence of various hyperparameters and the size of the training data. M. S. M. Prasanna et al. [8] presents an innovative method for detecting sarcasm in user reviews or opinions by analyzing text patterns at the clause level. The sentences are formulated using rules derived from a decision tree and a neuro-fuzzy system. T. Yue et al. [4] discusses the identification of the relationship between images and text and introduces a novel approach called KnowleNet. It utilizes the ConceptNet database to access relevant information and employs similarity-based semantic parallelism detection. R. A. Potamias et al. [9] present a novel approach that utilizes a transformer-based methodology, incorporating the pre-trained RoBERTa model in conjunction with a recurrent convolutional neural network, to address the challenge of analyzing figurative language in social media. K. Sundararajan and A. Palanisamy [5] gives a selection method utilizing an ensemble approach to identify the effective set of features for sarcasm detection in online comments. D. Zhang et al. [10] examine the practical applications of weakly supervised object localization and detection techniques. S. Suzuki et al. [11] present a novel approach for detecting sarcasm and emotional content in phrases, utilizing the underlying emotions conveyed by the phrases. present a novel corpus for the purpose of sarcasm detection. It includes both unbalanced and self-annotated labels. P. Goel et al. [12] seeks to address the disparity between human and machine intelligence in the identification of sarcastic behavior and patterns. It employs a combination of neural techniques, i.e., Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Baseline Convolutional Neural Networks (CNN), in an ensemble model for the purpose of detecting sarcasm on the internet. M. Abulaish et al. [13] provides an overview of figurative language, encompassing both literal and non-literal structures. It explores the incorporation of humor, simile, metaphor, sarcasm, irony, and other elements within figurative language. Additionally, it delves into the utilization of these linguistic devices within language systems. R. Misra and P. Arora [3] outline the distinctive features of dataset and conduct a comparative analysis of its various attributes in relation to other established benchmark datasets in the field of sarcasm detection. Additionally, they employ a Hybrid Neural Network architecture to gain a deeper understanding of the elements that define sarcasm within a given text. T. Saha et al. [14] presents a novel approach for multimodal classification by proposing a multitask ensemble adversarial learning framework. It incorporated, bidirectional constraints to effectively capture and integrate shared semantic relationships across different modalities. O. Vitman et al. [15] capture context features, a pre-trained transformer, and CNN it employed. It possesses emotion detection and sentiment analysis tasks.

A. Problem Definition

The goal of our study is to create a multitasking method for identifying sarcasm and its numerous subtypes on social media. Each situation in this investigation comprises linked text that is categorized using a variety of techniques, involving sarcasm detection (the primary task), sentiment analysis, emotion recognition, and further sarcasm-type categorization. Depending on context, some of the expressions within these terms are to be taken both literally and sarcastically. $\{K, K_{sarc}, K_{nsarc}, D_S, E_A, L_T\}_{i=1}^{i=m}$ here $K = \{k_1, k_2, k_3, \ldots, k_n\}$ represents the collection of sentences extracted from the dataset. K_{sarc} refers to the sarcastic text in the dataset, K_{nsarc} refers to the non-sarcastic text, D_S indicates the sentiment tag, E_A represents the class of emotion analysis, and L_T to denote the type of sarcasm, whether it is implicit or apparent in nature.

III. The Proposed Methodology

Our innovative approach, known as LDL-SDTE (Logistic Distribution Learning based Sarcasm Detection using Tsallis Entropy) approach, aims to address the gap by offering an easily accessible yet effective unsupervised mathematical optimization technique for sarcasm detection. We know in the context of sarcasm detection the authorship of a comment is known. Each individual possesses a unique communication style, shaped by personal preferences, experiences, and semantic preferences. First, we harness three key aspects: the target of offensive language using the textacy Python library, the emotional scores of user comments, through Tex2emotion Python library, and the tendencies of their sentiments, as revealed by the VADER Sentiment lexicon

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:00:37 UTC from IEEE Xplore. Restrictions apply.

tool. Our proposed algorithm seeks to extrapolate the probability of sarcasm for each user's comment, leveraging their expressed emotions and sentiment polarities.

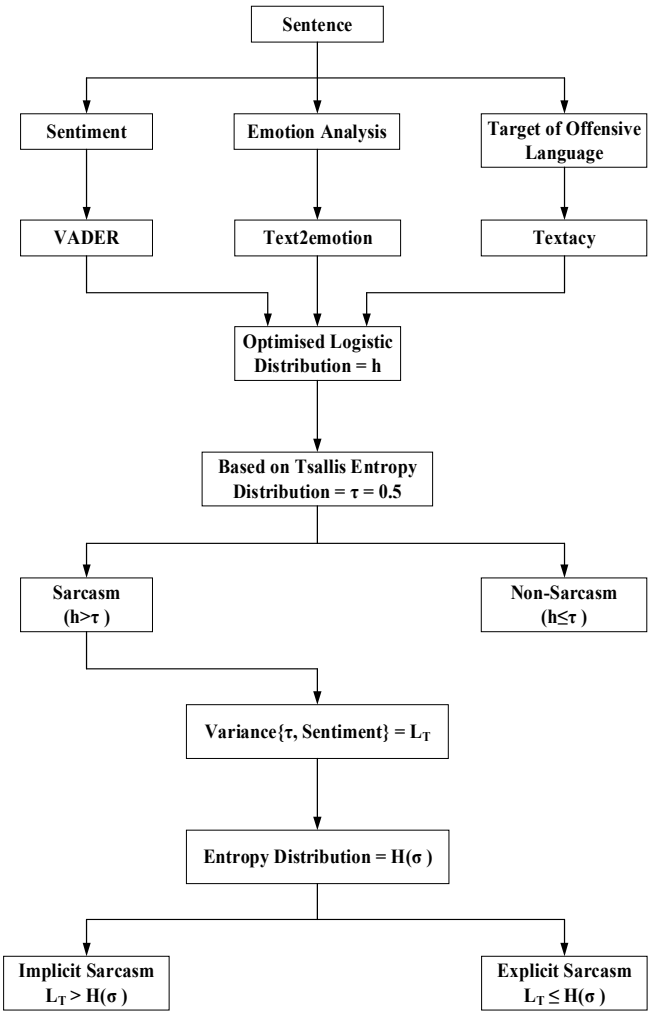


Fig. 1. Flowchart of the suggested approach

This information helps in prediction tasks related to sarcasm detection. This strategic approach aligns with the notion of treating speaker information as a form of probability calculated using Logistic distribution. It becomes an optimized probability condition, subsequently integrated into the model output with the support of Tsallis entropy and entropy distribution. Our method involves addressing a multi-classification task encompassing sarcasm, sentiment, emotion analysis, and sarcasm type. With the help of these metrics, helps in enriching the model's predictive capabilities. The schematic representation of this approach is depicted in Fig. 1.

A. Emotion analysis of text

Emotion The fundamental component of our strategy is the capture of emotions from texts. It is facilitated by the utilization of the "Tex2emotion" Python library. Using this library, we can identify emotions that are expressed in statements and provide matching ratings, enabling us to comprehend the true emotional context more thoroughly. In this first step of our system, texts are processed to provide a more precise summary of emotions. Let's refer to a series of sentences where each expression represents a distinct feeling. We do an evaluation of the emotional scores termed *emotive score analysis* denoted as δ_E, for each unique feeling. Five emotions are present in our library: surprise, fear, anger, sadness, and happiness. We derive two emotions i.e., Positive Emotion which involves an average score of $\frac{Surprise+Happy}{2}$ and Negative Emotion which involves a mean score of $\frac{Angry+Fear}{2}$, $\frac{Sad+Fear}{2}$ and $\frac{Sad+Angry}{2}$.

B. Sentiment analysis

The VADER sentiment, a sentiment lexicon analyzer relies on Natural Language Processing (NLP). Both positive and negative components are used to determine the mood of the provided text. The strength of the contrasts is also shown by the text's compound rating. The *sentiment compound score* from S_T is then imported. The suggested framework for sarcasm recognition uses the standard deviation of the compound score, designated as α_S.

C. Target of Offensive Language

Through the aid of the target of offensive language obtained using keyword extraction method i.e., textacy a Natural Language Processing (NLP) python library. The targeted offensive word is marked by the highest score value it gives to the proposed framework for sarcasm identification

D. Proposed model (LDL-SDTE) for sarcasm detection

The Logistic Distribution, which represents the likelihood distribution of a variable is autonomous of the entire distribution model. It is sinusoidal or S-shaped, with an ongoing range of possibilities. It is used in numerous applications, such as analysis of data, transport costs, and enhancing dynamics models. It has here undertaken three parameters depicted by (1).

$$h(k_i;\ \delta_E, \alpha_{S_i}) = \frac{e^{-\frac{(k-\delta_E)}{\alpha_{S_i}}}}{\alpha_{S_i}\left(1+e^{-\frac{(k-\delta_E)}{\alpha_{S_i}}}\right)^2} \quad (1)$$

Here, h is termed as the final score of the defined distribution called on the basis of which sarcasm detection is to be done. The basic parameters in this study involve k, δ_E, α_s and τ which is sentence *sentiment compound score*, average of *emotive score analysis*, standard deviation of sentiment compound score, and sarcasm *threshold value* respectively. It can be inferred that the situational intuitive emotive qualities of these factors are what detectors for sarcasm rely on. The sarcasm *threshold value* calculated is on the basis of the Tsallis entropy distribution comes up with a value of 0.5 with a fixed score.

The statement was deduced to be a sarcastic label i.e., K_{sarc} if the *final score* 'h' is greater than 0.5 and non-sarcastic if the *final score* 'h' is less than equal to 0.5. Equation (2) depicts a numerical elaboration of it.

$$K_{sarc} \rightarrow h > \tau = 0.5;\ K_{nsarc} \rightarrow h \leq \tau = 0.5 \quad (2)$$

The sarcasm *threshold value* denoted as τ in (3) is calculated on the basis of using the Tsallis entropy structure such as with $q = 3.7$ as a fixed value metric in it.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:00:37 UTC from IEEE Xplore. Restrictions apply.

$$\tau_q(K) = \frac{1}{1-q}\left(1 - \sum_i \left(P(h_i)\right)^q\right) = 0.5 \quad (3)$$

Algorithm for sarcasm detection using LDL-SDTE

Input: δ_E (average emotive analysis score), α_S (Standard deviation of sentiment compound score), k (sentiment compound score), τ (sarcasm threshold score)

Output: sarcasm (K_{sarc}) and non-sarcasm (K_{nsarc})

1. Calculate the sentiment and compound score of sentences K using the VADER sentiment library
2. Using Tex2emotion library to obtain emotion analysis score
3. Calculate the Logistic distribution based on the given parameters
   p1 = $e^{-\frac{(k-\delta_E)}{\alpha_S}}$
   p2 = $\left(1 + e^{-\frac{(k-\delta_E)}{\alpha_S}}\right)^2$
   p3 = α_S
4. Define h :
   $h = \frac{p1}{p2 \times p3}$
   return h
5. Calculate the Logistic distribution score
   h = calculate_logistic_distribution (k; δ_E, α_S)
6. Compare the calculated score with the threshold τ to determine sarcasm
   if $h > \tau$:
   return "K_{sarc}" (sarcasm)
   else:
   return "K_{nsarc}" (non-sarcasm)
   End

where K resembles the total number of sentences over the dataset and h_i denoted as normalized average sentiment sarcasm *final score*. The Algorithm gives the perfect visualization of the proposed methodology of sarcasm detection with the help of sentiment, emotion analysis, and target of offensive words in the text within it.

E. Implicit and explicit sarcasm analysis

The analysis of sarcastic sentences is further assessed to give its nature. Implicit sarcasm deals with an indirect kind of intended mockery. Explicit sarcasm deals with direct implied intended mockery of another person. Applying the statistical measure i.e., calculating variance of final score and average of positive and negative sentiments of each sentence. Another defined parameter is the sarcasm type threshold value denoted as σ whose value is fixed at 0.7. This is calculated using another entropy distribution system as depicted in (4). Here, $p(\sigma_i)$ is the conditional probability of the variance score. Based on this structure if $L_T(\sigma) > H(\sigma) \rightarrow$ Implicit sarcasm else it is explicit sarcasm.

$$H(\sigma) = -\left(\sum_i \left(p(\sigma_i)\log\left(p(\sigma_i)\right)\right)\right) = 0.7 \quad (4)$$

IV. Results and Discussion

The proposed model has been compared with the existing state-of-the-art methods on Dataset-1 and 2 respectively. These methods are Bi-LSTM [1], Pattern Based Approach (PBA) [2], RoBERTa+XLNet [6], BERT [7], Neuro-Fuzzy Set (NFS) [8], KnowleNet [4], RCNN+RoBERTa [9] and Ensemble Neuro Fuzzy Method (ENFM) [5] are used to analyze performance evaluation on the two datasets.

A. Dataset Description

The authors created and assessed two datasets of online comments from Twitter [16], News headlines [3], and Reddit [17] datasets for sarcasm detection and other multitasking functions. Dataset-1 involves an aggregate of 7,776 comments, with 3,618 sarcastic and 4,158 non-sarcastic remarks. Dataset-2 encompasses 6,237 records, with 2,698 sarcastic and 3,539 non-sarcastic headlines.

B. Performance Indicators

In order to evaluate the effectiveness of our model's use of the confusion matrix, which includes True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values, while other indicators utilize such as Accuracy (A), Precision (P), Recall (R), and F1 score (F1), refer to (5)–(9). It acts as the entity's primary criteria for tasks associated with classification.

$$A = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$R = \frac{TP}{TP+FN} \quad (6)$$

$$P = \frac{TP}{TP+FP} \quad (7)$$

$$F1 = \frac{2 \times P \times R}{P+R} \quad (8)$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(\mathrm{TP+FP}) \times (\mathrm{TP+FN}) \times (\mathrm{TN+FP}) \times (\mathrm{TN+FN})}} \quad (9)$$

For Dataset-1 and 2, the simulation model received an accuracy score of 95.57 and 95.23, respectively. These findings outperform those of earlier research on recognizing sarcasm. The analysis table also compares the Matthews Correlation Coefficient (MCC) on Dataset-1 and Dataset-2, as well as accuracy, precision, recall, and F1 score. On two different datasets, we evaluated our proposed approach alongside other predetermined models, and we thoroughly compared our results to state-of-the-art methods. The results shown in TABLE I and II convincingly show how well our suggested strategy performs. Particularly, with rates of 91.44 and 90.63, respectively, the correctness of Matthews Correlation Coefficient (MCC) scores demonstrates considerable gains over the two sets of data.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:00:37 UTC from IEEE Xplore. Restrictions apply.

TABLE I. COMPARISON OF PROPOSED MODEL (LDL-SDTE) WITH OTHER PRE-DEFINED MODELS ON THE DATASET-1

Methods	A	P	R	F1	MCC
BiLSTM [1]	0.8564	0.8096	0.8321	0.8045	0.7812
PBD [2]	0.7345	0.7891	0.7471	0.7012	0.6992
RoBERTa+XLNet[6]	0.9012	0.8876	0.8913	0.9002	0.8362
BERT [7]	0.9156	0.9013	0.8873	0.8991	0.8451
NFS [8]	0.7012	0.6982	0.7102	0.7031	0.6723
KnoweleNet [4]	0.8745	0.852	0.8824	0.879	0.8294
RCNN+RoBERTa[9]	0.9125	0.913	0.9724	0.9025	0.8991
ENFM [5]	0.6924	0.6813	0.6634	0.7012	0.6014
LDL-SDTE	**0.9557**	**0.9171**	**0.9947**	**0.9543**	**0.9144**

TABLE II. COMPARISON OF PROPOSED MODEL (LDL-SDTE) WITH OTHER PRE-DEFINED MODELS ON THE DATASET-2

Methods	A	P	R	F1	MCC
BiLSTM [1]	0.8066	0.7596	0.7821	0.7544	0.7302
PBD [2]	0.6971	0.7411	0.7002	0.6492	0.6556
RoBERTa+XLNet[6]	0.9122	0.8997	0.9036	0.9128	0.8479
BERT [7]	0.8954	0.8964	0.876	0.8829	0.8347
NFS [8]	0.7468	0.7338	0.7551	0.7483	0.7163
KnoweleNet [4]	0.8832	0.8616	0.8991	0.8826	0.8382
RCNN+RoBERTa[9]	0.9099	0.9037	0.9874	0.8988	0.8806
ENFM [5]	0.6778	0.6636	0.6482	0.8824	0.5883
LDL-SDTE	**0.9523**	**0.9107**	**0.9866**	**0.9471**	**0.9063**

TABLE III. CONFUSION MATRIX OF SARCASM DETECTION ON DATASET-1 AND 2 RESPECTIVELY

Proposed Method	No. of sentences	TP	TN	FP	FN
Identifying sarcasm on Dataset-1	7776	3599	3833	325	19
Identifying sarcasm on Dataset-2	6237	2662	3278	261	36

When our proposed algorithm is compared to BERT and transformer-based methods like RCNN+RoBERTa and KnowleNet, its effectiveness becomes very clear. Our strategy successfully beats existing well-known techniques across a range of measures. Our paradigm is strengthened even more by the inclusion of a speaker vocabulary as well as sentiment and emotion embeddings. Notably, none of these improvements result in a negative rise in processing time. The revised situation best highlights our advantages. In essence, our work revolves around the proposition of an unsupervised mathematical optimization methodology for textual sarcasm detection. The confusion matrix for Dataset-1 and 2 is shown in TABLE III. This confusion matrix lists the samples of sarcasm and non-sarcasm that are present in datasets. These improvements substantially enhance the functionality of current models while having little to no effect on computing speed.

A linear trendline for the suggested framework on Dataset-1 and 2 is shown in Fig. 2 and 3 as the best-fit straight line for the linear analysis of the data. A straight trendline shows that the accuracy, precision, recall, and F1 score for the proposed technique on both datasets are all increasing gradually. Across two datasets with accuracy scores of 0.9557 and 0.9523, respectively, it consistently rose over the prior level. This shows the ability of the dashed line's fit to the data is good given the monotonic increase in the methodology's initial settings.

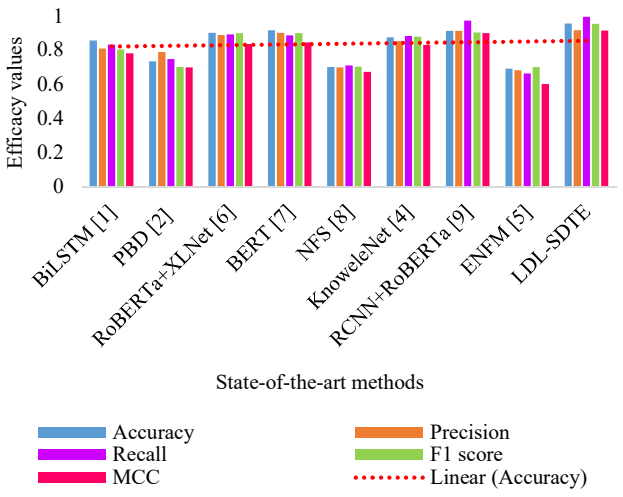


Fig. 2. Comparison graph of proposed model (LDL-SDTE) with the predefined models on the Dataset-1

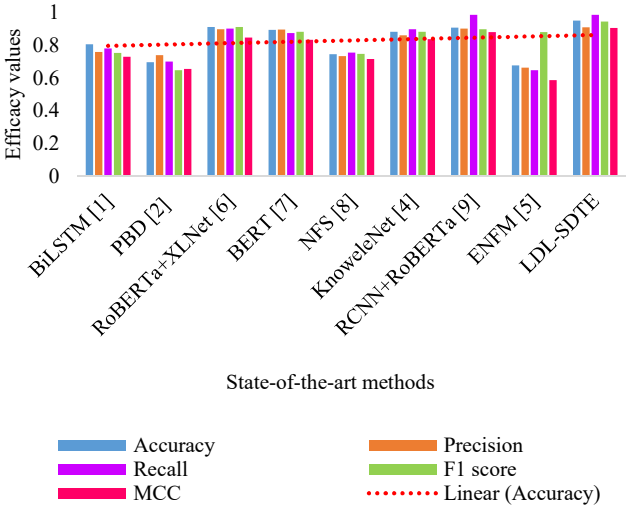


Fig. 3. Comparison graph of proposed model (LDL-SDTE) with the predefined models on the Dataset-2

The statistical validation of the proposed model is confirmed by consistent accuracies observed across a sample of datasets. The validity of the results is assessed using a z-test proportion, as depicted in Fig. 4. This indicates that the proposed model, LDL-SDTE, consistently achieves accurate results when applied to various datasets. The time complexity of the algorithm can be approximated as $O(n)$, where 'n' represents the number of sentences being processed for sentiment and emotion analysis. Concluding, we affirm that the time complexity of the proposed model primarily depends on the number of sentences, 'n' that are being analyzed for sentiment and emotion helps in sarcasm detection.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:00:37 UTC from IEEE Xplore. Restrictions apply.

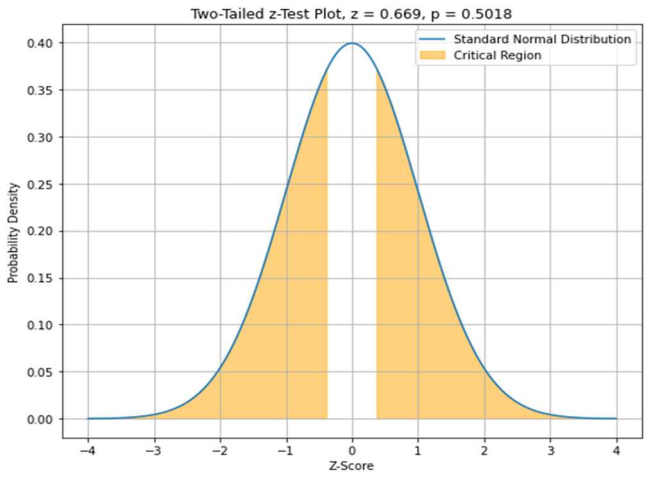


Fig. 4. Statistical z-test for the proposed model LDL-SDTE

V. CONCLUSION AND FUTURE WORK

This paper proposed a novel algorithm, called LDL-SDTE, which is designed to detect sentiment, emotion, sarcasm, and offensive language in online comments on social networking sites. The model is formulated to identify and categorize instances of subjective emotional sarcasm that include foul language it. To improve sarcasm detection performance, the proposed algorithm involves optimized mathematical formulations. The results of the proposed model LDL-SDTE demonstrate the superiority of the algorithm compared to other existing models. This highlights its enhanced detection capabilities and nuanced categorization. For future research in the enhancement of the proposed model performance, we explore the integration of word embeddings and the use of ensemble language models. These defined refinements have the potential to improve the algorithm's effectiveness and contribute to its ongoing development for future applications.

REFERENCES

[1] Y. Cai, H. Cai, and X. Wan, "Multi-modal sarcasm detection in twitter with hierarchical fusion model," in *Proceedings of the 57th annual meeting of the association for computational linguistics*, 2019, pp. 2506–2515.

[2] M. Bouazizi and T. O. Ohtsuki, "A pattern-based approach for sarcasm detection on twitter," *IEEE Access*, vol. 4, pp. 5477–5488, 2016.

[3] R. Misra and P. Arora, "Sarcasm detection using news headlines dataset," *AI Open*, vol. 4, pp. 13–18, 2023.

[4] T. Yue, R. Mao, H. Wang, Z. Hu, and E. Cambria, "KnowleNet: Knowledge fusion network for multimodal sarcasm detection," *Inf. Fusion*, vol. 100, p. 101921, 2023.

[5] K. Sundararajan and A. Palanisamy, "Multi-rule based ensemble feature selection model for sarcasm type detection in twitter," *Comput. Intell. Neurosci.*, vol. 2020, 2020.

[6] H. Gregory, S. Li, P. Mohammadi, N. Tarn, R. Draelos, and C. Rudin, "A transformer approach to contextual sarcasm detection in twitter," in *Proceedings of the second workshop on figurative language processing*, 2020, pp. 270–275.

[7] Y. Liu *et al.*, "Roberta: A robustly optimized bert pretraining approach," *arXiv Prepr. arXiv1907.11692*, 2019.

[8] M. S. M. Prasanna, S. G. Shaila, and A. Vadivel, "Polarity classification on twitter data for classifying sarcasm using clause pattern for sentiment analysis," *Multimed. Tools Appl.*, pp. 1–37, 2023.

[9] R. A. Potamias, G. Siolas, and A.-G. Stafylopatis, "A transformer-based approach to irony and sarcasm detection," *Neural Comput. Appl.*, vol. 32, no. 23, pp. 17309–17320, 2020.

[10] D. Zhang, J. Han, G. Cheng, and M.-H. Yang, "Weakly supervised object localization and detection: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 5866–5885, 2021.

[11] S. Suzuki, R. Orihara, Y. Sei, Y. Tahara, and A. Ohsuga, "Sarcasm Detection Method to Improve Review Analysis.," in *ICAART (2)*, 2017, pp. 519–526.

[12] P. Goel, R. Jain, A. Nayyar, S. Singhal, and M. Srivastava, "Sarcasm detection using deep learning and ensemble learning," *Multimed. Tools Appl.*, vol. 81, no. 30, pp. 43229–43252, 2022.

[13] M. Abulaish, A. Kamal, and M. J. Zaki, "A survey of figurative language and its computational detection in online social networks," *ACM Trans. Web*, vol. 14, no. 1, pp. 1–52, 2020.

[14] T. Saha, A. Upadhyaya, S. Saha, and P. Bhattacharyya, "A multitask multimodal ensemble model for sentiment-and emotion-aided tweet act classification," *IEEE Trans. Comput. Soc. Syst.*, vol. 9, no. 2, pp. 508–517, 2021.

[15] O. Vitman, Y. Kostiuk, G. Sidorov, and A. Gelbukh, "Sarcasm detection framework using context, emotion and sentiment features," *Expert Syst. Appl.*, p. 121068, 2023.

[16] A. Ghosh and T. Veale, "Fracking sarcasm using neural network," in *Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis*, 2016, pp. 161–169.

[17] T. Dadu and K. Pant, "Sarcasm detection using context separators in online discourse," in *Proceedings of the Second Workshop on Figurative Language Processing*, 2020, pp. 51–55.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:00:37 UTC from IEEE Xplore. Restrictions apply.

Improved Performance of Oppositely Doped Core-Shell Junctionless Nanowire FET Using Gate Material Engineering

Bharti
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
bhartigndit@gmail.com

Poornima Mittal
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
poornimamittal@dtu.ac.in

2024 11th International Conference on Signal Processing and Integrated Networks (SPIN) | 979-8-3503-0843-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/SPIN60856.2024.10511527

***Abstract*— The impact of work function variation on Dual Material Gate Oppositely Doped Core-Shell Junctionless Nanowire FET (DMG-ODCS-JLNWFET) is explored in this paper. The variation in surface potential, electric field, band energy, drain current (I_D), leakage current (I_{OFF}), subthreshold slope (SS), transconductance (g_m), and switching ratio (I_{ON}/I_{OFF}) are analyzed. The performance of DMG-ODCS-JLNWFET is also compared with Single Material Gate (SMG) ODCS-JLNWFET. The results demonstrate that the proposed device shows significant advancements, boasting improvements of 18.9%, 73.7%, 26.8%, and 10.8% in I_D, I_{OFF}, g_m, and SS respectively when compared to SMG-ODCS-JLNWFET. These noteworthy enhancements underscore the device's exceptional suitability for both analog and digital applications within the confines of nano-scale dimensions.**

Keywords— Core-Shell (CS), Dual material Gate (DMG), Nanowire FET (NWFET), Work function.

I. INTRODUCTION

The notable Short Channel Effects (SCEs) observed at scaled dimensions are a consequence of ineffective gate control and physical restrictions in achieving ultra-scaled dimensions. This has surpassed the capabilities of traditional planar Field Effect Transistors (FET), prompting researchers to redirect their attention towards multiple-gate 3-D topologies, particularly Nanowires FET (NWFET), as the prospective foundation for technologies [1]. However, it faces major hurdles in achieving an ultra-sharp doping gradient at the junction during fabrication [2]. Considerable research efforts have been dedicated to exploring flexible organic devices, primarily due to their cost-effective fabrication process and their capacity to operate effectively at lower temperatures. Despite these advantages, the reliability of organic devices poses a notable challenge, making them less suitable for applications requiring high-speed analog and digital functionalities [3], [4].

Junctionless Nanowire FET (JLNWFET) has emerged as a noteworthy alternative exhibiting a low thermal budget during the manufacturing process due to a lack of p-n junctions [5]. Furthermore, to achieve better electrostatic control of gate voltage over the channel, Junctionless Transistors (JLTs) have also embraced cylindrical architecture [6]. Despite several advantages, the high doping concentration inherent in JLNWFET introduces challenges, notably resulting in reduced carrier mobility, leading to low ON-state current (I_{ON}) and transconductance. Additionally, insufficient volume depletion and electron tunneling during OFF-state contribute to an increase in leakage current. The incorporation of an oppositely doped core within the Core-Shell (CS) architecture of JLNWFET helps in volume depletion in the OFF state which tends to reduce I_{OFF} [7], [8]. In 2014, Lee *et al.* [9] demonstrated the feasibility of CS-JLNWFET marking the first experimental validation of the innovative approach. Previously, multiple researchers have employed Dual Material Gate (DMG) to improve carrier mobility and mitigate the impact of SCEs by altering the channel potential and electric field distributions as compared to Single Material Gate (SMG) devices [10]-[13].

The DMG architecture comprises two gate materials, M_1 and M_2, characterized by distinct work functions Φ_{M1} and Φ_{M2}, with the condition $\Phi_{M1} > \Phi_{M2}$ [10]. To harness the benefits of DMG in conjunction with an oppositely doped core-shell structure, a novel device structure named Dual Material Gate Oppositely Doped Core-Shell (DMG-ODCS) JLNWFET is proposed in this paper. Moreover, the effect of workfunction variation of M_2 (Φ_{M2}) by keeping Φ_{M1} constant, on the performance of the DMG-ODCS-NWFET device. One challenge associated with implementing work function variation in practical device fabrication is the reliance on the DMG approach. This method utilizes a single Reactive Ion Etching (RIE) process to pattern dissimilar gate metals simultaneously, ensuring high selectivity to the ultrathin gate dielectric. An alternative method entails leveraging molybdenum (Mo) as a commonly utilized gate material, with its work function susceptible to adjustment through careful control of nitrogen implantation within the Mo substrate, offering a potential avenue for achieving desired work function variations in devices [14], [15].

This paper is systematized into four sections, commencing with Section I, which serves as the introduction. In Section II, a detailed account of the device structure parameters is presented. Section III is dedicated to presenting the results and facilitating a comprehensive discussion, which includes a comparative analysis between the proposed device and SMG-ODCS-JLNWFET. The final section, Section IV, encapsulates the conclusion drawn from the analysis and findings presented throughout the paper.

II. DEVICE STRUCTURE

The devices are simulated using the SILVACO TCAD simulator [16]. A 3-D view of the proposed DMG-ODCS-JLNWFET device is presented in Fig. 1 (a) whereas its 2-D schematic view is illustrated in Fig. 1 (b). Additionally, a 2-D cross-sectional view of SMG-ODCS-JLNWFET is depicted in Fig. 2. Table I provides the structural parameters of both devices. These parameters are chosen as per experimental work [17]. The channel thickness (t_{si}) of DMG-ODCS-JLNWFET is given as [7]

$$t_{si} = (2 \text{ X } t_{shell}) + t_{core} \quad (1)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:59:42 UTC from IEEE Xplore. Restrictions apply.

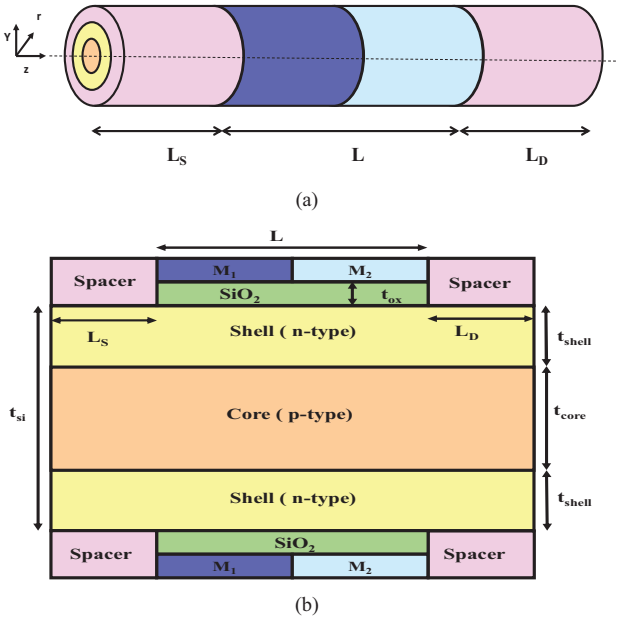


Fig. 1. (a) 3-D and (b) 2-D view of DMG-ODCS-JLNWFET.

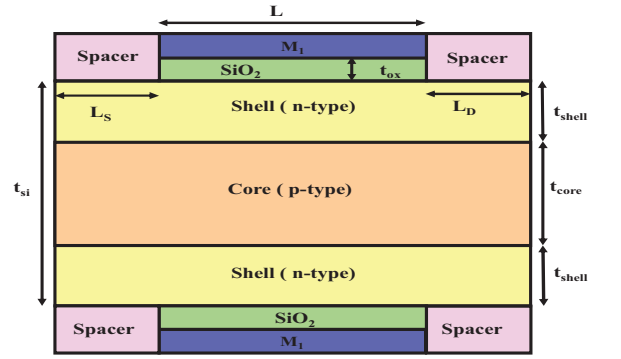


Fig. 2. 2-D cross-sectional view of SMG-ODCS-JLNWFET.

TABLE I. DEVICE PARAMETERS

Parameters	SMG-ODCS-JLNWFET	DMG-ODCS-JLNWFET	
Channel Length, L (nm)	60	L_{M1}	L_{M2}
		30	30
Channel thickness, t_{si} (nm)	20	20	
Length of S/D, L_S/L_D(nm)	1	1	
Gate work function (eV)	5.1	M_1	M_2
		5.1	4.2 - 5.0
Channel doping, N_D (cm^{-3})	1 X 10^{19}	1 X 10^{19}	
Core doping N_A, (cm^{-3})	1 X 10^{19}	1 X 10^{19}	
Gate oxide thickness, t_{ox} (nm)	25	25	

The HfO_2 serves as the spacer material. The default ATLAS TCAD models are calibrated with the experimental results documented in [18] for V_{DS} = 1 V. Fig. 3 visually represents the corresponding outcomes. These models incorporate carrier mobility models that are contingent on both the electric field and carrier concentration.

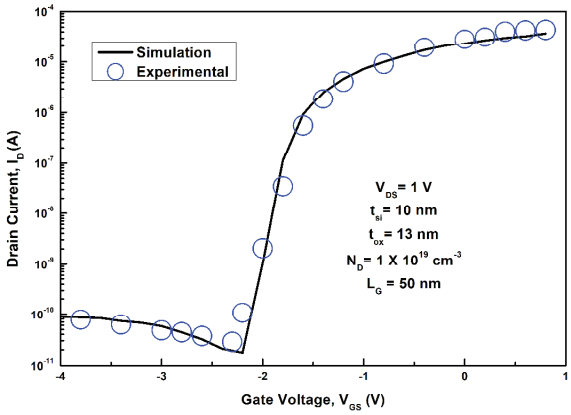


Fig. 3. Simulation models calibration with the experimental work reported in [18].

The non-local BTBT model is employed to account for electron tunneling. Additionally, the Lombardi CVT model and BGN model are activated. Gummel and Newton methods are employed to derive numerical solutions. To address recombination effects, SRH and Auger recombination models are accommodated.

III. RESULTS AND DISCUSSION

To investigate the influence of work function variation on the performance of DMG-ODCS-JLNWFET, Φ_{M1} is kept at 5.1 eV and Φ_{M2} is varied from 4.2 eV to 5.0 eV. The corresponding results are compared with SMG-ODCS-JLNWFET. The illustration of surface potential, electric field and energy band diagram is implemented by taking a horizontal cutline at 1 nm below the oxide-channel interface in the SILVACO TCAD simulator. The potential of DMG-ODCS-JLNWFET with varied Φ_{M2} and SMG-ODCS-JLNWFET along device length is demonstrated in Fig. 4. In the DMG-ODCS-JLNWFET, a noticeable enhancement in the transition at the interface of two gate materials becomes more pronounced with a decrease in Φ_{M2}. This abrupt change is caused by gate work function difference ($\delta\Phi = \Phi_{M1} - \Phi_{M2}$). The electric profile of DMG-ODCS-JLNWFET displays an additional peak, attributed to the abrupt transition in its potential profile, as depicted in Fig. 5.

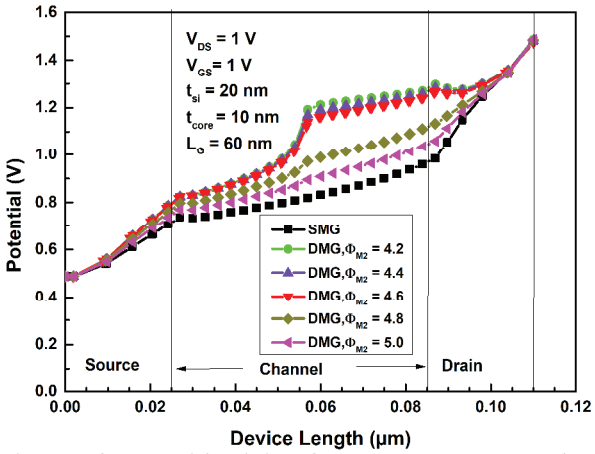


Fig. 4. Surface potential variation of DMG-ODCS-JLNWFET and SMG-ODCS-JLNWFET along device length.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:59:42 UTC from IEEE Xplore. Restrictions apply.

Due to the potential abruptness in Fig 4, the DMG-ODCS-JLNWFET exhibits two electric field peaks at the M1 and M2 interfaces. These supplementary electric field peaks, situated near the source region, accelerate the movement of electrons within the channel, consequently boosting the drain current. As Φ_{M2} is increased up to 4.6 eV in the DMG-ODCS-JLNWFET, the electric peak near the source region experiences a rise, influencing the carrier mobility within the channel [10]. Beyond this point, the increase becomes marginal. Furthermore, the electric peak degrades with the increase of Φ_{M2} near the channel-drain region which tends to decrease the tunneling of electrons during the OFF state.

Fig 6. exhibits the influence of work function variation on the energy band diagram of DMG-ODCS-JLNWFET. The observation indicates that an increase in Φ_{M2} results in the bending of the Conduction Band (CB) and Valence Band (VB). This pronounced band bending has enhanced the tunneled width, thereby significantly reducing electron tunneling at the interface between the channel and drain. The transfer characteristic of DMG-ODCS-JLNWFET with varying Φ_{M2} and SMG-ODCS-JLNWFET is exhibited in Fig.7. It is observed that drain current (I_D) increases as Φ_{M2} decreases till 4.6 eV beyond which there is marginal enhancement due to improved carrier mobility in the channel as explained in Fig. 5.

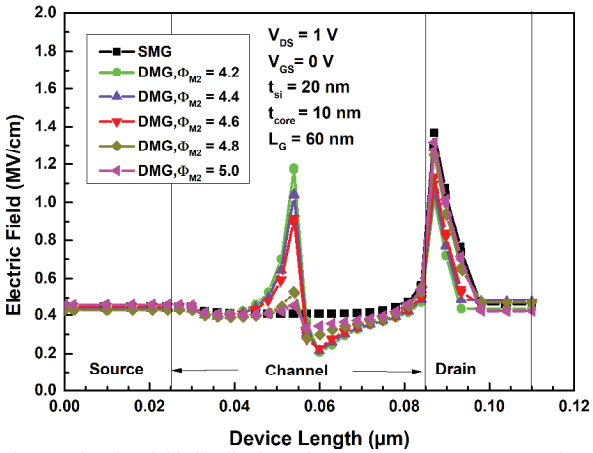


Fig. 5. Electric Field distribution of DMG-ODCS-JLNWFET and SMG-ODCS-JLNWFET along device length.

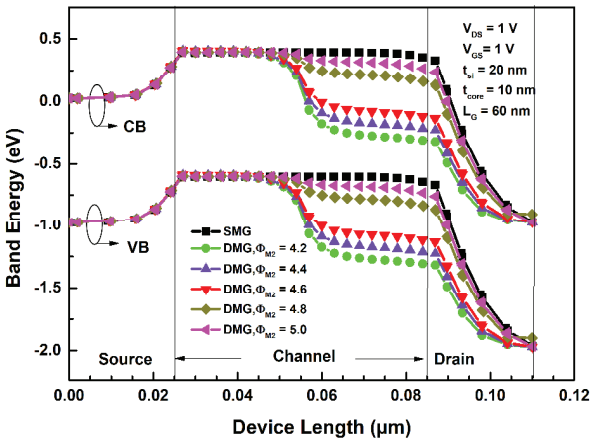


Fig. 6. Band energy profile of DMG-ODCS-JLNWFET and SMG-ODCS-JLNWFET along device length.

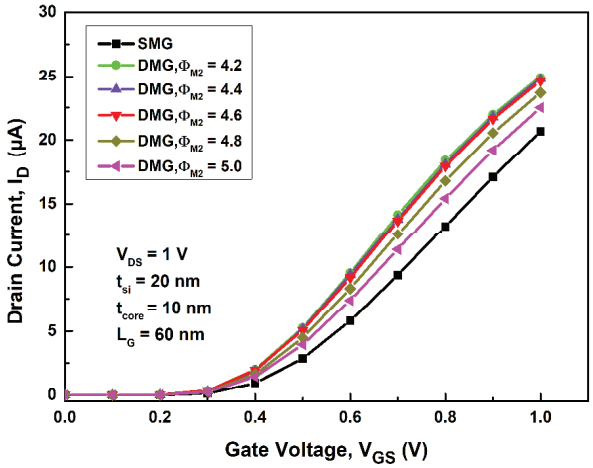


Fig. 7. Transfer characteristics of DMG-ODCS-JLNWFET and SMG-ODCS-JLNWFET with variation of Φ_{M2}.

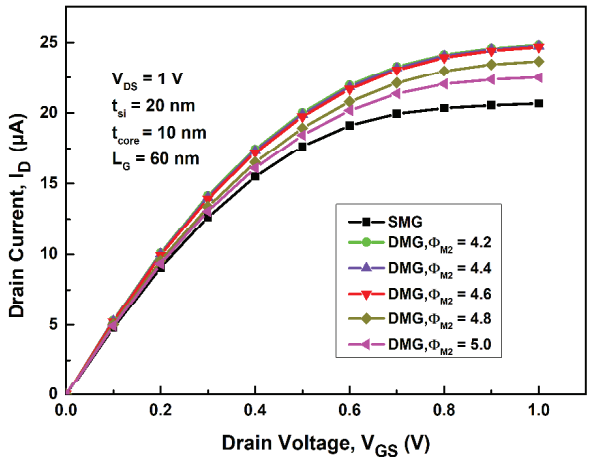


Fig. 8. Output characteristics of DMG-ODCS-JLNWFET and SMG-ODCS-JLNWFET.

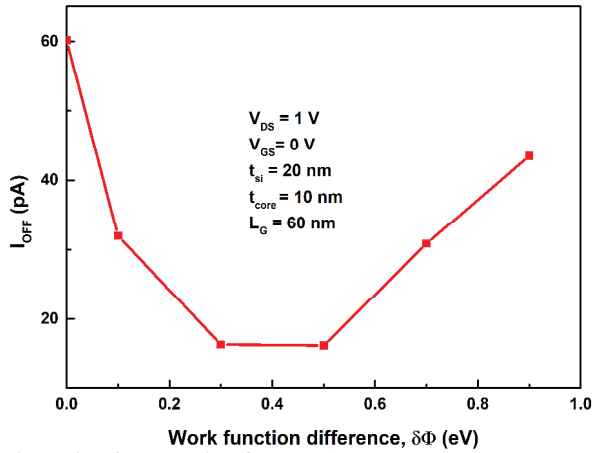


Fig. 9. Plot of I_{OFF} at various δΦ.

Furthermore, an enhancement of 20.8% in I_D is observed in DMG-ODCS-JLNWFET as compared to SMG-ODCS-JLNWFET when Φ_{M2} is varied from 5.0 eV to 4.2 eV. Fig. 8 illustrates the influence of work function variation on the output characteristics of DMG-ODCS-JLNWFET and SMG-ODCS-JLNWFET. It is perceived from the plot that I_D increases in DMG-ODCS-JLNWFET as Φ_{M2} decreases due to improved carrier mobility as compared to SMG-ODCS-JLNWFET. In Fig. 9, I_{OFF} with varying δΦ is plotted where

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:59:42 UTC from IEEE Xplore. Restrictions apply.

$\delta\Phi$ = 0 corresponds to SMG-ODCS-JLNWFET. It is observed that I_{OFF} decreases with increasing workfunction difference due to an enhancement in tunneling width as discussed in Fig. 6. However, when the $\delta\Phi$ surpasses 0.5 eV, the increase in I_{OFF} becomes more prominent, attributed to the increased SCEs. Additionally, a notable reduction of 27.7% is evident in DMG-ODCS-JLNWFET compared to SMG-ODCS-JLNWFET when Φ_{M2} undergoes variation from 5.0 eV to 4.2 eV. The Switching Ratio (I_{ON}/I_{OFF}) as a function of $\delta\Phi$ is demonstrated in Fig. 10. This metric is crucial in evaluating the performance of electronic devices, particularly in digital applications. This is formulated as [8]

$$\frac{I_{ON}}{I_{OFF}} = \frac{I_{DS} \text{ at } V_{DS} = 1 \text{ V and } V_{GS} = 1 \text{ V}}{I_{DS} \text{ at } V_{DS} = 1 \text{ V and } V_{GS} = 0 \text{ V}} \quad (2)$$

As $\delta\Phi$ begins to rise, the I_{ON}/I_{OFF} ratio enhances due to the simultaneous reduction in I_{OFF} and enhancement in I_{ON}, as previously discussed. However, when $\delta\Phi$ exceeds 0.5 eV, there is only a marginal increase in I_D, and I_{OFF} begins to rise. Consequently, the I_{ON}/I_{OFF} ratio decreases during this phase. The dependence of the Subthreshold Slope (SS) on the work function variation is presented in Fig.11. It is referred to as [18]

$$SS = \frac{\partial V_{GS}}{\partial \log_{10}(I_D)} \quad \text{mV/decade} \quad (3)$$

SS starts to degrade when $\delta\Phi$ surpasses 0.5 eV due to the reduction of I_{OFF}. Fig. 12 illustrates the transconductance (g_m) as a function of V_{GS} of SMG-ODCS-JLNWFET and DMG-ODCS-JLNWFET at various Φ_{M2}. The ability to modulate g_m is crucial in determining the efficiency of the transistor in signal processing and amplification applications. The g_m is expressed as [12]

$$g_m = \frac{\partial I_D}{\partial V_{GS}}\Big|_{V_{DS}=\text{constant}} \quad (4)$$

The graph illustrates a 23.8% improvement in g_m of DMG-ODCS-JLNWFET as Φ_{M2} decreases to 4.2 eV as compared to SMG-ODCS-JLNWFET. This enhancement is attributed to the superior drain characteristics exhibited by DMG-ODCS-JLNWFET.

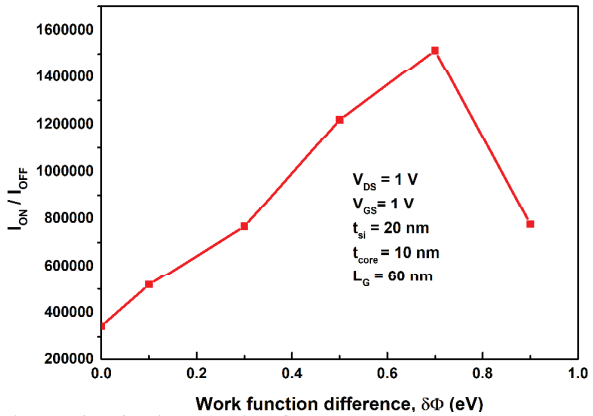


Fig. 10. Plot of I_{ON}/I_{OFF} at various $\delta\Phi$.

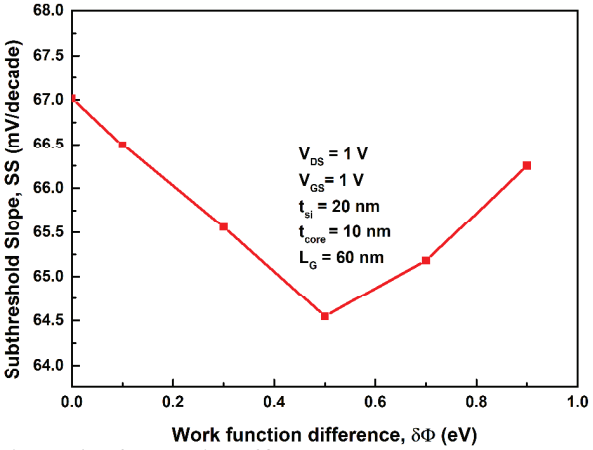


Fig. 11. Plot of SS at various of $\delta\Phi$

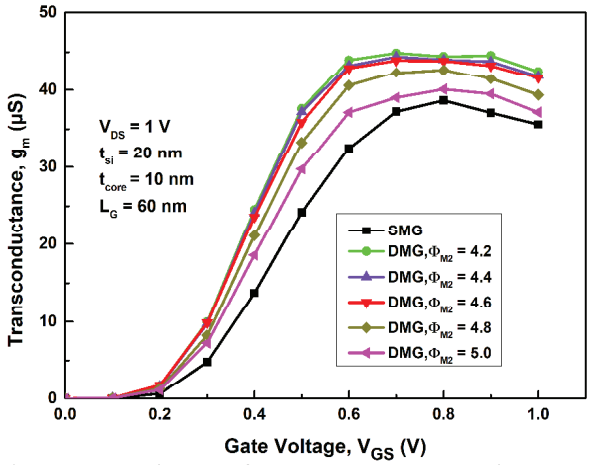


Fig. 12. Transconductance of DMG-ODCS-JLNWFET and SMG-ODCS-JLNWFET with variation of Φ_{M2}.

TABLE II. THE PERFORMANCE COMPARISON OF THE PROPOSED DEVICE WITH OTHER DEVICES

Device	L (nm)	SS (mV/dec)	I_{ON} (μA)	I_{OFF} (μA)	I_{ON}/I_{OFF}
JL-BS-Tapered FinFET [19]	14	68.65	2.15	2.08 X 10^{-9}	1.03 X 10^{9}
JGAA [20]	20	66.43	12	7.3X 10^{-7}	1.6 X 10^{6}
Cylindrical Gate-all-around JLFET [21]	30	73.25	20	3.05 X 10^{-4}	6.58 X 10^{4}
Proposed Device	60	64.8	24.5	1.68 X 10^{-5}	1.81 X 10^{6}

The aforementioned analysis underscores the superior performance of DMG-ODCS-JLNWFET in comparison to SMG-ODCS-JLNWFET across various key parameters such as I_D, I_{OFF}, I_{ON}/I_{OFF} ratio, SS, and g_m. Additionally, a comprehensive exploration of internal device characteristics, including potential, electric field, and band energy, have been presented to offer a clear insight into the underlying device physics. Notably, the extracted values of I_D, I_{ON}/I_{OFF}, I_{OFF}, SS, and g_m for DMG-ODCS-JLNWFET exhibit enhanced performance at Φ_{M1} = 5.1 eV and Φ_{M2} = 4.6 eV when compared to SMG-ODCS-JLNWFET. The performance of the proposed device in this paper is further compared with

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:59:42 UTC from IEEE Xplore. Restrictions apply.

other reported devices as illustrated in Table II. The smaller value of SS and higher value of I_{ON} highlight the utilization of DMG-ODCS-JLNWFET for both digital and analog applications, showcasing its potential for advanced semiconductor device applications.

IV. Conclusion

This paper explored the impact of workfunction variation of DMG-ODCS-JLNWFET device, focusing on a 60 nm channel length. The results revealed that the DMG-ODCS-JLNWFET outperformed the SMG-ODCS-JLNWFET device regarding overall performance by modulating the work function of gate material. Consequently, the proposed device exhibited an improvement of 18.9%, 73.7%, 26.8% and 10.8% in I_D, I_{OFF}, SS, and g_m, respectively $\Phi_{M1} = 5.1$ eV and $\Phi_{M2} = 4.6$ eV when compared to SMG-ODCS-JLNWFET demonstrating its excellent suitability for analog and digital application in nano-scale dimensions.

References

[1] K. J. Kuhn, "Considerations for Ultimate CMOS Scaling," *IEEE Trans. Electron Devices*, vol. 59, no. 7, pp. 1813-1828, Jul. 2012.

[2] S. Gundapaneni, M. Bajaj, R. K. Pandey, K. V. R. Murali, S. Ganguly, and A. Kottantharayil, "Effect of band-to-band tunneling on junctionless transistors," *IEEE Trans. Electron Devices*, vol. 59, no. 4, pp. 1023-1029, Apr. 2012.

[3] P. Mittal, Y. S. Negi, and R. K. Singh, "A depth analysis for different structures of organic thin film transistors: Modeling of performance limiting issues," *Microelectron. Eng.*, vol. 150, pp. 7-18, Jan. 2016.

[4] P. Mittal, S. Yadav, and S. Negi, "Advancements for organic thin film transistors: Structures, materials, performance parameters, influencing factors, models, fabrication, reliability and applications," *Mater. Sci. Semicond. Process.*, vol. 133, p. 105975, Oct. 2021.

[5] J. P. Colinge *et al.*, "Nanowire transistors without junctions," *Nature Nanotechnol.*, vol. 5, no. 3, pp. 225-229, Mar. 2010.

[6] J. P. Duarte, S. -J. Choi, D. -I. Moon, and Y. -K. Choi, "A Nonpiecewise Model for Long-Channel Junctionless Cylindrical Nanowire FETs," *IEEE Electron Device Lett.*, vol. 33, no. 2, pp. 155-157, Feb. 2012.

[7] S. Sahay and M. J. Kumar, "Controlling L-BTBT and Volume Depletion in Nanowire JLFETs Using Core-Shell Architecture," *IEEE Trans. Electron Devices*, vol. 63, no. 9, pp. 3790-3794, Sept. 2016.

[8] Bharti and P. Mittal, "Oppositely-Doped Core-Shell Junctionless Nanowire FET: Design and Investigation," *ECS J. Solid State Sci. Technol.*, vol. 13, no. 1, p. 013004, Jan. 2024, doi: https://doi.org/10.1149/2162-8777/ad1c90.

[9] Y. J. Lee *et al.*, "High performance poly Si junctionless transistors with sub-5 nm conformally doped layers by molecular monolayer doping and microwave incorporating CO2 laser annealing for 3D stacked ICs applications," *IEDM Tech. Dig.*, pp. 6.2.1-6.2.4, Dec. 2015.

[10] H. Lou *et al.*, "A junctionless nanowire transistor with a dual-material gate," *IEEE Trans. Electron Devices*, vol. 59, no. 7, pp. 1829-1836, Jul. 2012.

[11] M. Kumar, S. Haldar, M. Gupta, and R. S. Gupta, "Impact of gate material engineering(GME) on analog/RF performance of nanowire Schottky-barrier gate all around (GAA) MOSFET for low power wireless applications: 3D T-CAD simulation," *Microelectronics Journal*, vol. 45, no. 11, pp. 1508-1514, Nov. 2014.

[12] R. K. Baruah and R. P. Paily, "A Dual-Material Gate Junctionless Transistor With High- k Spacer for Enhanced Analog Performance," *IEEE Trans. Electron Devices*, vol. 61, no. 1, pp. 123-128, Jan. 2014.

[13] S. Mathew, K. N. Bhat, Nithin, and R. Rao, "Performance Enhancement of Dual Material Gate Junctionless FinFETs using Dielectric Spacer," *IETE Journal of Research*, pp. 1-12, Nov. 2023.

[14] R. Lin, Q. Lu, P. Ranade, T. J. King and C. Hu, "An adjustable work function technology using Mo gate for CMOS devices," *IEEE Electron Device Lett.*, vol. 23, no. 1, pp. 49-51, Jan. 2002.

[15] P. Ranade, H. Takeuchi, T. J. King, and C.-M. Hu, "Work Function Engineering of Molybdenum Gate Electrodes by Nitrogen Implantation," *Electrochem. Solid-State Lett.*, vol. 4, no. 11, pp. G85-G85, Jan. 2001.

[16] *Atlas User's Manual: Device Simulation Software,* Silvaco Inc., Santa Clara, CA, USA, 2018.

[17] J. Fan, M. Li, X. Xu, Y. Yang, H. Xuan, and R. Huang, "Insight into gate-induced drain leakage in silicon nanowire transistors," *IEEE Trans. Electron Devices*, vol. 62, no. 1, pp. 213-219, Jan. 2015.

[18] S. J. Choi, D. I. Moon, S. Kim, J. P. Duarte, and Y. K. Choi, "Sensitivity of threshold voltage to nanowire width variation in junctionless transistors," *IEEE Electron Device Lett.*, vol. 32, no. 2, pp. 125-127, Feb. 2011.

[19] S. Bhukya and B. R. Nistala, "Design optimization of junctionless bottom spacer tapered FinFET: Device to circuit level implementation," *Microelectronics Journal*, vol. 139, p. 105907, Sep. 2023.

[20] P. Srivastava, A. Upadhyaya, S. Yadav, and C. M. S. Negi, "Performance Evaluation of Junctionless Cylindrical Gate-All-Around FET for Low Power Applications," *Semiconductor Science and Information Devices*, vol. 5, no. 2, pp. 1-10, 2023.

[21] A. Kumar, A. P. Singh, T. K. Gupta and B. P. Shrivastav, "Electrostatic Performance Enhanced in a Dielectric Pocket Junctionless Gate All around MOSFET," *2023 International Conference on Next Generation Electronics (NEleX)*, Vellore, India, 2023, pp. 1-5, doi: 10.1109/NEleX59773.2023.10421725.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2024 at 05:59:42 UTC from IEEE Xplore. Restrictions apply.

Educational Administration: Theory and Practice
2024, 30(5), 56-66
ISSN: 2148-2403
https://kuey.net/

Research Article

Information Visualization and Documentation for themes of Visual Culture

Nanki Nath[1*][0000-0003-2647-6691], Ranganath M Singari[2][0000-0003-4981-3857]

[1]University School of Design and Innovation (USDI), Guru Gonbind Singh Indraprastha University, East Delhi Campus, Vishwas Nagar Extension, Surajmal Vihar, Delhi 110032, nankinath@ipu.ac.in
[2]Department of Design, Delhi Technological University (DTU), Shahbad Daulatpur, Main Bawana Road, Delhi-110042, ranganath@dce.ac.in

Citation: Nanki Nath, et al. (2024), Information Visualization and Documentation for themes of Visual Culture, *Educational Administration: Theory and Practice*, *30*(5), 56-66, Doi: 10.53555/kuey.v30i5.2769

ARTICLE INFO

ABSTRACT

Visual Culture envisages a trajectory of creations with information, visual styles, concepts and practices by both the makers and receivers. This paper presents the applications of popular cognitive theories during the information visualization stage followed by the graphic design students in the course of Documentation of Visual Culture at a design school in India. The baseline approach in the course has been to impart the theories and models from the fields of information visualization, cognition, history, social space and visual imagery. The course contributes a conceptual framework of 5-Pillars of Cognition: 1) What and How to Observe, 2) How to See, 3) How to Read, 4) How to Plan a Theme and 5) How to Create a Visual Language for a Visual Space/Culture. Each student applied these five pillars along with the theories of cognitive semiotics to form, reform and transform the documentation designs. The research methodologies tap the styles and trends in the context of readers' preferences, social milieus and the linked technological scenarios. Designs showcase the styles in documentations with respect to the information content, readers, social milieus, heritage scenarios and communication. The insights articulate the rationale behind preserving form-content associations in the visual culture themes.

Keywords: Cognition theories, Conceptual framework, Documentation, Visualization, Semiotics, Visual Culture

Introduction

The field of cognition with semiotics[1] is a trans-disciplinary field that integrates methods and theories developed in the disciplines of cognitive science with methods and theories developed in humanities. The aim is to provide new insights into the realm of information perception, human signification and manifestation in cultural practices. Practice is not viewed as a 'context for' but as constitutive and constituent of the cognitive process itself. It is a constitutive aspect of all forms of thinking, from the simplest to the most complex (Trasmundi and Toro, 2023; Böhme et. al., 2022; Poizat, et. al., 2022).

At a systemic scale, semiotic interactions include specific media channels, with the term 'medium' understood in its broadest sense, including bio-entities along with artefacts and technologies that can be material, and thus physically perceived by humans and animals, but also digital, which can lie beyond the scope of perception (López-Varela Azcárate, 2022). Cognition and thinking approaches utilize the use of language in communicating the surrounding social situations involving senders and receivers in the circle of information dissemination, reception and retrievals. This makes up for the fundamental premise of a two-way communication for meaning making in information studies. In context of this objective, the present paper highlights the learning experience that aimed to capture how the visual language developments based on cognitive thinking could become instrumental in crafting a visual experience of cultures, spaces and people associated in the documentation designs done by the undergraduate graphic design students, 2018 -19 at the respective national institute in India.

[1] Zlatev, Jordan, "Cognitive Semiotics," in *International Handbook of Semiotics,* ed. Peter Pericles Trifonas (Germany, Berlin: Springer Science+Business Media Dordrecht, 2015), 1043.

Copyright © 2024 by Author/s and Licensed by Kuey. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Visualization and cognition in everyday life helps in building a continuous experience to draw things together, decipher meanings and their essential role in developing a language that undergoes transitions over time. Taking this as a guiding light to document novel experiences of visual media in a hybrid of texts, images, and sounds became the vision for the course of Documentation of Visual Culture [DVC] in the academic year of 2018 – 2019 as full 3-week units, 6.0 credits continuous education module. The course gave design students opportunities to build and communicate the identity, processes and meaningful perspectives that unpack a range of themes of visual culture in India (Figure 1). For instance, the story of the sawantwadi wooden toys from Maharashtra, India visual documentation contains detailed process involved in the making of these special sawantwaadi wooden toys through illustrations and pictures – clicked by the student himself as part of the fieldwork done in the city of Mumbai. Inclusive from all aspects; the book emphasizes its special place and/or use by children. They get an opportunity to make theor own toys in the instructions provided in the ending sections of the documented book – using cut-outs and easy dismantling of pop-up shapes. For parents and their children as primary users, this documentation invites participatory design approach in the most playful and experiential way (Figure 1 and 2). Of all the documented books,oOnly one DVC documented publication covered a global trend in a foreign city as a theme of visual culture as part of the course (Figure 7).



Figure 1. Sawantwadi Toys theme from Maharashtra, India (left) and Chitrakaar making the Pattachitra Art with Techniques from Odisha, India (right) © 2018-19 course documentation by Author Name (hidden for peer review).

Sawantwadi is famous for its art and culture, popularly known for Ganjifa playing cards and wooden toys made from mango tree. The craft is traditionally done by the Chitari or Chitrakar community recognized by the king. These communities migrated from Karwada and settled in Sawantwadi. These families survived mainly by supplying handicrafts items to the royal families. Later other communities also adopted this craft because of its commercial success. Queen Satwasheela Devi continues the tradition even today, where many of the families are involved in Ganjifa cards and wooden toys making. The toys have been made by assembling all parts to shape solid wood. Seasoned mango wood is chiseled or carved into desired toy shape. These examples have been the practice-driven content covered in the dvc documentation by the student (Figure 1 and 2).



Figure 2. Sawantwadi Toys documented visual culture theme from Maharashtra, India (left and right), © 2018-19 course documentation by Author Name (hidden for peer review).

India is a cultural melting pot of visually rich cultural identities, events, festivals, customs, arts, crafts, practices, etc. To deconstruct with a novel lens this rich visual culture in India, the present paper presents the course of DVC and the research methodologies applied by the graphic design students developed applied learnings around visualization and cognitive semiotics.

Course Relevance

This course is important in the context of changing societal and technological scenarios. Visual approaches in everyday life has been changing frequently. In this context, documentation of special experiences around marked visual culture and the themes – that imbibes visual media in a hybrid as texts, images, analog

techniques and tactile surfaces; sound symbolisms, etc. as part of the course has been aimed to offer graduation students of graphic design to build rich, multi-perspective points of view of a range of themes as part of visual culture as a field of depth study and documentation.

Course Objectives

- To provide understanding of visual sensibility in connection to several aspects - personal, national, ethnic.
- To provide understanding of strategic analysis on cultural visual forms.
- To understand the Rhetoric of the Image.[2] Rhetoric is a creative means to design and organize discourse in order for it to be persuasive but not manipulative (D'Armenio, 2022; Greco and Mazzalli-Lurati, 2022; Barthes, 1988).
- To understand comparison with popular visual culture for attention and traditional Visual cultures.
- To understand the image making in context of everyday life experiences.

Teaching Methodology and Course Detailing

1) Instructional Process

- Lectures/ sessions included visual media-based presentations describing striking visual culture theories, models and concepts with respect to communication design, cognitive semiotics, visualization and publication designed for visual cultures. The specific aim helped teacher-learner as well as student-learners to plan and develop researched visual strategies for the creation of distinctive and reader-friendly documentations as part of the larger outcome-based education approach aided by the Bloom's taxonomy model.
- Demonstrative modes of orientating the students' batch to different and innovative ways of book binding, formats involved and the suitable fusion of printing methods have been imparted. The resulting documentations took those inputs as practice-based prompts at different stages of the ideation of final dvc documentations.

2) Student-centric approach

- The teacher-learner as well as each student-learner interacted over a series of studio-based feedback sessions by displaying parts of the dvc documents crafted by the student learners' for continuous evaluations and review feedback sessions organized three times (each time per week unit) by the teacher-learner in the total of 3 week units allotted for the course instruction.
- Honing of specific analog and digital techniques by student-learners'; alongwith participatory design approaches were a springboard for out-of-the-box ideas and final identity of the documentations designed by each student. Students' coordination of the head, the heart and the hand became unique, with respect to the strength of hand skills alongwith other soft skills of designing and representing 'content' alongwith 'form' for new meaning-making (in the context of cognitive semiotics).

Theories and Models of Cognition, Semiotics and Visualization

Generally, Cognition alongwith semiotics Semiotics has been a transdisciplinary study that comprises modes of meaning-making, applications of methods and theories from semiotics, linguistics, psychology, cognitive science, computational modeling, anthropology, philosophy and other sciences. Cognition entails complex networks of biological processes and actions that encompass perception, attention, manipulation of objects, memory mechanisms, and the formation of knowledge by means of direct experience as well as by learning from others, for which forms of communication and comprehension are also necessary. It has often been assumed that one but not both of these two types of representations underlie processing in specific domains of cognition – for instance, perception, mental imagery, and language (Kaup, et. al, 2023; Azcárate, 2022). Participatory sense-making – a cognitive act of humans during interactions with the 'medium', addresses how our intersubjective interactions enter into meaning constitution as well as co-constitution of a meaningful world. It provides an answer to the question: How do we, together, interacting with each other, constitute the meaning of the world? (Gallagher, 2023; Shapiro and Spaulding, 2021).

Within the purview of DVC course objectives that imbibed a Special Focus within the context of Visual Cultures and Visual Language Developent(s) with Communication as the raison d'être; Cognitive Semiotic Theories and Models were selectively collated, analyzed and implemented in the final print publication designs submitted for the DVC course documentations. Following are the prime theories and models applied by students in their DVC documents:

1) *Theory of the Actions of Signs and the Dialogical Model of Interspecific Communication*

Defined as the 'action of signs', semiotics has recognized the 'inter-actions' and 'intra-actions' of anything acting as a sign; present also within the framework of the 4Es -Embodied, Embedded, Enacted and Extended. The E-fields share the view, after all, that the brain-centrism of traditional cognitive science, as well as its dependence on the computer for inspiration, stands in the way of a correct understanding of cognition (Azcárate, 2023; Konderak, 2021; Shapiro and Spaulding, 2021). At a systematic scale, semiotic interactions

[2] Barthes, Roland, "The Rhetoric of the Image", from Image - Music - Text. Sei. and Trans. Stephen Heath. In: *Elements of Semiology*. New York: Hill and Wang, 1,977.32-51, (1977): 160-163.

include specific media channels, with the term 'medium' understood in its broadest sense, including bio-entities along with artifacts and technologies that can be materials, and thus physically perceived by humans and animals (Figure 3), but also digital, which can lie beyond the scope of perception.

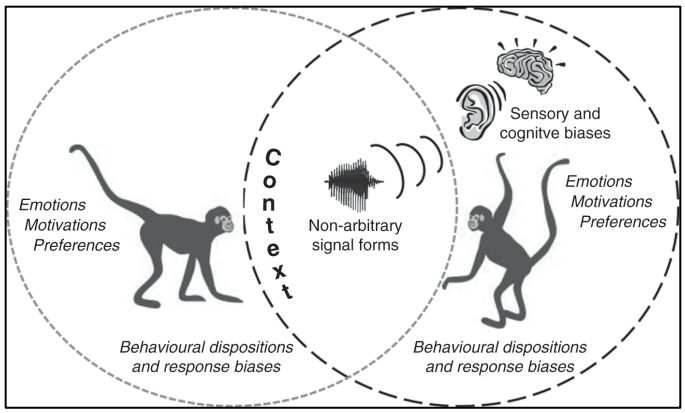


Figure 3. The Dialogical Model of Interspecific Communication © 2013 Drew Rendall and Michael J. Owren. Source: Drew Rendall and Michael J. Owren, "Communication without meaning or information: Abandoning language-based and informational constructs in animal communication theory," *Animal Communication Theory: Information and Influence,* ed. Ulrich Stegmann © Cambridge University Press 2013, pp.173, http://dx.doi.org/10.1017/CBO9781139003551.010

2) *Bio-cultural theory and the Semiotic Hierarchy model*

There has been an attempt made to define a unified bio-cultural theory of meaning and lay context of Cognitive Semiotics for developing later hierarchical models (Wiggins, 2020; Silvera-Roig, 2020; Zlatev, 2019). One of the models expresses that:

"Our conception of meaning has become increasingly fragmented, along with much else in the increasing post modernization of our worldview. The trenches run deep between different kinds of meaning theories: mentalist, behaviourist, (neural) reductionist, (social) constructivist, function- a list, formalist, and computationalist deflationist... And they are so deep that a rational debate between the different camps seems impossible. The concept is treated not only differently but incommensurably within the different disciplines."

The Semiotic Hierarchy: that each subsequent level somehow replaces the earlier one(s), for instance, a (minimal/enculturated/linguistic) self is independent of the 'underlying biological self', the organism (Mendoza-Collazos, et. al., 2021; Mendoza-Collazos, 2020). This model represents an affinity with 'layered models' of development with each higher level subsuming earlier ones (Figure 4).

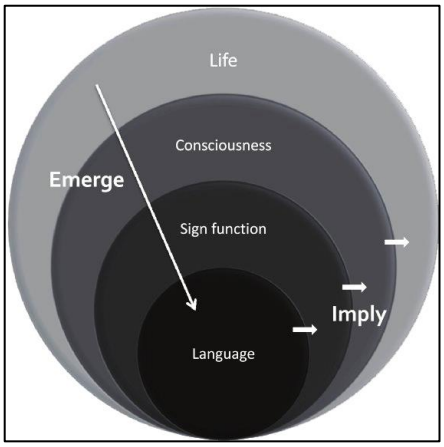


Figure 4. The Semiotic Hierarchy Model © 2009 Jordan Zlatev. Source: Jordan Zlatev, "The Semiotic Hierarchy: Life, consciousness, signs and language," *Cognitive Semiotics*, no. 4 (Spring 2009), pp. 189 https://doi.org/10.3726/81608_169

Research Methodology

Students' secondary research-based design language developments of their document designs envisaged applied learnings from two methodologies, namely: Phenomenology and Field Work. These applied learnings helped build the building blocks of a conceptual framework of reference to guide final ideations of a publication form for each documentation of visual culture per student.

Phenomenology or the systematic study of lived experience(s)

Phenomenology is grounded in certain traditions of philosophy and the humanities, and that aims to reflect on pre-reflective human experience (Manen, 2023; Mendoza-Collazos and Zlatev, 2022; Køster and Fernandez, 2021; Vasterling, 2021; Stew, 2020). In this course, the students have applied the phenomenological methods of qualitative data collections in their fieldworks on selective visual cultures - that helped them accumulate their self-experiences of the history, material culture, social spaces, visualizations and cognitions (Figure 5 and 6). Individual students and quick orientations did fieldwork and warm-up exercises were given to students (Figure 4 and 5).

A brief collection of the ways of doing fieldwork were imparted to students covers following (in a nutshell):

- How people navigate to witness a visual culture?
- How and what to observe as they navigate?
- How to design a visual language for a given space incorporating the visual culture?
- What codes govern a special visual language to communicate that visual culture?
- What discourses are happening in these spaces of special visual culture themes?
- How a public interest gets constructed around a visual culture theme?
- Why is it important to consider this public interest in final DVC documents?

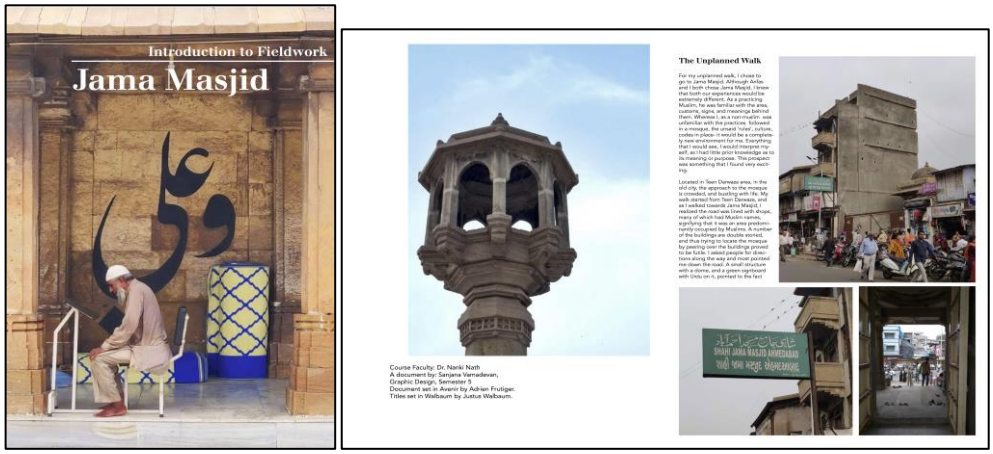


Figure 5. DVC documentation cover (left) and inside spread (right) of Jama Masjid as a visual culture theme © 2018-19 course documentation by Author Name (hidden for peer review).

Special Content Developments for DVC course

Students and the faculty instructor decided to document a range of distinct visual culture themes from the Indian land - aiming to draw things together at the intersections of bringing forth new knowledge and representation of: the People, the Environment, the Geographical Location, the tools and skills used for producing the visual material, any additional resources that are enablers in crafting special design language and communication of these themes (Figure 6). The content developments also include observations and depth analysis about how documenting the connection with larger communication goals alongwith identification of sources from everyday life, such as Cinema, Calendar, Persuasive Communication, Traditional Visual Influences, Cultural Influences, and Mythological Symbol, etc. carve out a meaning-making system of visual design elements for the final DVC documentations.

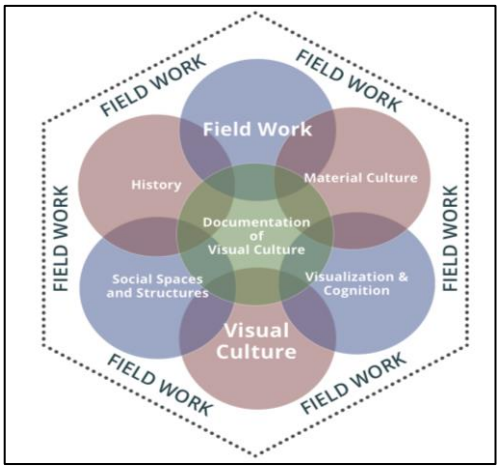


Figure 6. Information Content Mapping, Linking and Analytical approach in DVC documentations © 2018-19 course documentation by Author Name (hidden for peer review).

DVC course: the five-pillar conceptual framework

Analysing Bruno Latour's research and for course methodology to apply visualization and cognition for drawing things together to bridge Content, Form and a Visual Language Development became a pertinent series of in-classes, discourse oriented actionable tasks to ideate for the final design outcomes for each student's DVC publication cum documentations (Solberg, 2021; Wilde, 2020). The tone of voice in the descriptions below has a deliberate conversational tone of voice that has been an effort to bring forth the essence of this in-classes activity-oriented methodology. Following are the series of those activities that involved students' as well as the researcher of this paper, who as the faculty instructor implemented the method in five pillars, as follows:

1) *What and How to Observe?*

Observation is a powerful method to understand human behaviour and patterns of living. In the context of space, students were required to apply observation in following ways:

- Learn to Listen: Developing the skill of *learning to listen* would help you to go, explore, interpret and infer the space of study within the larger space.
- Temporal Discourse: Space Discourse is always temporal. It changes with time. This discourse that you record will assume different identities; and hence interpretations with time. For instance, one of the B.Des student's DVC theme of 'Revitalizing Traditional Indian Toys' involving children experiencing the space with the document design as an enabler for the temporal discourse in real-time (Figure 7).
- This publication titled, 'Khel Khoj', talks about India's rich cultural heritage of crafts in toy making; based on the books, such as 'The Joy of making Indian Toys' by Sudarshan Khanna (2002). In the context of visual culture in India, 'toys' add an innovation-based dimension of experience with the resource-scarce tools and equipments that have been utilized keeping in mind the affordability and the low-cost parameter for better access to masses. This pulsating identity associated with Indian toys as a craft heritage has been the spine to structure the publication – including 'magic toys' with respect to the Indian land. The world knows it by the name, 'Jakob's Ladder'. The ladder form becomes the structure of opening, folding, reading as well as experiencing the visual culture of magic toys as part of this interesting experiential document design done as part of the course (Figure 7).



Figure 7. Revitalizing Traditional Indian Toys visual culture theme for DVC documentations © 2018-19 course documentation by Author Name (hidden for peer review).

2) *How to See?*

Learning to See is the most essential quality that we as designers need to learn. It is because of those qualities that we will eventually learn in order to ideate designs (Peng, 2022; Vikas et. al., 2022; Cascini, 2021; Larraz-Rábanos, 2021).

- This is a blessing and a curse—once we have learned to see design, both good and bad, we cannot un-see. The downside is that the more you learn to see, the more you lose the distinct eye that you design with.
- For instance, one of the B.Des student's DVC theme of 'Documenting the city of Melbourne's Graffiti' involving the tactile and ephemeral quality as seen by the designer to become the core design identity in naming, visual language and compositional developments (Figure 8).
- Stencil City project as represented; has been a comprehensive document the depicts the unique beauty and aesthetics, modes/methods of creations, etc. as an integral part of graffiti and street art visual culture in the city of Melbourne (Figure 8). The detailing included in the documentation as field works included decoding the elements of visual aspects and cultural context that encapsulates the city's dynamic visual language. It equips the reader with graphical cues that narrate the social, economic and physical terrains of the city in an informed manner. The documentation celebrates this captivating visual culture theme of graffiti street art from

three perspectives that become chapters in the documentation, namely: a. The Fact, b. The Feel and, c. The Form. This special approach in content-form association build for the publication has been instrumental in painting a holistic picture of the graffiti scenarios in Melbourne city with technicalities and personalized travelogue section on personal interactions, observations and reflections with the graffiti painters and makers.
Developing an Eye:
Good designers try to keep their ability to see things as a customer would. You need a designer's eye to design; and a non-designer eye to feel what you designed.
"See with one eye, feel with the other." —Paul Klee



Figure 8. The Stencil City–Melbourne City's Graffiti visual culture theme for DVC documentations © 2018-19 course documentation by Author Name (hidden for peer review).

3) *How to Read?*
Learning to See is the most essential quality that we as designers need to learn. It is because of those qualities that we will eventually learn in order to ideate designs (Peng, 2022; Vikas et. al., 2022; Cascini, 2021; Larraz-Rábanos, 2021).

- In the development of design skills, theory can get in the way of practice, but only until the theory becomes practice. With practice your intuition evolves, and the better you understand what you do, the deeper your intuition. Only once you do not consciously think about the theory anymore are you achieving mastery.
- How to Read, involves that approach of graphic design that gives to the viewer a range of images or indirect meaning related to the theme/subject matter. This is largely known to be image making in graphic design; wherein, a graphic designer through images communicates direct or indirect messages (Zhang, 2022; Xu, 2022; Günay, 2021). Denotations are direct, but connotations are indirect.

For instance, one of the B.Des student's DVC theme of 'Odisha's Pattachitra' involving the insightful content for augmenting readers' interests to know the detailed product and processes involved in the making of 'chitra', that means a piece of an artwork or simply put, a painting, Pattachitra may literally translate to cloth painting (Figure 8). Derived from the Sanskrit word 'patta', that means 'cloth' or 'canvas'; and, 'chitra' that means 'a piece of an artwork' / simply put; 'a painting' – is born the 'pattachitra'. Famous as 'Cloth paintings', these pattachitras depict and celebrate the themes of ras lila, vastra haran, kaliya dalan, ten incarnations of the Lord Vishnu, nava grahas (nine planets), kanchi avijana of Lord Jagannath, panchamukhi Lord Hauman as well as the panchmukhi Lord Ganesha (Figure 9).



Figure 9. Painted Prayers –Odisha's Pattachitra visual culture theme for DVC documentations © 2018-19 course documentation by Author Name (hidden for peer review).

4) *How to Plan a Theme ?*

Brainstorm the ideas you wish to select as themes of Visual culture for your final document, as follows:

- Building an image library: Select different kinds of visual images. Some may be connotative (indirect) and others may be denotative (direct) representations of the main theme of your visual culture. Collect these by photo-documenting your space entirely. Eliminate any repetitive or incomplete pictures to do away with any confusing messages they convey.
- Construct a public interest: By including those texts from your notes, data collections, etc. that help define the identity, essence, cultural ethos and meaning of a particular visual theme for the public who would be the direct receivers.

In this context, customary practices within a space must be archived to the last detail (all micro and macro process steps). For instance, the festival of Holi imbibes a good prominence of customary practices as part of it as a visual culture in India. The larger context includes interesting associations with folklore and folk cultural practices that aim to build communities from socio-cultural as well as socio-political perspectives within the context of India.

Another instance to establish the unique significance of customary practices; when seen through the lens of rich visual cultures in India – is the dynamic Chhau dance from Bihar. The said dance form is a tradition – deeply embedded in the easter fabric of India; especially Bihar. The dance form is used to enact episodes from the epics (including Ramayana and Mahabharata), locl folklore as well as abstract forms of stories and storytelling. It's three distinct styles hail from the popular region of Seraikella, Purulia, Mayurbhanj in East India. Chhau, as a visual culture to be celebrated in the region – has been intimately connected with regional festivities; notably the spring festival Chaitra Parva. Its origin has been traced back to the indigenous forms of dance and martial practices. The vocabulary of movement includes mock combat techniques, stylized gates of birds and animals; and the movements modeled on the chores of village housewives. The effect of the dance form stand for the development of unsophisticated forms of folk dance to highly stylized forms. Chaau, symbolic of the earliest indigenous forms of dance in India – has demonstrated a living continuity with the past in order to acknowledge and preserve heritage within the silos of visual cultural practices and preservations.

While doing this as part of the fieldworks by each student, it should be kept in mind that conventions/rules define a certain cultural group's interaction and impact the related customary practice accordingly. For instance, the 'Paithani style of wearing saree from Maharashtra' is a tradition from the Western part of India as documented and designed in a coffee table book form to reach as well as touch resonance with a good section of public residing in Western India by one of the B.Des students (Figure 10). This DVC course documentation is in the form of a coffee table booklet – that visually celebrates the heritage process of wearing the iconic paithani sarees by the women of the concerned ethnic culture.



Figure 10. The Paithani: Maharashtra's pride as a visual culture theme for DVC documentations © 2018-19 course documentation by Author Name (hidden for peer review).

5) *What to create as a Visual Language for a Space/Living Culture ?*

- Message selection and communication strategies for content developments:

In last three days of the first week of the course, rigorous discussions over each visual culture theme selected by these students. They were expected to share a two -page write-up that would cover as much data and information from following perspectives. These articles that they finally submitted on the week's end became individual blueprints for their targeted visual culture documents. These write-ups covered following in a nutshell.

For each header in the content, students would pick up 3-5 keywords that will associate the purpose of your chunk to be written under that header. The text of this chunk must imbibe following:

- Cognitive underpinnings of all the observations made.
- Continuity in-text of all the observations and the sections of the space covered. The

sequence in text should ideally be linear and ordered.

- Composition of all Visual elements.
- Identity understanding and articulation of visual elements must be done with depth and sensitivity.
- Depending on your selected space, say a line as a visual element communicates a sense of alignment; or it may mean a boundary or lines in your space are facilitators of segmentation/divisions/ distributions, etc.

- Visual culture theme discussions with students:

The first two days of the second week of the course was time put in extensively to discuss with each student his/her visual culture theme selection. Larger visual culture domains got defined from the thematic information collections the students shared in these sessions. Briefly, they collected thematic information under following categories of considering 'Space as a Social means of Culture and Communication'. The criteria are as follows:

1) Space and Associations with Class
2) Space and Associations with Materials
3) Space and Associations with History
4) Space and Associations with Multicultural Artifacts
5) Space and associations with Artifacts
6) Space and Associations with a City
7) Space and Associations with an Event / Festival / Space
8) Space and Associations with a Practice
9) Space and Associations with Seeing and Reading (observations and articulations or even conjectures)
10) Space and associations with Self-Image

Insights and Discussion

Field work based research and the applied five-pillared framework for design outcomes as part of the DVC course works in this paper reveal guiding lights for future. Among them, some pertinent insights from teaching-learning perspectives can incorporate following visions for pedagogical developments around information visualization, cognition and semiotics, are as follows:

- Field works should include surveys and detailed interviews with the main stakeholders; in order to archive the processes employed by the concerned visual artists. Special story records of developing their arts, craftsmanship and entrepreneurship-based scope could be documented – for encompassing the complete circle of all the know-hows that students could experience in a better way at the grass root level.
- Ethnography, Participatory research modes of inquiry, co-creation research techniques are imparted to them as part of course and could be showcased as special chapter(s) in future documentations as part of such DVC courses. This would help them learn the importance of developing a mind and mode of operating as a design researcher.
- Students would tend to assume the role of a visual archeologist by filtering and harnessing content (text and graphics) to document varied kinds of visual cultures – in-depth through the lens of the people, the things (objects/artefacts, etc.) and the spaces in future DVC documentations. This would enhance information visualization, storage as well as retrieval in context of both socio-cultural and socio-spatial perspectives; bringing forth unknown facts and/or innovation aspects that represent the identity of a new kind of visual culture.
- Students need to understand the visual design aspects, print production aspects (selection of paper makes / quality and print production process for final outcomes). They are able to speculate a workable budget for production of their document to learn making a mini business model around their book design – and this as production content could become part of the larger information content planned as a chapter for future DVC documentations.
- Innovation in future documentations may also include explorations of outside-field digital archival methods to gather research for interesting content developments; in consideration of the recent new normal future scenarios. Such an approach could be operated in a workshop-mode for the course.
- Visualization and experiments with book design formats for future DVC documentations would add responsiveness to their designs with respect to digital modes of visual representations.

Conclusion

An in-depth and field work based research and documentation-cum-publication design courses and the applied five-pillared framework for design outcomes as part of the DVC course works in this paper reveal guiding lights for future that could incorporate following visions for pedagogical developments around 'Visualization' and 'Cognitive Semiotics'.

Visual culture may also be decoded and analyzed in a focused way with respect to the skills and craft of the makers alongwith the unique legacy testimonials in their creations that imprint the relevance of a particular visual culture ahead of its time. In the realm of preserving the heritage of the analog or special manual capabilities of such makers/artists – the form-content associations implemented in the

documentations may cater to out-of-the box approaches in research, ideation and communication aspects needed to unravel the unknown stories of a good range of visual cultures from the present and the near future scenarios. Therefore, the researcher has summarized arguments in the form of research questions that would be needed to develop pedagogy around the said DVC course, as follows:-

1) Can these documentations by students be positioned as an essential source of knowledge creation about the visual arts, design and communication design?
2) Can they be produced as 'Visual Anthologies' or tangible collection of stories of a visual culture open to anyone to read and experience those - regardless of age, experience, ethnicity, or income ?
3) What sort of visual language design approaches would be needed in these special anthologies ?

In a nutshell, the above questions earmark the potential areas of scope that may regulate discovery of new qualitative methods for information content developments and information content to form associations for experimental research. Experimentations would aid reformation and transformation of existing practices related with documentations that celebrate the essence of a particular visual culture.

References

1. Barthes, Roland. "Rhetoric of the Image." In: *Image - Music - Text. Sei. and Trans. Elements of Semiology*. Stephen Heath. New York: Hill and Wang, 1,977.32-51, (1977): 152-163.
2. Böhme, Jessica , Zack Walsh, and Christine Wamsler. "Sustainable lifestyles: towards a relational approach." *Sustainability Science*, 17 (2022): 2063–2076, https://doi.org/10.1177/10597123211072352.
3. Cascini, Gaetano et. al. "Perspectives on design creativity and innovation research: 10 years later." *International Journal of Design Creativity and Innovation*, 10, no. 1 (2022): 1-30, https://doi.org/10.1080/21650349.2022.2021480.
4. D'Armenio, Enzo. The rhetorical dimension of images: identity building and management on social
5. networks. *Semiotica: Journal of the International Association for Semiotic Studies / Revue de l*
6. *Association Internationale de Sémiotique*, 246, (2022): 87-115,
7. https://doi.org/10.1515/sem-2020-0113.
8. Gallagher, Shaun. "Embodied and Enactive Approaches to Cognition." In: *Cambridge Elements: Elements in Philosophy of Mind*, edited by Keith Frankish, The University of Sheffield, 1-65. UK: Cambridge University Press, 2023.
9. Greco, Sara and, Sabrina Mazzalli-Lurati. "Rhetoric." In: (eds) The Palgrave Encyclopedia of the Possible, edited by Glăveanu, 1431–1439. V.P. Palgrave Macmillan. Cham: Springer, 2022, https://doi.org/10.1007/978-3-030-90913-0_8.
10. Günay, Mustafa. "Design in Visual Communication." *Art and Design Review*, 9 (2021): 109-122, https://doi.org/10.4236/adr.2021.92010.
11. Kaup, Barbara et. al. "Modal and amodal cognition: an overarching principle in various domains of psychology." *Psychological Research: An International Journal of Perception, Attention, Memory, and Action,* 88, (2024)*:* 307–337, https://doi.org/10.1007/s00426-023-01878-w.
12. Khanna, Sudrashan. "Joy of Making Indian Toys." New Delhi: National Book Trust, (2002): 1-125. ISBN-13978-8123722443 :.
13. Konderak, Piotr. "Towards an integration of two aspects of semiosis – A cognitive semiotic perspective." *Sign Systems Studies*, 49*,* no.1/2, (2021): 132–165, https://doi.org/10.12697/SSS.2021.49.1-2.06.
14. Køster, Allan and, Anthony Vincent Fernandez. "Investigating modes of being in the world: an introduction to Phenomenologically grounded qualitative research." *Phenomenology and the Cognitive Sciences,* 22*,* (2021): 149–169, https://doi.org/10.1007/s11097-020-09723-w.
15. Larraz-Rábanos, Natalia."Development of Creative Thinking Skills in the Teaching-Learning Process." In: *Teacher Education-New Perspectives,* edited by Ulas Kayapinar. IntechOpen Books, United Kingdom, 2021. PRINT ISBN978-1-83969-288-8, https://doi.org/ 10.5772/intechopen.94952.
16. López-Varela Azcárate, Asun. "Posthuman Intermedial Semiotics and Distributed Agency for Sustainable Development." In: *The Palgrave Handbook of Intermediality,* edited by Bruhn, J., Azcárate, A.LV., de Paiva Vieira, M., 1213–1240. Palgrave Macmillan, Springer Cham, 2023, https://doi.org/10.1007/978-3-031-28322-2_61.
17. López-Varela Azcárate, Asun. "Mind and Matter - Challenges and Opportunities in Cognitive Semiotics and Aesthetics." IntechOpen Books, United Kingdom, (2022): 1-19. PRINT ISBN: 978-1-83969-935-1, http://doi.org/10.5772/intechopen.95671.
18. Manen, Max van. "Phenomenology of Practice: Meaning-Giving Methods in Phenomenological Research and Writing." 2nd Edi. Routledge, London, (2023): 1-520. https://doi.org/10.4324/9781003228073.
19. Mendoza-Collazos, Juan and, Jordan Zlatev."A Cognitive-Semiotic Approach to Agency: Assessing Ideas from Cognitive Science and Neuroscience." *Biosemiotics*, 15, (2022): 141–170, https://doi.org/10.1007/s12304-022-09473-z.
20. Mendoza-Collazos, Juan and, Sonesson, G. "Revisiting the life of things: A cognitive semiotic study of the agency of artefacts in Amazonia." *Public Journal of Semiotics*, 9, no. 2, (2021): 30-52, https://doi.org/10.37693/pjos.2020.9.22012.

21. Mendoza-Collazos, Juan 2020). "On the importance of things: a relational approach to agency." *Cognitive Semiotics*, 13, no. 2, (2020): 1-11, https://doi.org/10.1515/cogsem-2020-2034.
22. Franke, Peng. "Design Thinking: From Empathy to Evaluation." In: *Foundations of Robotics: A Multidisciplinary Approach with Python and ROS*, edited by Damith Herath and David St-Onge
23. D., 63–81. Springer, Singapore, 2022. https://doi.org/10.1007/978-981-19-1983-1_3.
24. Poizat, Germain., Simon Flandin and, Jacques Theureau. "A micro-phenomenological and semiotic approach to cognition in practice: a path toward an integrative approach to studying cognition-in-the-world and from within." *Adaptive Behaviour*, 31, *no.* 2 (2022): 109-125, https://doi.org/10.1177/10597123211072352.
25. Shapiro, Lawrence and, Shannon Spaulding. "Embodied Cognition." In: *Stanford Encyclopedia of Philosophy Winter (2021 Edition),* edited by Edward N. Zalta. The Metaphysics Research Lab, Stanford University, USA, 2021. https://plato.stanford.edu/archives/win2021/entries/embodied-cognition/
26. Silvera-Roig, Marta. "Cognitive Semiotics and Conceptual Blend: A Case Study from The Crying of Lot 49." In: *Cognitive and Intermedial Semiotics*, edited by Silvera-Roig, M. and López-Varela Azcárate, A., 1-178. UK: IntechOpen Books, (2020): PRINT ISBN978-1-83880-552-4, https://doi.org/10.5772/intechopen.92606.
27. Solberg, Mads (2021). "A Cognitive Ethnography of Knowledge and Material Culture: Cognition, Experiment, and the Science of Salmon Lice." In Series: *Culture, Mind and Society*, 1st edi., 2637-6806. Switzerland AG: Springer Nature, 2021. Hardcover ISBN: 978-3-030-72510-5, https://doi.org/10.1007/978-3-030-72511-2.
28. Stew, Graham (2020). "Phenomenology – questioning consciousness and experience." In: *Critical Qualitative Health Research*, 1st Edi., 1-17. Routledge, London., 2020. https://doi.org/10.4324/9780429432774.
29. Trasmundi, Sarah Bro and, Juan Toro. "Mind wandering in reading: An embodied approach." *Frontiers in Human Neuroscience*, 17 (2023): 1-15, https://doi.org/10.3389%2Ffnhum.2023.1061437.
30. Vasterling, Veronica. "Phenomenology." In: *The Routledge Companion to Historical Theory*, edited by Chiel van den Akker, 1st Edi., 1-636. UK: Routledge, 2021. https://doi.org/10.4324/9780367821814
31. Vikas, T. N., et.al. "Design Thinking: A Review Paper." *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*, 2, no. 2, (2022): 405-412. https://doi.org/ 10.48175/IJARSCT-2893
32. Wiggins, Geraint A. "Creativity, information, and consciousness: The information dynamics of thinking."*Physics of Life Reviews*, 34–35, (2020):1-39. https://doi.org/10.1016/j.plrev.2018.05.001
33. Wilde, Danielle. "Design Research Education and Global Concerns." *She Ji: The Journal of Design, Economics, and Innovation***,** 6, no. 2, (2020): 170-212. https://doi.org/10.1016/j.sheji.2020.05.003.
34. Xu, Dan. "Application Research of Graphic Design Based on Information Resource-Sharing and Big Data Technology." *Big Data Modelling of Engineering and Management 2022 [Special Issue],* 2022. https://doi.org/10.1155/2022/7321073
35. Zlatev, Julian J. "I May Not Agree With You, but I Trust You: Caring About Social Issues Signals Integrity." *Psychological Science*, 30, no. 6, 2019. https://doi.org/10.1177/0956797619837948.
36. Zhang, Wei-wei. "Research on computer graphic and image design method based on visual communication technology." In Book: *MATEC Web of Conferences*, 365, 01043 (2022). ISC-BE 2022. https://doi.org/10.1051/matecconf/202236501043

International Journal of Innovative Research in Computer Science and Technology (IJIRCST)
ISSN(Online): 2347-5552, Volume-12, Issue-2, March 2024
https://doi.org/10.55524/ijircst.2024.12.2.25
Article ID IJIRD-1287, Pages 140-148
www.ijircst.org

Machine Learning Approaches in Spatial Data Mining

Shalini Bhaskar Bajaj[1], Ashima Narang[2], and Priyanka Vashisth[3]

[1] Professor, Department of Computer Science & Engineering, Amity University, Gurugram, Haryana, India
[2]Assistant Professor, Department of Computer Science & Engineering, Amity University, Gurugram, Haryana, India
[3]Associate Professor, Department of Computer Science & Engineering, Amity University, Gurugram, Haryana, India

Correspondence should be addressed to Shalini Bhaskar Bajaj; shalinivimal@gmail.com

Received 9 March 2024; Revised 22 March 2024; Accepted 31 March 2024

Copyright © 2024 Made Shalini Bhaskar Bajaj et al. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT- This review paper surveys the integration of machine learning techniques in spatial data mining, a crucial intersection of geographic information systems and data mining. It examines the application of various machine learning algorithms such as classification, regression, clustering, and deep learning in spatial data analysis. The paper discusses challenges like data preprocessing, feature selection, and model interpretability, alongside recent advancements including spatial-temporal analysis and heterogeneous data integration. Through critical analysis of existing literature, it identifies trends, methodologies, and future research directions. Practical implications and applications across domains like urban planning, environmental monitoring, and epidemiology are explored. As a comprehensive resource, this review facilitates understanding and utilization of machine learning approaches for extracting insights from spatial data, benefiting researchers, practitioners, and policymakers alike.

KEYWORDS- Spatial data mining, machine learning, geographic information systems, classification, clustering, spatial-temporal analysis, data integration

I. INTRODUCTION

Spatial data mining is a specialized field within data mining that focuses on extracting valuable patterns, relationships, and insights from spatial datasets. These datasets contain geographic information associated with specific locations on Earth's surface, such as satellite imagery, GPS coordinates, land cover data, and environmental measurements. The integration of machine learning techniques with spatial data mining has revolutionized the field, enabling more efficient and accurate analysis of complex spatial phenomena. This introduction provides an overview of spatial data mining, discusses the importance of machine learning in this context, outlines the objectives of the review paper, and describes its structure.

A. Overview of Spatial Data Mining

Spatial data mining involves the application of data mining techniques to spatial datasets for knowledge discovery and decision-making. By analyzing spatial relationships, trends, and patterns, spatial data mining enables insights into various phenomena, such as land cover changes, urban sprawl, environmental degradation, and disease spread. Spatial data mining tasks include spatial pattern recognition, spatial clustering, predictive modeling, and anomaly detection. These tasks are crucial for addressing challenges in diverse domains, including environmental science, urban planning [2, 6, 7, 9, 12, 13], public health, agriculture, and disaster management. The integration of advanced analytical techniques, such as machine learning, enhances the effectiveness and efficiency of spatial data mining processes.

B. Importance of Machine Learning in Spatial Data Mining

Machine learning plays a vital role in spatial data mining by offering powerful tools to analyze and interpret complex spatial datasets. Traditional spatial analysis methods often struggle to handle the high dimensionality, complexity, and heterogeneity of spatial data. Machine learning algorithms, however, excel at processing large volumes of spatial data, identifying patterns, and making predictions. They can handle spatial autocorrelation, non-linearity, and complex relationships inherent in spatial datasets, enhancing the accuracy and efficiency of spatial data mining tasks. Moreover, machine learning enables the automation of spatial analysis processes, reducing human effort and time required for data interpretation. By leveraging machine learning techniques, researchers and practitioners can unlock the full potential of spatial data, leading to improved decision-making in various domains.

This review paper: (i) Provide a comprehensive overview of machine learning approaches utilized in spatial data mining; (ii) Analyze and compare the effectiveness of different machine learning algorithms in addressing various spatial data mining tasks; (iii) Explore the applications of machine learning techniques in real-world scenarios across diverse domains; (iv) Identify key challenges and limitations associated with the integration of machine learning in spatial data mining and propose potential solutions or avenues for future research; (v) Offer insights into emerging trends and future directions in the field of machine learning approaches in spatial data mining, highlighting opportunities for innovation and advancement.

The paper has been organized as follows: Section 2 discusses fundamentals of spatial data mining; section 3 focusses on machine learning techniques for spatial data mining; section 4 discusses spatial pattern recognition; section 5 gives details on predictive modelling in spatial data mining; section 6 elaborates on spatial data clustering;

section 7 mentions anomaly detection and outlier analysis; section 8 discusses challenges and future directions, and last section gives conclusions.

II. FUNDAMENTALS OF SPATIAL DATA MINING

Spatial data mining is a specialized field within data mining that deals with extracting valuable insights and patterns from spatial datasets. This section explores the definition and scope of spatial data mining, key concepts, techniques, and challenges inherent in spatial data analysis.

A. Definition and Scope of Spatial data Mining

Spatial data mining involves the application of data mining techniques to spatial datasets, which contain geographic information associated with specific locations on Earth's surface. The scope of spatial data mining extends across various domains, including environmental science, urban planning [2, 6, 7, 9, 12, 13], public health, agriculture, and disaster management. By analyzing spatial relationships, trends, and patterns, spatial data mining enables insights into various phenomena, such as land cover changes, urban sprawl, environmental degradation, and disease spread.

B. Key Concepts in Spatial Data Mining

In spatial data mining, the key concepts are spatial data representation, spatial autocorrelation, spatial clustering, spatial regression, and spatial pattern recognition. Spatial data can be represented in different formats, including raster and vector. Raster data represent spatial information as a grid of cells, while vector data represent spatial features as points, lines, and polygons. Spatial autocorrelation refers to the degree of correlation between spatially adjacent observations within a dataset. It indicates the extent to which similar values or attributes cluster together in space. Spatial clustering involves grouping spatial objects or locations based on their spatial proximity or similarity. This task aims to identify spatially coherent groups of features or phenomena within a dataset. Spatial regression is a statistical technique used to model the relationship between a dependent variable and one or more independent variables, while accounting for spatial dependencies. Spatial pattern recognition involves identifying and categorizing spatial features or structures within a dataset. It aims to detect and analyze spatial patterns such as clusters, hotspots, trends, and anomalies.

C. Challenges in Spatial data Mining

Challenges in spatial data mining include data heterogeneity, scalability, spatial autocorrelation, interpretability, and domain-specific considerations. Spatial datasets often comprise heterogeneous data types, formats, and resolutions. Integrating and analyzing diverse data sources require specialized techniques for data preprocessing, transformation, and fusion. Spatial datasets can be massive in size, posing challenges in terms of scalability and computational efficiency. Analyzing large-scale spatial data requires scalable algorithms and parallel processing techniques. Spatial data often exhibit spatial autocorrelation, meaning that nearby locations tend to have similar values or attributes. Ignoring spatial autocorrelation can bias statistical analysis results and lead to inaccurate interpretations of spatial patterns and relationships. Interpreting spatial data mining results and understanding the underlying spatial patterns can be challenging, particularly with complex machine learning models. Ensuring the interpretability and explainability of spatial data mining methods is crucial for gaining insights and informing decision-making processes. Spatial data mining tasks often involve domain-specific considerations and constraints, such as environmental regulations, land use policies, and socio-economic factors. Integrating domain knowledge and expertise into spatial data mining workflows is essential for producing meaningful and actionable results.

In summary, understanding the fundamentals of spatial data mining, including key concepts, techniques, and challenges, is essential for effectively analyzing spatial datasets and extracting valuable insights for decision-making in various domains. By addressing challenges such as data heterogeneity, scalability, spatial autocorrelation, interpretability, and domain-specific considerations, researchers and practitioners can unlock the full potential of spatial data mining and derive actionable insights from complex spatial datasets.

III. MACHINE LEARNING TECHNIQUES FOR SPATIAL DATA MINING

Machine learning techniques have revolutionized spatial data mining by providing powerful tools to analyze complex spatial datasets and extract valuable insights. This section provides an overview of machine learning in spatial data mining, covering supervised and unsupervised learning techniques, as well as deep learning approaches.

A. Overview of Machine Learning in Spatial Data Mining

Machine learning encompasses a diverse set of algorithms and techniques that enable computers to learn from data and make predictions or decisions without being explicitly programmed. In the context of spatial data mining, machine learning algorithms can analyze spatial datasets to identify patterns, relationships, and anomalies, facilitating decision-making processes across various domains.

B. Supervised Learning Techniques

Supervised learning involves training a model on a labeled dataset, where the input data is paired with corresponding output labels. Supervised learning techniques are commonly used for regression and classification tasks in spatial data mining. (a) Regression Models: Regression models aim to predict a continuous target variable based on one or more input features. Examples of regression techniques used in spatial data mining include linear regression, polynomial regression, and geographically weighted regression; (b) Support Vector Machines (SVM): SVM is a powerful supervised learning algorithm used for classification and regression tasks. It works by finding the optimal hyperplane that separates different classes or predicts continuous outcomes while maximizing the margin between data points; (c) Decision Trees: Decision trees are tree-like structures where each internal node represents a decision based on a feature, and each leaf node represents the outcome or prediction. Decision trees are intuitive and easy to interpret, making them suitable for spatial data mining tasks such as land cover classification and vegetation mapping; (d) Random Forests**:** Random

forests [15] are an ensemble learning technique that combines multiple decision trees to improve predictive performance and reduce overfitting. Random forests are robust and versatile, making them well-suited for spatial data mining tasks with high-dimensional and noisy datasets; (e) Neural Networks: Neural networks [28] are a class of machine learning models inspired by the structure and function of the human brain. They consist of interconnected layers of neurons that process input data and produce output predictions. Neural networks, including feedforward neural networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs), are increasingly used in spatial data mining for tasks such as image classification, time-series analysis, and spatial prediction.

C. Unsupervised Learning Techniques

Unsupervised learning involves training a model on an unlabeled dataset to find hidden patterns or structure in the data. Unsupervised learning techniques are commonly used for clustering and dimensionality reduction tasks in spatial data mining. (a) Clustering Algorithms: Clustering algorithms group similar data points together based on their distance or similarity measures. Common clustering algorithms used in spatial data mining include k-means clustering, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), and hierarchical clustering. These algorithms are used for tasks such as hotspot detection, land use classification, and urban growth analysis; (b) Dimensionality Reduction: Dimensionality reduction techniques aim to reduce the number of features or variables in a dataset while preserving its essential information. Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE) are popular dimensionality reduction techniques used in spatial data mining to visualize high-dimensional data, identify important features, and improve computational efficiency.

D. Deep Learning Approaches

Deep learning is a subset of machine learning that utilizes artificial neural networks with multiple layers of abstraction to learn complex patterns from data. Deep learning approaches have gained popularity in spatial data mining due to their ability to handle large-scale, high-dimensional spatial datasets and extract hierarchical representations of spatial features. (a) Convolutional Neural Networks (CNNs): CNNs are a class of deep neural networks designed for processing spatial data, such as images and grids. CNNs use convolutional layers to automatically learn spatial hierarchies of features from raw input data, making them well-suited for tasks such as image classification, object detection, and land cover mapping; (b) Recurrent Neural Networks (RNNs): RNNs are a class of neural networks designed for processing sequential data with temporal dependencies. RNNs are commonly used in spatial data mining for tasks such as time-series analysis, trajectory prediction, and spatial sequence modelling; (c) Graph Neural Networks (GNNs): GNNs are a class of neural networks designed for processing graph-structured data, such as social networks, transportation networks, and spatial networks. GNNs leverage graph convolutional layers to learn spatial dependencies and relational patterns from graph-structured data, enabling tasks such as spatial reasoning, graph-based clustering, and network analysis in spatial data mining.

In summary, machine learning techniques, including supervised and unsupervised learning algorithms, as well as deep learning approaches, have revolutionized spatial data mining by enabling the analysis of complex spatial datasets and extraction of valuable insights for decision-making across various domains. By leveraging machine learning techniques, researchers and practitioners can unlock the full potential of spatial data mining and address real-world challenges in environmental science, urban planning, public health, agriculture, and disaster management.

IV. SPATIAL PATTERN RECOGNITION

Spatial pattern recognition is a fundamental task in spatial data mining, aiming to identify and categorize spatial features or structures within a dataset. This section explores the application of machine learning for spatial pattern recognition, including case studies and applications such as land cover classification, object detection in satellite imagery, and urban scene understanding [2, 6, 7, 9, 12, 13].

A. Applications of Machine Learning for Spatial Pattern Recognition

Machine learning techniques play a crucial role in spatial pattern recognition by enabling the automated detection and characterization of spatial features and structures in complex datasets. Traditional spatial analysis methods often struggle to handle the high dimensionality and complexity of spatial data, making machine learning approaches particularly valuable for spatial pattern recognition tasks. Machine learning algorithms, including supervised, unsupervised, and deep learning techniques, are used for spatial pattern recognition tasks such as: (a) Classification: Classifying spatial features into predefined categories or classes based on their attributes and spatial characteristics; (b) Detection: Detecting and localizing specific objects or phenomena within spatial datasets, such as buildings, roads, vegetation, and land cover types; (c) Segmentation: Partitioning spatial datasets into homogeneous regions or segments based on their spatial and spectral properties; (d) Prediction: Forecasting future spatial patterns or trends based on historical data and spatial relationships.

These machine learning techniques leverage various types of spatial data, including satellite imagery, aerial photographs, LiDAR data, and geographic information systems (GIS) layers [19], to extract meaningful spatial patterns and structures.

B. Case Studies and Applications

Land Cover Classification: Land cover classification is a common application of spatial pattern recognition, aiming to classify different land cover types (e.g., forests, water bodies, urban areas) within a geographic region. Machine learning algorithms [18, 26], such as random forests, support vector machines (SVM), and convolutional neural networks (CNNs), are used to classify land cover types based on spectral, spatial, and temporal features extracted from satellite imagery.

C. Object Detection in Satellite Imagery

Object detection in satellite imagery [21] involves identifying and locating specific objects or features of interest within large-scale spatial datasets. Machine learning techniques, such as region-based convolutional neural networks (R-CNN), You Only Look Once (YOLO), and Single Shot Multibox Detector (SSD) [8], are commonly used for object detection tasks in satellite imagery. For instance, a study conducted by Lin et al. (2017) applied a deep learning approach based on SSD to detect and localize vehicles in high-resolution satellite images. The SSD model achieved accurate detection of vehicles in urban and rural areas, demonstrating the potential of deep learning for object detection in satellite imagery.

D. Urban Scene Understanding

Urban scene understanding involves analyzing and interpreting complex spatial patterns and structures within urban environments, such as building footprints, road networks, and land use patterns. Machine learning techniques, including deep learning approaches such as graph convolutional neural networks (GCNs), are used for urban scene understanding tasks, such as semantic segmentation, road extraction, and building footprint detection. For example, a study conducted by Chen et al. (2019) applied a graph-based deep learning approach to perform semantic segmentation of urban scenes using high-resolution aerial imagery. The GCN model effectively segmented urban scenes into semantic classes, such as buildings, roads, vegetation, and water bodies, enabling comprehensive urban scene understanding and analysis.

In summary, spatial pattern recognition is a critical task in spatial data mining, enabling the identification and characterization of spatial features and structures within complex datasets. Machine learning techniques [18, 26], including supervised, unsupervised, and deep learning approaches, play a crucial role in spatial pattern recognition, facilitating automated analysis and interpretation of spatial data across various domains. Through case studies and applications such as land cover classification, object detection in satellite imagery, and urban scene understanding, machine learning demonstrates its effectiveness in extracting meaningful spatial patterns and insights from diverse spatial datasets.

V. PREDICTIVE MODELLING IN SPATIAL DATA MINING

Predictive modeling is a fundamental aspect of spatial data mining, enabling the forecasting of future spatial patterns and phenomena based on historical data and spatial relationships. This section provides an overview of predictive modeling in spatial data mining, covering regression models for spatial prediction, time-series analysis, and forecasting, as well as ensemble learning techniques. Additionally, it explores case studies and applications of predictive modeling in environmental monitoring, public health, and urban planning [2, 6, 7, 9, 12, 13].

A. Overview of Predictive Modeling in Spatial Data Mining

Predictive modeling in spatial data mining involves developing mathematical models that can predict spatial patterns, trends, and events based on input variables and spatial relationships. These models utilize historical data to learn patterns and relationships and then apply them to new spatial datasets to make predictions. Regression models [29], time-series analysis techniques, and ensemble learning methods are commonly used for predictive modeling in spatial data mining [1, 16]. These techniques leverage spatial data such as satellite imagery, sensor data, geographic information systems (GIS) layers, and socioeconomic variables to make spatial predictions.

- **Regression Models for Spatial Prediction**

Regression models are used to predict a continuous target variable based on one or more input features. In spatial data mining, regression models are often used for spatial prediction tasks such as land cover mapping, environmental modeling, and spatially explicit forecasting. Techniques such as linear regression, spatial autoregressive models (SAR), and geographically weighted regression (GWR) are commonly used for spatial prediction tasks. These models consider spatial dependencies between observations and incorporate spatially varying coefficients to account for spatial heterogeneity.

- **Time-Series Analysis and Forecasting**

Time-series analysis involves analyzing sequential data points collected over time to identify patterns, trends, and seasonality. In spatial data mining, time-series analysis techniques are used for forecasting spatial phenomena such as environmental variables, disease outbreaks, and urban growth. Methods such as autoregressive integrated moving average (ARIMA), exponential smoothing, and seasonal decomposition are commonly used for time-series analysis and forecasting. These techniques capture temporal dependencies and seasonality patterns in spatial data to make accurate predictions.

- **Ensemble Learning Techniques**

Ensemble learning techniques combine multiple base models to improve predictive performance and generalization ability. In spatial data mining, ensemble learning methods are used to aggregate predictions from diverse models and reduce prediction errors. Approaches such as random forests, gradient boosting machines (GBM), and ensemble neural networks are commonly used for ensemble learning in spatial data mining. These methods combine predictions from multiple regression models, decision trees, or neural networks to produce more robust and accurate predictions.

B. Case Studies and Applications

- **Environmental Monitoring**

Predictive modeling is widely used in environmental monitoring to forecast environmental variables such as air quality, water quality, and land surface temperature. For example, a study conducted by Li et al. (2018) used machine learning techniques to predict air pollution levels in urban areas based on satellite imagery and meteorological data. The predictive model accurately forecasted air pollution levels, enabling proactive

measures for pollution control and public health protection.

- **Public Health**

Predictive modeling plays a crucial role in public health for forecasting disease outbreaks, identifying high-risk areas, and guiding resource allocation for disease prevention and control. For instance, a study conducted by Yang et al. (2019) used machine learning techniques to predict the spatial spread of dengue fever in urban areas based on environmental factors and historical disease data. The predictive model identified high-risk areas for dengue fever transmission, informing targeted intervention strategies and public health interventions.

- **Urban Planning**

Predictive modeling is essential in urban planning for forecasting urban growth, land use changes, and transportation demand. For example, a study conducted by Shen et al. (2020) used machine learning techniques to predict future land use changes in urban areas based on socioeconomic variables, land cover data, and demographic trends. The predictive model provided insights into future urban development patterns, supporting informed decision-making in urban planning and land management [2, 6, 7, 9, 12, 13].

In summary, predictive modeling is a crucial aspect of spatial data mining, enabling the forecasting of future spatial patterns and phenomena based on historical data and spatial relationships. Regression models [29], time-series analysis techniques, and ensemble learning methods are commonly used for predictive modeling in spatial data mining, facilitating applications in environmental monitoring, public health, urban planning, and other domains. Through case studies and applications, predictive modeling demonstrates its effectiveness in predicting spatial phenomena and informing decision-making processes in various fields.

VI. SPATIAL DATA CLUSTERING

Spatial data clustering is a fundamental task in spatial data mining [14, 17], aiming to group spatial objects or locations based on their spatial proximity or similarity. This section explores the role of machine learning in spatial data clustering, compares different clustering algorithms, and examines case studies and applications such as land use classification, urban growth analysis, and hotspot detection.

A. Role of Machine Learning in Spatial Data Clustering

Machine learning plays a crucial role in spatial data clustering by providing automated and scalable techniques for grouping spatial objects based on their spatial characteristics. Traditional clustering algorithms often struggle to handle the high dimensionality and complexity of spatial datasets, making machine learning approaches particularly valuable for spatial data clustering tasks. Machine learning algorithms [5, 10, 22, 24, 25, 27], including both supervised and unsupervised techniques, are used for spatial data clustering: (a) Supervised Clustering: Supervised clustering techniques use labeled training data to learn clustering patterns and relationships. Examples include fuzzy clustering algorithms, self-organizing maps (SOM), and hierarchical clustering with labeled data; (b) Unsupervised Clustering: Unsupervised clustering techniques group spatial objects without using labeled training data. Common unsupervised clustering algorithms used in spatial data mining include k-means clustering, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [11], and spectral clustering. These machine learning techniques leverage spatial data such as satellite imagery, geographic information systems (GIS) layers, and sensor data to identify spatial patterns and structures in complex datasets.

B. Comparison of Clustering Algorithms

Several clustering algorithms are commonly used in spatial data mining, each with their strengths and limitations. A comparison of clustering algorithms can help researchers and practitioners choose the most suitable approach for their specific spatial data clustering [14, 17] tasks. (a) K-means Clustering: K-means clustering is a simple and widely used algorithm that partitions data into k clusters based on the mean distance between data points and cluster centroids. It is computationally efficient and works well for globular-shaped clusters but may struggle with non-linear or irregularly shaped clusters; (b) DBSCAN: DBSCAN [11] is a density-based clustering algorithm that groups spatial objects based on their density distribution. It can identify clusters of arbitrary shape and is robust to noise and outliers. However, DBSCAN requires careful parameter tuning and may struggle with datasets of varying density or clusters with different shapes and sizes; (c) Spectral Clustering: Spectral clustering is a graph-based clustering algorithm that partitions data based on the eigenvalues and eigenvectors of a similarity matrix. It can identify clusters with complex shapes and is robust to noise and outliers. However, spectral clustering may be computationally intensive and require pre-processing steps such as affinity matrix construction and eigenvector decomposition.

C. Case Studies and Applications

- **Land Use Classification**

Spatial data clustering is widely used in land use classification to identify and categorize different land cover types within a geographic area. Machine learning techniques [5, 10, 22, 24, 25, 27] such as k-means clustering, spectral clustering, and fuzzy clustering are applied to satellite imagery and remote sensing data to classify land cover types such as forests, water bodies, urban areas, and agricultural land. For example, a study conducted by Liu et al. (2019) used k-means clustering to classify land cover types in urban areas based on multispectral satellite imagery. The clustering algorithm accurately identified different land cover classes, enabling comprehensive land use mapping and analysis.

- **Urban Growth Analysis**

Spatial data clustering is essential for urban growth analysis to identify spatial patterns and trends in urban development over time. Machine learning techniques such as DBSCAN, hierarchical clustering, and self-organizing maps are used to cluster urban features such as buildings, roads, and infrastructure to analyze urban growth patterns and trajectories. For instance, a study conducted by Li et al. (2020) used hierarchical clustering to analyze urban growth patterns in rapidly developing regions based on time-series satellite imagery. The clustering algorithm identified distinct spatial clusters representing different

stages of urban development, providing insights into urban growth dynamics and land use change processes [2, 6, 7, 9, 12, 13].

- **Hotspot Detection**

Spatial data clustering is employed for hotspot detection to identify areas with unusually high or low concentrations of spatial phenomena such as disease outbreaks, crime incidents, and environmental hazards. Machine learning techniques such as DBSCAN, spatial scan statistics, and kernel density estimation are used to cluster spatial events and detect hotspots based on their spatial distributions. For example, a study conducted by Wang et al. (2018) used DBSCAN clustering to detect hotspots of dengue fever transmission based on spatially aggregated disease incidence data. The clustering algorithm identified clusters of high disease transmission risk, enabling targeted intervention strategies and public health interventions.

In summary, spatial data clustering plays a crucial role in spatial data mining by enabling the identification of spatial patterns and structures within complex datasets. Machine learning techniques such as k-means clustering, DBSCAN, and spectral clustering are commonly used for spatial data clustering tasks, facilitating applications such as land use classification, urban growth analysis, and hotspot detection. Through case studies and applications, spatial data clustering demonstrates its effectiveness in extracting meaningful spatial patterns and insights from diverse spatial datasets.

VII. ANOMALY DETECTION AND OUTLIER ANALYSIS

Anomaly detection and outlier analysis are essential tasks in spatial data mining, aiming to identify unusual patterns or data points that deviate significantly from the norm. This section explores machine learning approaches for anomaly detection, applications in various domains such as cybersecurity and environmental monitoring and provides case studies and examples.

A. Machine Learning Approaches for Anomaly Detection

Machine learning techniques [5, 10, 22, 24, 25, 27] play a crucial role in anomaly detection by providing automated and scalable methods for identifying anomalous patterns or outliers in spatial datasets. These techniques leverage statistical models, clustering algorithms, and supervised learning methods to detect deviations from expected patterns or behaviors. Common machine learning approaches for anomaly detection include:

- **Statistical Methods**

Statistical methods such as z-score, Dixon's Q-test, and Grubbs' test are used to detect anomalies based on statistical properties of the data, such as mean, standard deviation, and distribution. These methods compare observed data points to expected values and flag outliers that deviate significantly from the norm.

- **Clustering Algorithms**

Clustering algorithms such as k-means, DBSCAN, and hierarchical clustering can be used for anomaly detection by identifying clusters of normal behavior and flagging data points that do not belong to any cluster or are isolated from the rest of the data. Outliers are often defined as data points that are distant from their nearest neighbors or have low cluster membership probabilities.

- **Supervised Learning Methods**

Supervised learning methods such as support vector machines (SVM), decision trees, and random forests [15] can be trained on labeled data to distinguish between normal and anomalous patterns. These methods learn patterns of normal behavior from training data and classify new data points as either normal or anomalous based on their learned models.

- **Deep Learning Approaches**

Deep learning approaches [3, 4, 23] such as autoencoders, recurrent neural networks (RNNs), and convolutional neural networks (CNNs) are increasingly used for anomaly detection in spatial data. These methods learn hierarchical representations of spatial features and identify anomalies based on deviations from learned representations of normal behaviors.

B. Applications in Cybersecurity, Environmental Monitoring, Healthcare and Finance

Anomaly detection and outlier analysis have diverse applications across various domains, including: (a) Cybersecurity: Anomaly detection is critical for cybersecurity to identify malicious activities, intrusions, and cybersecurity threats in network traffic, system logs, and user behavior. Machine learning techniques are used to analyze network traffic patterns, detect anomalies indicative of cyberattacks, and prevent security breaches; (b) Environmental Monitoring: Anomaly detection is essential for environmental monitoring to identify abnormal patterns or events in environmental variables such as air quality, water quality, and climate data. Machine learning techniques are used to detect anomalies indicative of environmental hazards, pollution incidents, and ecological disruptions; (c) Healthcare: Anomaly detection is employed in healthcare to identify unusual patterns or deviations from normal physiological parameters in patient data such as vital signs, medical images, and electronic health records. Machine learning techniques are used to detect anomalies indicative of medical conditions, disease outbreaks, and adverse health events; (d) Finance: Anomaly detection is utilized in finance to detect fraudulent transactions, money laundering activities, and financial anomalies in transaction data, stock prices, and market behavior. Machine learning techniques are used to analyze financial data patterns, detect anomalies indicative of fraudulent activities, and mitigate financial risks.

C. Case Studies and Examples

- **Cybersecurity**

In cybersecurity, anomaly detection is used to identify suspicious network traffic patterns indicative of cyberattacks. For example, a study conducted by Gao et al. (2020) used machine learning techniques to detect distributed denial-of-service (DDoS) attacks based on anomalous network traffic patterns. The anomaly detection model accurately identified DDoS attacks and helped mitigate cybersecurity threats in real-time.

- **Environmental Monitoring**

In environmental monitoring, anomaly detection is used to identify abnormal patterns or events in environmental

variables such as air quality, water quality, and climate data. For example, a study conducted by Jia et al. (2019) used machine learning techniques to detect anomalies in water quality data collected from river monitoring stations. The anomaly detection model identified abnormal water quality events such as pollution incidents and helped improve water resource management and environmental protection efforts.

- **Healthcare**

In healthcare, anomaly detection is used to identify unusual patterns or deviations from normal physiological parameters in patient data. For example, a study conducted by Choi et al. (2021) used machine learning techniques to detect anomalies in electrocardiogram (ECG) signals indicative of cardiac arrhythmias. The anomaly detection model accurately identified abnormal ECG patterns and helped diagnose cardiovascular diseases in patients.

In summary, anomaly detection and outlier analysis are essential tasks in spatial data mining, enabling the identification of unusual patterns or data points that deviate significantly from the norm. Machine learning approaches such as statistical methods, clustering algorithms, supervised learning methods, and deep learning approaches are used for anomaly detection in various domains such as cybersecurity, environmental monitoring, healthcare, and finance. Through case studies and examples, anomaly detection demonstrates its effectiveness in identifying anomalies and mitigating risks in diverse spatial datasets.

VIII. CHALLENGES AND FUTURE DIRECTIONS

Machine learning techniques have revolutionized spatial data mining, offering powerful tools for extracting valuable insights from complex spatial datasets. However, several challenges persist, and emerging trends present new opportunities for research and development in this field. This section explores the current challenges in machine learning for spatial data mining, emerging trends and opportunities, future research directions, and ethical considerations and implications.

A. Current Challenges in Machine Learning for Spatial Data Mining

- **Data Quality and Heterogeneity**

Spatial datasets often exhibit varying levels of quality, resolution, and accuracy, posing challenges in data integration, preprocessing, and analysis. Handling data heterogeneity and ensuring data quality are crucial for accurate and reliable spatial data mining.

- **Scalability**

Analyzing large-scale spatial datasets can be computationally intensive and resource-intensive, posing challenges in terms of scalability and efficiency. Developing scalable machine learning algorithms and parallel processing techniques is essential for handling massive spatial datasets effectively.

- **Spatial Autocorrelation**

Spatial data often exhibit spatial autocorrelation, where nearby locations tend to have similar values or attributes. Ignoring spatial autocorrelation can bias analysis results and lead to inaccurate spatial predictions. Incorporating spatial dependencies into machine learning models is necessary for addressing this challenge.

- **Interpretability**

Complex machine learning models such as deep neural networks [28] often lack interpretability, making it challenging to understand and explain model predictions. Ensuring the interpretability and transparency of machine learning models is crucial for gaining trust and acceptance in spatial data mining applications.

- **Domain-specific Considerations**

Spatial data mining tasks often involve domain-specific considerations and constraints, such as environmental regulations, land use policies, and socio-economic factors. Integrating domain knowledge and expertise into machine learning workflows is essential for producing meaningful and actionable results.

B. Emerging Trends and Opportunities

- **Geospatial Deep Learning**

Deep learning approaches, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and graph neural networks (GNNs), are increasingly applied to spatial data mining tasks. Geospatial deep learning [3, 4, 23] enables the automatic extraction of hierarchical spatial features and representations from raw spatial data, facilitating advanced analysis and prediction tasks.

- **Transfer Learning and Domain Adaptation**

Transfer learning techniques, which leverage knowledge learned from one domain to improve performance in another domain, are gaining popularity in spatial data mining. Transfer learning and domain adaptation methods enable the transfer of spatial patterns and representations learned from one geographic region to another, overcoming data scarcity and domain shift challenges.

- **Explainable AI(XAI)**

Explainable AI techniques aim to enhance the interpretability and transparency of machine learning models, enabling users to understand and trust model predictions. XAI methods such as model interpretability techniques, model-agnostic explanations, and interactive visualization tools are increasingly applied to spatial data mining tasks to facilitate decision-making and knowledge discovery.

- **Spatial-Temporal Modeling**

Spatial-temporal modeling techniques integrate spatial and temporal dimensions into machine learning models, enabling the analysis of dynamic spatial processes and phenomena over time. Spatial-temporal modeling approaches are applied to various applications such as environmental monitoring, urban planning, and transportation management, facilitating spatiotemporal prediction, forecasting, and decision-making.

C. Future Research Directions

- **Addressing Data Quality and Heterogeneity**

Future research should focus on developing robust techniques for handling data heterogeneity, ensuring data quality, and integrating heterogeneous spatial datasets from diverse sources. Methods for uncertainty quantification and error propagation in spatial data mining should also be explored.

- **Advancing Scalable and Efficient Algorithms**

Future research should aim to develop scalable machine learning algorithms and parallel processing techniques capable of handling large-scale spatial datasets efficiently. Distributed computing frameworks and cloud-based platforms can be leveraged to improve scalability and computational efficiency in spatial data mining applications.

- **Enhancing Interpretability and Transparency**

Future research should focus on enhancing the interpretability and transparency of machine learning models in spatial data mining. Developments in explainable AI techniques, model interpretability methods, and interactive visualization tools can facilitate model understanding and trust in spatial data analysis tasks.

- **Integrating Domain Knowledge and Expertise**

Future research should emphasize the integration of domain knowledge and expertise into machine learning workflows in spatial data mining. Collaborative research efforts between machine learning researchers and domain experts can lead to the development of domain-specific models, algorithms, and tools tailored to specific application domains.

D. Ethical Considerations and Implications

- **Privacy and Data Security**

Spatial data mining raises concerns about privacy and data security, particularly when dealing with sensitive or personal location-based data. Ensuring data anonymization, encryption, and secure data handling practices are essential for protecting individuals' privacy and preventing unauthorized access to spatial data.

- **Bias and Fairness**

Machine learning models used in spatial data mining may exhibit biases and unfairness, leading to discriminatory outcomes or decisions. Addressing biases and ensuring fairness in machine learning models require careful consideration of data sampling, feature selection, and model evaluation methods to mitigate bias and promote fairness in spatial data analysis tasks.

- **Transparency and Accountability**

Transparency and accountability are essential principles in spatial data mining to ensure that machine learning models are transparent, accountable, and explainable. Providing clear explanations of model predictions, documenting model development processes, and establishing mechanisms for model auditing and oversight can enhance transparency and accountability in spatial data mining applications.

- **Environmental and Societal Impact**

Spatial data mining has significant environmental and societal implications, affecting land use, natural resource management, urban planning, public health, and transportation. Ethical considerations such as environmental sustainability, social equity, and community engagement should be integrated into spatial data mining research and practice to ensure responsible and ethical use of spatial data and machine learning techniques.

In conclusion, while machine learning has greatly advanced spatial data mining, several challenges remain, and new opportunities are emerging. Addressing these challenges and embracing emerging trends will shape the future of spatial data mining, enabling researchers and practitioners to unlock the full potential of machine learning for addressing complex spatial challenges and promoting sustainable and equitable development. Moreover, ethical considerations and implications are essential to ensure that spatial data mining is conducted responsibly, ethically, and for the benefit of society.

IX. CONCLUSIONS

In this review paper, we have explored the role of machine learning approaches in spatial data mining, covering various aspects such as spatial data clustering, predictive modeling, anomaly detection, and outlier analysis. We discussed the fundamentals of spatial data mining, including spatial data representation, spatial autocorrelation, spatial clustering, spatial regression, and spatial pattern recognition. Moreover, we examined the challenges and future directions in machine learning for spatial data mining, as well as emerging trends and opportunities in the field. Machine learning techniques have shown significant potential in addressing various challenges in spatial data mining, including data heterogeneity, scalability, spatial autocorrelation, and interpretability. These techniques have been applied to a wide range of applications, including environmental monitoring, urban planning, public health, cybersecurity, and finance. Through case studies and examples, we demonstrated the effectiveness of machine learning in extracting meaningful spatial patterns and insights from diverse spatial datasets.

A. Implications for Research and Practice

The findings of this review paper have several implications for both research and practice in spatial data mining: (a) Research Implications: Researchers in spatial data mining should focus on addressing current challenges such as data heterogeneity, scalability, spatial autocorrelation, and interpretability. Future research directions include advancing scalable and efficient algorithms, enhancing model interpretability and transparency, integrating domain knowledge and expertise, and addressing ethical considerations and implications in spatial data mining; (b) Practical Implications: Practitioners in spatial data mining can leverage machine learning techniques to tackle real-world challenges and applications such as environmental monitoring, urban planning, public health, cybersecurity, and finance. By adopting state-of-the-art machine learning approaches, practitioners can extract valuable insights from spatial datasets, improve decision-making processes, and address societal and environmental challenges effectively.

In conclusion, machine learning approaches play a crucial role in spatial data mining, enabling the extraction of valuable insights from complex spatial datasets. Despite various challenges, emerging trends and opportunities in machine learning offer promising avenues for advancing research and practice in spatial data mining. By addressing current challenges, embracing emerging trends, and considering ethical implications, researchers and practitioners can harness the power of machine learning to unlock the full potential of spatial data mining for addressing societal, environmental, and economic challenges in the future.

CONFLICT OF INTERREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] L. Anselin, “Spatial Econometrics: Methods and Models,” Springer, 2019.
[2] Y. Chen et al., “Urban Scene Understanding via Graph Convolutional Networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
[3] J. Choi et al., “Deep Learning-Based Arrhythmia Detection from Ambulatory Electrocardiography Data,” Sensors, vol. 21, no. 3, p. 842, 2021.
[4] J. Gao et al., “A Deep Learning Approach to Detect DDoS Attacks Using Convolutional Neural Networks,” Future Generation Computer Systems, vol. 102, pp. 1122-1131, 2020.
[5] X. Jia et al., “Anomaly Detection in Water Quality Monitoring Data Using Machine Learning Techniques,” Environmental Science and Pollution Research, vol. 26, no. 31, pp. 31680-31691, 2019.
[6] X. Li et al., “Deep Learning-Based Air Quality Prediction in Urban Areas,” Environmental Pollution, vol. 242, pp. 608-618, 2018.
[7] Y. Li et al., “Hierarchical Clustering Analysis of Urban Growth Patterns Using Remote Sensing Data,” Remote Sensing, vol. 12, no. 7, p. 1146, 2020.
[8] T. Y. Lin et al., “Single Shot Multibox Detector,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
[9] W. Liu et al., “Land Cover Classification Using K-Means Clustering in Urban Areas,” Remote Sensing, vol. 11, no. 8, p. 896, 2019.
[10] Z. Shen et al., “Predicting Future Land Use Changes Using Machine Learning Techniques,” Land Use Policy, vol. 90, p. 104305, 2020.
[11] H. Wang et al., “Anomaly Detection of Dengue Fever Transmission Using DBSCAN Clustering,” International Journal of Environmental Research and Public Health, vol. 15, no. 11, p. 2434, 2018.
[12] S. Yang et al., “Machine Learning-Based Prediction of Dengue Fever Spread in Urban Areas,” International Journal of Environmental Research and Public Health, vol. 16, no. 3, p. 443, 2019.
[13] J. Zhang et al., “Deep Learning-Based Land Cover Classification in Urban Areas Using High-Resolution Satellite Imagery,” Remote Sensing, vol. 12, no. 6, p. 932, 2020.
[14] M. Ankerst et al., “OPTICS: Ordering Points to Identify the Clustering Structure,” ACM SIGMOD Record, vol. 28, no. 2, pp. 49-60, 1999.
[15] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
[16] N. Cressie, “Statistics for Spatial Data,” Wiley, 2015.
[17] B. S. Everitt et al., “Cluster Analysis,” Wiley, 2011.
[18] J. H. Friedman, “Greedy Function Approximation: A Gradient Boosting Machine,” Annals of Statistics, vol. 29, no. 5, pp. 1189-1232, 2001.M. F.
[19] Goodchild et al., “Foundations of Geographic Information Science,” CRC Press, 2007.Hastie, T., et al., The Elements of Statistical Learning: Data Mining, Inference, and Prediction”, Springer, 2009
[20] K. He et al., "Deep Residual Learning for Image Recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[21] B. Lantz, "Machine Learning with R," Packt Publishing Ltd., 2013.
[22] Y. LeCun et al., "Deep Learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.
[23] T. M. Mitchell, "Machine Learning," McGraw Hill, 1997.
[24] K. P. Murphy, "Machine Learning: A Probabilistic Perspective," MIT Press, 2012.
[25] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.
[26] S. Raschka et al., "Python Machine Learning," Packt Publishing Ltd., 2015.
[27] B. Schölkopf et al., "Estimating the Support of a High-Dimensional Distribution," Neural Computation, vol. 13, no. 7, pp. 1443-1471, 2001.
[28] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," Journal of the Royal Statistical Society: Series B (Methodological), vol. 58, no. 1, pp. 267-288, 1996.
[29] I. H. Witten et al., "Data Mining: Practical Machine Learning Tools and Techniques," Morgan Kaufmann, 2016.

ABOUT THE AUTHORS



Dr. Shalini Bhaskar Bajaj completed her Ph.D, from IIT Delhi, M.Tech from Delhi College of Engineering, Delhi University. She has more than 20 years of experience in teaching. She has authored numerous technical research papers and book chapters in international conferences and journals of repute. Her research interests are Databases, Data Mining, Predictive Analytics, Pattern Recognition, Cloud Computing. She is member of IEEE, CSI and IAENG and fellow of Institute of Engineers (India).



Dr. Ashima Narang has completed her Btech, Mtech and PhD in the field of computer Science from various prestigious institutes of India. Her areas of expertise include Cloud Computing, Security, Software Engineering and Virtualization. She has rich teaching experience of 12 years. She has authored numerous research papers in reputed international journals and conferences and has guided students for projects from undergraduate and graduate courses. She is also an active member in the various professional bodies like IAASSE, internet society, SCIEI etc. She is the reviewer to various journals from her expertise field.



Dr. Priyanka Vashisht has received her Ph.D from ThaparUniversity, Patiala, M.Tech from Banasthali Vidyapeeth, Rajsthan. She has 20 years of teaching and industry experience. Her areas of expertise include Cloud Computing, Fog Computing and Virtualization. She has authored numerous technical research papers and book chapters in international conferences and journals of repute. She has authored a book. She has 4 patents in her name. She is member of IETE, CSI and ACM

Check for updates

Received: 14 September 2023 | Revised: 12 February 2024 | Accepted: 17 March 2024

DOI: 10.1111/hequ.12515

RESEARCH ARTICLE

Higher Education Quarterly WILEY

Modelling student issues with MOOCs using TISM-P linkages

Shikha N. Khera | **Himanshu Pawar**

Delhi Technological University, Delhi School of Management, Rohini, India

Correspondence
Himanshu Pawar, Delhi Technological University, Delhi School of Management, Bawana Road, Shahbad Daulatpur Village, Rohini, New Delhi, 110042, India.
Email: himanshu_phd2k18@dtu.ac.in, himanshuihm1@gmail.com

Abstract

To date, student issues with Massive Open Online Courses (MOOCs) have only been explored in context-specific environments. Mainstream problems such as declining student motivation during a course, massive student dropout rates, accountability, user experience, etc., persist due to the permutations and combinations of these issues. Literature is replete with a deep understanding of such problems, but the causal relationships among these issues are less focused upon. We delve into these problems by studying the interrelations among student issues that cause such problems. Garnering insights from students ($N = 149$) and using Total Interpretive Structural Modelling with Polarity (TISM-P), the study has established direct and transitive relations among nine detrimental MOOC-related student issues. The results of the study depict clear positive, negative and transitive relationships between the student issues. Matrice d'Impacts croises-multipication applique' an classment (MICMAC) analysis was also used to assess the driving and dependence power of all issues that further allowed the model to trace out negative and positive pathways of influence. The model constructed in the study will provide a platform for future research to test these interconnections as independent factors affecting problems such as dropout rates, motivation, etc. Therefore, the TISM-P model could further be explored to understand the behaviour of such issues, which might have far-reaching consequences on major existing problems with MOOCs.

 © 2024 John Wiley & Sons Ltd.

1 | INTRODUCTION

The speed with which Massive open online courses (MOOCs) have gained momentum in the last decade and the stupendous increase in enrolments across different platforms has legitimised the place of online learning in mainstream higher education (Jitpaisarnwattana et al., 2022). Learning from the exploits of early MOOCs offering platforms such as MITx (transforming to Edx), various countries such as China (XuetangX), India (Swayam), the UK (Future learn) and Spain (Miríadax), etc., have now developed indigenous platforms for offering MOOCs. Albeit the positive aspects and benefits of MOOCs, such as support in career progression, the gain of new knowledge, learning new skills (Goglio & Bertolini, 2021; Littlejohn et al., 2016), reaching the educationally underserved, delivering low-cost quality higher education (Abu-Shanab & Musleh, 2018), etc. MOOCs have successfully stood the rigour and test of time. Analogous to the new technology adoption graph, that is the Gartner cycle, MOOCs are now beginning to enter the phase of productivity (Corbeil et al., 2018). The spread is massive in terms of the number of student enrolments and a culmination of both global and indigenous platforms is transforming the fabric of online education (Ruipérez-Valiente et al., 2022) Undoubtedly with such a worldwide expansion and collaboration of universities and governments, MOOCs are bound to face criticism on numerous facets of their growth (Chiappe-Laverde et al., 2015; Laaser, 2023). The integration of MOOCs in higher education systems is different across all levels of global social stratification. It is very much dependent upon factors such as early usage, technology adoption rate, university and government support etc. (Wang, Hall, & Wang, 2019; Wang, Liu, & Li, 2019). Thus, the stakeholder issues with such differentiating adoption rates are bound to vary with time and place. Furthermore, being a global phenomenon, it becomes necessary to muster a comprehensive understanding of issues associated with MOOCs. Such an approach might not only shed light on the bigger picture but, can also help understand the interconnections and causal relations among said issues. In the context of education, students will always be considered crucial primary stakeholders, which invariably justifies the need for research encompassing their perceptions, thoughts and opinions, revealing both positive and negative insights. Thus, our study aims to focus on a broader understanding of student issues with MOOCs, which are still as persistent and equally disruptive as they were during the advent of the concept. It is observed that some of the issues had already been resolved in peculiar settings of time and place. However, the majority of them still pose as a hurdle in the effective integration of MOOCs in mainstream higher education systems across the world. For example, language might not be a barrier to online learning in most developed countries. Still, for underdeveloped and developing countries, English as the lingua franca might not be well accepted as the appropriate medium of instruction (Meina et al., 2021). Over the last decade, literature has consolidated and is now replete with how student issues with MOOCs lead to bigger problems in the online education system. Massive dropout rates (Borrella et al., 2022; Goopio & Cheung, 2021; Hew & Cheung, 2014); lack of student motivation and intention to adapt to new technology (Ma & Lee, 2018a); asynchronous nature of majority MOOCs (Alario-Hoyos et al., 2014); low perceived value compared to university degrees (Rosendale, 2017) and doubts regarding the credibility of such courses in the employment market (Farrow et al., 2021; Trehan et al., 2017) are some of the manifestations of these issues. A comprehensive and fair level of understanding of basic issues and their impact is ubiquitous among researchers, but studies on interrelations and causal analysis seem to remain lacking.

Our study aims to draw interconnecting lines among student issues using a total interpretive structural modelling with polarity (TISM-P) model, which is an extended and improvised arm of interpretive structural modelling (ISM) developed by J. Warfield in 1974 (Warfield, 1974). The use of ISM has been voluminous over the years, and it has successfully assisted theory building in answering the 'what' and 'how' questions whilst explaining the driver-dependence relationships among factors. What the causal relationship among factors and how it is formed is well captured by ISM, whereas TISM, as purported by Sushil (2012), emphasises answering the 'why' question (why one factor influences another?), yielding in more interpretive power to complex hierarchical models. TISM has readily gained acceptance in the last decade and researchers have extensively used it for theory building, hypothesis testing, modelling and conceptualising ill-structured mental models

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

and ideas from real-life scenarios. As human beings, we often tend to create mental models of problems and issues surrounding us but, with an increase in the number of influencing factors, the ability of an individual to comprehend the value of such interrelations gets fatigued. TISM helps researchers streamline that process and follows up with a hierarchical model of our cognitive connections, emphasising interpretations from direct and transitive relationships (Kumar & Barua, 2022). The applicability of TISM is increasing with time, and the scale of its horizon is expanding to various fields such as supply chain management, operations and information management, strategic performance management, sustainable supply chain development and emotional intelligence, etc. In the context of higher education TISM has been previously used to understand the challenges of management education (Mahajan et al., 2015) and to create a model of benchmarking for higher education (Yeravdekar & Behl, 2017). Talib and Rahman (2020) used the technique to model the barriers in the growth of higher education institutions (HEIs), whereas Fathi et al. (2019) focused on teamwork training in HEIs. Coming down to the latest development in the TISM technique, an element of polarity was introduced by Sushil (2018) to extend the functioning and novelty of its application; in other words, an aggregated view was presented incorporating positive and negative influence of factors over one another. The extension, that is (TISM-P), allows the individual to capture the positive and/or negative influence of one factor over the other, furnishing a better picture of the behaviour of a model or a system. The application of TISM-P is relatively new, and the literature is still in its emerging forms; hence, as our primary research aim, we look forward to applying the extended version of TISM from students' perspective, which would encourage instructors, e-learning platforms and policymakers to holistically analyse the behaviour of student issues related to MOOCs.

2 | REVIEW OF RELATED LITERATURE

2.1 | Language barrier

One of the oldest issues within the MOOC scenario is the barrier of language (Gul et al., 2018) associated with the medium of instruction and teaching pedagogy. Even though English might be considered the lingua franca, there are countries such as Brazil or China where English is not used in major fashion at universities and colleges (Liu et al., 2010). At present, all major global platforms such as Edx® or Coursera® have begun translating MOOCs with sub-scripts of other languages; however, the level of comfort among the learners is bleak. Students have highlighted their discomfort with the accent and pronunciation of foreign instructors, which makes them feel alienated at certain times (Liyanagunawardena et al., 2014). They have raised concerns over the struggle to cope with the speed, short abbreviations and tone of the language, which results in students watching the same video repeatedly to grasp the concepts (Jones, 2014), thus hampering productivity.

2.2 | Technological barriers

Massive problems arise when technology starts acting as a deterrent for something it was supposed to enhance in the first place (Rhoads, 2015). The use of technology has been well-researched in the context of MOOCs and the challenges it brings with its advancements. For students, it is a major challenge because, even though it is safe to assume that students accessing MOOCs would be competent with new technology, there might be grey areas in developing countries where students lack skills and awareness of such mediums of teaching and learning (Ma & Lee, 2018b). In one way or another other it affects students' intention to use and accept MOOCs as an aid to traditional learning methods (Henderikx et al., 2018; Zulkifli et al., 2020). The lack of appropriate technological infrastructure refrains students from getting connected with the latest advancements in education on the global level (Leontyeva, 2018). MOOCs require high-speed internet connections for buffering and downloading, but the

lack of such connectivity affects students' motivation to access e-learning platforms for learning and development purposes.

2.3 | Overload of information

Students have often complained about the overload and segmentation of information from MOOCs (Chen et al., 2011; Khalid et al., 2021). There is no standardised framework for MOOC development, and the instructors are generally free to design and prepare course information and material based on their own individual experiences and comfort. An issue of expectation mismatch becomes quite real between MOOCs and students when there is information overload on discussion forums along with hefty course readings (Chen et al., 2011). For asynchronous MOOCs, the repetition of ideas on such discussion forums is common via multiple threads of discussions, leading to the segmentation of the core message (Fischer, 2014). The issues are still not deeply researched and often neglected when addressing student issues related to MOOCs.

2.4 | Discussion forum

MOOC discussion forums are mainly in asynchronous mode, depending upon the course and service provider (Li et al., 2018). They are expected to encourage peer learning and improve commitment in a lax learning environment (Galikyan et al., 2021; Zhang et al., 2016). Teaching hundreds and thousands of students online at the same time brings challenges in all sorts of forms, shapes and figures. Students have often failed to engage in the discussion forums due to the overwhelming participation of their peers (Onah et al., 2014), thus restraining themselves from such indulgence. A feeling of alienation in such discussion forums is quite robust, and students' tendency to actively participate is dependent upon instructor-led engagements only (Bali, 2014). Little or no regard towards peer-to-peer and instructor-student associations severely affects the effectiveness of MOOCs in the long run (Naidu, 2017).

2.5 | Student support and feedback system

Feedback is crucial for every learning endeavour to become fruitful and is considered as the bedrock of all systems to function effectively. Without appropriate student support and quality feedback, students are often found to be losing interest in MOOCs (Gregori et al., 2018; Kasch et al., 2021) adding to higher attrition rates (Veletsianos, 2013). An effective student support system is crucial in improving MOOC quality and instructional design, whereas timely feedback helps to augment student interest over a period of time (Alario-Hoyos et al., 2014). The central premise is staying connected to students whilst responding to their queries in an appropriate manner which is supportive and conducive to student learning. The lack of such support systems eventually hampers MOOCs' effectiveness and students' motivation to continue with MOOCs (Gregori et al., 2018).

2.6 | Ambiguous assignments and peer assessment

Course assessments are built on multiple factors such as the type and nature of the course, the duration of the course, the complexities involved in the course and, majorly, the number of student enrolments in the MOOC (Gamage et al., 2021; Muñoz-Merino et al., 2015). Reilly et al. (2014) examined the automated essay scoring (AES) tool for grading assignments and found inefficiencies in its accuracy when compared to instructor-led grading. Peer review in itself is highly ambiguous, and it is hard to track the quality of subjective grading and interpretations.

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

Students often do not agree with peer-reviewed grades (Admiraal et al., 2015) and are found to skip assessments (Eriksson et al., 2016). Ambiguity in designing assessments is also a major concern for students because of misleading and unclear course expectations (Khan et al., 2017), which will lead to more dropouts in the future.

2.7 | Design and quality of MOOCs

Much has been said about how a MOOC should be designed and what quality parameters are good enough to judge upon the effectiveness of a MOOC (Quality Matters, 2014). Trying to sum up quality is a complicated process that demands the right selection of epistemological, methodological and theoretical pre-conditions (Ghislandi, 2016; Stracke & Trisolini, 2021). The quality of MOOCs could be bifurcated into concepts, that is quality assurance and quality enhancement. In the recent past, there has been an increase in the number of universities offering MOOCs as part of credit programmes (Bordoloi et al., 2020), but the awareness and adoption rates are still low in terms of the total number of universities offering credit exchange. Students have also raised concerns about the quality of MOOCs, encompassing the effectiveness of instructional design, course material and appropriateness of content since it directly influences their satisfaction levels towards a MOOC (Albelbisi & Yusop, 2019). Quality enhancement is the umbrella term for parameters concerned with MOOC instructional design, development and delivery at an institutional level (Ghislandi, 2016). Student perception about the quality of a MOOC could take various forms where instructional quality is associated with the effectiveness of teaching methods and overall course design (Fianu et al., 2018) and is a significant predictor of student satisfaction (Pilli & Admiraal, 2017). The issue with the design of MOOCs makes it harder for students to select the right type of MOOC for their academic development. Thus, designing an effective MOOC is hard and requires a balance of appropriate pedagogical approaches, understanding student motivation, investment in material resources and creating active learning environments (Wang, Hall, & Wang, 2019; Wang, Liu, & Li, 2019).

2.8 | Credibility of MOOCs certificates in job market

Studies are now beginning to emerge which are focusing on instilling the credibility of MOOC certificates in the employment market (Rivas et al., 2020). The debate around the value of a MOOC certificate in the eyes of the recruiter and a student's perceived value of that certificate in the job market is leading us to several unanswered questions (Rivas et al., 2020). The distinction is still not clear since the credibility of MOOCs in the employment market mainly affects university students, and there is an existing gap between industry-university expectations (Wells et al., 2009). Furthermore, the value of using online certifications for the upskilling of the workforce is a topic which also lacks serious investigation by researchers (Santandreu Calonge et al., 2019).

2.9 | Teaching and pedagogy

Students have often shown discomfort with the teaching approach of instructors who deliver online classes. Effective teaching reflects a holistic learning experience; instructor intervention at the stages of attraction, orientation, interaction and consolidation is deemed necessary to improve effectiveness in teaching MOOCs (Wong, 2016). Innovative use of technology is found to stimulate new approaches for online learning and student collaboration (Myers et al., 2014). Learning via a connectivist approach in the case of cMOOCs allows for cognitive connections to be made from available learning materials and peer-to-peer interactions (Mackness et al., 2013). On the other hand, xMOOCs offer a more synchronised form of learning where instructions are ordered in the mainstream fashion of a learn-test-learn repeat cycle. Instructors often get confused in selecting the right pedagogical

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

approach for their course (Armellini & Padilla Rodriguez, 2016), and a lack of pedagogical innovations is leading MOOCs towards a passive medium of knowledge acquisition (Parr, 2013).

3 | METHOD

3.1 | Factor validation

Once the student issues were identified from the literature, it was deemed necessary to validate them from a group of experts from the academia and education-technology industry. Judgemental sampling was used to identify ($N=40$) such experts. The experts from the academia ($N=35$) were seasoned faculties who were running at least one MOOC on a global e-learning platform or a government-funded MOOC offering platform. In contrast, the industry experts ($N=5$) were developers of indigenous e-learning platforms. A *t* test for hypothesis testing was applied to capture the experts' opinions.

H0. Null: No significant difference exists between the observed mean and specified mean for attribute Fn.

HA. Alternate: Significant difference exists between the observed mean and specified mean for attribute Fn. ($N=9$).

A questionnaire was prepared using a 5-point Likert scale and exercised online. Whilst testing the value of 3.5 was used (Singh et al., 2019) for factor verification and the results are displayed in (Table 1).

3.2 | TISM-P application

The data for TISM-P was collected from a technical university where MOOCs are an integral part of their higher education curriculum ordinance, bearing transferable credits towards successful completion. The students selected for filling out the questionnaire were taken from a cohort of 3rd year engineering students who had completed at least one MOOC from any e-learning platform (Appendix A) as part of their credit transfer programme. The

TABLE 1 *t* test analysis of factors ($N=40$).

		Test value = 3.5			
S. no.	Factors	Mean	SD	*t*-value	Result[a]
1	Language barrier	3.87	0.563	3.368	Significant
2	Technological barrier	4.47	0.505	12.13	Significant
3	Overload of information	3.87	0.822	2.845	Significant
4	Discussion forum	4.32	0.474	10.933	Significant
5	Student support and feedback systems	4.45	0.503	11.925	Significant
6	Ambiguous assignments and peer assessments	3.8	0.516	3.674	Significant
7	Design and quality of MOOCs	4.65	0.483	15.057	Significant
8	Credibility of MOOCs certificates in job market	3.57	0.5	0.884	Significant
9	Teaching and pedagogy	4.45	0.503	11.925	Significant

[a]Significant if the calculated *t*-statistic value > *t*-critical value (2-tailed).

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

participants were voluntarily asked to fill out the survey because of the complexity involved in the process, and convenience sampling was used. However, due to the hybrid mode, initially, the response rate was extremely low, but upon formal request from their mentors, we were able to get hold of ($N = 149$) students who were willing to fill up the questionnaire under the authors' supervision. We used the TISM questionnaire for simultaneous transitivity checks coupled with polarity approximations to gather student responses. An online meeting was scheduled, and the students were briefed about the process in detail, and it took approximately 30 min for everyone to fill out the questionnaire. Adding polarity is an extension to the original TISM technique, and the steps involved in the advanced version are mentioned in Figure 1 and the illustrations for our research are shortly followed.

3.2.1 | Step I

The contextual definition and brief explanation of the issues were laid down after a thorough review of the associated literature. The issues identified from students' perspectives affect MOOC integration, effectiveness, dropouts and student motivation etc. in a number of ways. Table 2 offers insight into establishing the contextual relationship for paired comparisons with polarity (+ve or −ve) and for interpreting the influence of one factor over the other.

3.2.2 | Step II

A TISM-P questionnaire was developed for paired comparison and simultaneous transitivity checks (Sushil, 2017). The paired comparisons were made in the order 1,2; 2,3; 3,4…; 8,9 by individual students. To collate all responses for the reachability matrix, we followed with the work of Prasad and Suri (2011) to set up a threshold of 70% for direct comparisons and 50% for transitive comparisons. If more than 70% of the responses (direct) were 'Yes' for a particular comparison, then the response was taken as 1; otherwise, 0 in the final reachability matrix (Table 3). The process was followed similarly for transitive comparisons too. The questionnaire captures the direction of the relationship as forward ($i \rightarrow j$), backward ($j \rightarrow i$), both ($i = j$) and no relationship (0). The polarity of direct relations (+ve or −ve) was also taken into consideration at this stage.

3.2.3 | Step III

Transitivity is defined as the relation between any such three elements such that if i influence j and j influences k then i is bound to influence k. In the case of transitive relations, the polarity was justified according to the logic of polarity from the previous relationship. For example: if $i \rightarrow j$ is +ve and $j \rightarrow k$ is −ve, then $i \rightarrow k$ would become −ve (Sushil, 2018). As mentioned in the previous step, the direct pair comparisons were done by students, and afterwards, the transitivity check was performed by authors in MS-Excel®. The polarity of the relationship was specified before the creation of the transitive reachability matrix and successive comparison digraph.

3.2.4 | Step IV

To visualise direct pair comparisons and transitivity checks with polarity, a successive comparison digraph in Figure 2 is portrayed along with a transitive reachability matrix with polarity in Table 4. For easy identification and effective visualisation, the comparisons were marked on the digraph and colour-coded along with the entries of +1, −1 or 0 in the reachability matrix with polarity. Finally, all the entries in the reachability matrix (+1, −1 or 0) were converted into 1 or 0 entries as presented in Table 5 in the form of a transitive reachability matrix without polarity.

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

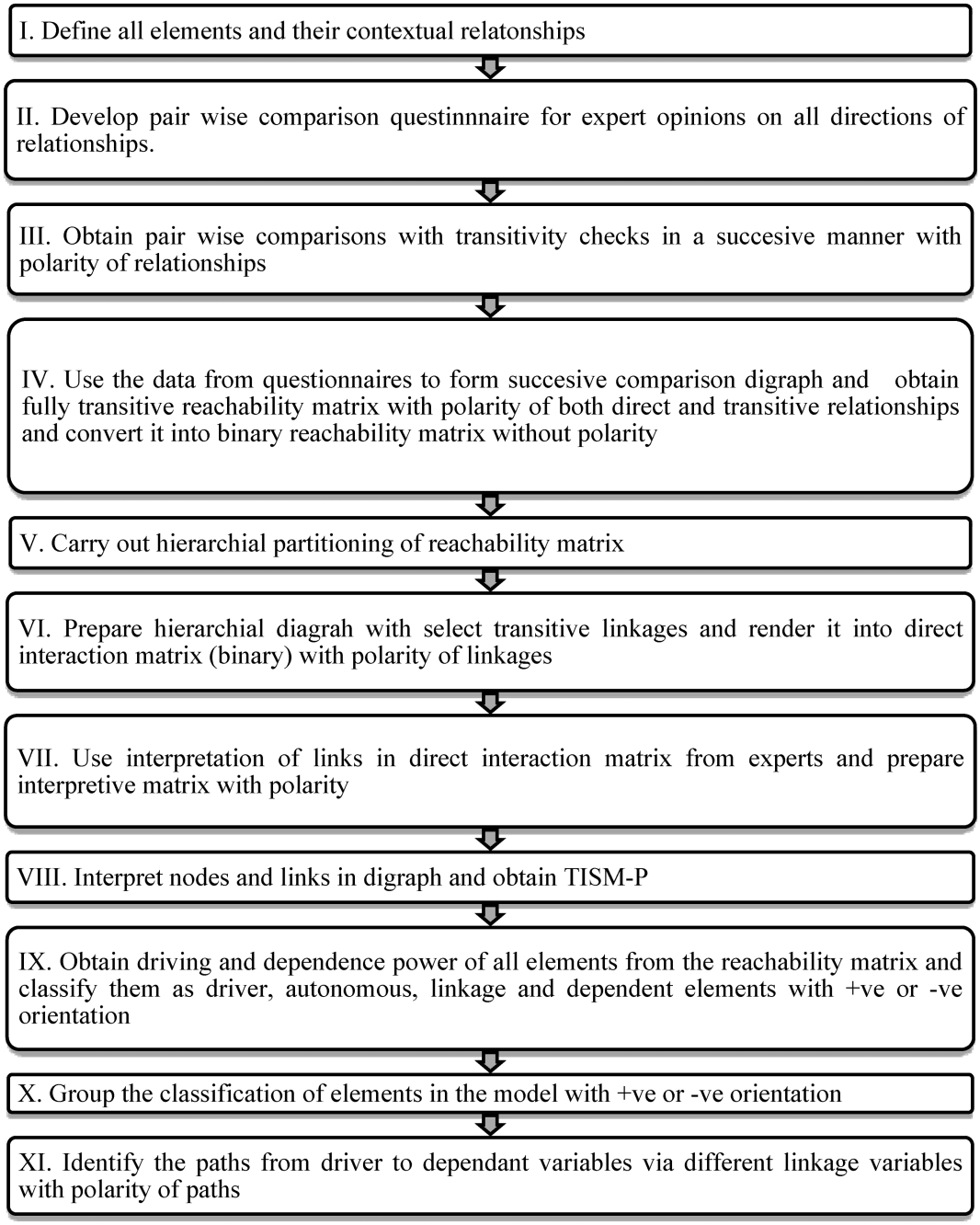


FIGURE 1 Steps involved in TISM-P adapted from Sushil (2018).

3.2.5 | Step V

The matrix from Table 5 is used as the final matrix for hierarchical partitioning. The original ISM/TISM methodology was followed to derive the reachability set, antecedent set and intersection set. The process is iterative and allows us to divide and categorise the student issues into different levels systematically. The student issues having the same reachability and antecedent set are selected for the highest level, and the rest of the levels are attained

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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 Defining the contextual relationship between student issues.

Issue no.	Student issues with MOOCs	Contextual relationship	Interpretation
1	Language barrier	Issue A will influence issue B (positively or negatively	How or in What way will the issue A influence issue B?
2	Technological barrier		
3	Overload of information		
4	Discussion forum		
5	Student support and feedback system		
6	Ambiguous assignments and peer assessments		
7	Design and quality of MOOCs		
8	Credibility of MOOCs certificates in job market		
9	Teaching and pedagogy		

TABLE 3 Final reachability matrix.

Student issues	1	2	3	4	5	6	7	8	9
1	1	0	0	1	0	0	0	0	0
2	0	1	0	0	1	0	1	0	0
3	0	0	1	1	0	0	0	0	0
4	0	0	1	1	1	0	0	0	0
5	0	0	0	1	1	1	0	0	0
6	0	0	0	0	0	1	1	0	0
7	0	0	0	0	0	1	1	1	0
8	0	0	0	0	0	0	1	1	1
9	0	0	0	0	0	0	0	1	1

after iteratively removing previous issues. The illustration of the process is given in Appendix B and the results from level partitioning of student issues are given in Table 6.

3.2.6 | Step VI

The student issues numbering '1' to '9' are arranged in a hierarchical order after level partitioning in the form of a digraph in Figure 3. It is important to note that only prominent transitive links having distinct influence are retained for the final digraph (after model validation by experts). The rest of the links were dropped out to avoid unnecessary crowding. Now, a binary interaction matrix with polarity (Table 7) is made using the final digraph from Figure 3.

3.2.7 | Step VII

Student opinions were used to convert all +1 and −1 entries from the binary matrix into an interpretive matrix in Table 8. The interpretation for each link is given. For example, language barrier (1) has a direct and negative influence on students' intention to participate in discussion forums (4), etc.

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

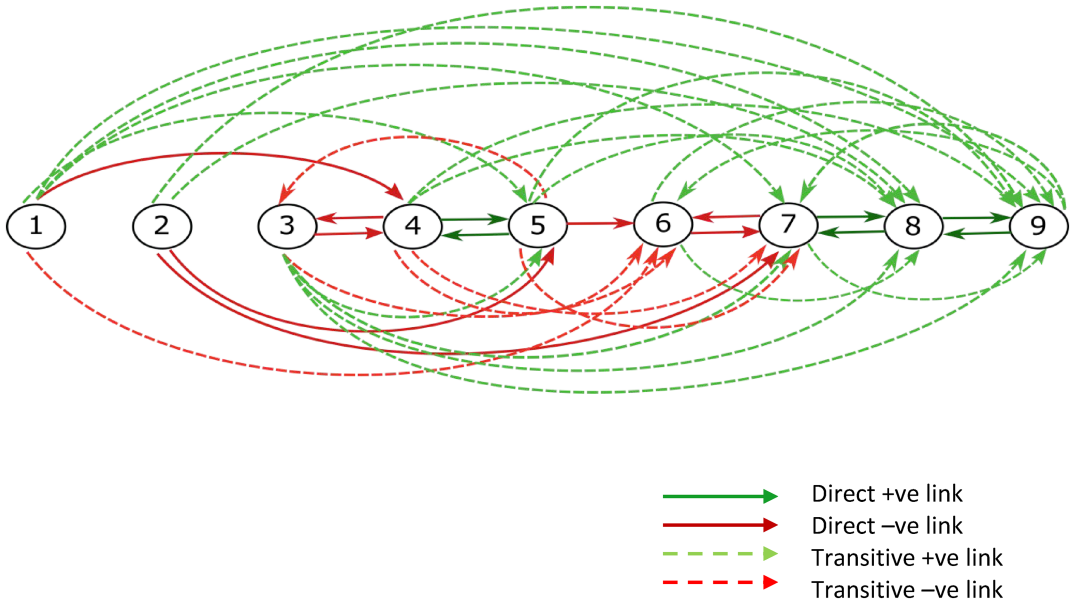


FIGURE 2 Successive comparison digraph with direct and transitive links with polarity of relationships.

TABLE 4 Transitive reachability matrix with polarity of relationships.

Student issues		1	2	3	4	5	6	7	8	9	Driving power +ve	Driving power −ve
1		1	0	0	−1	+1*	−1*	+1*	+1*	+1*	4	2
2		0	1	0	0	−1	0	−1	+1*	+1*	2	2
3		0	0	1	−1	+1*	−1*	+1*	+1*	+1*	4	2
4		0	0	−1	1	+1	−1*	−1*	+1*	+1*	3	3
5		0	0	−1*	+1	1	+1	−1*	+1*	+1*	3	3
6		0	0	0	0	0	1	−1	+1*	+1*	2	1
7		0	0	0	0	0	−1	1	+1	+1*	2	1
8		0	0	0	0	0	0	+1	1	+1	2	0
9		0	0	0	0	0	+1*	+1*	+1	1	3	0
Dependence power	+ve	0	0	0	1	3	2	4	8	8		
	−ve	0	0	2	2	1	4	4	0	0		

*Transitive Relationships.

3.2.8 | Step VIII

The interpretation of both nodes and links from the hierarchical digraph and binary interaction matrix with polarity gives the final TISM-P model of student issues with MOOCs in Figure 4.

3.2.9 | Step IX

The driving and dependence power from the transitive reachability matrix (Table 5) are part of a technique called *Matrice d'Impacts croises-multipication applique' an classment* **MICMAC** analysis (Jothimani et al., 2015). MICMAC

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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 Transitive reachability matrix.

Student issues	1	2	3	4	5	6	7	8	9	Driving power
1	1	0	0	1	1[a]	1[a]	1[a]	1[a]	1[a]	7
2	0	1	0	0	1	0	1	1[a]	1[a]	5
3	0	0	1	1	1[a]	1[a]	1[a]	1[a]	1[a]	7
4	0	0	1	1	1	1[a]	1[a]	1[a]	1[a]	7
5	0	0	1[a]	1	1	1	1[a]	1[a]	1[a]	7
6	0	0	0	0	0	1	1	1[a]	1[a]	4
7	0	0	0	0	0	1	1	1	1[a]	4
8	0	0	0	0	0	0	1	1	1	3
9	0	0	0	0	0	1[a]	1[a]	1	1	4
Dependence power	1	1	3	4	5	7	9	9	9	

[a]Transitive relations.

TABLE 6 List of student issues and their levels in TISM-P.

Issue no.	Student issues with MOOCs	Levels in TISM-P
1	Language barrier	III
2	Technological barrier	III
3	Overload of information	II
4	Discussion forum	II
5	Student support and feedback system	II
6	Ambiguous assignments and peer assessment	I
7	Design and quality of MOOCs	I
8	Credibility of MOOCs certificates in job market	I
9	Teaching and pedagogy	I

analysis is used to analyse the driver-dependence relationship among factors over one another (Figure 5). The two-dimensional graph is divided into four quadrants. The first quadrant consists of ***autonomous*** criteria, which has a low driver and dependence power; the second quadrant represents the ***dependent criteria*** which has low driving power and high dependence power; the third quadrant is made up of ***linkage criteria*** which has high driving and dependence powers, and the last quadrant represents the ***driver criteria*** which has high driving power and low dependence powers. The polarity of each student issue was interpreted from the transitive reachability matrix in Figure 4.

3.2.10 | Step X

The results from MICMAC analysis (Figure 4) are overlayed upon the TISM-P model as depicted in Figure 6. The model briefs upon the classification of student issues with MOOCs along with +ve and −ve orientation in the form of driver-dependent relationships.

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

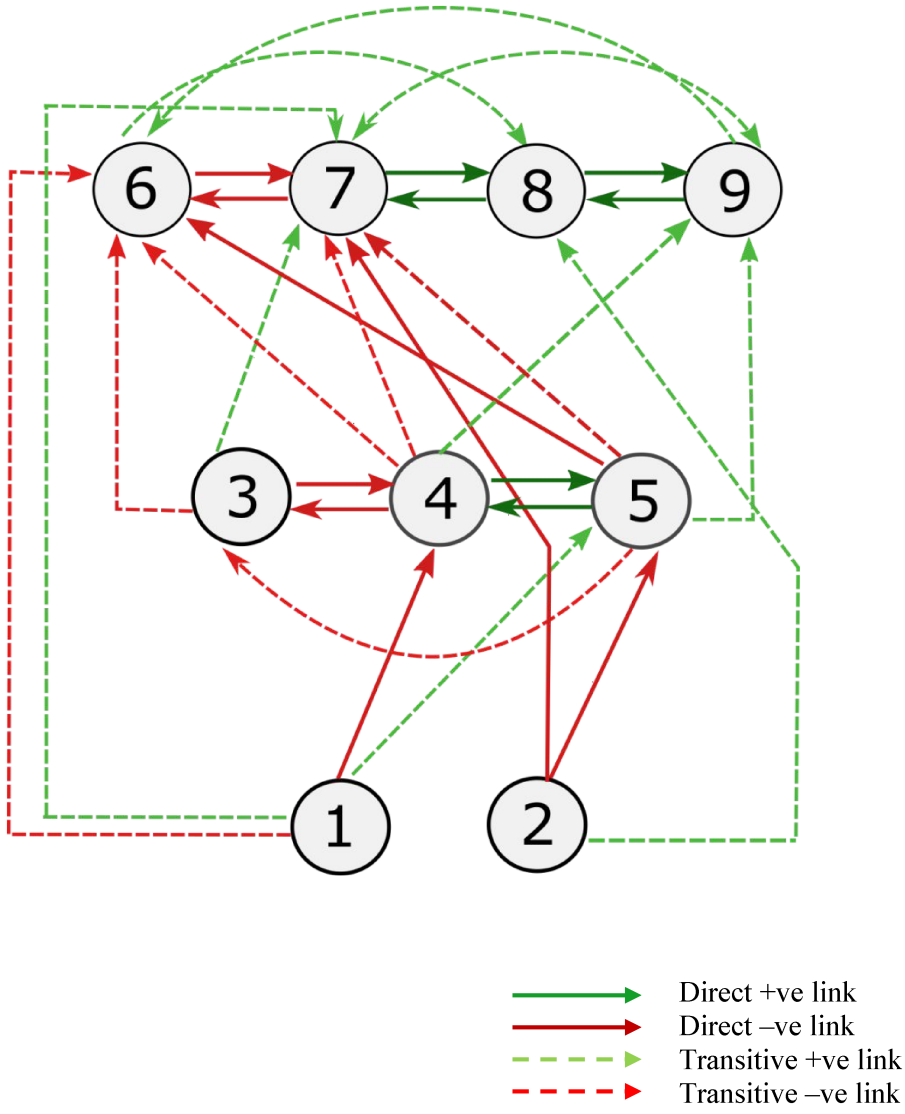


FIGURE 3 Digraph after hierarchical partitioning with polarity linkages.

TABLE 7 Binary interaction matrix.

Student issues	1	2	3	4	5	6	7	8	9
1	–	0	0	−1	+1*	−1*	+1*	0	0
2	0	–	0	0	−1	0	−1	+1*	0
3	0	0	–	−1	0	−1*	+1*	0	0
4	0	0	−1	–	+1	−1*	−1*	0	+1*
5	0	0	−1*	+1	–	+1	−1*	0	+1*
6	0	0	0	0	0	–	−1	+1*	0
7	0	0	0	0	0	−1	–	+1	+1*
8	0	0	0	0	0	0	+1	–	+1
9	0	0	0	0	0	+1*	+1*	+1	–

*Transitive Relationships.

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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 8 Interpretive matrix.

Student issues	1	2	3	4	5	6	7	8	9
1	–	–	–	Affects students' ability to participate	Affects students' ability to seek support	Increased ambiguity due to lack of discussions	Quality gets affected due to inefficient discussion forums, student support and ambiguous assignments	–	–
2	–	–	–	–	Technological issues might affect feedback systems effectiveness	–	Technological issues will affect quality of MOOC	Technological issues will affect the quality in turn lowering the credibility of the course	–
3	–	–	–	Reduced efficiency	–	Information overload will reduce discussion efficiency thus increasing ambiguity	Information overload will lower the quality of MOOC via multiple issues	–	–
4	–	–	Multiple discussion threads	–	Effective feedback is possible	Reduced ambiguity in assignments via support	Quality of MOOC will increase with lower ambiguity in assignments and better understanding of peer assessment	–	Inputs to teaching pedagogy via effective fedback systems
5	–	–	Effective student support system will help reduce information overload	Easy to approach and attend to student queries and issues	–	Effective feedback systems and student support will help reduce ambiguity in assignments	Reduced ambiguity in assignments and updated information about peer assessment will help increase quality of MOOC	–	Effective feedback will help instructors improve teaching pedagogy with time
6	–	–	–	–	–	–	Lowers	Increased ambiguity of assignments will lower the quality of MOOC thus hampering credibility	–

(Continues)

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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 8 (Continued)

Student issues	1	2	3	4	5	6	7	8	9
7	–	–	–	–	–	Reduces	–	Higher the quality better the credibility	Effective quality of MOOC leads to improved teaching pedagogy
8	–	–	–	–	–	–	Better credibility perceives a high-quality MOOC	–	Credible MOOC will have effective teaching pedagogy
9	–	–	–	–	–	Effective teaching pedagogy leads to reduction in ambiguities	Effective teaching pedagogy leads to improved quality of MOOC	Positive influence on credibility	–

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

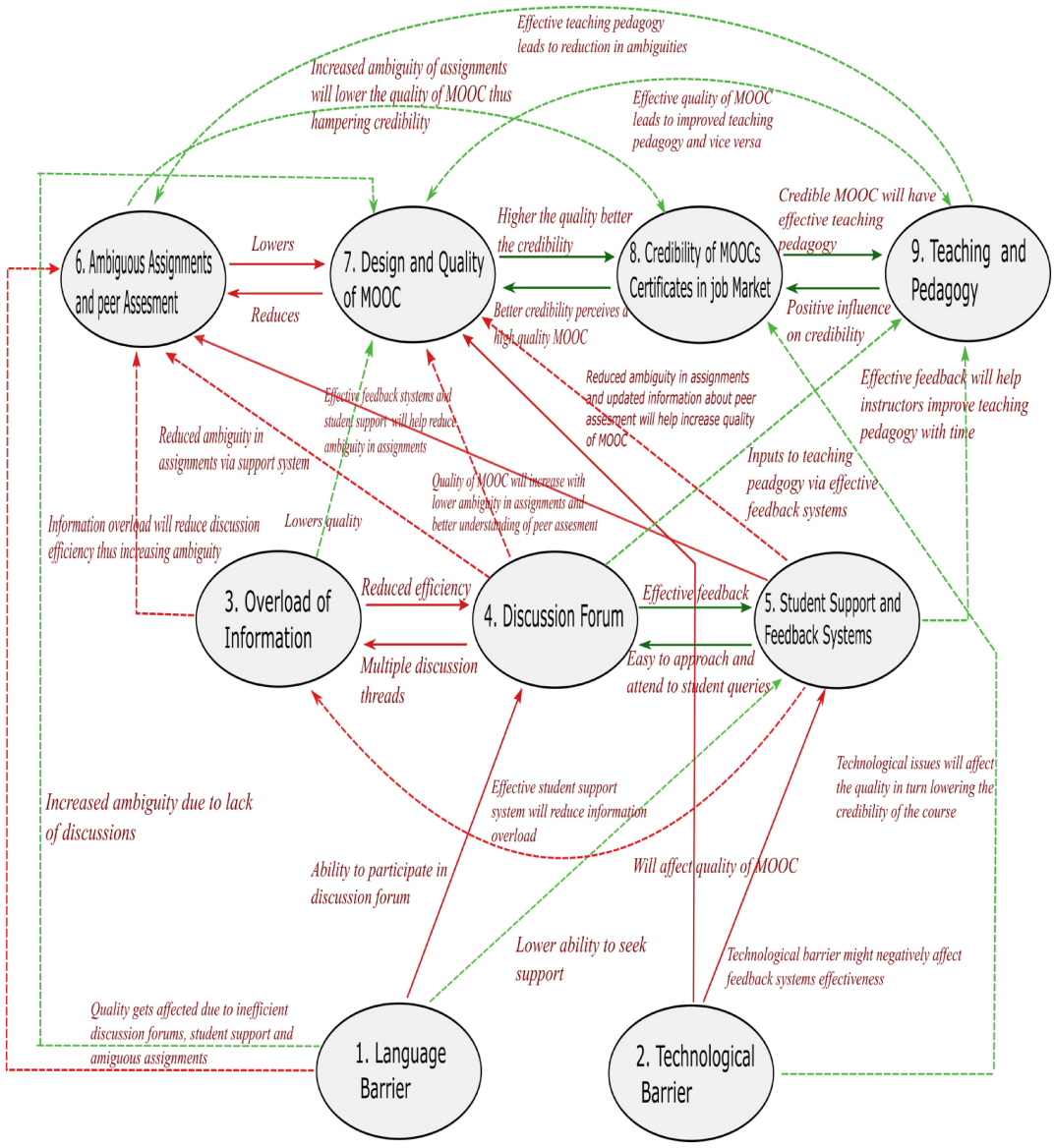


FIGURE 4 TISM-P model for student issues with MOOCs.

3.2.11 | Step XI

The driver factors are used to trace out the +ve and −ve paths of student issues in reaching the outcome issue through intermediate factors (Table 9). It will aid in reflecting upon the need to control the impact of driver factors on outcome variables.

3.3 | Model validation

Once the hierarchical model was developed, it was important to vet the information in the model with the panel of experts contacted earlier in factor validation. They had already taken part in the study and thus

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

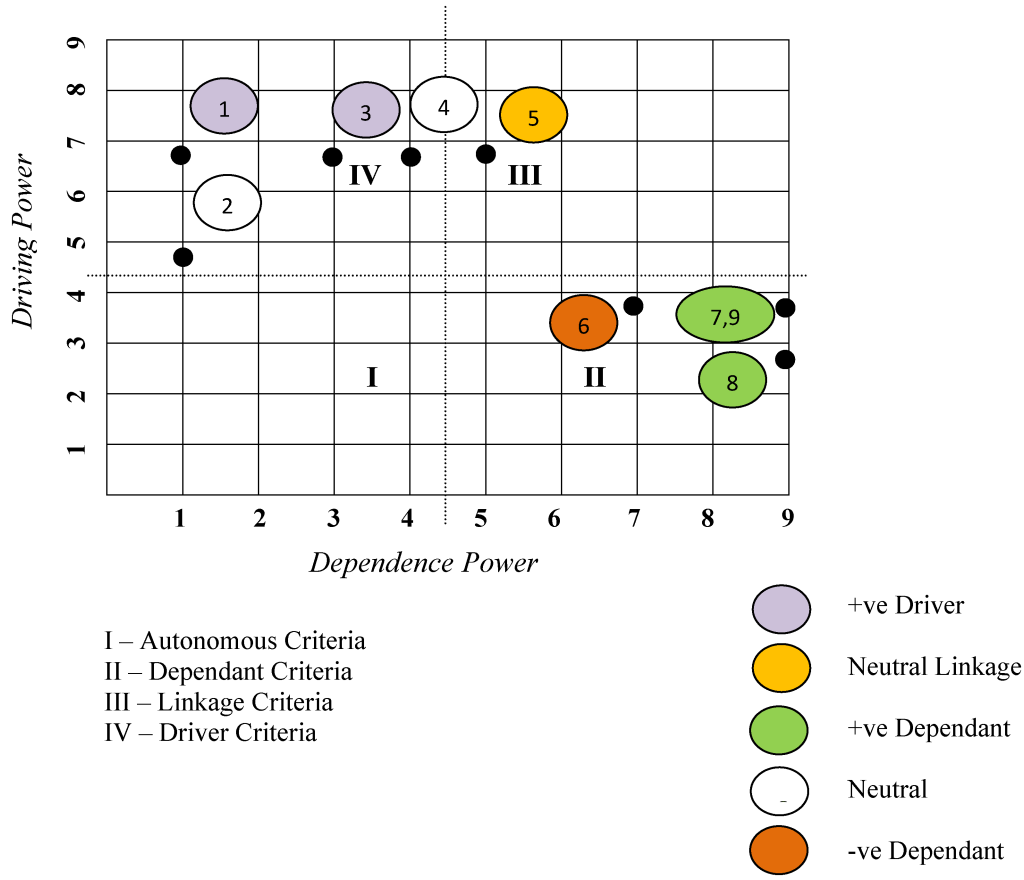


FIGURE 5 MICMAC analysis of student issues with MOOCs.

had a better idea of comprehending the nature of the work. A total of 39 linkages (Table 10) were formed in the original model and were tested for validation. It ideally depends upon the researchers' discretion to remove unnecessary transitive linkages to avoid overcrowding in the model (Agrawal, 2020; Sushil, 2018). We, however, consulted with expert opinions, and all 39 linkages were validated using hypothesis testing with a *t*-test. The final model was constructed whilst retaining the 30 most influential linkages. 39 questions were developed on a 5-point Likert scale questionnaire from strongly disagree to strongly agree for the test.

Hi0. Null: No significant difference exists between the observed mean and specified mean in respective factor linkage.

HiA. Alternate: Significant difference exists between the observed mean and specified mean in respective factor linkage.

Ms-Excel® was used to collate the responses, and a test value of 3.5 was used; results are shown in Table 10.

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

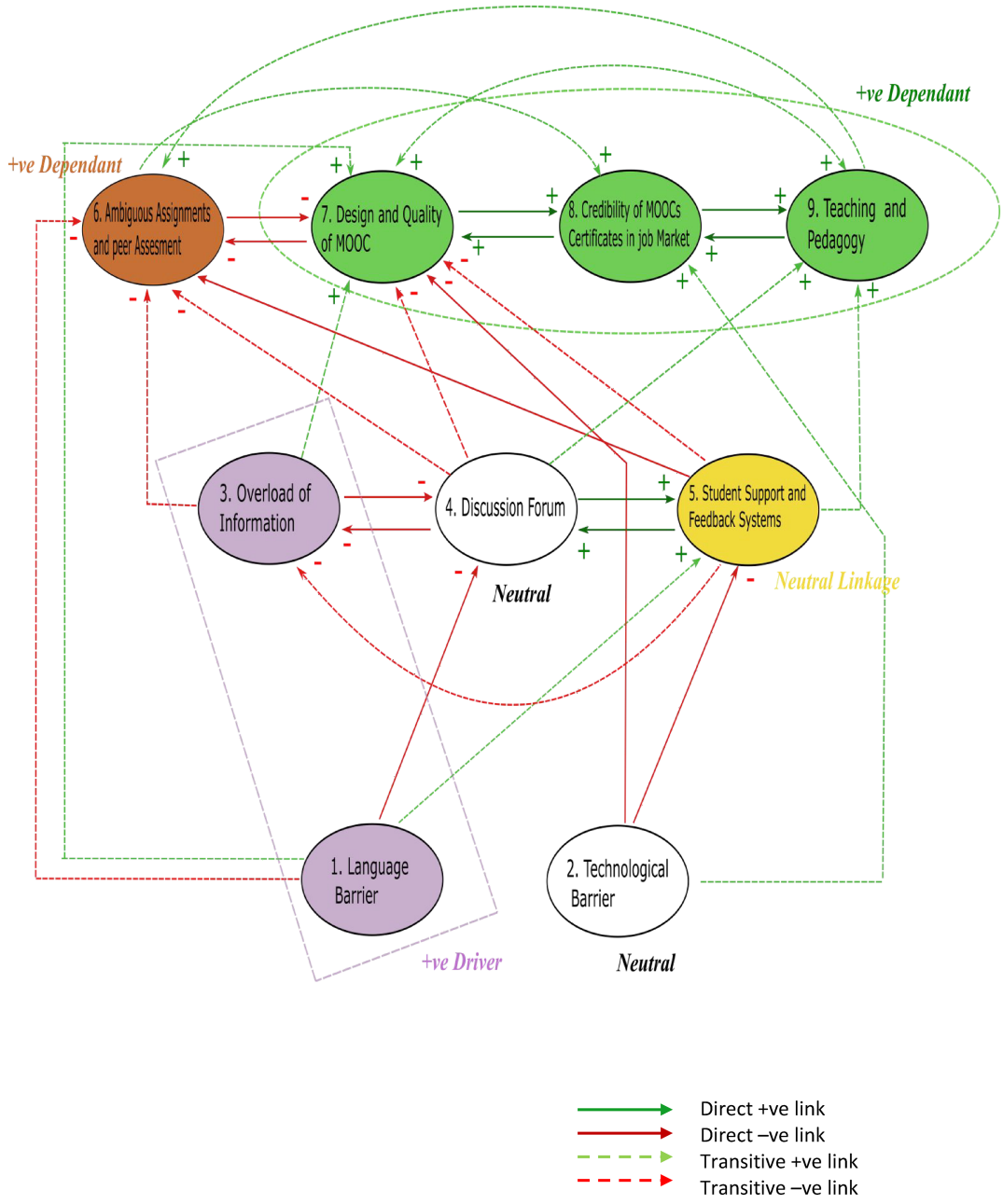


FIGURE 6 TISM-P model overlayed upon MICMAC analysis.

TABLE 9 Flow and nature of paths.

Driver factors	Path through factors	Polarity of path
Language barrier	Discussion forum	+ve
	Student support and feedback services	+ve
Technological barrier	Student support and feedback services	−ve
Overload of information	Discussion forum	+ve
Discussion forum	Student support and feedback services	+ve

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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 10 Hypothesis testing result ($N = 40$).

S. no.	Factor link	Mean	SD	*t*-value	Accept/Reject[a]
1	1–4	4.08	0.730	4.939	Accept
2	1–5	4.70	0.464	1.635	Accept
3	1–6	3.58	0.747	5.925	Accept
4	1–7	3.85	0.662	3.343	Accept
5	1–8	1.63	0.490	−2.425	Reject
6	1–9	1.20	0.405	−3.591	Reject
7	2–5	4.65	0.483	1.506	Accept
8	2–7	3.98	0.660	4.506	Accept
9	2–8	4.45	0.504	1.193	Accept
10	2–9	1.35	0.700	−1.1943	Reject
11	3–4	4.03	0.530	6.2	Accept
12	3–5	1.50	0.506	−2.498	Reject
13	3–6	4.05	0.904	3.846	Accept
14	3–7	4.35	0.770	6.985	Accept
15	3–8	1.68	0.474	−2.44	Reject
16	3–9	1.65	0.483	−2.422	Reject
17	4–3	4.55	0.504	1.318	Accept
18	4–5	4.20	0.564	7.851	Accept
19	4–6	4.45	0.504	1.193	Accept
20	4–7	4.18	0.594	7.13	Accept
21	4–8	1.18	0.385	−3.83	Reject
22	4–9	4.50	0.506	1.249	Accept
23	5–3	4.55	0.504	1.318	Accept
24	5–4	4.60	0.496	1.402	Accept
25	5–6	4.23	0.620	7.349	Accept
26	5–7	4.28	0.554	8.789	Accept
27	5–8	1.80	0.464	−2.317	Reject
28	5–9	4.43	0.501	1.162	Accept
29	6–7	4.13	0.648	6.052	Accept
30	6–8	4.28	0.554	8.789	Accept
31	6–9	1.05	0.221	−7.02	Reject
32	7–6	4.35	0.622	8.64	Accept
33	7–8	4.25	0.588	8.062	Accept
34	7–9	4.50	0.555	1.14	Accept
35	8–7	4.33	0.572	9.061	Accept
36	8–9	4.35	0.622	8.64	Accept
37	9–6	4.43	0.501	1.162	Accept
38	9–7	4.23	0.768	5.993	Accept
39	9–8	4.40	0.632	9	Accept

[a]Accept if the calculated *t*-statistic value > *t*-critical value (2-tailed).

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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 | RESULTS AND DISCUSSION

4.1 | TISM-P model

The model exhibits substantial insights into the interrelations between the student issues with MOOCs due to the effect of the positive and negative influence of factors over one another. The first factor, which is the language barrier, exhibits a direct and negative relationship with the factor discussion forum. If a country doesn't have an indigenously developed platform for offering MOOCs and English is not the medium of instruction in colleges and universities, then the students might not feel comfortable on global platforms (Ding & Shen, 2020), as was found in the case of Turkish MOOCs (Kurt, 2019). Student's ability to participate in discussion forums on global MOOC platforms might be affected due to language acting as a barrier and an issue itself. In such circumstances, students often lack the motivation to participate and might not be comfortable seeking answers to their queries. Furthermore, the language barrier was found to have a negative transitive effect on maintaining, if not increasing, the ambiguity of assignments. This effect flows through two factors, that is discussion forums and student support and feedback services. It is important to note whilst interpreting the model that not all transitive linkages will make sense in the real-world application (Sushil & Dinesh, 2022). Only those links and nodes are to be interpreted which carry substantial influence over another factor. Thus, all transitive linkages were consulted with expert opinions. Therefore, an increasing effect of language as a barrier will bring down the effectiveness of discussion forums, which in turn will increase the ambiguity of student assignments. The presence of a positive transitive link with the design and quality of a MOOC reflects the importance of recognising language as a major barrier since the link is mediated by discussion forums. A high language barrier will negatively affect a student's ability to participate in discussion forums which will eventually lower the quality of any MOOC (Sanchez-Gordon & Luján-Mora, 2020). The Technological barrier has shown a direct and negative relationship with factors 5 and 7. It is important to note that high-speed internet connections and appropriate bandwidth might not be within the reach of all current and prospective students in developing or underdeveloped countries (Ma & Lee, 2018b; Ramij & Sultana, 2020). The obstructive effects of lack of internet penetration and net neutrality are quite real in tier 2 and 3 cities in such countries (Sambuli, 2016). Students opined that technological barriers have a direct negative effect on reducing the efficiency of feedback systems. Also, the lack of appropriate bandwidth will affect students' ability to access MOOCs since they require high-speed internet connections to load videos and access supporting materials, thus lowering the overall quality. Technology as a student issue also has a positive transitive effect on maintaining the credibility of a course in the market. Increased barriers to technology will lower the quality and, thus, the credibility of the course in the market. An overload of information has a direct negative relation with discussion forums. The results are not new from our findings since, it is already well supported by literature (Chen et al., 2011; Khalid et al., 2021; Zeng et al., 2022). However, a transitive positive relation was found with the design and quality of MOOCs. An overloaded discussion forum is constrained and ineffective, which might have a negative effect on the quality of MOOCs. The negative transitive relationship with factor 6 reiterates the fact that an increase in information overload will negatively affect the discussion forum, which in turn might increase the ambiguity in assignments and negatively affect peer assessment as well. Students felt that the discussion forum had a direct and positive effect on factor 5 and a direct negative effect on factor 3. A well-regulated discussion forum could also serve as a legitimate platform to deliver effective feedback to the instructors and vice-versa (Lowenthal & Dunlap, 2020). However, it often becomes too overcrowded in the presence of multiple discussion threads, which could further increase the information overload. The negative transitive effect on factor 6 implies that effective discussion forums will support and ensure more active student feedback and help in reducing ambiguities in understanding assignments via fruitful discussions. The presence of a negative transitive effect on factor 7 tacitly expresses the fact that increased knowledge sharing helps to bank upon the basic premise of MOOCs, that is learning via a connectivist approach; thus, a

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

well-regulated discussion forum would also contribute towards the better quality of MOOCs. Finally, the positive transitive relationship with factor 9 states that a well-maintained discussion forum could support instructors to garner real-time student feedback, which might be useful for improving teaching pedagogy. Student support and feedback services are crucial elements within MOOC issues and were found to have a direct positive influence in increasing the effectiveness of a discussion forum (Gamage et al., 2020). Additionally, this effect might lead to a reduction in the overload of information, thereby establishing a negative transitive relationship between factors 5 and 3. Moreover, a direct negative effect of factor 5 on 6 and a negative transitive effect on factor 7 exist. Lack of student support in MOOCs has been shown to have a negative effect on students dropping out of the course (Aldowah et al., 2020). Having an efficient support and feedback system will delimit any barriers between student-teacher interactions, thus, assisting in understanding assignments and peer assessment, further improving the quality of a MOOC. The presence of a positive transitive influence over factor 9 suggests that timely feedback from students might help teachers to make improvements in their teaching pedagogies and also adds to the completeness of a MOOC. Ambiguous assignments and peer assessment both have a direct negative effect on lowering the quality of a MOOC, which also has a positive transitive effect in lowering the credibility of such courses in the employment market. The Design and quality of MOOCs have a direct positive influence on factor 8 and a direct negative influence on factor 6. Generally, a high-quality and seasoned MOOC is famous and credible among students and in the job market as well, whereas better quality reduces ambiguity in assignments and peer assessments. Moreover, the positive transitive link with factor 9 highlights the importance of a well-designed MOOC via a direct positive relationship with credibility, that is an increase in both will serve as a driving factor for instructors to improve teaching pedagogy over a period of time (Julia & Marco, 2021). The Credibility of MOOCs in the job market has a direct positive relationship both ways forward and backwards. An increase in the credibility of a MOOC implies high-quality and effective teaching pedagogy and vice-versa. Teaching and Pedagogy have a direct positive influence on the credibility of MOOCs, and the positive transitive relationship with the quality of a MOOC reveals that an effective teaching pedagogy will improve the quality of a MOOC (Jamebozorg et al., 2022) and also gather credibility in the employment market. Finally, the positive transitive relationship with factor 6 via improved quality of a MOOC will help in reducing any ambiguities in understanding assignments.

4.2 | MICMAC analysis

No autonomous factors were identified, which communicates to us the fact that all the factors have some relationship with each other. In quadrant (II) Factors 6, 7, 8 and 9 are major dependant factors in the system. The arrangement points towards the outcome of the model and needs factors with high driving power to trigger them. The credibility of a MOOC certificate in the employment market might not be of concern to other stakeholders, but in certain education systems' where securing a job is hyper-competitive (Goglio & Bertolini, 2021; Pingle & Sharma, 2013) upskilling via MOOCs speaks for itself. Also, the teaching pedagogy and design of a MOOC are intertwined with each other and receive the maximum amount of effect from drivers. It is important to note that factor 6 has a negative dependence power due to the nature of the issue itself. The quadrant (IV) is home to four major driving factors, that is 1, 2, 3 and 4. Factors 1 and 3 are positive drivers in the system with very low dependence. Factors 2 and 4 are neutral in nature, that is they have equal amounts of positive and negative driving powers behind them. They will drive the influence but will not take a positive or negative stand. Any change in these issues will bring negative repercussions to other student issues. Finally, the third quadrant makes up the linkage criteria, which consists of factor 5 as a positive linkage. This factor is considered to be of utmost importance and volatility, that is they are highly influential and endangered at the same time. It flows both ways, that is any change in them will influence the remaining issues and the results of the model, creating a feedback effect (Mahajan et al., 2015).

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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 overlay of MICMAC analysis over the TISM-P model provides us with a clear picture of corresponding factor influence and flow as well. The driver factors 1 and 3 do not have any relationship with each other but are responsible for positively driving the path through the linkage factor in order to reach the dependent factors. Factors 2 and 4 are neutral in nature, that is they may influence positively or negatively the linkage factor. Factor 6 is of crucial importance because it is the sole negative dependent factor. For all direct relations, the model has to flow through the dedicated strict paths in order to influence or reach the dependent factors, but for transitive relationships, the paths could be complicated. For example, factor 4 (driver) has a positive transitive relationship with factor 9 (dependent); thus, in order to reach the outcome factor, the path could move via multiple configurations, that is 4→5→6→7→8→9; 4→5→9; 4→6→9; 4→7→9; 4→8→9. The issues interact with each other via multiple influences; thus, the nature of the path shall also have multiple inferences depending on the situation in the real world.

5 | IMPLICATIONS

The current study has key theoretical implications that could aid researchers in understanding the behaviour of student issues related to MOOCs. The study finds interrelations among student issues which interact with one another either positively or negatively depending upon the context. The model presented in the study also sheds light on the cause-and-effect relationship among student issues that can take multiple forms due to the nature of effect and transitivity. All transitive relationships could be interpreted according to the positive or negative effect that flows throughout the model. However, not examined in the current study, it is observed that the TISM-P model behaves similarly to a system dynamics causal loop model where balancing and reinforcing loops could be identified. The student issues, in this sense, are open to interpretation if examined from a systems perspective. Past research on student issues is focused on models related to factors that affect dropout rates, quality, motivation, user experience and perception (Aldowah et al., 2020; Peng & Xu, 2020; Shah et al., 2022). These studies deal with larger problems in the e-learning system but do not address the interrelationships among factors or issues. Prior literature focuses on individual issues and has studied them in depth at a micro level. The present study allows for exploration and a holistic understanding of a model dedicated to understanding interrelationships and the influence of student issues on one another.

6 | CONCLUSION AND LIMITATIONS

A MOOC must not be treated and experienced just as an online certification course but, rather, it has the potential to offer a holistic learning and development experience. High student dropout rates, lack of motivation, difficult evaluation, etc., are some major problems of online learning systems which are still difficult to address. A huge number of intrinsic and extrinsic factors exist that affect these problems and are difficult to study comprehensively. Therefore, the current study brings an extensive model of issues related to MOOCs which affect these problems. The problems arising out of the permutation and combination of these issues are well-read and explored in the literature, but a holistic understanding of major prominent issues was missing. A look into the interrelationships between these issues is achieved by exploring the causal relationships in the study. The results have highlighted the pitfalls that are ignored when we talk about student issues and their interactions among themselves. Controlling the effect of one issue has a ripple effect on other issues when viewed from a holistic perspective. A cause-and-effect relation helps to understand better the reasons behind the behaviour of a particular element in the system. Therefore, the current study carefully explores the negative or positive effects of these issues on one another. For further research, our model could be used to generate simulations and modelling exercises that can interpret the behaviour of these issues in the real world. The study is currently limited from the student's perspective. It has the potential to involve different student levels, that is High school, post-graduate and doctorate to garner multifarious perceptions about MOOC issues. In order to use the

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

current results effectively, it is recommended that further studies could use qualitative methods of inquiry to gather detailed information about such issues in the system.

CONFLICT OF INTEREST STATEMENT

There was no conflict of interest between the authors.

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.

ORCID


Himanshu Pawar https://orcid.org/0000-0002-9661-3797

REFERENCES

Abu-Shanab, E., & Musleh, S. (2018). The adoption of massive open online courses: Challenges and benefits. *International Journal of Web-Based Learning and Teaching Technologies*, *13*, 62–76.

Admiraal, W., Huisman, B., & Pilli, O. (2015). Assessment in massive open online courses. *The Electronic Journal Of E-Learning*, *13*(4), 207–216.

Agrawal, A. (2020). Modified Total interpretive structural model of corporate financial flexibility. *Global Journal of Flexible Systems Management*, *21*(4), 369–338. https://doi.org/10.1007/s40171-020-00253-7

Alario-Hoyos, C., Perez-Sanagustin, M., Cormier, D., & Delgado-Kloos, C. (2014). Proposal for a conceptual framework for educators to describe and design MOOCs. *Journal of Universal Computer Science*, *20*(1), 6–23. https://doi.org/10.3217/jucs-020-01-0006

Albelbisi, N. A., & Yusop, F. D. (2019). Factors influencing learners' self–regulated learning skills in a massive open online course (MOOC) environment. *Turkish Online Journal of Distance Education*, *20*(3), 1–16. https://doi.org/10.17718/tojde.598191

Aldowah, H., Al-Samarraie, H., Alzahrani, A. I., & Alalwan, N. (2020). Factors affecting student dropout in MOOCs: A cause and effect decision-making model. *Journal of Computing in Higher Education*, *1-26*, 429–454. https://doi.org/10.1007/s12528-019-09241-y

Armellini, A., & Padilla Rodriguez, B. C. (2016). Are massive open online courses (MOOCs) pedagogically innovative? *Journal of Interactive Online Learning*, *14*(1), 17–28.

Bali, M. (2014). MOOC pedagogy: Gleaning good practice from existing MOOCs. *MERLOT Journal of Online Learning and Teaching*, *10*(1), 44.

Bordoloi, R., Das, P., & Das, K. (2020). Lifelong learning opportunities through MOOCs in India. *Asian Association Of Open Universities Journal*, *15*(1), 83–95. https://doi.org/10.1108/aaouj-09-2019-0042

Borrella, I., Caballero-Caballero, S., & Ponce-Cueto, E. (2022). Taking action to reduce dropout in MOOCs: Tested interventions. *Computers & Education*, *179*, 104412. https://doi.org/10.1016/j.compedu.2021.104412

Chen, C.-Y., Pedersen, S., & Murphy, K. L. (2011). Learners' perceived information overload in online learning via computer-mediated communication. *Research in Learning Technology*, *19*(2), 101–116. https://doi.org/10.1080/21567069.2011.586678

Chiappe-Laverde, A., Hine, N., & Martínez-Silva, J. (2015). Literature and practice: A critical review of MOOCs. *Comunicar: Revista Científica de Comunicación y Educación*, *22*(44), 9–18. https://doi.org/10.3916/c44-2015-01

Corbeil, J., Khan, B., & Corbeil, M. (2018). MOOCs revisited: Still transformative or passing fad? *Asian Journal of University Education*, *12*(2).

Ding, Y., & Shen, H. Z. (2020). English language MOOCs in China: Learners' perspective. *The EuroCALL Review*, *28*(2), 13–22.

Eriksson, T., Adawi, T., & Stöhr, C. (2016). "Time is the bottleneck": A qualitative study exploring why learners drop out of MOOCs. *Journal of Computing in Higher Education*, *29*(1), 133–146. https://doi.org/10.1007/s12528-016-9127-8

Farrow, R., Ferguson, R., Weller, M., Pitt, R., Sanzgiri, J., & Habib, M. (2021). Assessment and recognition of MOOCs: The state of the art. *Journal of Innovation in Polytechnic Education*, *3*(1), 15–26.

Fathi, M., Ghobakhloo, M., & Syberfeldt, A. (2019). An interpretive structural modeling of teamwork training in higher education. *Education in Science*, *9*(1), 16. https://doi.org/10.3390/educsci9010016

Fianu, E., Blewett, C., Ampong, G., & Ofori, K. (2018). Factors affecting MOOC usage by students in selected Ghanaian Universities. *Education in Science*, *8*(2), 70. https://doi.org/10.3390/educsci8020070

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

Fischer, G. (2014). Beyond hype and underestimation: Identifying research challenges for the future of MOOCs. *Distance Education*, *35*(2), 149–158. https://doi.org/10.1080/01587919.2014.920752

Galikyan, I., Admiraal, W., & Kester, L. (2021). MOOC discussion forums: The interplay of the cognitive and the social. *Computers & Education*, *165*, 104133. https://doi.org/10.1016/j.compedu.2021.104133

Gamage, D., Perera, I., & Fernando, S. (2020). MOOCs lack interactivity and collaborativeness: Evaluating MOOC platforms. *International Journal of Engineering Pedagogy (iJEP)*, *10*(2), 94–111.

Gamage, D., Staubitz, T., & Whiting, M. (2021). Peer assessment in MOOCs: Systematic literature review. *Distance Education*, *42*(2), 268–289. https://doi.org/10.1080/01587919.2021.1911626

Ghislandi, P. (2016). "The fun they had" or about the quality of MOOC. *Journal of E-Learning and Knowledge Society*, *12*(3), 99–114. https://doi.org/10.20368/1971-8829/1178

Goglio, V., & Bertolini, S. (2021). The contribution of MOOCs to upskilling the labor force. *Journal of Workplace Learning*, *33*(7), 561–574. https://doi.org/10.1108/JWL-10-2020-0159

Goopio, J., & Cheung, C. (2021). The MOOC dropout phenomenon and retention strategies. *Journal of Teaching in Travel & Tourism*, *21*(2), 177–197. https://doi.org/10.1080/15313220.2020.1809050

Gregori, E., Zhang, J., Galván-Fernández, C., & Fernández-Navarro, F. (2018). Learner support in MOOCs: Identifying variables linked to completion. *Computers & Education*, *122*, 153–168. https://doi.org/10.1016/j.compedu.2018.03.014

Gul, S., Mahajan, I., Shafiq, H., Shafi, M., & Shah, T. (2018). Massive open online courses: Hype and hope. *DESIDOC Journal of Library & Information Technology*, *38*(1), 63–66. https://doi.org/10.14429/djlit.38.1.11141

Henderikx, M., Kreijns, K., & Kalz, M. (2018). A classification of barriers that influence intention achievement in MOOCs. In V. Pammer-Schindler, M. Pérez-Sanagustín, H. Drachsler, R. Elferink, & M. Scheffel (Eds.), *Lifelong technology-enhanced learning*. EC-TEL 2018. Lecture Notes in Computer Science (p. 11082). Springer. https://doi.org/10.1007/978-3-319-98572-5_1

Hew, K., & Cheung, W. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review*, *12*, 45–58. https://doi.org/10.1016/j.edurev.2014.05.001

Jamebozorg, Z., Sarkeshikiyan, S. M., & Sadat Sarkeshikiyan, M. (2022). The developing pedagogical and psychological trends and MOOCs: A systematic review. *Journal of Pharmaceutical Negative Results*, *13*(3), 401.

Jitpaisarnwattana, N., Darasawang, P., & Reinders, H. (2022). Defining success in a language MOOC from learner perspectives. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, *12*(1), 1–16. https://doi.org/10.4018/IJCALLT.291108

Jones, R. (2014). Global reach and local practice: The promise of MOOCs. *Language Learning & Technology*, *18*(3), 5–15.

Jothimani, D., Bhadani, A. K., & Shankar, R. (2015). Towards understanding the cynicism of social networking sites: An operations management perspective. *Procedia-Social and Behavioral Sciences*, *189*, 117–132. https://doi.org/10.1016/j.sbspro.2015.03.206

Julia, K., & Marco, K. (2021). Educational scalability in MOOCs: Analysing instructional designs to find best practices. *Computers & Education*, *161*, 104054. https://doi.org/10.1016/j.compedu.2020.104054

Kasch, J., van Rosmalen, P., Löhr, A., Klemke, R., Antonaci, A., & Kalz, M. (2021). Students' perceptions of the peer-feedback experience in MOOCs. *Distance Education*, *42*(1), 145–163. https://doi.org/10.1080/01587919.2020.1869522

Khalid, A., Lundqvist, K., Yates, A., & Ghzanfar, M. (2021). Novel online recommendation algorithm for massive open online courses (NoR-MOOCs). *PLoS One*, *16*(1), e0245485. https://doi.org/10.1371/journal.pone.0245485

Khan, M., Nambobi, M., & Ali, M. (2017). MOOCs in TVET sectors of developing countries: Benefits, possibilities, and challenges. In F. Giuseffi (Ed.), *Emerging self-directed learning strategies in the digital age*. IGI Global.

Kumar, S., & Barua, M. K. (2022). A modeling framework of green practices to explore their interrelations as a conduit to policy. *Journal of Cleaner Production*, *335*, 130301. https://doi.org/10.1016/j.jclepro.2021.130301

Kurt, S. (2019). The case of Turkish university students and MOOCs. *American Journal of Distance Education*, *33*(2), 120–131. https://doi.org/10.1080/08923647.2019.1582284

Laaser, W. (2023). The rise and fall of the "massively open online courses". *South Eastern European Journal of Public Health*, *3*. https://doi.org/10.56801/seejph.vi.56

Leontyeva, I. A. (2018). Modern distance learning technologies in higher education: Introduction problems. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(10), em1578. https://doi.org/10.29333/ejmste/92284

Li, J., Tang, Y., Cao, M., & Hu, X. (2018). The moderating effects of discipline on the relationship between asynchronous discussion and satisfaction with MOOCs. *Journal of Computers in Education*, *5*(3), 279–296. https://doi.org/10.1007/s40692-018-0112-2

Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *The Internet and Higher Education*, *29*, 40–48. https://doi.org/10.1016/j.iheduc.2015.12.003

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

Liu, X; Liu, S; Lee, S.H. & Magjuka, R.J. (2010). Cultural differences in online learning: International student perceptions. *Educational Technology & Society*, *13*(3), 177–88. http://www.ifets.info/journals/13_3/16.pdf. Accessed 29 January 2020.

Liyanagunawardena, T. R., Williams, S., & Adams, A. A. (2014). The impact and reach of MOOCs: A developing countries' perspective. *eLearning Papers*, *33*, 38–46.

Lowenthal, P. R., & Dunlap, J. C. (2020). Social presence and online discussions: A mixed method investigation. *Distance Education*, *41*(4), 490–514. https://doi.org/10.1080/01587919.2020.1821603

Ma, L., & Lee, C. (2018a). Investigating the adoption of MOOCs: A technology-user-environment perspective. *Journal of Computer Assisted Learning*, *35*(1), 89–98. https://doi.org/10.1111/jcal.12314

Ma, L., & Lee, C. (2018b). Understanding the barriers to the use of MOOCs in a developing country: An innovation resistance perspective. *Journal of Educational Computing Research*, *57*(3), 571–590.

Mackness, J., Waite, M., Roberts, G., & Lovegrove, E. (2013). Learning in a small, task–oriented, connectivist MOOC: Pedagogical issues and implications for higher education. *The International Review of Research in Open and Distributed Learning*, *14*(4), 140–159. https://doi.org/10.19173/irrodl.v14i4.1548

Mahajan, R., Agrawal, R., Sharma, V., & Nangia, V. (2015). Analysis of challenges for management education in India using total interpretive structural modelling. *Quality Assurance in Education*, *24*(1), 95–122. https://doi.org/10.1108/qae-07-2013-0030

Meina, Z. H. U., Sabir, N., Annisa, S. A. R. I., Shuya, X. U., & Minkyoung, K. I. M. (2021). Addressing learner cultural diversity in MOOC design and delivery: Strategies and practices of experts. *Turkish Online Journal of Distance Education*, *22*(2), 1–25. https://doi.org/10.17718/tojde.906468

Muñoz-Merino, P. J., Ruipérez-Valiente, J. A., Moreno, J. L. S., & Kloos, C. D. (2015). Assessment activities in massive open on-line courses: Assessment activities in MOOCs. In *Furthering higher education possibilities through massive open online courses* (pp. 165–192). IGI Global.

Myers, T., Blackman, A., Andersen, T., Hay, R., Lee, I., & Gray, H. (2014). Cultivating ICT students' interpersonal soft skills in online learning environments using traditional active learning techniques. *Journal of Learning Design*, *7*(3), 38–53. https://doi.org/10.5204/jld.v7i3.194

Naidu, S. (2017). Open educational practice: Caveat emptor. In D. Singh & C. Stueckelberger (Eds.), *Ethics in higher education: Values-driven leaders for the future*. Globethics.net.

Onah, D., Sinclair, J., & Boyatt, R. (2014). Exploring the use of MOOC discussion forums. In *London international conference on education (LICE-2014)* (pp. 1–4). LICE. Proceedings of London international conference on education.

Parr, C. (2013). *MOOC creators criticise courses' lack of creativity*. Times Higher Education Retrieved from http://www.timeshighereducation.co.uk/news/mooc-creatorscriticise-courses-lack-of-creativity/2008180.fullarticle

Peng, X., & Xu, Q. (2020). Investigating learners' behaviors and discourse content in MOOC course reviews. *Computers & Education*, *143*, 103673. https://doi.org/10.1016/j.compedu.2019.103673

Pilli, O., & Admiraal, W. (2017). Students' learning outcomes in massive open online courses (MOOCs): Some suggestions for course design. *Yuksekogretim Dergisi*, *7*(1), 46–71. https://doi.org/10.2399/yod.17.001

Pingle, S. S., & Sharma, A. (2013). External employer attractiveness: A study of management students in India. *Journal of Contemporary Management Research*, *7*(1), 78–95.

Prasad, U., & Suri, R. (2011). Modeling of continuity and change forces in private higher technical education using total interpretive structural modeling (TISM). *Global Journal of Flexible Systems Management*, *12*(3–4), 31–39. https://doi.org/10.1007/bf03396605

Quality Matters. (2014). *Quality matters rubric standards* (5th ed.). Quality Matters.

Ramij, M. G., & Sultana, A. (2020). Preparedness of online classes in developing countries amid COVID-19 outbreak: A perspective from Bangladesh. Available at SSRN 3638718.

Reilly, E., Stafford, R., Williams, K., & Corliss, S. (2014). Evaluating the validity and applicability of automated essay scoring in two massive open online courses. *The International Review of Research in Open and Distributed Learning*, *15*(5), 83–98. https://doi.org/10.19173/irrodl.v15i5.1857

Rhoads, R. (2015). *Moocs, high technology, and higher learning*. John Hopkins University Press. https://doi.org/10.1353/book.42319

Rivas, M., Baker, R., & Evans, B. (2020). Do MOOCs make you more marketable? An experimental analysis of the value of MOOCs relative to traditional credentials and experience. *AERA Open*, *6*(4), 233285842097357. https://doi.org/10.1177/2332858420973577

Rosendale, J. A. (2017). Gauging the value of MOOCs: An examination of American employers' perceptions toward higher education change. *Higher Education, Skills and Work-Based Learning*, *7*(2), 141–154. https://doi.org/10.1108/HESWBL-09-2016-0065

Ruipérez-Valiente, J. A., Staubitz, T., Jenner, M., Halawa, S., Zhang, J., Despujol, I., Maldonado-Mahauad, J., Montoro, G., Peffer, M., Rohloff, T., Lane, J., Turro, C., Li, X., Pérez-Sanagustin, M., & Reich, J. (2022). Large scale analytics of global and regional MOOC providers: Differences in learners' demographics, preferences, and perceptions. *Computers & Education*, *180*, 104426. https://doi.org/10.1016/j.compedu.2021.104426

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

Sambuli, N. (2016). Challenges and opportunities for advancing internet access in developing countries while upholding net neutrality. *Journal of Cyber Policy*, *1*(1), 61–74. https://doi.org/10.1080/23738871.2016.1165715

Sanchez-Gordon, S., & Luján-Mora, S. (2020). Design, implementation and evaluation of MOOCs to improve inclusion of diverse learners. In *Accessibility and diversity in education: Breakthroughs in research and practice* (pp. 52–79). IGI Global. https://doi.org/10.4018/978-1-7998-1213-5.ch004

Santandreu Calonge, D., Aman Shah, M., Riggs, K., & Connor, M. (2019). MOOCs and upskilling in Australia: A qualitative literature study. *Cogent Education*, *6*(1), 1687392. https://doi.org/10.1080/2331186x.2019.1687392

Shah, V., Murthy, S., Warriem, J., Sahasrabudhe, S., Banerjee, G., & Iyer, S. (2022). Learner-centric MOOC model: A pedagogical design model towards active learner participation and higher completion rates. *Educational Technology Research and Development*, *70*(1), 263–288. https://doi.org/10.1007/s11423-022-10081-4

Singh, A., Sushil, K., & Pamucar, D. (2019). Stakeholder role for developing a conceptual framework of sustainability in organization. *Sustainability*, *11*(1), 208. https://doi.org/10.3390/su11010208

Stracke, C. M., & Trisolini, G. (2021). A systematic literature review on the quality of MOOCs. *Sustainability*, *13*(11), 5817. https://doi.org/10.3390/su13115817

Sushil. (2012). Interpreting the interpretive structural model: Organization research methods. *Global Journal of Flexible Systems Management*, *13*(2), 87–106. https://doi.org/10.1007/s40171-012-0008-3

Sushil. (2017). Modified ISM/TISM process with simultaneous transitivity checks for reducing direct pair comparisons. *Global Journal of Flexible Systems Management*, *18*(4), 331–351. https://doi.org/10.1007/s40171-017-0167-3

Sushil. (2018). Incorporating polarity of relationships in ISM and TISM for theory building in information and organization management. *International Journal of Information Management*, *43*, 38–51. https://doi.org/10.1016/j.ijinfomgt.2018.06.003

Sushil, & Dinesh, K. K. (2022). Structured literature review with TISM leading to an argumentation based conceptual model. *Global Journal of Flexible Systems Management*, *23*(3), 387–407. https://doi.org/10.1007/s40171-022-00309-w

Talib, F., & Rahman, Z. (2020). Modeling the barriers toward the growth of higher education institutions. *Qualitative Research Journal*, *20*(2), 243–264. https://doi.org/10.1108/qrj-09-2019-0070

Trehan, S., Sanzgiri, J., Li, C., Wang, R., & Joshi, R. (2017). Critical discussions on the massive open online course (MOOC) in India and China. *Critical Discussions on the Massive Open Online Course (MOOC) in India and China*, *13*(2), 141–165. Retrieved 22 January 2022, from.

Veletsianos, G. (2013). *Learner experiences with MOOCs and open online learning*. Hybrid Pedagogy.

Wang, T., Liu, J., & Li, T. (2019). Design variables for self-directed learning in MOOC environment. *Journal of Educational Technology Development and Exchange*, *12*(1), 6. https://doi.org/10.18785/jetde.1201.04

Wang, X., Hall, A., & Wang, Q. (2019). Investigating the implementation of accredited massive online open courses (MOOCs) in higher education: The boon and the bane. *Australasian Journal of Educational Technology*, *35*(3), 1–14. https://doi.org/10.14742/ajet.3896

Warfield, J. (1974). Toward interpretation of complex structural models. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-4*(5), 405–417. https://doi.org/10.1109/tsmc.1974.4309336

Wells, P., Gerbic, P., Kranenburg, I., & Bygrave, J. (2009). Professional skills and capabilities of accounting graduates: The New Zealand expectation gap? *Accounting Education: An International Journal*, *18*(4–5), 403–420. https://doi.org/10.1080/09639280902719390

Wong, B. (2016). Factors leading to effective teaching of MOOCs. *Asian Association of Open Universities Journal*, *11*(1), 105–118. https://doi.org/10.1108/aaouj-07-2016-0023

Yeravdekar, S., & Behl, A. (2017). Benchmarking model for management education in India. *Benchmarking: An International Journal*, *24*(3), 666–693. https://doi.org/10.1108/bij-06-2016-0082

Zeng, S., Zhang, J., Gao, M., Xu, K. M., & Zhang, J. (2022). Using learning analytics to understand collective attention in language MOOCs. *Computer Assisted Language Learning*, *35*(7), 1594–1619. https://doi.org/10.1080/09588221.2020.1825094

Zhang, J., Skryabin, M., & Song, X. (2016). Understanding the dynamics of MOOC discussion forums with simulation investigation for empirical network analysis (SIENA). *Distance Education*, *37*(3), 270–286. https://doi.org/10.1080/01587919.2016.1226

Zulkifli, N., Hamzah, M. I., & Bashah, N. H. (2020). Challenges to teaching and learning using MOOC. *Creative Education*, *11*(3), 197–205. https://doi.org/10.4236/ce.2020.113014

How to cite this article: Khera, S. N., & Pawar, H. (2024). Modelling student issues with MOOCs using TISM-P linkages. *Higher Education Quarterly*, *00*, 1–27. https://doi.org/10.1111/hequ.12515

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

APPENDIX A

Profile of e-learning platforms.

Profile of platforms	Number of participants
Coursera	53
Udemy	22
Coding Ninjas	17
Udacity	14
EdX	9
Coding blocks	8
Pep coding	6
Microsoft	5
NPTEL	5
SWAYAM	3
Unwired learning	2
Qwiklabs	2
Internshala	2
Stanford	1

Note: The data was collected from a number of engineering streams, namely Mechanical, Computer Science, Information Technology, Electrical, Environmental, Automobile and Production.

APPENDIX B

Hierarchical partitioning of reachability matrix for MOOCs issues.

Student issues	Reachability set	Antecedent set	Intersection set	Level
Iteration-1				
1	1,4,5,6,7,8,9	1	1	
2	2,5,7,8,9	2	2	
3	3,4,5,6,7,8,9	3,4,5	3,4,5	
4	3,4,5,6,7,8,9	1,3,4,5	3,4,5	
5	3,4,5,6,7,8,9	1,2,3,4,5	3,4,5	
6	6,7,8,9	1,3,4,5,6,7,9	6,7,9	
7	6,7,8,9	1,2,3,4,5,6,7,8,9	6,7,8,9	I
8	7,8,9	1,2,3,4,5,6,7,8,9	7,8,9	
9	6,7,8,9	1,2,3,4,5,6,7,8,9	6,7,8,9	I
Iteration-2				
1	1,4,5	1	1	
2	2,5	2	2	
3	3,4,5	3,4,5	3,4,5	II
4	3,4,5	1,3,4,5	3,4,5	II
5	3,4,5	1,2,3,4,5	3,4,5	II

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

Student issues	Reachability set	Antecedent set	Intersection set	Level
Iteration-3				
1	1	1	1	III
2	2	2	2	III

14682273, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/hequ.12515 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

Modified Six State Cryptographic Protocol with Entangled Ancilla Component States

Rashi Jain* and Satyabrata Adhikari†
Department of Applied Mathematics,
Delhi Technological University, Delhi-110042, India

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 sitting in between the sender and the receiver. In this work, we address the question of whether the sender and the receiver establish a secret key between them even though Eve possesses quantum power with her. This means that we probe here the existence of the possibility of the situation of generation of a secret key even if the eavesdropper is able to construct an entangled ancilla state in such a way that she can extract information from the intercepted qubit. Thus, our primary goal here is to discuss conditions for the establishment of a secret key between two distant parties in the presence of an eavesdropper. To achieve this task, we consider and modify the six-state QKD protocol in which Eve can construct the unitary transformation that may make all ancilla components entangled at the output. Then 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 here we have shown that it also depends on the concurrence of the ancilla component states. 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 are able to show that a secret key can be generated between Alice and Bob even if the disturbance is large enough. Moreover, we compare our result with Bruss's six-state QKD protocol and find that due to the occurrence of entangled ancilla states, our modified QKD protocol can generate a secret key in some region of both the cases where we consider mutual information of Alice and Eve with disturbance and disturbance free mutual information of Alice and Eve. In these cases, we find Bruss's six state QKD protocol failed to generate the secret key in the above mentioned region where the modified six-state QKD protocol is able to generate the secret key.

arXiv:2405.17046v1 [quant-ph] 27 May 2024

I. INTRODUCTION

One of the most important challenges of the information era is the ability to communicate secretly. Quantum key distribution (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 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

* rashijain_23phdam07@dtu.ac.in
† satyabrata@dtu.ac.in

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 of 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 motivate 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 aim to show that the eavesdropper Eve construct the unitary transformation in such a way that it may transform her input ancilla state to the different ancilla component states at the output, which are entangled. 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 component states. Secondly, our aim is to show that the generation of the secret key also depends on the entanglement of ancilla component states.

The rest of the paper is organized as follows: In Section II, we will discuss in detail the unitary transformation constructed by Eve that may generate the entangled ancilla component states at the output of the transformation. In section III, we derive the mutual information of Alice and Eve and show that it depends on the concurrence of the entangled ancilla component states. Moreover, we also derive the mutual information of Alice and Bob, which solely depends on the disturbance D. In section IV, 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 V.

II. 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.

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 component state before interaction with the input qubit while

$|\phi_i\rangle$ and $|\psi_i\rangle$ $(i = 0, 1)$ refers to normalized ancilla component states after interaction. D denote 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

$$F + D = 1 \tag{4}$$

A. Constraints to be satisfied by the ancilla component states at the output

The constraints that can be satisfied by the ancilla component 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

$$\langle\psi_0|\psi_1\rangle = 0 \tag{6}$$

$$\mathrm{Re}\langle\phi_1|\phi_0\rangle = \frac{1-2D}{1-D} \tag{7}$$

$$\langle\phi_0|\psi_0\rangle + \overline{\langle\phi_1|\psi_1\rangle} = 0 \tag{8}$$

B. Structure of the two-qubit ancilla component states

We are now in a position to determine the structure of the two-qubit ancilla component states generated at the output when the initial ancilla component states of Eve interact with the qubit sent by the sender, Alice. The specific form of the ancilla component states may be obtained by using the constraints given in (6-8).

1. *Specific form of $|\psi_0\rangle$ and $|\psi_1\rangle$*

To start with, let us first choose the ancilla component 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 component states $|\psi_0\rangle$ and $|\psi_1\rangle$ represents 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. *Specific form of $|\phi_0\rangle$ and $|\phi_1\rangle$*

Let us now move on to the ancilla component 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

$$|\phi_0\rangle = \alpha_{\phi_0}|00\rangle + \beta_{\phi_0}|01\rangle + \gamma_{\phi_0}|10\rangle + \delta_{\phi_0}|11\rangle \quad \text{with } |\alpha_{\phi_0}|^2 + |\beta_{\phi_0}|^2 + |\gamma_{\phi_0}|^2 + |\delta_{\phi_0}|^2 = 1 \tag{13}$$

and

$$|\phi_1\rangle = \alpha_{\phi_1}|00\rangle + \beta_{\phi_1}|01\rangle + \gamma_{\phi_1}|10\rangle + \delta_{\phi_1}|11\rangle \quad \text{with } |\alpha_{\phi_1}|^2 + |\beta_{\phi_1}|^2 + |\gamma_{\phi_1}|^2 + |\delta_{\phi_1}|^2 = 1 \tag{14}$$

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 component states $|\phi_0\rangle$ and $|\phi_1\rangle$ given in (13) and (14) reduces to

$$|\chi_0\rangle = \beta_{\phi_0}|01\rangle + \gamma_{\phi_0}|10\rangle, \ |\beta_{\phi_0}|^2 + |\gamma_{\phi_0}|^2 = 1 \tag{17}$$

$$|\chi_1\rangle = \beta_{\phi_1}|01\rangle + \gamma_{\phi_1}|10\rangle,\ |\beta_{\phi_1}|^2 + |\gamma_{\phi_1}|^2 = 1 \tag{18}$$

It may be easily seen that the ancilla component 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

$$\begin{aligned} C_{\chi_0} &= 2|\mathrm{Re}(\beta_{\phi_0}\gamma^*_{\phi_0})| \\ &= 2|b_1 g_1 + b_2 g_2| \end{aligned} \tag{19}$$

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

$$C^2_{\chi_0} = 4\Big(|\beta_{\phi_0}|^2(1 - |\beta_{\phi_0}|^2)\Big) \tag{20}$$

Solving (20), we can obtain the value of $|\beta_{\phi_0}|^2$ in terms of the concurrence C_{χ_0} as

$$|\beta_{\phi_0}|^2 = \frac{1 \pm \sqrt{1 - C^2_{\chi_0}}}{2}, 0 \le C_{\chi_0} \le 1 \tag{21}$$

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

$$|\beta_{\phi_0}|^2 = \frac{1 - \sqrt{1 - C^2_{\chi_0}}}{2} \tag{22}$$

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

$$|\beta_{\phi_0}|^2 = \frac{1 + \sqrt{1 - C^2_{\chi_0}}}{2} \tag{23}$$

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 component state $|\chi_1\rangle$ can be denoted by C_{χ_1}, which is given by

$$C^2_{\chi_1} = 4\Big(|\beta_{\phi_1}|^2(1 - |\beta_{\phi_1}|^2)\Big) \tag{24}$$

Therefore, $|\beta_{\phi_1}|^2$ can be expressed in terms of the concurrence C_{χ_1} as

$$|\beta_{\phi_1}|^2 = \frac{1 \pm \sqrt{1 - C^2_{\chi_1}}}{2}, 0 \le C_{\chi_1} \le 1 \tag{25}$$

Following the similar discussion above, we can have

$$\begin{aligned} &(i)\ \text{ If }\ |\beta_{\phi_1}|^2 = \frac{1 - \sqrt{1 - C^2_{\chi_1}}}{2}\ \text{ then }\ |\beta_{\phi_1}|^2 \in \left[0, \frac{1}{2}\right] \\ &(ii)\ \text{ If }\ |\beta_{\phi_1}|^2 = \frac{1 + \sqrt{1 - C^2_{\chi_1}}}{2}\ \text{ then }\ |\beta_{\phi_1}|^2 \in \left[\frac{1}{2}, 1\right] \end{aligned} \tag{26}$$

III. 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]

$$I^{XY} = \sum_x \sum_y p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \tag{27}$$

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 component states to extract information from the intercepted qubit, which was sent by the sender, Alice.

A. Mutual information of Alice and Eve

Let Alice prepares a two-qubit maximally entangled state of the form

$$|\psi^{(B_1)}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{28}$$

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

component 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 component 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

$$(i)\quad I_1^{AE}(C_{\chi_0},C_{\chi_1},\tau_1,D) = F\left(\frac{1-\sqrt{1-C_{\chi_0}^2}}{2}+\frac{1-\sqrt{1-C_{\chi_1}^2}}{2}\right)+D+\epsilon\left[F\left(\frac{1-\sqrt{1-C_{\chi_0}^2}}{2}\right)+\eta, F\left(\frac{1-\sqrt{1-C_{\chi_1}^2}}{2}\right)+\eta\tau_1^2\right],$$
$$|\beta_{\phi_0}|^2\in\left[0,\frac{1}{2}\right], |\beta_{\phi_1}|^2\in\left[0,\frac{1}{2}\right] \tag{35}$$

$$(ii)\quad I_2^{AE}(C_{\chi_0},C_{\chi_1},\tau_1,D) = F\left(\frac{1-\sqrt{1-C_{\chi_0}^2}}{2}+\frac{1+\sqrt{1-C_{\chi_1}^2}}{2}\right)+D+\epsilon\left[F\left(\frac{1-\sqrt{1-C_{\chi_0}^2}}{2}\right)+\eta, F\left(\frac{1+\sqrt{1-C_{\chi_1}^2}}{2}\right)+\eta\tau_1^2\right],$$
$$|\beta_{\phi_0}|^2\in\left[0,\frac{1}{2}\right], |\beta_{\phi_1}|^2\in\left[\frac{1}{2},1\right] \tag{36}$$

$$(iii)\quad I_3^{AE}(C_{\chi_0},C_{\chi_1},\tau_1,D) = F\left(\frac{1+\sqrt{1-C_{\chi_0}^2}}{2}+\frac{1-\sqrt{1-C_{\chi_1}^2}}{2}\right)+D+\epsilon\left[F\left(\frac{1+\sqrt{1-C_{\chi_0}^2}}{2}\right)+\eta, F\left(\frac{1-\sqrt{1-C_{\chi_1}^2}}{2}\right)+\eta\tau_1^2\right],$$
$$|\beta_{\phi_0}|^2\in\left[\frac{1}{2},1\right], |\beta_{\phi_1}|^2\in\left[0,\frac{1}{2}\right] \tag{37}$$

$$(iv)\quad I_4^{AE}(C_{\chi_0},C_{\chi_1},\tau_1,D) = F\left(\frac{1+\sqrt{1-C_{\chi_0}^2}}{2}+\frac{1+\sqrt{1+C_{\chi_1}^2}}{2}\right)+D+\epsilon\left[F\left(\frac{1+\sqrt{1-C_{\chi_0}^2}}{2}\right)+\eta, F\left(\frac{1+\sqrt{1-C_{\chi_1}^2}}{2}\right)+\eta\tau_1^2\right]$$
$$|\beta_{\phi_0}|^2\in\left[\frac{1}{2},1\right], |\beta_{\phi_1}|^2\in\left[\frac{1}{2},1\right] \tag{38}$$

B. Mutual information of Alice and Bob

Eve extract the information from the intercepted qubit by using her ancilla component 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

$$I^{AB}(D) = 1 + D\log_2 D + (1-D)\log_2(1-D) \quad (40)$$

IV. 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 component 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.

A. Relationship between $I^{AE}(C_{\psi_0}, C_{\chi_0}, D)$ and the concurrences C_{ψ_0} and C_{χ_0} of the ancilla component 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:

$$|\beta_{\phi_0}|^2 + |\beta_{\phi_1}|^2 = 1 \quad (41)$$

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

$$\begin{aligned} I^{AE}(|\beta_{\phi_0}|^2, \tau_1, D) = 1 &+ \left[(1-D)|\beta_{\phi_0}|^2 + \eta\right] \times \\ &\log_2\left[(1-D)|\beta_{\phi_0}|^2 + \eta\right] + \\ &\left[(1-D)(1-|\beta_{\phi_0}|^2) + \eta\tau_1^2\right] \times \\ &\log_2\left[(1-D)(1-|\beta_{\phi_0}|^2) + \eta\tau_1^2\right] \end{aligned} \quad (42)$$

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 $C_{\chi_0}^2$ of the ancilla component 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

$$\begin{aligned} I^{AE}(C_{\psi_0}, C_{\chi_0}, D) = 1 &+ \frac{1}{2}\left((1-\sqrt{1-C_{\chi_0}^2}) + \frac{D\left(C_{\psi_0}^2 - (1-\sqrt{1-C_{\chi_0}^2})(1-\sqrt{1-C_{\psi_0}^2})\right)}{1-\sqrt{1-C_{\psi_0}^2}}\right) \\ &\log_2\left[\frac{1}{2}\left((1-\sqrt{1-C_{\chi_0}^2}) + \frac{D\left(C_{\psi_0}^2 - (1-\sqrt{1-C_{\chi_0}^2})(1-\sqrt{1-C_{\psi_0}^2})\right)}{1-\sqrt{1-C_{\psi_0}^2}}\right)\right] + \\ &\frac{1}{2}\left((1+\sqrt{1-C_{\chi_0}^2}) + \frac{D\left(2-C_{\psi_0}^2 - 2\sqrt{1-C_{\psi_0}^2} - \left(1+\sqrt{1-C_{\chi_0}^2}\right)\left(1-\sqrt{1-C_{\psi_0}^2}\right)\right)}{1-\sqrt{1-C_{\psi_0}^2}}\right) \\ &\log_2\left[\frac{1}{2}\left((1+\sqrt{1-C_{\chi_0}^2}) + \frac{D\left(2-C_{\psi_0}^2 - 2\sqrt{1-C_{\psi_0}^2} - \left(1+\sqrt{1-C_{\chi_0}^2}\right)\left(1-\sqrt{1-C_{\psi_0}^2}\right)\right)}{1-\sqrt{1-C_{\psi_0}^2}}\right)\right] \end{aligned} \quad (43)$$

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

$$I^{AB}(D) - I^{AE}(C_{\psi_0}, C_{\chi_0}, D) > 0, \quad (44)$$

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

$$I^{AE}(C_{\psi_0},D) = 1+\frac{1}{2}\left(1-(1-2D)\sqrt{1-C_{\psi_0}^2}\right)\log_2\left[\frac{1}{2}\left(1-(1-2D)\sqrt{1-C_{\psi_0}^2}\right)\right]+$$
$$\frac{1}{2}\left(1+(1-2D)\sqrt{1-C_{\psi_0}^2}\right)\log_2\left[\frac{1}{2}\left(1+(1-2D)\sqrt{1-C_{\psi_0}^2}\right)\right] \quad (45)$$

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 the Fig.1.

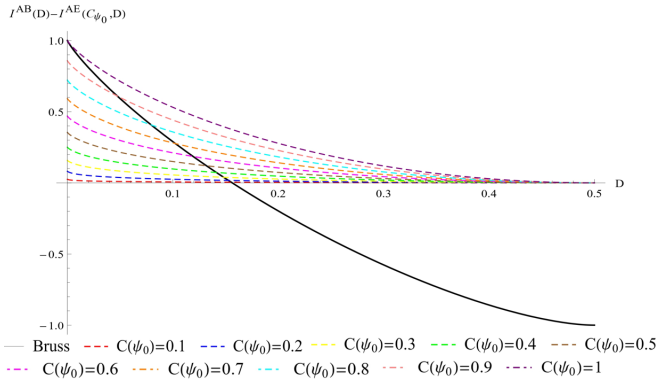


FIG. 1. Secret key generation for the six-state modified QKD protocol with each ancilla component state as entangled states compared to the six-state protocol adopted by [29]. The D-axis represents the amount of disturbance $D \in (0, 0.5)$ in the system as a result of Eve's interference and the $I^{AB}(D) - I^{AE}(C_{\psi_0}, D)$ axis represents the possibility for generation of the secret key between Alice and Bob in Eve's presence.

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.

B. Relationship between the concurrence C_{ψ_0} of the ancilla component 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 component states.
Recalling (42), let's re-express it in the following way

$$I^{AE}(|\beta_{\phi_0}|^2, \tau_1^2, D) = 1 + \left[(1-D)|\beta_{\phi_0}|^2 + \eta\right] \times \log_2\left[(1-D)|\beta_{\phi_0}|^2 + \eta\right] + \left[(1-D)|\beta_{\phi_1}|^2 + \eta\tau_1^2\right] \times \log_2\left[(1-D)|\beta_{\phi_1}|^2 + \eta\tau_1^2\right] \quad (46)$$

If we now choose the parameters $|\beta_{\phi_0}|^2$ and $|\beta_{\phi_1}|^2$ in such a way that

$$|\beta_{\phi_0}|^2 = \frac{1}{1+\tau_1^2}, |\beta_{\phi_1}|^2 = \frac{\tau_1^2}{1+\tau_1^2} \quad (47)$$

then (41) holds and the mutual information of Alice and Eve will now become independent of D, and it may be given as

$$I^{AE}(|\beta_{\phi_0}|^2, |\beta_{\phi_1}|^2) = 1 + |\beta_{\phi_0}|^2 \log_2 |\beta_{\phi_0}|^2 + |\beta_{\phi_1}|^2 \log_2 |\beta_{\phi_1}|^2 \quad (48)$$

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 so 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) gives

$$\left(1-\sqrt{1-C_{\chi_0}^2}\right)\left(1-\sqrt{1-C_{\psi_0}^2}\right) = C_{\psi_0}^2$$
$$\implies \left(\sqrt{1-C_{\psi_0}^2}-1\right)\left(\sqrt{1-C_{\psi_0}^2}+\sqrt{1-C_{\chi_0}^2}\right) = 0 \quad (49)$$

Since $0 < C_{\psi_0} \leq 1$ and $0 \leq C_{\chi_0} < 1$ so (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

$$C_{\psi_0} = C_{\chi_0} \quad (50)$$

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 $(0.5, 1]$.
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

$$\begin{aligned} I^{AE}(C_{\psi_0}) =& 1 + \frac{1+\sqrt{1-C_{\psi_0}^2}}{2} \log_2 \frac{1+\sqrt{1-C_{\psi_0}^2}}{2} \\ &+ \frac{1-\sqrt{1-C_{\psi_0}^2}}{2} \log_2 \frac{1-\sqrt{1-C_{\psi_0}^2}}{2} \end{aligned}$$

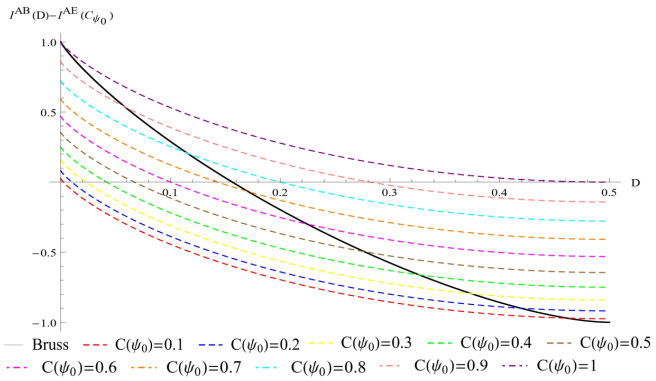


FIG. 2. The figure represents that in the modified six-state QKD protocol, the secret key is generated when the curves are lying above $D-$ axis in $(D, I^{AB}(D) - I^{AE}(C_{\psi_0}))$-plane and the secret key is not generated when the curves lying below the $D-$ axis. The different curves in the graph are drawn for different values of the concurrence of the ancilla component states. The black curve represents when the secret key is generated or is not generated in Bruss's six-state QKD protocol.

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 1.
(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 component state $|\psi_0\rangle$ is maximally entangled state.

V. 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 component 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 component 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 component 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 component 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 by [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 component 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 component 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.

[1] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Quantum cryptography, Rev. Mod. Phys. **74**, 145 (2002).
[2] C. Portmann and R. Renner, Security in quantum cryptography, 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. John Wiley & Sons (1999).
[5] C. Arndt, Information measures: information and its description in science and engineering. Springer Science & Business Media (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 ed.), New Jersey: Prentice Hall (1995).
[8] R. Horodecki, P. Horodecki, M. Horodecki, and 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, and 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, and A. Zeilinger, Quantum information: An introduction to basic theoretical concepts and experiments, **173**. Springer (2003).
[13] M. Hillery, V. Buzek, and Berthiaume, Phys. Rev. A **59**, pp. 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, and A. K. Fedorov, Quantum Sci. Technol. **3**, 035004 (2018).
[15] C. H. Bennett and 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, and H. Zbinden, Appl. Phys. Lett. **87**, 194108 (2005).
[19] B. Kraus, N. Gisin, and R. Renner, Phys. Rev. Lett. **95**, 080501 (2005).
[20] R. Renner, N. Gisin, and B. Kraus, Phys. Rev. A **72**, 012332 (2005).
[21] D. Gottesman, and H. K. Lo, IEEE Transactions on Information Theory, **49**, 457-475 (2003).
[22] N. J. Beaudry, M. Lucamarini, S. Mancini, and R. Renner, Phys. Rev. A **88**, 062302 (2013).
[23] D. Mayers, J. ACM. **48**, 351-406 (2001).
[24] H. K. Lo and H. F. Chau, Science **283**, 2050 (1999).
[25] H. K. Lo, Quantum Inf. Comput. **1**, 2 (2001).
[26] P. W. Shor and J. Preskill, Phys. Rev. Lett. **85**, 441 (2000).
[27] C. A. Fuchs, N. Gisin, R. B. Griffiths, C. S. Niu, and 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 and A. Peres, Phys. Rev. Lett. **85**, 3313 (2000).
[31] D. Bruß and C. Macchiavello, Phys. Rev. Lett. **88**, 127901 (2002).
[32] Z. Shadman, H. Kampermann, T. Meyer, and 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. and Peev, Rev. Mod. Phys. **81**, 1301 (2009).
[35] N. Gisin, and S. Wolf, Phys. Rev. Lett. **83**, 4200 (1999).
[36] I. Csiszar and 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 (2002).

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 $0 < \tau_1^2 < 1$, so we have

$$0 < \frac{2 - C_{\psi_0}^2 + 2\sqrt{1-C_{\psi_0}^2}}{C_{\psi_0}^2} < 1 \tag{A5}$$

We may observe that LHS inequality is always true but simplifying the RHS inequality, we get

$$\sqrt{1-C_{\psi_0}^2}\left(\sqrt{1-C_{\psi_0}^2}+1\right) < 0 \tag{A6}$$

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 component state $|\phi_0\rangle$ and $|\phi_1\rangle$ in the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ as

$$\begin{aligned} |\phi_0\rangle =& \alpha_{\phi_0}|00\rangle + \beta_{\phi_0}|01\rangle + \gamma_{\phi_0}|10\rangle + \delta_{\phi_0}|11\rangle, \\ \text{with } & |\alpha_{\phi_0}|^2 + |\beta_{\phi_0}|^2 + |\gamma_{\phi_0}|^2 + |\delta_{\phi_0}|^2 = 1 \end{aligned} \tag{B1}$$

and

$$\begin{aligned} |\phi_1\rangle =& \alpha_{\phi_1}|00\rangle + \beta_{\phi_1}|01\rangle + \gamma_{\phi_1}|10\rangle + \delta_{\phi_1}|11\rangle \\ \text{with } & |\alpha_{\phi_1}|^2 + |\beta_{\phi_1}|^2 + |\gamma_{\phi_1}|^2 + |\delta_{\phi_1}|^2 = 1 \end{aligned} \tag{B2}$$

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

$$|\chi_0\rangle = \beta_{\phi_0}|01\rangle + \gamma_{\phi_0}|10\rangle,\ |\beta_{\phi_0}|^2 + |\gamma_{\phi_0}|^2 = 1$$
$$|\chi_1\rangle = \beta_{\phi_1}|01\rangle + \gamma_{\phi_1}|10\rangle,\ |\beta_{\phi_1}|^2 + |\gamma_{\phi_1}|^2 = 1$$

Check for updates

Received: 23 March 2023 | Revised: 8 April 2024 | Accepted: 4 May 2024

DOI: 10.1111/exsy.13626

ORIGINAL ARTICLE

Expert Systems WILEY

Multi-armed bandit based online model selection for concept-drift adaptation

Jobin Wilson[1,2] | Santanu Chaudhury[2,3] | Brejesh Lall[2]

[1]R&D Department, Flytxt, Trivandrum, India

[2]Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, India

[3]Department of Computer Science and Engineering, Indian Institute of Technology Jodhpur, Jodhpur, India

Correspondence
Jobin Wilson, R&D Department, Flytxt, Trivandrum, Kerala, India.
Email: jobin.wilson@flytxt.com

Abstract

Ensemble methods are among the most effective concept-drift adaptation techniques due to their high learning performance and flexibility. However, they are computationally expensive and pose a challenge in applications involving high-speed data streams. In this paper, we present a computationally efficient heterogeneous classifier ensemble entitled OMS-MAB which uses online model selection for concept-drift adaptation by posing it as a non-stationary multi-armed bandit (MAB) problem. We use a MAB to select a single *adaptive learner* within the ensemble for learning and prediction while systematically exploring promising alternatives. Each ensemble member is made drift resistant using explicit drift detection and is represented as an arm of the MAB. An exploration factor ϵ controls the trade-off between predictive performance and computational resource requirements, eliminating the need to continuously train and evaluate all the ensemble members. A rigorous evaluation on 20 benchmark datasets and 9 algorithms indicates that the accuracy of OMS-MAB is statistically at par with state-of-the-art (SOTA) ensembles. Moreover, it offers a significant reduction in execution time and model size in comparison to several SOTA ensemble methods, making it a promising ensemble for resource constrained stream-mining problems.

KEYWORDS
concept-drift, ensemble methods, model selection, multi-armed bandits

1 | INTRODUCTION

With rapid advances in the Internet of Things (IoT), computing infrastructure and communication networks, data is increasingly generated, organized, and consumed as streams. In several real-world applications such as weather forecasting, e-commerce recommendations, and early warning systems, the underlying data generation process evolves over time causing *concept-drift*. Concept-drift means that the underlying data distribution of a stream changes over time in unforeseen ways (Lu et al., 2018). Offline storage and batch processing techniques would be ineffective in this context due to large data volume, fast arrival rate, dynamically changing data distribution, and computational resource constraints. Unlike learning techniques which assume data to be independent and identically distributed (i.i.d.), specialized algorithms that rapidly adapt to concept-drift to provide fast predictions and model updates while operating within the computational resource constraints are essential in this context (Alippi et al., 2013).

Ensemble learning is a very effective concept-drift adaptation strategy for data stream mining (Gomes, Barddal, et al., 2017). Depending on their composition, ensembles could be classified as heterogeneous or homogeneous. While the former employs different classifier types as base learners, the latter uses a single classifier type for all its members (Neto & Canuto, 2021). Ensembles adapt to concept-drift by adding new members, removing obsolete members, training multiple base models, and updating model combination rules. Though ensembles provide

 © 2024 John Wiley & Sons Ltd.

high predictive performance, computational efficiency is not their strong trait (Elwell & Polikar, 2011). This limits their usage in high-speed data stream mining (Alippi et al., 2013), motivating the need for computationally efficient ensembles with competitive predictive performance.

Diversity among members is a key factor for the effectiveness of ensembles. One approach to ensure diversity is to construct a heterogeneous ensemble with base models belonging to different model families (van Rijn et al., 2018). This eliminates the need to manually select the most appropriate base learner for a stream. Furthermore, heterogeneous ensembles often use much fewer members to provide comparable predictive performance to that of homogeneous ensembles with hundreds of members (Luong et al., 2021). Moreover, studies comparing batch-incremental and instance-incremental learning indicate equivalence in their overall performance, with instance-incremental methods requiring fewer computational resources (Read, Bifet, Pfahringer, & Holmes, 2012). For the aforementioned reasons, our proposal focuses on heterogeneous ensembles which are instance-incremental.

As data streams evolve, the predictive performance of models varies over time. Thus, a classifier that best models a certain stream segment may not be the best for a different segment of the same stream (van Rijn et al., 2014). By adaptively choosing the model best suited for each stream segment, heterogeneous ensembles could increase their predictive performance and adapt to concept-drift (Rossi et al., 2014; van Rijn et al., 2018). To this end, works such as Heterogeneous Dynamic Weighted Majority (HDWM) (Idrees et al., 2020), HEterogeneous Ensemble Selection (HEES) (Luong et al., 2021) and Best Last (BLAST) (van Rijn et al., 2018) present ensembles that simultaneously train multiple heterogeneous classifiers and use online model selection to adaptively choose a subset to make predictions. Despite being less resource intensive compared to homogeneous ensembles, the complexity of their model selection mechanism and the need to simultaneously train all the ensemble members introduce significant computational overheads. Moreover, meta-learning based heterogeneous ensembles such as (van Rijn et al., 2014) and (Rossi et al., 2014) which rely on periodic algorithm selection provide marginal performance improvements while requiring extraction of computationally expensive meta-features (van Rijn et al., 2018).

While multi-armed bandits (MAB) are well studied in the context of multivariate testing (Burtini et al., 2015), its utility as a computationally efficient online model selection framework for classification on streaming data affected by concept-drift is under-explored. Moreover, most studies on online model selection focus on choosing a subset of ensemble members that are best suited for making predictions (Idrees et al., 2020; Luong et al., 2021; van Rijn et al., 2018). Identifying which ensemble members are to be used for both learning and prediction at a time instant significantly reduces computational resource requirements as it eliminates the need to continuously train and evaluate all ensemble members. However, environment non-stationarity makes this non-trivial. Furthermore, despite several studies on enhancing the predictive performance of heterogeneous ensembles for data stream mining, little work has been dedicated to improving their computational efficiency. The above aspects motivate the present study.

In this paper, we present a heterogeneous ensemble named OMS-MAB (Online Model Selection Multi-Armed Bandit) which uses a non-stationary MAB for online model selection to adapt to concept-drift. Our algorithm operates at two levels. At the inner-level, a pool of diverse instance-incremental *adaptive learners* (ALs) each equipped with a drift detector is maintained to solve the original decision problem. At the outer level, a MAB selects the adaptive learner (AL) best suited for an instance based on past experience, while systematically exploring promising alternatives. The MAB switches between *exploration* to identify the current best AL and *exploitation* to train and evaluate the best AL. This significantly reduces computational resource requirements as only a single AL is trained and evaluated during *exploitation*, independent of the total number of ALs within the ensemble. Though all ALs undergo training and evaluation during *exploration*, the extent of exploration is controlled using the exploration factor ϵ. Furthermore, MAB based model selection is computationally lightweight compared to approaches requiring costly meta-feature extraction or complex ensemble selection. Moreover, explicit drift detection at the inner-level periodically resets ensemble members, preventing their unbounded growth in size and complexity. In summary, OMS-MAB addresses 3 challenges posed by several state-of-the-art (SOTA) instance-incremental ensembles: high computational cost, model selection complexity, and unbounded model growth. While computational efficiency is attained by adaptively training a single ensemble member for most parts of the stream, model selection is simplified using a non-stationary MAB. Furthermore, each ensemble member is equipped with a drift-detector and a backup learner to prevent their unbounded growth. Our experiments on OMS-MAB covering 20 benchmark streams and 9 classification algorithms indicate that its predictive performance is statistically at par with SOTA ensembles, while with a significant reduction in execution time and model size, making it attractive for resource constrained environments.

This paper makes three contributions:

1. It poses online model selection in non-stationary environments as a bandit problem and uses a non-stationary MAB to adaptively choose the classifier member best suited for learning and prediction from time to time. To the best of our knowledge, our proposal is the first to approach concept-drift adaptation in a classification setting as a bandit based online ensemble selection problem.
2. It combines MAB based online model selection and ALs with explicit drift detection for computational efficiency. Unlike popular instance-incremental ensembles like (Bifet, Holmes, & Pfahringer, 2010), (Frías-Blanco et al., 2016), and (van Rijn et al., 2018) that continuously train all ensemble members, our approach optimizes computational needs by adaptively identifying a single member for training by using a MAB, while preventing unbounded model growth using drift-detectors.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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. Complexity analysis and empirical evaluation of the proposed approach indicates that it is faster and consumes less computational resources compared to several SOTA ensembles, while providing statistically equivalent predictive performance.

The rest of the paper is organized as follows. In Section 2, we review the related work. Our proposed method is described in Section 3. In Section 4 and Section 5 we present the experimental methodology and results respectively. Finally, Section 6 concludes the paper.

2 | RELATED WORKS

2.1 | Online ensemble selection

Several heterogeneous ensembles simultaneously train multiple classifiers and choose a subset to make predictions. BLAST simultaneously trains a group of heterogeneous classifiers, while using an online performance estimation framework to dynamically select a single classifier for prediction on each test instance (van Rijn et al., 2018). While window-based and fading factor-based performance estimation have been proposed, empirical evaluations favor the fading factor-based variant. Since the performance of BLAST depends on the number of base learners and their diversity, it becomes computationally expensive for large ensembles as each member needs to be trained and evaluated continuously. Moreover, the absence of explicit drift detection may result in unbounded growth in model size and complexity.

HDWM algorithm adaptively switches between different types of base classifiers within an ensemble to increase predictive performance (Idrees et al., 2020). It uses both an active and passive approach to handle concept-drift, while using seed learners from different model families to ensure ensemble diversity. Weaker learners are temporarily removed from the ensemble when their weights drop below a threshold. Although HDWM reduces the manual effort required to select base classifiers for ensembles, its performance is lower compared to some of the SOTA homogeneous ensembles (Luong et al., 2021).

HEterogeneous Ensemble Selection (HEES) identifies a competent subset of base classifiers with high predictive performance and high confidence, for making predictions (Luong et al., 2021). The predictive performance of classifiers is estimated from their prequential accuracy whereas confidence estimation uses adaptive reliability thresholds. HEES uses ADWIN (Bifet & Gavalda, 2007) to track the performance of each base classifier and resets the respective classifier whenever a change is detected. Despite the competitive performance of HEES, computational resources for training and evaluation of multiple classifiers will still be incurred.

Diversity Pool (DP) algorithm maintains a pool of classifiers and selects a single representative member for making predictions. If a single member cannot be confidently chosen for prediction, the pool is used as a weighted majority vote ensemble (Chiu & Minku, 2018). New models are introduced into the pool once concept-drift is detected, enabling each pool member to hold the knowledge corresponding to a concept. It uses a diversity metric to maximize the difference between pool members for allowing the reuse of relevant past knowledge when recurring concepts are encountered. Though the adaptation strategy of DP enables handling recurring concepts well, it is not designed for handling diverse drift types.

Dynamic Ensemble Selection for Drift Detection (DESDD) (Albuquerque et al., 2019) creates a population of classifier ensembles each with a different level of diversity and dynamically selects a single classifier ensemble for making predictions based on global prequential accuracy. It uses explicit drift detection to reset the population of classifiers upon encountering concept-drift. Though DESDD provides competitive accuracy, its computational complexity is very high as the entire population of classifier ensembles are trained and evaluated continuously.

2.2 | Meta-learning based adaptive ensembles

Few heterogeneous ensembles use meta-learning to identify learning algorithms best suited for a specific window of instances within data streams (Rossi et al., 2014; van Rijn et al., 2014). MetaStream periodically chooses among regression algorithms based on meta-features extracted from past and incoming data windows (Rossi et al., 2014). A meta-classifier is regularly induced to map meta-features to the actual performance of regression algorithms, for predicting the best algorithm in the subsequent window.

Meta-learning based algorithm selection for data stream classification was first proposed in (van Rijn et al., 2014). Their approach determines the best classifier among several candidates for predicting the next window of instances by considering data characteristics of the previous window and meta-knowledge from streams processed previously. Limitations of the above approaches include determining the optimal window size over which meta-features need to be extracted, and the methods inherently being batch-incremental which limits their applicability on high-speed instance-incremental streams (Luong et al., 2021). Furthermore, meta-feature extraction introduces additional computational overheads (van Rijn et al., 2018).

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

Fast Adaptive Stacking of Ensembles (FASE) (Frías-Blanco et al., 2016) uses an adaptive meta-learner to combine the predictions of base classifiers within the ensemble for countering concept-drift. Base classifiers themselves are adaptive learners that track the error rate of the induced classifier using the Hoeffding-based Drift Detection Method (HDDM) (Frias-Blanco et al., 2014). Upon encountering a warning level, an alternate classifier is trained by the adaptive learner. In case the warning is followed by a drift detection, the adaptive learner replaces the original classifier with the alternate classifier. Each adaptive learner has at most two classifiers whose predictions are combined by weighted voting to generate a single prediction. The meta-level adaptive learner combines the predictions of all base-level adaptive learners in the ensemble to make the final prediction. Though FASE provides competitive accuracy, computational complexity due to its three-level ensemble structure will still be incurred.

2.3 | Multi-armed bandits

MAB provides a framework for trade-off among exploring the environment to discover profitable actions and exploiting the existing knowledge from past experience, to maximize cumulative rewards (Garivier & Moulines, 2008; Sutton & Barto, 2018). At each time step t, an agent pulls an arm $I_t \in \{1,\ldots,K\}$ according to a policy π learned from the sequence of past arm pulls and the associated rewards, to receive a reward $R_t(I_t)$. In the case of a stationary MAB, rewards $\{R_t(i)\}_{t\geq 1}$ associated with each arm $i \in \{1,\ldots,K\}$ is assumed to be i.i.d., with $\mu(i)$ as the expected reward corresponding to arm i. The optimal policy π^* is the one which always pulls the best arm $i^* \in \{1,\ldots,K\}$ with the largest expected reward.

$$\mu(*) = \max_{1\leq i\leq K} \mu(i), \quad i^* = \arg\max_{1\leq i\leq K} \mu(i) \tag{1}$$

The agent improves its policy π by minimizing cumulative *regret*, which is the expected difference between the total rewards obtained if the agent would have followed the optimal policy π^* and the total rewards that was obtained by following the policy π (Garivier & Moulines, 2008).

The expected regret after T time steps would be:

$$\mathbb{E}_\pi\left[\sum_{t=1}^{T}\{\mu(*)-\mu(I_t)\}\right] = \sum_{i\neq i^*}\{\mu(*)-\mu(i)\}\mathbb{E}_\pi[N_T(i)], \tag{2}$$

where $N_T(i)$ is the number of times the arm i was pulled in T time steps and $\mathbb{E}_\pi$ is the expectation under policy π. Popular strategies for solving stationary MABs include ϵ-greedy action selection, Upper Confidence Bound (UCB) and Thompson sampling (Burtini et al., 2015; Sutton & Barto, 2018). In the $\epsilon-$greedy approach, the agent selects an arm $i \in \{1,\ldots,K\}$ at random for $0 \leq \epsilon \leq 1$ fraction of the time, and the arm i^* with the highest estimated action value $Q(i^*)$ otherwise.

For non-stationary MABs, the underlying reward distributions evolve, resulting in the best arm getting changed from time to time. Specialized algorithms such as Discounted UCB, Sliding-Window UCB, Exp3.R, and Kalman Filtered Bandits have been proposed for this more challenging setting (Burtini et al., 2015). However, a simple solution to handle non-stationarity is to allow a decay of the rewards such that the contribution of long-past rewards fades away while the effect of recent rewards reflect in the action value estimation (Burtini et al., 2015; Sutton & Barto, 2018). This can be incorporated into the incremental update rule for action value estimation of ϵ-greedy algorithm by using a constant step-size parameter as per the Equation (3).

$$Q_{t+1}(i) = Q_t(i) + \alpha[R_t(i) - Q_t(i)] \tag{3}$$

Here $Q_t(i)$ is the estimated action value for arm i at time step t, $R_t(i)$ is the reward obtained by pulling arm i at time step t, and $\alpha \in (0,1]$ is the step-size parameter.

3 | PROPOSED METHOD

We introduce OMS-MAB, a heterogeneous ensemble that uses online model selection for concept-drift adaptation. Our framework has a two-level learning architecture as presented in Figure 1. At the inner level, a pool of diverse instance-incremental ALs is maintained for solving the original decision problem. Each AL is a *Drift detection method classifier* (Gama et al., 2004) which induces a main classifier for prediction and uses HDDM_A-test (Frias-Blanco et al., 2014) to decide if an alternate classifier needs to be trained. The alternate classifier replaces the main classifier

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

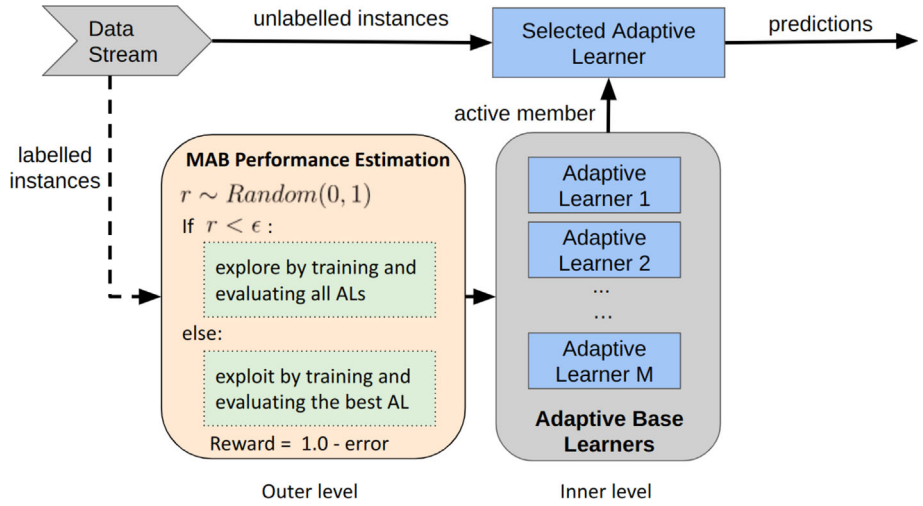


FIGURE 1 OMS-MAB ensemble architecture. A pool of heterogeneous ALs each with a builtin HDDM_A drift detector is maintained to solve the original prediction problem. A non-stationary MAB periodically selects the best AL for training and prediction based on past experience, while systematically exploring promising alternatives.

TABLE 1 Summary of notations.

Notations	Description
S	Data stream
x_t, y_t	x_t is the data instance and y_t is the corresponding data label at time t
M	Number of ALs in the ensemble
Ψ_i	Ensemble member at the index position i
$\hat{y}_t^i$	Prediction of Ψ_i for the unlabelled instance x_t
L_t^i	Loss of the ith ensemble member at time t
R_t^i	Reward for the ith ensemble member at time t
P_t^j	Performance estimate for the jth ensemble member at time t
θ	Fading factor
Ψ	Ensemble with M heterogeneous learners
ϵ	Exploration factor of the MAB
L	Loss function
t	A specific time step

once concept-drift is detected. Any instance-incremental classifier, irrespective of being stand-alone or ensemble-based, could be used to construct an AL. Outer level MAB identifies a single inner level AL best suited for an instance. It tracks the competency of each AL by translating inner level prediction errors into rewards and uses a non-stationary MAB with $\epsilon-$greedy action selection for a cost-effective but accurate performance estimation of ALs. Greedy action selection is used to choose a single best AL for prediction on a particular instance (behavior policy), whereas exploration required to identify and train the best AL is controlled by $\epsilon-$greedy action selection (exploration policy).

Let S be a data stream consisting of a sequence of instances $\{x_1,x_2,...x_t...\}$ where $x_t \in \mathbb{R}^d$ denotes the instance arriving at time step t, represented as a d-dimensional feature vector. The true label corresponding to x_t is denoted by $y_t \in \{y_1,y_2...y_K\}$, which belongs to a classification problem having a finite set of known class labels. We assume y_t to arrive subsequently, after which the pair (x_t,y_t) is available for incremental learning. These assumptions pertain to one of the common operational settings studied in data stream mining (Albuquerque et al., 2019; Luong et al., 2021) and are relevant for applications such as electricity price movement prediction (Harries et al., 1999) and weather prediction (Elwell & Polikar, 2011).

OMS-MAB maintains a pool of M ALs forming a heterogeneous ensemble $\Psi = \{\Psi_1,\Psi_2,..\Psi_M\}$. The outer level MAB models each ensemble member Ψ_j as an arm and adaptively selects a single *active member* Ψ_i best suited for an instance to make predictions. At each training time step, the MAB chooses among *exploration* to identify the current best AL within Ψ and *exploitation* to train and evaluate the best AL. In case of

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

exploitation, the *active member* $\Psi_i \in \Psi$ makes a prediction $\hat{y}_t^i$ for the unlabelled instance x_t. Once y_t becomes available, the corresponding loss $L_t^i = L\left(\hat{y}_t^i, y_t\right)$ is used to generate a reward $R_t^i = 1 - L_t^i$, and (x_t, y_t) is used to update Ψ_i. The reward R_t^i acts as a feedback signal for the MAB decision to use Ψ_i as the *active member* at t. On the other hand, *exploration* involves evaluating and training all ensemble members $\Psi_j \in \Psi$ to derive rewards R_t^j corresponding to their respective predictions, to update performance estimates for all arms. Though this approach is different from that of general bandit algorithms which evaluate a single arm during exploration, it provides faster estimation of expected rewards for all arms due to their simultaneous evaluation. As our prediction procedure follows a greedy policy, the above exploration strategy results in better predictions since the best performing AL can be quickly identified by the MAB. Moreover, the computational overhead associated with evaluating all arms as opposed to a single arm during *exploration* is relatively low in this case, as arms represent instance-incremental learners which are fast to train and evaluate. Our complete learning procedure is presented in Algorithm 1. Table 1 summarizes the notations used.

OMS-MAB optimistically initializes performance estimates P_t^j associated with each ensemble member Ψ_j as 1 and randomly selects a member Ψ_i as the *active member* upon initialization. The MAB treats P_t^j as the action value estimate corresponding to the jth arm at time t. Optimistic initialization ensures frequent evaluation of each ensemble member in the early stages of the stream, until they are proven to be suboptimal. Ensemble members require training and evaluation for sufficient time steps for reliable performance estimation, a prerequisite for the MAB to correctly identify the best AL for a particular instance. To this end, an ϵ-greedy exploration policy is used. A small value for ϵ (e.g., 0.01) results in inaccurate performance estimates for ALs, whereas large values introduce the computational overhead of evaluating and training all the ensemble members during the associated larger exploration. Thus, ϵ provides a trade-off between model accuracy and computational resource requirements, a desirable property for high-speed stream mining algorithms operating in resource constrained environments. The values for ϵ range from 0 to 1.

Each AL monitors the error rate of the main classifier using the HDDM_A-test. If the error rate of an AL begins to increase, a concept-drift is anticipated and a warning is triggered. In the warning phase, the corresponding AL trains an alternate classifier. In case the warning escalates to an out-of-control level, concept-drift is detected and the main classifier is replaced with the alternate classifier. This procedure is based on (Gama et al., 2004) and is summarized in Figure 2. Each AL can have two classifiers at most whose predictions are combined using weighted voting as described in (Frías-Blanco et al., 2016). We use HDDM_A for drift detection as it is a SOTA method with mathematical guarantees on the rate of false positives and false negatives while processing each instance with a constant space and time complexity (Frias-Blanco et al., 2014; Frías-Blanco et al., 2016). Though any instance-incremental classifier can be used to construct ALs, standalone learners such as Hoeffding Tree (HT) (Domingos & Hulten, 2000), Extremely Fast Decision Tree (EFDT) (Manapragada et al., 2018), NaiveBayes (NB) (John & Langley, 1995) and Perceptron (Rosenblatt, 1958) are preferable as they provide reasonable accuracy at a lower computational cost. Though the learning strategy of ALs causes periodic model resets which prevent historical knowledge reuse during recurring concepts, it provides computational savings by avoiding unbounded growth of model size and complexity.

While evaluating an AL Ψ_j, either during *exploration* or *exploitation*, the loss L_t^j associated with each prediction $\hat{y}_t^j$ is used to calculate the reward $R_t^j = 1 - L_t^j$. We make use of the zero/one loss function as proposed in (van Rijn et al., 2018). This translates to obtaining a reward of 1 for correct predictions and 0 otherwise. Performance estimate P_t^j of the corresponding AL Ψ_j at time t is updated using the Equation (4). We observe that Equation (4) takes the form of Equation (3) which represents the update rule for action value estimation in non-stationary contexts by decaying rewards. However, it is to be noted that the fading factor $\theta = 1 - \alpha$, where α is the step size. Thus, the MAB ensures that accurate performance estimates of each member is maintained by combining their historical performance with recent accuracy using the fading factor θ. The values for θ range from 0 to 1.

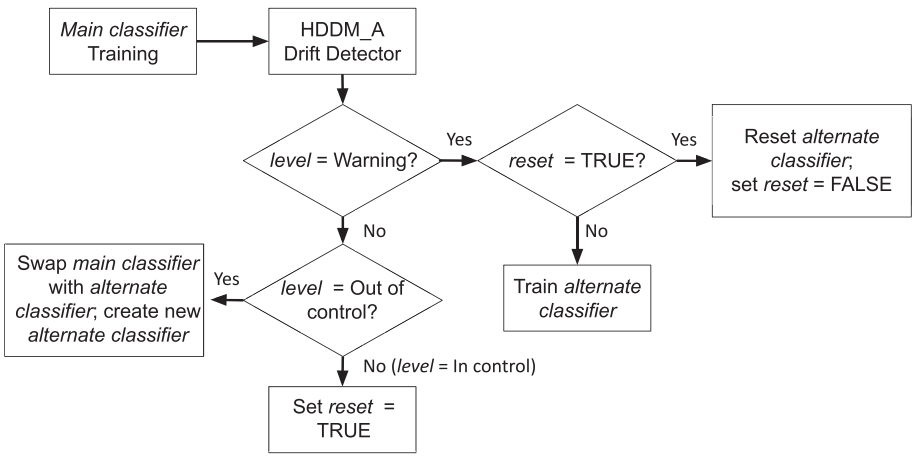


FIGURE 2 Learning procedure for adaptive learners within OMS-MAB.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

ALGORITHM 1 OMS-MAB training

Input:

Set of adaptive learners Ψ, fading factor θ, loss function L

exploration factor ϵ, data stream S

Output:

active member index $i \in \{1,2,...|\Psi|\}$

1: Initialize all adaptive learners in Ψ

2: $P_t^j \leftarrow 1 \quad \forall j \in \{1,2,...|\Psi|\}, t = 1$

3: $active_member \leftarrow random.uniform(1,|\Psi|)$

4: **for all** training samples $(x_t, y_t) \in S$ **do**

5: $r \sim random.uniform(0,1)$

6: **if** $r < \epsilon$ **then** //exploration move

7: **for all** $j \in \{1,2,...|\Psi|\}$ **do**

8: $\hat{y}_t^j \leftarrow$ predicted label of x_t by Ψ_j

9: $R_t^j \leftarrow \left(1 - L\left(\hat{y}_t^j, y_t\right)\right)$

10: $P_t^j \leftarrow \theta^* P_{t-1}^j + (1-\theta)^* R_t^j$ //update estimates

11: update Ψ_j with (x_t, y_t) using procedure in Figure 2

12: **end for**

13: **else** //exploitation move

14: $i \leftarrow active_member$

15: $\hat{y}_t^i \leftarrow$ predicted label of x_t by Ψ_i

16: $R_t^i \leftarrow \left(1 - L\left(\hat{y}_t^i, y_t\right)\right)$

17: $P_t^i \leftarrow \theta^* P_{t-1}^i + (1-\theta)^* R_t^i$ //update estimates

18: update Ψ_i with (x_t, y_t) using procedure in Figure 2

19: **end if**

20: $active_member \leftarrow \underset{1 \le j \le |\Psi|}{\arg\max} P(j)$ //greedy selection

21: **end for**

22: **return** *active_member*

$$P_t^j = \theta^* P_{t-1}^j + (1-\theta)^* R_t^j \tag{4}$$

The prediction procedure of OMS-MAB is straightforward. The MAB uses a greedy behavior policy to choose the *active member* which makes the prediction at each time step. As the MAB decision relies on the action value estimates derived using the exploration policy, the AL best suited for the current instance is automatically selected to be the *active member* as the stream progresses. The prediction corresponding to each unlabelled sample x_t is made by the *active member* at that time instance.

We note that unlike ensembles such as BLAST which uses a passive approach for drift adaptation, OMS-MAB uses both an active and passive strategy to handle concept-drift, enabling it to effectively deal with both gradual and abrupt drift. Our passive strategy involves using the best performing AL for prediction at each instant, while eliminating the need to continuously train and evaluate all ensemble members. On the other hand, the active strategy uses explicit drift detection within each AL to keep them up to date with the latest concept. Moreover, as ensemble members are trained from limited samples based on MAB's stochastic decisions, explicit drift detection is essential to quickly discard obsolete knowledge. The two-level architecture of OMS-MAB enables both scalability and adaptability in handling large volume-high speed data streams affected by diverse drift types. From a scalability perspective, OMS-MAB capitalizes on the inherent diversity among heterogeneous learners within the ensemble, eliminating the need for computationally expensive techniques like online bagging for diversity generation. Moreover, the MAB framework further enhances computational efficiency, as only a single ensemble member is adaptively trained for most parts of the stream, independent of the total ensemble size. This makes OMS-MAB well-suited for large data volumes. On the adaptability front, explicit drift detection and backup learners in OMS-MAB facilitate swift adaptation to streams with both abrupt and gradual drifts, even at high speeds. Though explicit drift detection discards obsolete knowledge, choosing a lower ϵ results in members getting trained only on a subset of samples. This enables OMS-MAB to reuse historical knowledge during recurring drift, in case some of the members are not already reset.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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 | Space and time complexity

We analyze the space and time complexity of OMS-MAB and compare it with BLAST, a popular heterogeneous ensemble that trains all member classifiers continuously while deterministically using the best performer at each time step to make predictions.

3.1.1 | Space complexity

OMS-MAB maintains M ALs at the inner-level and a MAB at the outer-level. Let $S(N)$ be the space required by a base classifier within an AL to store a model learned from N samples. We note that $S(N)$ depends on the base classifier. For instance, $S(N) = O(ldvc)$ for HT where l represents the number of leaves, and $S(N) = O(ndvc)$ for EFDT where n denotes the number of nodes (Manapragada et al., 2018). In both cases, d denotes the number of attributes, v the values per attribute, and c the number of classes per instance.

The space complexity of OMS-MAB depends on the total number of classifiers within the ensemble. In the worst case when all ALs are in the *warning* phase, each AL maintains a *main classifier* and an *alternate classifier*, resulting in a total of $2M$ classifiers within the ensemble. With an exploration factor of ϵ, $2M$ classifiers will be trained on $(\epsilon.N)$ samples, whereas 2 classifiers corresponding to the *active_member* will be trained on $((1-\epsilon).N)$ samples. Outer-level MAB maintains performance estimates for each AL, resulting in a space complexity of $O(M)$. On similar lines, HDDM_A detectors within each AL result in a space complexity of $O(M)$ as their individual complexity is $O(1)$ (Frias-Blanco et al., 2014). Thus, the final space complexity of OMS-MAB comes out to be $O(2M.S(\epsilon.N) + 2.S((1-\epsilon).N) + 2M)$.

On the other hand, space complexity for BLAST which trains all ensemble members wherein each member is a standalone classifier as opposed to an AL would be $O(M.S(N) + M)$. We observe that as N grows large while ϵ being set to a small value, the gain in space complexity for OMS-MAB becomes more pronounced. Furthermore, our empirical study reconfirms this analysis as indicated by the lower model size growth rate for OMS-MAB when compared to BLAST, in Figure 7. Moreover, the average model size across datasets is much lower for OMS-MAB when compared to BLAST, as captured in Table 6. For practical large-scale data stream mining applications, this translates to lower memory requirements suitable for resource constrained settings.

3.1.2 | Time complexity

Let $O(T_t)$ be the time complexity to train a base classifier on an instance and $O(T_c)$ be the time complexity for a base classifier to classify an instance. T_t and T_c depend on the base classifier. For instance, $T_t = O(dvc)$ for HT, where d is the number of attributes, v the values per attribute and c the number of classes per instance (Manapragada et al., 2018).

The time complexity of OMS-MAB with M ALs depends on the total number of standalone classifiers within the ensemble, which could be up to $2M$ in the worst case, as explained in the space complexity analysis. We consider OMS-MAB being trained on N samples. With exploration factor ϵ, $2M$ classifiers are trained on $(\epsilon.N)$ samples, whereas 2 classifiers corresponding to the *active_member* are trained on $((1-\epsilon).N)$ samples. Performance estimation updates of ALs (Algorithm 1, steps 10 and 17) and *active_member* selection (Algorithm 1, step 20) by the MAB results in a time complexity of $O(2M)$. Similarly, HDDM_A detectors within ALs contribute to a time complexity of $O(M)$ as their individual time complexity is $O(1)$ (Frias-Blanco et al., 2014). Since each training step of OMS-MAB involves the evaluation and training of ALs, its overall training time complexity is $O((2M.\epsilon.N.(T_t + T_c)) + (2.(1-\epsilon).N.(T_t + T_c)) + 3M)$. In the case of prediction, only the *active_member* classifies an instance, resulting in a time complexity of $O(N.T_c)$.

On the other hand, the training time complexity for BLAST would be $O((M.N.(T_t + T_c)) + 2M)$, wherein all ensemble members are trained and evaluated on all samples. As the prediction procedure is the same as that of OMS-MAB, its prediction time complexity is $O(N.T_c)$ itself. The gain in time complexity for OMS-MAB becomes evident as N grows large while ϵ being set to a small value. Our experiments reconfirm the lower execution time growth rate for OMS-MAB in comparison to BLAST, as visualized in Figure 6. Moreover, the average total execution time across datasets is much lower for OMS-MAB when compared to BLAST, as captured in Table 5. For practical large-scale data stream mining applications, this translates to faster model updates and decisioning.

4 | EXPERIMENTAL METHODOLOGY

Our experiments evaluate the effectiveness of OMS-MAB in concept-drift adaptation using MAB based online model selection. For a rigorous evaluation, we select 10 synthetic and 10 real-world datasets pertaining to binary and multi-class problems from diverse domains affected by a variety of drift types. We compare our method with SOTA homogeneous, heterogeneous, and pool based ensembles. We perform 10 runs for all algorithms using different random seeds in all experiments to account for randomness and study the average results. Non-parametric Friedman

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

test is used to verify if a statistically significant difference exists among the algorithms based on their performance. In cases where such differences exist, post-hoc Nemenyi test is used to determine which ensembles perform significantly better than others. For comparing multiple algorithms based on their performance across different datasets, the aforementioned tests are considered to be SOTA (Demšar, 2006; van Rijn et al., 2018). We use implementations of these tests from the autorank package (Herbold, 2020).

OMS-MAB is implemented in java, on top of the *Massive Online Analysis* (MOA) (Bifet, Holmes, Pfahringer, Kranen, et al., 2010) framework. Implementations of DP and FASE as released by their respective authors were used. For all other algorithms, their implementations within MOA 21.07.0 was used. Hyperparameters of all algorithms are set to their default values within MOA unless specified otherwise. Experiments were conducted on an Ubuntu 16.04 machine with a 10-core Intel(R) Xeon(R) CPU E5-2687W v3, 3.10GHz processor and 64 GB of RAM.

Our comparisons primarily focus on instance-incremental algorithms due to their merits presented in Section 1. Overall, we compare OMS-MAB with the following 9 algorithms belonging to 3 categories:

- *Homogeneous ensembles*: We include FASE and Adaptive Random Forest (ARF) (Gomes, Bifet, et al., 2017) as they are among the SOTA methods (de Barros & de Carvalho Santos, 2019; Gomes, Bifet, et al., 2017). Leveraging Bagging (LevBag) (Bifet, Holmes, & Pfahringer, 2010) with HT as the base learner is also considered based on its competitive performance (de Barros & de Carvalho Santos, 2019; Read, Bifet, Pfahringer, & Holmes, 2012). OzaBag (OBag), Online Boosting (OBoost), and DWM are also included considering their popularity and usage in several contemporary studies (Idrees et al., 2020; Zhao et al., 2020). Learn++.NSE (Elwell & Polikar, 2011) is considered as it is a highly acclaimed batch-incremental ensemble. The number of component classifiers is set to 10 for all the algorithms that require it. FASE, LevBag, OBag, OBoost, DWM and Learn++.NSE use HT as the base learner, whereas ARF uses its built in ARFHoeffdingTree as the base learner. As the ensemble time of Learn++.NSE increases exponentially with the number of base classifiers (Chen et al., 2021), we use its pruning enabled version within MOA for execution efficiency.
- *Heterogeneous ensembles*: We include BLAST, as it is a SOTA technique based on online algorithm selection (Luong et al., 2021). Fading factor based variant of BLAST is used, due to its merits reported by the authors (van Rijn et al., 2018). We choose NB, HT, EFDT, Perceptron (Percp), HoeffdingAdaptiveTree (HAT) (Bifet & Gavalda, 2009) and ASHoeffdingTree (ASHT) (van Rijn et al., 2018) as base learners to ensure diversity. The fading factor was set to 0.99.
- *Pool based ensembles*: We include DP due to its competitive performance and unique ability to handle recurring drifts. We use Hoeffding Tree with Naive Bayes (HTNB) at the leaves as the base learner and an ensemble size of 10, which is the default configuration for DP.

For a consistent evaluation, OMS-MAB is configured with the same base learners and fading factor ($\theta = 0.99$) as in the case of BLAST. Our choice for θ is based on the fading factor variational study presented in (van Rijn et al., 2018). The best value for ϵ is dependent on the stream. However, to identify a meaningful default value, initial experiments were conducted on 6 synthetic streams[1] by varying ϵ and measuring the corresponding average predictive performance and computational cost across these streams. It was observed that $\epsilon = 0.25$ provided reasonable accuracy at a low computational cost. Consequently, 0.25 was selected as the default value for ϵ to ensure sufficient exploration at a reasonable computational cost. The effect of other values of ϵ and its impact on model performance under different scenarios is studied in Section 5.4. Table 2 summarizes the hyperparameter configuration of the proposed and benchmark methods. All other hyperparameter values for all algorithms are set to their defaults in MOA.

TABLE 2 Hyperparameter configuration summary of the proposed and benchmark methods.

Method	Description
ARF	*ensembleSize* = 10
BLAST	*alpha* = 0.99, *baseClassifiers* = {*NB*, *HT*, *EFDT*, *Percp*, *HAT*, *ASHT*}
DP	*maxPoolSize* = 10 (default)
DWM	*baseLearner* = *HT*
FASE	*ensembleSize* = 10, *metaLearner* = *AdaptiveClassifier*(*HT*), *baseLearner* = *AdaptiveClassifier*(*HT*)
LevBag	*ensembleSize* = 10, *baseLearner* = *HT*
Learn++.NSE	*ensembleSize* = 10, *baseLearner* = *HT*, *pruningStrategy* = *ERROR*
OBag	*baseLearner* = *HT*, *ensembleSize* = 10
OBoost	*baseLearner* = *HT*, *ensembleSize* = 10
OMSMAB	*alpha* = 0.99, *baseClassifiers* = {*NB*, *HT*, *EFDT*, *Percp*, *HAT*, *ASHT*}, *epsilon* = 0.25

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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 | Predictive performance

We use classification accuracy as the evaluation metric for comparing the predictive performance of algorithms. Prequential evaluation is used in all experiments wherein a model first makes a prediction on an instance before using it for learning (Lu et al., 2018). A fading factor of 0.999 was used to decay the effect of past performance towards the current accuracy, and sampling was performed for prequential accuracy at every 500th example (Chiu & Minku, 2020).

Algorithms are ranked based on their average prequential accuracy on each dataset such that higher predictive performance is indicated by lower ranks. The average rank corresponding to each algorithm across datasets is reported as an overall indicator of its predictive performance. Additionally, we report the average prequential accuracy across datasets corresponding to each algorithm. Both mean and standard deviation of average prequential accuracy from 10 runs are reported for all algorithms on all datasets. However, for deterministic algorithms such as DP, DWM, Learn++.NSE and BLAST, the standard deviation is zero. Furthermore, the statistical significance of accuracy differences among algorithms is also analyzed.

4.2 | Overall execution time

Since OMS-MAB targets learning problems involving high-speed data streams with potentially infinite samples, it is critical to ensure that training and prediction time requirements are minimal. To this end, we study the total execution time which is the sum of the time required to train the model and the time taken by the model to make predictions. Total execution time provides an insight into the computational complexity of the model, wherein lower values are preferable. We analyze the execution time growth with increase in the number of instances processed, to better understand the runtime complexity of algorithms. Also, we report the average value for total execution time across datasets for each algorithm. Further, statistical significance of timing differences among top performing algorithms in terms of accuracy across all the datasets considered is also analyzed.

4.3 | Memory requirements

To analyze the memory requirements of algorithms, we report the size of classifier models in megabytes (MB). To gain more insights, the average size of classifiers across datasets for each algorithm is analyzed. The variation in model size as the stream progresses is also studied for a deeper understanding of the memory consumption patterns of algorithms. Additionally, statistical significance of model size differences among top performing algorithms in terms of accuracy across all the datasets considered is also reported.

4.4 | Hyperparameter relevance

The overall performance of OMS-MAB depends on the hyperparameter ϵ which provides a trade-off among model accuracy and computational resource requirements. To select a meaningful value for ϵ, we analyze the effect of systematically varying ϵ on the model performance. To this end, we vary ϵ among {0.01, 0.1, 0.25, 0.5, 1.0}. This gradual variation from minimal to complete exploration allows us to closely study the aforementioned trade-off. Corresponding to each value of ϵ, we report average values for accuracy, total execution time, and model size of OMS-MAB from 10 runs on each dataset, to account for MAB stochasticity.

4.5 | Relationship between model complexity and model accuracy

Complexity analysis of OMS-MAB in Section 3.1 indicates that two factors influence its complexity: the ensemble size (M), and the exploration factor (ϵ). We vary M among {2, 3, 4, 5, 6} and ϵ among {0.01, 0.1, 0.25, 0.5, 1.0} to examine the relationship between model complexity and model accuracy. For a consistent evaluation, we keep the order of learners within the ensemble as fixed across experiments: {NB, HT, EFDT, Percp, HAT, and ASHT}, and include the first M learners in the ensemble corresponding to each size M. For each combination of M and ϵ, we run OMS-MAB ten times on each dataset, calculate the average accuracy across the datasets for each run, and then visualize the mean and standard deviation of average accuracies across the ten runs as a heatmap. This helps to analyze the interplay between model complexity and accuracy.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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 Datasets.

Dataset	# Instances	# Features	# Classes	Type	Drift	# Drifts
SEA_A	200000	3	2	S	A	3
SEA_G	200000	3	2	S	G	3
LED_M	200000	24	10	S	M	3
TREE_F	100000	10	6	S	A & R	15
HYP_S	200000	10	2	S	I & S	1
HYP_F	200000	10	2	S	I & F	1
AGR_GR	200000	9	2	S	G & R	4
AGR_AR	200000	9	2	S	A & R	4
RBF_F	200000	15	5	S	I & F	1
RBF_IN	200000	15	5	S	I & IN	1
Elec	45312	6	2	R	?	?
Power	29928	2	24	R	?	?
Airlines	539383	7	2	R	?	?
Weather	18159	8	2	R	?	?
Poker	829201	10	10	R	?	?
Covtype	581012	54	7	R	?	?
Sensor	200000	5	54	R	?	?
Rialto	82250	27	10	R	?	?
Adult	48842	14	2	R	?	?
KDD99	494021	41	23	R	?	?

Note: [Type] S: synthetic, R: real world; [Drift] A: abrupt, G: gradual, I: incremental, R: recurring, M: mixed, S: slow, F: fast, IN: intermediate, ?: unknown.

4.6 | Datasets

Table 3 presents a summary of the datasets used in our experiments, out of which 10 are generated synthetically and 10 are real-world. These are commonly used benchmarks in data stream mining literature. Synthetic streams are generated using SEA (Street & Kim, 2001), LED (Brzezinski & Stefanowski, 2013), Tree (Brzezinski & Stefanowski, 2013), Hyperplane (Hulten et al., 2001), Agrawal (Aggarwal et al., 2003) and RBF (Brzezinski & Stefanowski, 2013) generators available in MOA framework. MOA commands for generating all the synthetic datasets considered in our study are available at https://github.com/jobinwilson/syntheticData. For further details on the datasets considered in our experiments, we refer the reader to (Brzezinski & Stefanowski, 2013) for LED_M, TREE_F, HYP_S and HYP_F; to (Chiu & Minku, 2018) for AGR_GR and AGR_AR; to (Xu et al., 2017) for RBF_F and RBF_IN; to (MOA, datasets 2019) for Forest Covertype (Covtype), Electricity (Elec), Poker-Hand (Poker), Airlines Dataset (Airlines); to (Stream Data Mining Repository 2021) for Sensor Stream (Sensor), Powersupply Stream (Power), Kddcup99 (KDD99); to (Losing et al., 2016) for Rialto; to (Elwell & Polikar, 2011) for Weather and to Luong et al. (2021) for Adult. For the datasets Covtype, Elec and Weather, their normalized versions were used. For the Sensor dataset, we used the first 200000 instances in all experiments.

5 | EXPERIMENTAL RESULTS

5.1 | Predictive performance

The prequential accuracy of all algorithms considered is presented in Table 4. The best results are in bold. We observe that the predictive performance of algorithms varies across datasets, justifying the usage of model selection to enhance predictive performance. Though a single top performing algorithm is missing, we see that LevBag, FASE, and BLAST are the top performers based on their average rank, as observed in similar studies (de Barros & de Carvalho Santos, 2019; van Rijn et al., 2018). In terms of average accuracy across datasets, OMS-MAB is close to the aforementioned top performers. Moreover, it falls in the top tier and immediately follows BLAST, outperforming ARF and DWM in terms of average rank. Statistical analysis was conducted for 10 populations (algorithms) with 20 paired samples (datasets). Friedman test with family-wise significance level $\alpha = 0.05$ indicates the existence of statistically significant differences among the performance of algorithms. *Critical difference*

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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 Classification accuracies of algorithms.

Dataset	LevBag	ARF	FASE	DP	DWM	OBag	OBoost	Learn++.NSE	BLAST	OMSMAB
SEA_A	**88.69±0.13**	88.37±0.05	88.37±0.04	86.97	87.80	85.02±0.06	86.52±0.11	85.61	87.99	87.64±0.06
SEA_G	**87.61±0.05**	86.30±0.10	87.21±0.07	86.71	86.77	84.82±0.04	85.99±0.08	84.91	86.78	86.53±0.08
LED_M	52.24±0.08	51.31±0.37	**53.14±0.05**	52.98	52.18	52.53±0.07	52.03±0.17	40.38	52.48	53.12±0.03
TREE_F	52.86±0.45	**57.54±0.22**	46.00±0.16	43.26	44.73	41.81±0.11	13.57±0.15	38.34	47.26	45.68±0.83
HYP_S	87.09±0.07	83.30±0.32	90.19±0.09	87.44	90.34	86.66±0.10	86.82±0.12	86.02	**92.20**	91.51±0.10
HYP_F	88.33±0.03	86.05±0.09	88.11±0.11	86.90	89.34	84.85±0.07	86.90±0.05	85.74	**91.15**	90.62±0.10
AGR_GR	91.40±1.17	82.23±0.37	92.08±0.55	89.79	91.86	83.20±0.16	87.69±0.51	76.95	**95.64**	92.12±0.83
AGR_AR	94.48±1.62	85.54±0.32	93.16±0.21	93.73	95.34	89.76±0.39	90.69±0.98	78.08	**97.98**	93.58±0.83
RBF_F	**98.88±0.03**	98.56±0.03	98.32±0.06	97.80	97.97	94.24±0.20	96.02±0.08	55.07	97.09	98.16±0.07
RBF_IN	**99.46±0.04**	99.43±0.02	99.37±0.06	98.96	99.08	98.23±0.21	99.37±0.02	61.07	98.73	99.10±0.05
Sensor	**94.82±0.07**	89.60±0.29	88.47±0.38	90.91	91.05	73.72±0.78	79.63±2.13	12.74	88.51	89.95±0.22
Elec	89.76±0.17	**89.91±0.12**	86.40±0.15	83.42	84.35	82.47±0.25	86.55±0.20	72.74	86.28	86.03±0.17
Power	16.28±0.13	2.64±0.37	**20.01±0.50**	14.33	16.14	15.87±0.10	4.16±0.01	4.16	16.59	16.63±0.41
Airlines	63.17±0.06	65.91±0.04	66.57±0.07	65.15	**66.86**	64.95±0.03	63.50±0.06	63.32	66.02	66.30±0.12
Weather	78.15±0.15	**78.62±0.17**	72.48±0.30	71.28	72.10	74.69±0.26	76.62±0.26	68.72	75.45	72.60±0.36
Poker	87.07±0.53	**87.09±0.09**	76.26±0.04	74.94	73.41	82.03±1.12	86.43±2.33	60.14	81.59	76.56±0.07
Covtype	91.71±0.11	**94.42±0.05**	87.90±0.08	83.30	84.42	83.59±0.34	89.97±0.34	63.72	87.68	87.92±0.04
Rialto	60.40±0.18	**70.51±0.31**	63.20±0.36	41.10	43.23	30.94±0.47	9.96±0.00	28.12	58.25	54.15±1.11
Adult	**84.75±0.09**	83.92±0.11	84.38±0.15	72.51	82.92	84.64±0.14	82.88±0.28	79.89	84.08	83.32±0.17
KDD99	99.80±0.00	99.85±0.02	99.70±0.01	99.68	99.73	99.77±0.01	**99.86±0.01**	30.15	99.77	99.66±0.01
Avg Acc.	**80.35**	79.05	79.06	76.06	77.48	74.69	73.26	58.79	79.58	78.56
Avg Rank	**3.00**	4.58	3.95	6.58	5.15	7.33	6.43	9.53	3.98	4.50

Bold indicates best values.

(CD) came out to be 3.03. As depicted in Figure 3a, post-hoc Nemenyi test indicates that OMS-MAB is among the top tier performers, and there is no statistically significant accuracy difference among LevBag, FASE, BLAST, OMS-MAB, ARF, and DWM. In the figure, algorithms with statistically equivalent predictive performance are connected using a horizontal line.

Accurate estimation of the predictive performance of each inner-level AL is essential for OMS-MAB to intelligently switch the *active member*, enabling it to use the current best performing AL to make predictions. In Figure 4, we present the performance estimates of OMS-MAB corresponding to all inner-level ALs, along with the MAB's choice of the *active member* for each time step. This visualization helps to verify if the MAB is able to correctly identify the best ALs for the respective stream segments. To account for randomness in data streams and MAB stochasticity, performance estimates are visualized from 10 independent runs of the model, enclosed by its 95% confidence interval. On the other hand, *active member* selection of the MAB is visualized from the first run. Key observations from these visualizations are the following:

- In the case of *HYP_S* dataset, OMS-MAB correctly identifies Perceptron based AL as the best performer as depicted in Figure 4a,b, despite the drift being incremental and slow. We observe that the MAB switches among multiple ALs during the early parts of the stream as performance estimates of ALs are not well separated in that region. However, as the stream progresses, the MAB is more confident on Perceptron and correctly switches to it for most parts of the stream.
- In the case of *AGR_AR* dataset, OMS-MAB identifies Perceptron based AL to be a poor performer throughout the stream and correctly eliminates it from the choices for *active member* selection. It is interesting to observe that the MAB switches among ALs based on HT, EFDT and ASHT as the *active member* for most parts of the stream due to their high performance, as depicted in Figure 4c,d. Furthermore, this observation confirms that a single AL need not be the best performer throughout a stream, and adaptively switching among ALs is a useful strategy for concept-drift adaptation.
- In the case of *Rialto* dataset, OMS-MAB correctly selects EFDT based AL as the best performer for most parts of the stream. However, the MAB switches among multiple ALs in a few parts of the stream as the performance estimates of ALs are not well separated in those regions. This behavior is illustrated in Figure 4e,f.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

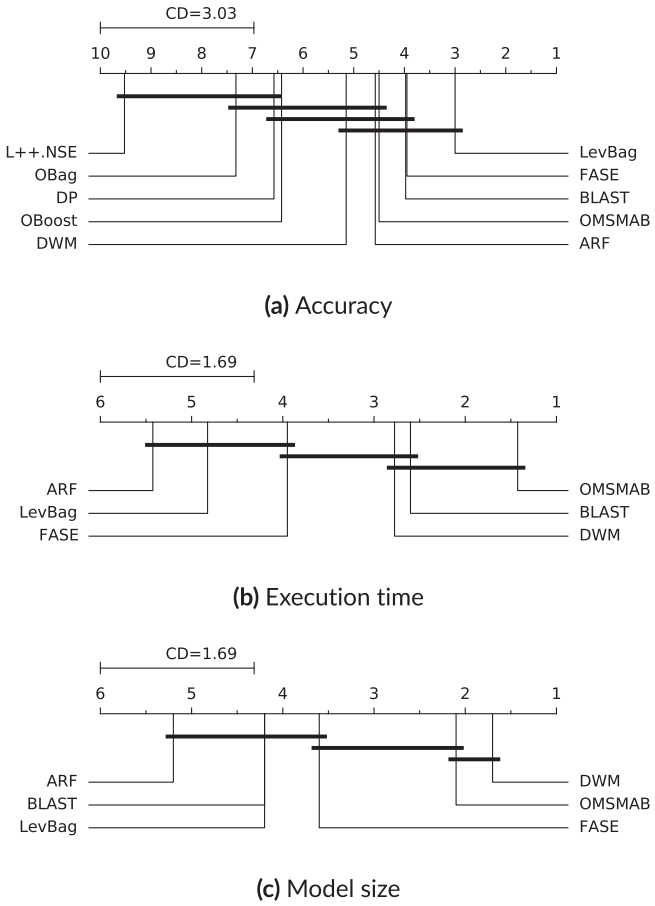


FIGURE 3 Results of Nemenyi post-hoc tests (95% confidence level).

The temporal variation in prequential accuracy of top tier algorithms on select datasets is visualized in Figure 5 to study their adaptability under diverse drift conditions. This enables us to verify that the proposed MAB based adaptation procedure allows OMS-MAB to quickly recover its predictive performance subsequent to experiencing a concept-drift. Below are the main observations:

- In the case of *LED_M* dataset having both abrupt and gradual drifts, models such as BLAST that slowly discard obsolete knowledge through implicit adaptation are outperformed by FASE and OMS-MAB. This can be attributed to the HDDM_A based explicit drift detection and backup learners associated with the latter models that enable maintaining up-to-date knowledge, allowing effective handling of both abrupt and gradual drifts. However, the superior performance of BLAST in comparison to homogeneous ensembles like LevBag and ARF indicates that heterogeneous learning is a beneficial diversity generation strategy for this stream. OMS-MAB exploits both explicit drift adaptation and heterogeneity while performing the second best on this stream, immediately following FASE.
- For *HYP_S* dataset with incremental and slow drifts, models like BLAST which use implicit adaptation to concept-drift by preserving relatively longer-term concept memory outshine models such as FASE, LevBag, and ARF. This can be attributed to the latter models rapidly eliminating outdated knowledge through explicit drift detection. While OMS-MAB also uses explicit detection, MAB based selective training prevents all members from getting updated with all samples, increasing the probability of knowledge retention. Furthermore, Figure 4a,b indicate superior performance of Perceptron based learners on this stream in comparison to Tree based learners. This is beneficial to ensembles like OMS-MAB and BLAST which maintain a pool of heterogeneous learners, allowing selection of the optimal learner matching the stream characteristics from time to time for prediction. Overall, OMS-MAB performs second best on this stream, immediately following BLAST.
- In the case of *RBF_F* dataset with incremental and fast drifts, superior performance of models like LevBag, ARF and FASE with explicit drift detection in comparison to implicit adaptation models like BLAST indicates the importance of quickly discarding obsolete knowledge. Prequential accuracy plot for this stream indicates that performance of BLAST drops significantly during drift in comparison to models equipped with explicit drift detection. OMS-MAB works reasonably well on this stream because on one hand it uses explicit drift detection to discard obsolete knowledge. On the other hand, MAB's ability to periodically select the best learner for prediction from a diverse pool of learners enables it to enjoy the benefits of heterogeneous ensembles like BLAST.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

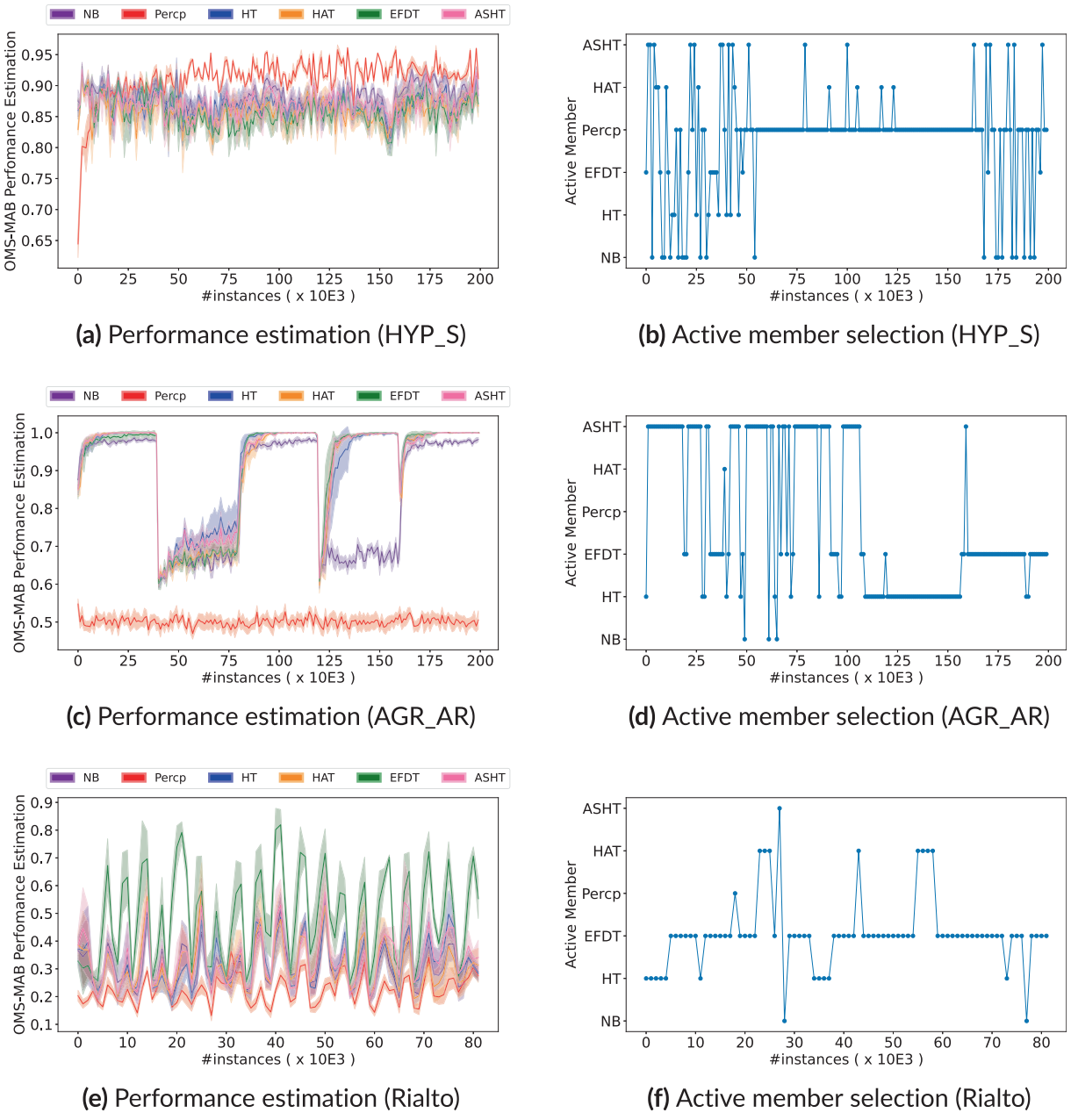


FIGURE 4 Performance estimation and active member selection of OMS-MAB.

- For *AGR_GR* dataset with gradual and recurring drifts, models such as LevBag, ARF and FASE which discard historical knowledge using explicit drift detection is outperformed by models like BLAST which uses implicit adaptation to slowly forget obsolete concepts. The adaptation strategy of former models focuses on exploiting short-term concept memory, preventing historical knowledge reuse during recurring drifts. Though OMS-MAB uses explicit drift detection to discard knowledge, lower values of ϵ ensure that members are trained on limited samples, enabling historical knowledge reuse in scenarios where they are not already reset. Moreover, OMS-MAB's strategy of alternating between *exploration* involving updates to all ensemble members and *exploitation* which solely updates the best performer results in all members getting partially trained on the latest concept at any point of time. This allows the MAB to activate suitable members from time-to-time, enabling faster adaptation. The aforementioned factors enable OMS-MAB to perform reasonably well on this stream.

It is interesting to observe that OMS-MAB exhibits a drift adaptation behavior that is comparable to other top tier ensembles that were part of our experiments. On one hand, this could be attributed to explicit drift detection and backup learners allowing obsolete knowledge to be discarded fast. On the other hand, MAB's ability to systematically estimate the performance of each AL enables selecting the most appropriate ensemble member from time to time for prediction and training, allowing adaptation to diverse drift types. However, it is to be noted that OMS-MAB improves

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

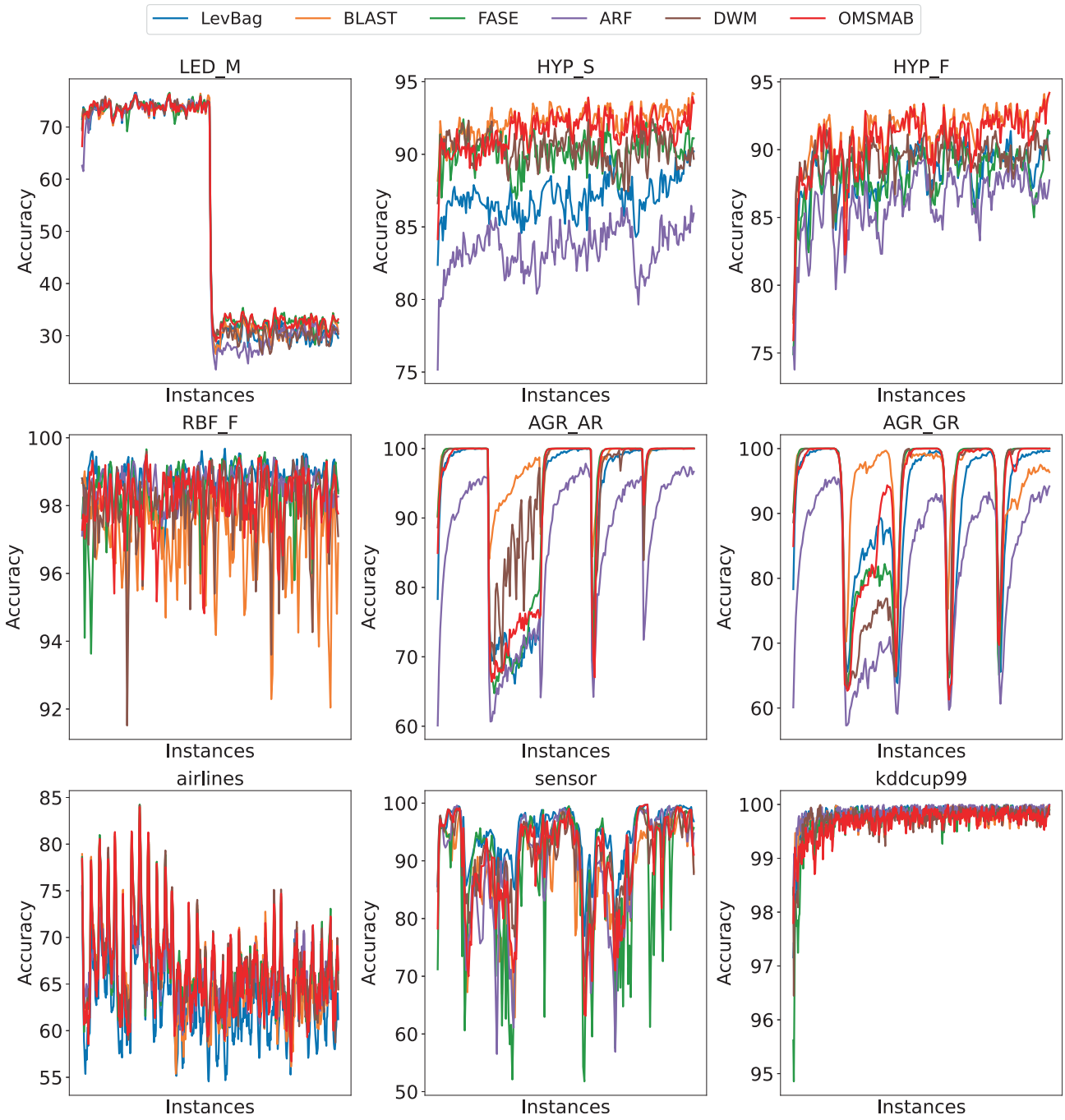


FIGURE 5 Prequential accuracy comparison of top tier ensembles on different datasets covering various drift types and speeds.

computational efficiency by selectively training promising ensemble members. Consequently, all ensemble members do not get trained on all samples, thereby obtaining a slightly lower predictive performance compared to ensembles that unconditionally train all members on all samples. Nonetheless, the exploration factor ϵ provides a tradeoff between computational resource requirements and predictive performance.

5.2 | Overall execution time

Table 5 presents total execution time along with the corresponding average values across datasets for all algorithms compared. The best results are in bold. To understand the computational efficiency provided by OMS-MAB, statistical analysis was conducted using 6 populations (top performing ensemble algorithms) with 20 paired samples (datasets). Friedman test with family-wise significance level $\alpha = 0.05$ indicates the existence

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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 Total execution time of algorithms in seconds (training and prediction).

Dataset	LevBag	ARF	FASE	DP	DWM	OBag	OBoost	Learn++.NSE	BLAST	OMSMAB
SEA_A	55±2.7	308±9.0	21±0.3	13±0.2	21±0.3	24±0.3	24±0.4	**11±0.3**	17±0.3	17±0.4
SEA_G	46±1.9	648±16.6	21±0.4	12±0.2	22±0.3	24±0.4	24±0.2	**10±0.2**	16±0.2	17±0.4
LED_M	54±1.5	112±3.1	52±0.7	**15±0.3**	138±0.8	38±0.7	38±0.8	20±0.5	38±0.5	22±1.0
TREE_F	41±0.8	49±0.7	22±0.4	12±0.1	32±0.4	29±0.6	30±1.0	**9±0.1**	15±0.4	11±0.3
HYP_S	137±5.2	1198±126.3	33±1.1	19±0.3	25±0.5	56±0.4	52±0.5	**15±0.4**	30±0.4	19±0.6
HYP_F	95±3.5	439±29.1	34±0.6	18±0.3	30±0.6	61±1.7	54±0.7	**14±0.3**	27±0.4	20±0.5
AGR_GR	137±4.7	422±4.8	34±0.6	23±0.4	23±0.4	184±4.5	267±8.4	**13±0.3**	57±0.5	20±0.8
AGR_AR	99±5.3	395±8.3	24±0.3	13±0.3	13±0.3	169±6.0	291±17.1	**13±0.3**	42±0.3	19±0.3
RBF_F	44±0.7	65±0.9	48±0.5	29±0.3	**19±0.4**	85±2.0	78±2.0	23±0.3	37±0.4	24±0.7
RBF_IN	51±0.6	85±4.4	47±0.8	24±0.6	**22±0.4**	43±0.9	67±1.1	**22±0.5**	27±0.5	25±0.4
Sensor	94±0.3	59±0.4	147±2.0	36±0.3	58±0.3	52±0.6	58±1.3	44±0.3	44±0.3	**31±0.3**
Elec	11±0.2	18±0.5	10±0.3	6±0.1	6±0.2	8±0.3	8±0.3	**4±0.2**	6±0.1	6±0.3
Power	6±0.2	3±0.2	8±0.2	**2±0.1**	16±0.3	3±0.1	3±0.1	3±0.1	3±0.1	5±0.1
Airlines	3162±31.4	2476±37.3	607±33.6	1191±6.9	406±2.8	4544±47.4	5005±63.0	**37±0.4**	1410±5.5	368±54.5
Weather	5±0.3	12±0.3	5±0.1	3±0.0	3±0.1	3±0.1	3±0.1	**2±0.0**	3±0.0	3±0.1
Poker	153±5.2	346±3.2	139±1.0	69±0.5	80±0.7	322±12.9	324±14.6	**45±0.4**	120±0.8	72±0.7
Covtype	334±7.8	567±14.8	333±3.9	231±1.8	226±1.7	899±19.2	992±32.5	95±2.0	544±3.8	**88±1.0**
Rialto	34±0.2	45±0.5	52±0.5	17±0.2	19±0.2	22±0.3	23±0.4	**14±0.1**	27±0.3	15±0.2
Adult	80±1.8	150±2.8	13±0.4	6±0.2	15±0.1	15±0.5	18±0.7	**5±0.2**	9±0.2	8±0.2
KDD99	1059±8.7	751±25.9	778±4.5	**111±0.8**	407±2.9	856±7.2	961±15.3	117±1.8	348±2.4	285±6.8
Avg Time (s).	285	408	121	93	79	372	416	**26**	141	54

Bold indicates total execution time and model size.

of statistically significant differences among the total execution time of top performing algorithms in terms of accuracy. *Critical difference* (CD) observed was 1.69. As depicted in Figure 3b, post-hoc Nemenyi test indicates that OMS-MAB, BLAST and DWM are the top tier performers in terms of execution time, and there is no statistically significant timing differences among them. In the figure, algorithms with statistically equivalent timing requirements are connected using a horizontal line. We observe that algorithms such as LevBag, ARF and FASE which are in the top accuracy league belong to the bottom tier when it comes to execution time, making them less attractive for high-speed stream mining. Furthermore, in terms of average rank for total execution time across datasets, OMS-MAB outperforms all the top performing ensembles considered.

Figure 6 compares the execution time growth of OMS-MAB with that of the top performing ensembles in terms of accuracy, to closely understand the differences in their runtime complexity. We observe that execution time growth for OMS-MAB is the least among all the top performing ensembles considered, indicative of its time efficiency. This gain is attributed to two factors. Firstly, our MAB based adaptation eliminates the need to continuously train and evaluate all ensemble members. Secondly, periodic reset of ALs upon encountering concept-drift prevents their unbounded growth, allowing faster learning and evaluation.

5.3 | Memory requirements

Table 6 presents the model size corresponding to all algorithms considered, along with their averages across datasets. The best results are in bold. To study the memory efficiency of ensembles based on their model size, statistical analysis was conducted using 6 populations (top performing ensemble algorithms) with 20 paired samples (datasets).

Friedman test with family-wise significance level $\alpha = 0.05$ indicates the existence of statistically significant differences among model size of algorithms. *Critical difference* (CD) observed was 1.69. As depicted in Figure 3c, post-hoc Nemenyi test indicates that DWM and OMS-MAB are the top tier performers in terms of memory efficiency, and no statistically significant differences exist between them based on model size. We observe that SOTA homogeneous ensembles such as ARF and LevBag are memory intensive. On the other hand, OMS-MAB immediately follows DWM in terms of average rank for model size across datasets, making it attractive in terms of memory savings. Figure 7 compares the memory growth pattern of OMS-MAB with that of the top performing ensembles in terms of accuracy. We observe that memory growth for OMS-MAB is

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

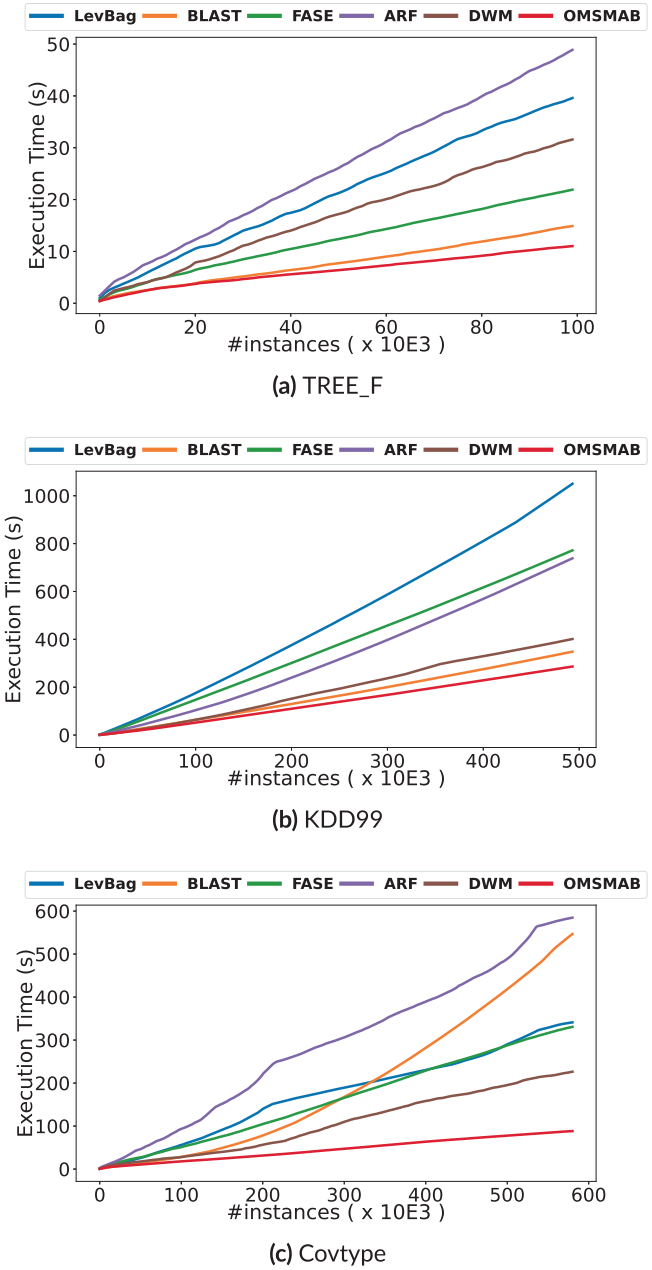


FIGURE 6 Execution time growth comparison (in seconds) across algorithms.

the least among all the top performing ensembles considered, indicative of its memory efficiency. As in the case of execution time, this gain is attributed to our adaptation strategy which eliminates the need to continuously train and evaluate all ensemble members, as well as the periodic reset of ALs upon encountering drift.

5.4 | Hyperparameter relevance

We study the effect of varying ϵ on the performance of OMS-MAB, as outlined in Section 4.4. Table 7 presents accuracy, total execution time and model size along with the corresponding average values across datasets for each ϵ. The best results are in bold. We observe that as ϵ increases, accuracy improves in most cases as expected. This is because a higher exploration rate results in more training and evaluation of all ALs, making the MAB more confident about their performance estimates and consequently improving its behavior policy. Though setting $\epsilon = 1$ is an extreme case of this scenario, it causes the drift detector to detect all drifts with high probability. This results in the corresponding AL getting reset, discarding its historical knowledge. On the other hand, lower ϵ values result in ALs getting trained on limited samples, allowing them to

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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 6 Model size comparison (in MBs) of algorithms.

Dataset	LevBag	ARF	FASE	DP	DWM	OBag	OBoost	Learn++.NSE	BLAST	OMSMAB
SEA_A	3.74±0.51	23.6±0.09	0.97±0.02	0.56	0.40	1.64±0.01	1.59±0.01	**0.25**	0.98	0.69±0.02
SEA_G	2.63±0.12	23.69±2.23	0.99±0.02	0.46	0.65	1.63±0.01	1.58±0.02	**0.25**	0.95	0.72±0.02
LED_M	0.51±0.03	3.44±0.94	0.77±0.05	0.55	1.12	2.04±0.07	2.20±0.10	**0.26**	1.73	0.67±0.00
TREE_F	2.10±0.38	1.56±0.27	0.69±0.03	1.35	0.47	3.85±0.09	4.08±0.16	**0.26**	1.87	0.68±0.09
HYP_S	8.05±0.76	76.01±14.43	1.28±0.10	0.76	1.00	4.01±0.07	3.98±0.07	**0.26**	2.03	0.75±0.11
HYP_F	4.61±1.23	13.25±3.16	1.57±0.14	0.68	0.74	5.01±0.08	4.24±0.03	**0.26**	1.79	0.81±0.06
AGR_GR	8.70±0.37	30.14±1.03	2.58±0.24	1.64	1.65	21.76±0.21	23.96±0.88	**0.26**	7.61	0.78±0.05
AGR_AR	3.96±1.00	26.46±0.51	0.80±0.03	0.58	0.31	23.01±0.99	27.33±1.35	**0.26**	5.80	0.70±0.01
RBF_F	0.87±0.10	1.89±0.26	1.06±0.09	0.94	0.40	7.15±0.23	5.80±0.21	**0.26**	2.92	0.86±0.12
RBF_IN	2.33±0.34	2.90±0.65	1.53±0.16	0.55	0.81	3.22±0.11	5.95±0.10	**0.26**	1.36	1.03±0.16
Sensor	0.51±0.00	0.82±0.08	0.92±0.01	0.75	0.47	3.13±0.09	2.50±0.06	**0.27**	1.87	0.67±0.00
Elec	1.07±0.07	1.65±0.24	0.61±0.00	0.62	0.71	1.86±0.05	1.57±0.06	**0.31**	0.96	0.60±0.02
Power	0.36±0.00	0.46±0.01	0.66±0.00	0.50	0.82	0.34±0.00	0.34±0.00	0.40	**0.40**	0.62±0.01
Airlines	30.89±13.48	30.71±8.15	25.08±6.47	52.34	5.13	150.78±0.64	161.79±1.28	**0.56**	64.61	17.41±7.14
Weather	1.19±0.21	3.29±0.72	0.61±0.01	0.54	0.52	0.59±0.01	0.57±0.01	**0.38**	0.43	0.60±0.01
Poker	3.08±0.83	3.43±0.86	0.64±0.00	0.94	**0.34**	7.18±0.30	7.61±0.40	0.44	2.84	0.61±0.01
Covtype	0.98±0.36	1.77±0.29	0.76±0.00	0.63	0.63	24.36±0.66	24.98±0.79	**0.33**	17.92	0.72±0.05
Rialto	0.65±0.05	0.70±0.10	0.85±0.01	1.41	0.5	1.73±0.07	2.27±0.24	**0.26**	1.53	0.74±0.03
Adult	24.94±0.58	46.02±0.81	3.99±0.40	1.28	1.06	4.28±0.28	5.70±0.27	**0.37**	2.60	1.54±0.11
KDD99	29.09±0.13	21.97±0.99	23.83±0.26	3.12	11.53	23.36±0.26	26.49±0.63	**0.36**	11.91	9.16±0.22
Avg Mem	6.51	15.69	3.51	3.51	1.46	14.55	15.73	**0.31**	6.61	2.02

Bold indicates total execution time and model size.

reuse historical knowledge in case if they are not already reset. For instance, in the case of AGR_AR dataset which is affected by recurring drift, intermediate values for ϵ results in better accuracy.

We observe that execution time increases with larger values of ϵ as expected, since a larger exploration factor results in the training and evaluation of all ALs on a larger set of samples, increasing the computational complexity. On similar lines, lower values of ϵ results in lower model size in most cases, as all ALs except the *active member* gets trained on fewer samples in this case. Moreover, periodic reset of ALs due to explicit drift detection is observed to restrict their unbounded growth, as indicated by the non-monotonic variation in model size with gradual increase in ϵ for some of the datasets such as *Rialto* and *Airlines*. Figure 8 visualizes the overall trend in performance variation of OMS-MAB across datasets as we gradually vary the exploration factor from minimal to complete exploration.

We note that though the best value for ϵ is stream dependent, choosing a value that is neither too small nor too large is desirable for OMS-MAB to provide optimal predictive and timing performance. It is observed that selecting $\epsilon = 0.25$ provides a reasonable trade-off between model accuracy and computational resource requirements, which is aligned with our initial experiments as discussed in Section 4.

5.5 | Relationship between model complexity and model accuracy

In this study, we assess the effect of model complexity on the accuracy of OMS-MAB, as described in Section 4.5. Figure 9 visualizes the overall trend in accuracy variation of OMS-MAB across datasets as the model complexity is gradually varied. At a broad level, we observe that accuracy increases with model complexity, as indicated by the darker shades as we move to the bottom-right of the heatmap. On one hand, an increase in ϵ is expected to improve ensemble accuracy due to the training and evaluation of all members on a larger set of samples as discussed in Section 5.4. On the other hand, an increase in M introduces more learners to the ensemble, which likely improves the accuracy subject to them being sufficiently diverse (van Rijn et al., 2018). It is interesting to note that average accuracy values along the principal diagonal of the heatmap are monotonically increasing, as both M and ϵ increase along that axis. We note that for a fixed ensemble size, average accuracy increases with ϵ in most cases, as observed in Section 5.4. It is observed that up to the ensemble size of 4, accuracy increases monotonically with M for a given ϵ.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

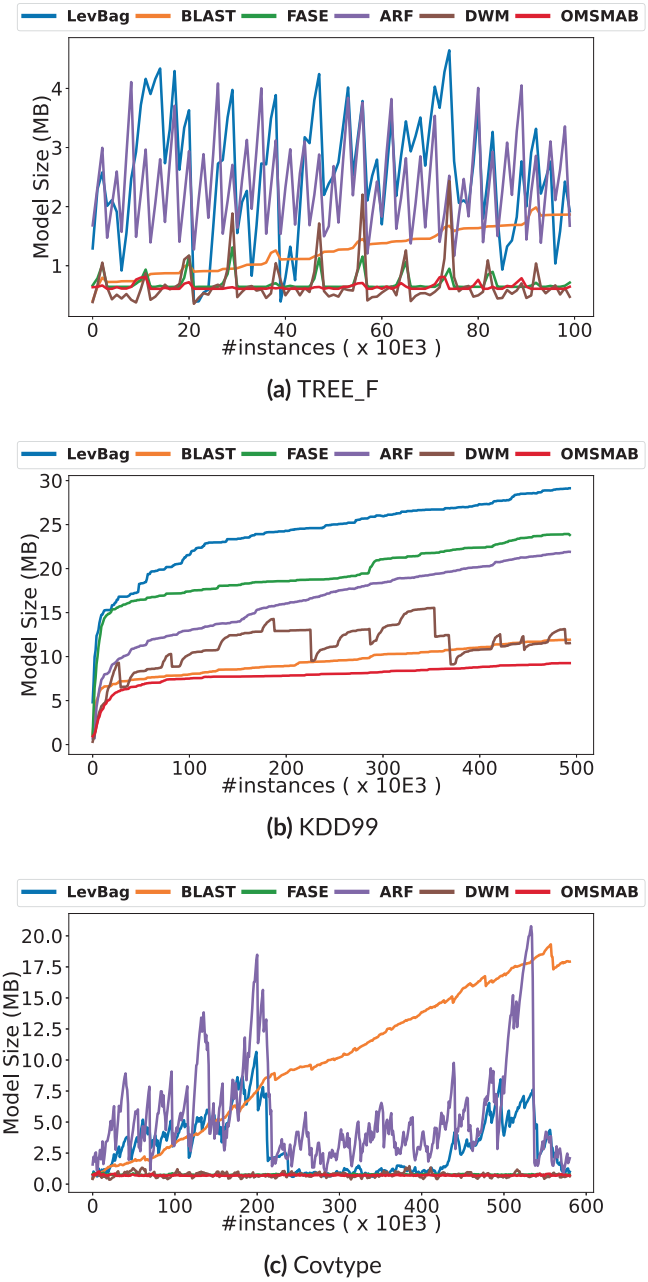


FIGURE 7 Model size growth comparison (in MBs) across algorithms.

Subsequently, it stagnates or even slightly decreases. This observation can be attributed to the ensemble diversity aspect, as the fifth and sixth members being added are respectively HAT and ASHT, both belonging to the Tree family as discussed in Section 4.5. Since the ensemble already has HT and EFDT as members representing the Tree family, introducing additional ALs from the same family will not significantly increase the diversity. In summary, a trade-off exists between model complexity and accuracy. However, it is critical to ensure sufficient diversity among members of the heterogeneous pool for superior predictive performance.

5.6 | Limitations and future research directions

Presently OMS-MAB is focused on standard non-stationary learning problems and does not address the more complex aspects of concept-drift adaptation like class-imbalance learning (CIL) (Cano & Krawczyk, 2022) and multilabel classification (MLC) (Alberghini et al., 2022). Moreover, we assume an online setting wherein labels are available for incremental learning immediately after prediction. However, several real-world streaming

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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 OMS-MAB performance variation with ϵ (averaged from 10 runs).

	Accuracy					Total execution time (s)					Model size (MB)				
Dataset	$\epsilon = 0.01$	$\epsilon = 0.10$	$\epsilon = 0.25$	$\epsilon = 0.50$	$\epsilon = 1.00$	$\epsilon = 0.01$	$\epsilon = 0.10$	$\epsilon = 0.25$	$\epsilon = 0.50$	$\epsilon = 1.00$	$\epsilon = 0.01$	$\epsilon = 0.10$	$\epsilon = 0.25$	$\epsilon = 0.50$	$\epsilon = 1.00$
SEA_A	87.56	87.35	87.64	87.92	**88.41**	16.92	**16.46**	16.81	17.15	18.99	0.67	**0.65**	0.69	0.75	0.86
SEA_G	86.37	86.27	86.53	86.68	**87.00**	**16.51**	16.74	17.18	17.90	19.00	**0.64**	0.68	0.72	0.76	0.89
LED_M	53.04	53.07	53.12	**53.19**	53.18	**19.67**	19.89	22.01	25.38	33.17	**0.67**	**0.67**	**0.67**	**0.67**	**0.67**
TREE_F	45.63	44.48	45.68	47.50	**49.99**	**10.17**	10.80	11.45	12.42	15.26	**0.62**	**0.62**	0.68	0.69	0.68
HYP_S	90.51	91.39	91.51	91.73	**92.09**	**17.23**	17.73	18.87	21.23	24.71	**0.61**	0.73	0.75	0.81	1.14
HYP_F	90.15	90.46	90.62	90.85	**91.25**	**16.35**	18.22	19.77	22.05	25.74	**0.64**	0.79	0.81	1.02	1.10
AGR_GR	91.21	91.51	92.12	92.58	**92.88**	**17.70**	18.85	19.80	21.91	25.21	**0.75**	0.81	0.78	1.01	1.40
AGR_AR	91.86	93.41	93.58	**94.65**	94.20	**16.33**	17.98	18.77	19.30	20.35	**0.70**	0.71	**0.70**	0.71	1.08
RBF_F	96.58	97.66	98.16	98.26	**98.36**	**20.53**	21.91	24.53	26.95	31.96	**0.67**	0.72	0.86	0.79	0.76
RBF_IN	98.76	**99.13**	99.10	99.08	99.11	**20.94**	22.71	24.50	26.92	32.45	**0.79**	0.86	1.03	1.08	0.95
Sensor	90.16	87.83	89.95	**92.16**	91.97	**23.08**	27.51	31.23	38.49	51.95	0.71	0.68	**0.67**	**0.67**	**0.67**
Elec	85.52	85.78	86.03	86.56	**87.51**	**5.28**	5.89	6.14	7.03	7.99	0.62	0.61	**0.60**	0.61	0.64
Power	15.22	15.98	16.63	16.48	**16.95**	**4.11**	4.25	4.52	4.91	5.03	0.62	**0.61**	0.62	0.62	0.63
Airlines	65.97	66.01	66.30	66.57	**66.89**	**146.86**	292.89	367.67	355.14	370.54	**5.97**	15.82	17.41	17.20	9.33
Weather	72.02	72.12	72.60	72.79	**73.76**	**2.76**	2.80	3.05	3.47	3.89	0.60	0.60	0.60	0.60	**0.59**
Poker	76.63	76.48	76.56	76.98	**77.66**	**60.93**	65.89	71.71	81.80	100.15	0.62	**0.61**	**0.61**	0.62	0.64
Covtype	87.54	87.73	87.92	87.95	**88.69**	**60.30**	72.31	88.02	120.24	164.83	**0.70**	0.73	0.72	0.72	0.72
Rialto	51.02	53.12	54.15	56.03	**60.31**	**12.26**	13.48	15.36	18.70	25.62	0.73	0.71	0.74	0.75	**0.70**
Adult	82.91	83.17	83.32	83.57	**83.90**	**6.43**	7.19	7.78	8.59	10.49	**1.03**	1.15	1.54	1.75	2.80
KDD99	99.53	99.61	99.66	99.72	**99.77**	**204.07**	266.25	286.48	324.75	381.45	**8.08**	8.69	9.16	10.41	12.18
Average	77.91	78.13	78.56	79.06	**79.69**	**34.92**	46.99	53.78	58.72	68.44	**1.32**	1.87	2.02	2.11	1.92

Bold indicates best results.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

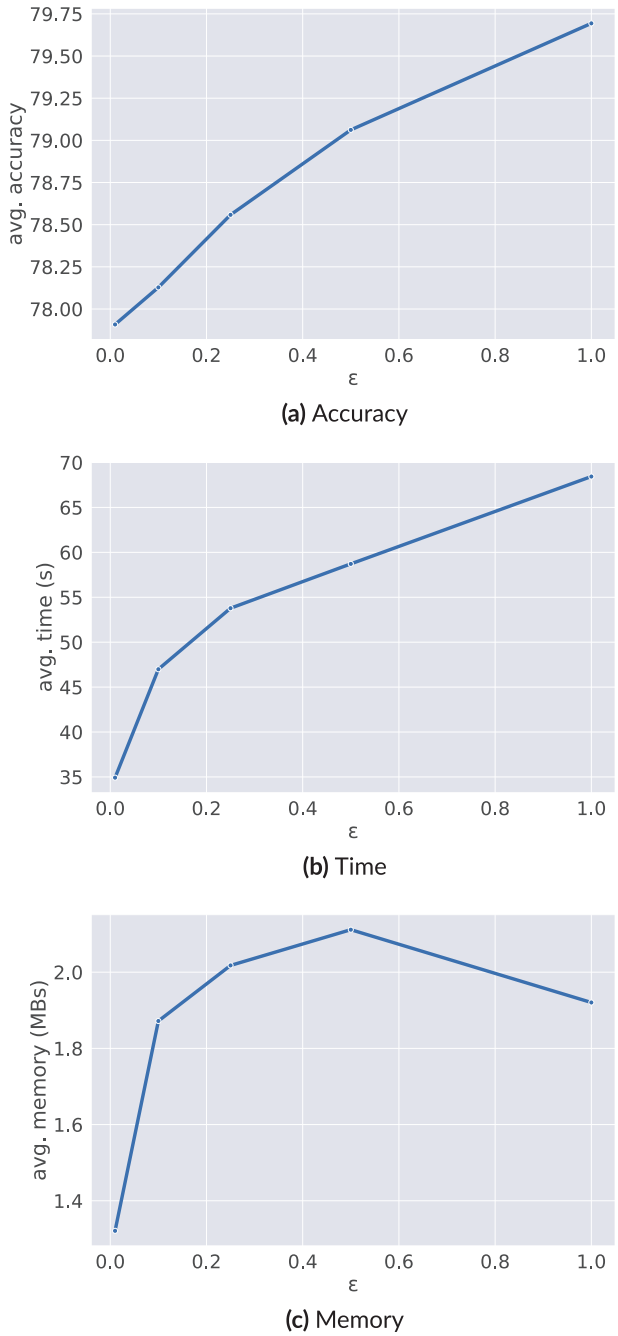


FIGURE 8 Variation in average performance of OMS-MAB across datasets with ϵ.

problems violate this assumption, as labels corresponding to previous instances arrive after a delay period (delayed setting) (Plasse & Adams, 2016). In this section, we discuss these aspects further and suggest potential approaches to extend OMS-MAB to such scenarios, along with future research directions.

CIL involves dealing with extremely skewed class distributions, with considerable underrepresentation of some classes (Cano & Krawczyk, 2022). Traditional classifiers are usually biased towards majority classes, resulting in poor predictions for minority classes which are typically more significant (Aguiar et al., 2023). Concept-drift makes CIL considerably harder as the imbalance ratio evolves with the stream, with classes possibly switching roles (Cano & Krawczyk, 2022). CIL is crucial in real-world applications such as medical data mining, social media analytics, and recommender systems with changing user preferences (Cano & Krawczyk, 2022). In contrast, MLC involves simultaneously mapping each data instance to multiple labels instead of a single class (Alberghini et al., 2022). MLC problems necessitate more intricate label and data representations to capture the interrelationships among instances and their labels, as well as among the labels themselves. Several real-world applications that involve text categorization, bioinformatics, and scene classification with objects mapped to multiple semantic meanings are well aligned to

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

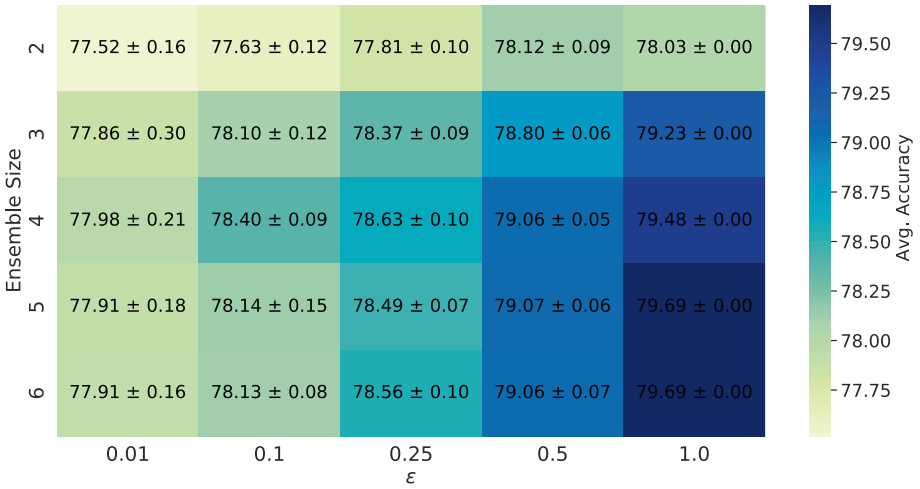


FIGURE 9 Variation in average accuracy across datasets with model complexity of OMS-MAB.

this setting (Zhu et al., 2023). Finally, handling delayed labels is an important aspect in streaming classification with practical applications such as malware classification, credit scoring for loans, and tweet labeling (Gomes et al., 2022). In such applications, the true class of some previously observed feature vector becomes available only after a delay, which could be fixed or varying.

Ensemble approaches have proven their effectiveness in CIL (Aguiar et al., 2023) and MLC (Zhu et al., 2023) problems in non-stationary data streams. Moreover, recent works highlight the use of different diversity induction strategies such as feature subspaces and bagging to build robustness in CIL (Cano & Krawczyk, 2022) and MLC (Alberghini et al., 2022). Due to its built-in diversity from heterogeneity and the instance-level diversity added by the MAB-based selective training process, OMS-MAB is a good candidate to be extended to tackle CIL and MLC.

According to the insights from (Cano & Krawczyk, 2022), one approach to extend OMS-MAB towards CIL is to maintain a per-class sliding window of instances globally for a balanced class representation by automatically downsampling majority classes. This sliding window can be used to periodically warm-start backup learners within ALs with a balanced class distribution, reducing majority class bias and managing evolving imbalance ratios. In order to extend OMS-MAB to MLC, we suggest capitalizing on the strengths of existing standalone MLC classifiers by employing them as ensemble base learners. Any standalone MLC classifier such as MLSAkNN (Roseberry et al., 2021), Multilabel Hoeffding Tree (Read, Bifet, Holmes, & Pfahringer, 2012), or i-SOUP Tree (Osojnik et al., 2017) could be suitable base learners in this context. To extend OMS-MAB to delayed label setting, we borrow from (Hu & Kantardzic, 2017) and propose keeping a fixed-size reservoir of labeled samples. The reservoir is updated with incoming data based on instance and label arrival times, while discarding excessively delayed samples which potentially represent obsolete concepts. Ensemble members could utilize this reservoir for periodic training and evaluation, ensuring access to relevant labeled data despite labeling delays. Furthermore, it is worth considering semi-supervised learning approaches in this context (Gomes et al., 2022).

In terms of additional future research directions, our approach can be extended to periodically select a best performing subset of members for training and prediction as opposed to choosing a single best performing member during ensemble selection. Individual predictions of the chosen subset can be combined using a voting scheme to obtain higher predictive performance while maintaining computational efficiency. Further, the approach can be generalized to support both homogeneous and heterogeneous ensembles. Finally, extending our approach to regression problems affected by concept-drift would be another interesting avenue.

6 | CONCLUSION

In this paper, we introduced a heterogeneous ensemble algorithm named OMS-MAB which uses both an active and passive strategy for concept-drift adaptation. A non-stationary MAB is used to select a single AL best suited for learning and prediction on an instance, from a pool of heterogeneous learners. The trade-off between predictive performance and computational resource requirements is controlled by an exploration factor ϵ. Our proposal provides competitive predictive performance compared to SOTA concept-drift ensembles at a significantly lower computational cost, by eliminating the need for continuous training and evaluation of all ensemble members. The above advantages make OMS-MAB a promising heterogeneous ensemble for mining high-speed data streams while operating in resource constrained environments.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. The first draft of the manuscript was written by Jobin Wilson and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All data generated or analysed during this study are included in this article.

ORCID

Jobin Wilson https://orcid.org/0000-0002-8923-0341

ENDNOTE

[1] https://github.com/jobinwilson/syntheticData6/

REFERENCES

Aggarwal, C. C., Philip, S. Y., Han, J., & Wang, J. (2003). A framework for clustering evolving data streams. Proceedings 2003 vldb conference. 81–92.

Aguiar, G., Krawczyk, B., & Cano, A. (2023). A survey on learning from imbalanced data streams: taxonomy, challenges, empirical study, and reproducible experimental framework. *Machine Learning*, 1–79. https://doi.org/10.1007/s10994-023-06353-6

Alberghini, G., Junior, S. B., & Cano, A. (2022). Adaptive ensemble of self-adjusting nearest neighbor subspaces for multi-label drifting data streams. *Neurocomputing*, *481*, 228–248.

Albuquerque, R. A. S., Costa, A. F. J., dos Santos, E. M., Sabourin, R., & Giusti, R. (2019). A decision-based dynamic ensemble selection method for concept drift. 2019 IEEE 31st international conference on tools with artificial intelligence (ictai). 1132–1139.

Alippi, C., Liu, D., Zhao, D., & Bu, L. (2013). Detecting and reacting to changes in sensing units: The active classifier case. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *44*(3), 353–362.

Bifet, A., & Gavalda, R. (2007). Learning from time-changing data with adaptive windowing. Proceedings of the 2007 siam international conference on data mining. 443–448.

Bifet, A., & Gavalda, R. (2009). Adaptive learning from evolving data streams. International symposium on intelligent data analysis. 249–260.

Bifet, A., Holmes, G., & Pfahringer, B. (2010). Leveraging bagging for evolving data streams. Joint european conference on machine learning and knowledge discovery in databases. 135–150.

Bifet, A., Holmes, G., Pfahringer, B., Kranen, P., Kremer, H., Jansen, T., & Seidl, T. (2010). Moa: Massive online analysis, a framework for stream classification and clustering. Proceedings of the first workshop on applications of pattern analysis. 44–50.

Brzezinski, D., & Stefanowski, J. (2013). Reacting to different types of concept drift: The accuracy updated ensemble algorithm. *IEEE Transactions on Neural Networks and Learning Systems*, *25*(1), 81–94.

Burtini, G., Loeppky, J., & Lawrence, R. (2015). A survey of online experiment design with the stochastic multi-armed bandit. *arXiv preprint arXiv: 1510.00757*.

Cano, A., & Krawczyk, B. (2022). Rose: robust online self-adjusting ensemble for continual learning on imbalanced drifting data streams. *Machine Learning*, *111*(7), 2561–2599.

Chen, Y., Zhu, Y., Chen, H., Shen, Y., & Xu, Z. (2021). A pruning optimized fast learn++ nse algorithm. *IEEE Access*, *9*, 150733–150743.

Chiu, C. W., & Minku, L. L. (2018). Diversity-based pool of models for dealing with recurring concepts. 2018 International Joint Conference on Neural Networks (ijcnn). 1–8.

Chiu, C. W., & Minku, L. L. (2020). A diversity framework for dealing with multiple types of concept drift based on clustering in the model space. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(3), 1299–1309.

de Barros, R. S. M., & de Carvalho Santos, S. G. T. (2019). An overview and comprehensive comparison of ensembles for concept drift. *Information Fusion*, *52*, 213–244.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, *7*(Jan), 1–30.

Domingos, P., & Hulten, G. (2000). Mining high-speed data streams. Proceedings of the sixth acm sigkdd international conference on knowledge discovery and data mining. 71–80.

Elwell, R., & Polikar, R. (2011). Incremental learning of concept drift in nonstationary environments. *IEEE Transactions on Neural Networks*, *22*(10), 1517–1531.

Frias-Blanco, I., del Campo-Ávila, J., Ramos-Jimenez, G., Morales-Bueno, R., Ortiz-Diaz, A., & Caballero-Mota, Y. (2014). Online and non-parametric drift detection methods based on hoeffding's bounds. *IEEE Transactions on Knowledge and Data Engineering*, *27*(3), 810–823.

Frías-Blanco, I., Verdecia-Cabrera, A., Ortiz-Díaz, A., & Carvalho, A. (2016). Fast adaptive stacking of ensembles. Proceedings of the 31st annual acm symposium on applied computing. 929–934.

Gama, J., Medas, P., Castillo, G., & Rodrigues, P. (2004). Learning with drift detection. Brazilian symposium on artificial intelligence. 286–295.

Garivier, A., & Moulines, E. (2008). On upper-confidence bound policies for non-stationary bandit problems. *arXiv preprint arXiv:0805.3415*.

Gomes, H. M., Barddal, J. P., Enembreck, F., & Bifet, A. (2017). A survey on ensemble learning for data stream classification. *ACM Computing Surveys (CSUR)*, *50*(2), 1–36.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

Gomes, H. M., Bifet, A., Read, J., Barddal, J. P., Enembreck, F., Pfharinger, B., … Abdessalem, T. (2017). Adaptive random forests for evolving data stream classification. *Machine Learning*, *106*(9), 1469–1495.
Gomes, H. M., Grzenda, M., Mello, R., Read, J., Le Nguyen, M. H., & Bifet, A. (2022). A survey on semi-supervised learning for delayed partially labelled data streams. *ACM Computing Surveys*, *55*(4), 1–42.
Harries, M., Wales, N. S., et al. (1999). Splice-2 comparative evaluation: Electricity pricing. Tech. Rep., The University of South Wales.
Herbold, S. (2020). Autorank: A python package for automated ranking of classifiers. *Journal of Open Source Software*, *5*(48), 2173.
Hu, H., & Kantardzic, M. M. (2017). Sliding reservoir approach for delayed labeling in streaming data classification. Hawaii international conference on system sciences. Retrieved from https://api.semanticscholar.org/CorpusID:37857226
Hulten, G., Spencer, L., & Domingos, P. (2001). Mining time-changing data streams. Proceedings of the seventh acm sigkdd international conference on knowledge discovery and data mining. 97–106.
Idrees, M. M., Minku, L. L., Stahl, F., & Badii, A. (2020). A heterogeneous online learning ensemble for non-stationary environments. *Knowledge-Based Systems*, *188*, 104983.
John, G. H., & Langley, P. (1995). Estimating continuous distributions in bayesian classifiers. Proceedings of the eleventh conference on uncertainty in artificial intelligence. 338–345.
Losing, V., Hammer, B., & Wersing, H. (2016). Knn classifier with self adjusting memory for heterogeneous concept drift. 2016 IEEE 16th international conference on data mining (icdm). 291–300.
Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., & Zhang, G. (2018). Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, *31*(12), 2346–2363.
Luong, A. V., Nguyen, T. T., Liew, A. W.-C., & Wang, S. (2021). Heterogeneous ensemble selection for evolving data streams. *Pattern Recognition*, *112*, 107743.
Manapragada, C., Webb, G. I., & Salehi, M. (2018). Extremely fast decision tree. Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining. 1953–1962.
Moa, datasets. (2019). Retrieved from. https://moa.cms.waikato.ac.nz/datasets/
Neto, A. F., & Canuto, A. M. (2021). Eocd: An ensemble optimization approach for concept drift applications. *Information Sciences*, *561*, 81–100.
Osojnik, A., Panov, P., & Džeroski, S. (2017). Multi-label classification via multi-target regression on data streams. *Machine Learning*, *106*, 745–770.
Plasse, J., & Adams, N. (2016). Handling delayed labels in temporally evolving data streams. 2016 IEEE international conference on big data (big data). 2416–2424.
Read, J., Bifet, A., Holmes, G., & Pfahringer, B. (2012). Scalable and efficient multi-label classification for evolving data streams. *Machine Learning*, *88*, 243–272.
Read, J., Bifet, A., Pfahringer, B., & Holmes, G. (2012). Batch-incremental versus instance-incremental learning in dynamic and evolving data. International Symposium on Intelligent Data Analysis. 313–323.
Roseberry, M., Krawczyk, B., Djenouri, Y., & Cano, A. (2021). Self-adjusting k nearest neighbors for continual learning from multi-label drifting data streams. *Neurocomputing*, *442*, 10–25.
Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386.
Rossi, A. L. D., de Leon Ferreira, A. C. P., Soares, C., De Souza, B. F., et al. (2014). Metastream: A meta-learning based method for periodic algorithm selection in time-changing data. *Neurocomputing*, *127*, 52–64.
Stream data mining repository. (2021). Retrieved from. https://www.cse.fau.edu/∼xqzhu/stream.html
Street, W. N., & Kim, Y. (2001). A streaming ensemble algorithm (sea) for large-scale classification. Proceedings of the seventh acm sigkdd international conference on knowledge discovery and data mining. 377–382.
Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
van Rijn, J. N., Holmes, G., Pfahringer, B., & Vanschoren, J. (2014). Algorithm selection on data streams. International conference on discovery science. 325–336.
van Rijn, J. N., Holmes, G., Pfahringer, B., & Vanschoren, J. (2018). The online performance estimation framework: heterogeneous ensemble learning for data streams. *Machine Learning*, *107*(1), 149–176.
Xu, J., Wang, G., Li, T., Deng, W., & Gou, G. (2017). Fat node leading tree for data stream clustering with density peaks. *Knowledge-Based Systems*, *120*, 99–117.
Zhao, P., Cai, L.-W., & Zhou, Z.-H. (2020). Handling concept drift via model reuse. *Machine Learning*, *109*(3), 533–568.
Zhu, X., Li, J., Ren, J., Wang, J., & Wang, G. (2023). Dynamic ensemble learning for multi-label classification. *Information Sciences*, *623*, 94–111.

AUTHOR BIOGRAPHIES

Jobin Wilson is a Principal Data Scientist - R&D at Flytxt, India. He holds an MS(Research) in Electrical Engineering from IIT Delhi, New Delhi, India. He is presently pursuing his Ph.D in Electrical Engineering at IIT Delhi. His research interests include Recommender systems, Artificial Intelligence and online non-stationary learning

Santanu Chaudhury is a Professor at the Department of Electrical Engineering, IIT Delhi, New Delhi, India. Recently he completed his tenure as the Director of IIT Jodhpur, Jodhpur, India. He has published more than 300 articles in international journals and conference proceedings. His research interests are in the areas of computer vision, robotics, embedded systems, and machine learning. Dr. Chaudhury is also a fellow of the Indian National Academy of Engineering and National Academy of Sciences, India, and the International Association of Pattern Recognition (IAPR).

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. 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

Brejesh Lall is a Professor at the Department of Electrical Engineering, IIT Delhi, India. His research areas include Multiscale Modeling of Stochastic Processing, Widesense Cyclostationary Process Representation, Physical layer in Wireless Communication. He holds B.E. (Electronics & Comm.) Delhi college of Engineering, M.E. Delhi college of Engineering and Ph.D. (Signal processing) - IIT Delhi.

How to cite this article: Wilson, J., Chaudhury, S., & Lall, B. (2024). Multi-armed bandit based online model selection for concept-drift adaptation. *Expert Systems*, e13626. https://doi.org/10.1111/exsy.13626

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13626 by Delhi Technological University, Wiley Online Library on [19/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI) | 979-8-3503-6052-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/IATMSI60426.2024.10503233

MULTIBAND SUNFLOWER FRACTAL ANTENNA FOR WIRELESS ENERGY HARVESTING

Neeraj Sharma
Department of Electronics and Communication Engineering
Delhi Technological University (DTU)
Delhi India
neeraj09ec@gmail.com

Asok De
Department of Electronics and Communication Engineering
Delhi Technological University (DTU)
Delhi India
asok.de@gmail.com

Anukul Pandey
Department of Electronics and Communication Engineering
Delhi Technological University (DTU)
Delhi India
anukul66@gmail.com

***Abstract*— This study presents a novel multiband fractal antenna design in the shape of a sunflower, specifically intended for wireless energy harvesting applications. The proposed antenna operates efficiently across four distinct frequency bands: 3.9 GHz to 4.1 GHz, 4.55 GHz to 4.66 GHz, 4.8 GHz to 5.05 GHz, and 7.3 GHz to 7.85 GHz. Each frequency band offers a respective bandwidth of 120 MHz, 110 GHz, 250 MHz, and 550 MHz. The antenna under consideration exhibits a peak gain of 5.1 dBi, 2.3 dBi, 2.2 dBi, and 4.9 dBi at frequencies of 4.1 GHz, 4.6 GHz, 4.9 GHz, and 7.45 GHz correspondingly. The antenna being presented exhibits high gain and multiband features, rendering it well-suited for RF energy harvesting applications. Specifically, it can effectively target the mid 5G band, WLAN, and 5G extended mid band, taking use of the available power in the surrounding environment. Subsequently, the acquired energy might be employed for the purpose of supplying power to low-power electronic sensors.**

***Keywords*— Multiband, Return Loss (RL), Fractal antenna, radio frequency energy harvesting, Gain.**

I. INTRODUCTION

Microstrip antennas are frequently employed in satellite and mobile applications due to their advantageous characteristics, including their lightweight nature, compact dimensions, cost-effectiveness, and ease of fabrication. Fractal-shaped antennas possess appealing characteristics that arise from their geometric aspects [1].

The term "fractal" was initially introduced by B. Mandelbrot, who is widely recognised as the progenitor of fractal geometry. The term "fractus" originates from the Latin language, specifically derived from the word "fractus," denoting a state of irregularity or fragmentation [2]. The aforementioned word was employed to characterise a novel assemblage of objects that surpass squares, circles, or triangles in terms of complexity, and are rigorously delineated within the framework of Euclidean geometry. Nevertheless, Euclidean geometry fails to adequately represent various everyday phenomena, including clouds, blood arteries, and irregular patterns like coastlines [3].

In addition to B. Mandelbrot, some esteemed mathematicians have made significant contributions to the discipline of fractal geometry. These include W. Sierpinski, D. Hilbert, N. F. H. von Koch, H. Minkowski, and G. Peano, and G. F. L. P. Cantor. The fractals served as a source of inspiration for the antenna engineering community, prompting them to investigate the potential utilisation of these geometries as antennas [4]–[9]. The current publications provide a historical context of their survives and introduce antenna influenced by the fractals.

In [10], the authors present a unique design that utilised the Fibonacci sequence and created a defective ground structure (DGS) in the lower part of the antenna. This particular design enables the antenna to operate effectively throughout the C-Band frequency range.

According to the source cited as [11], The ultimate configuration of the antenna is achieved by many iterations. This amalgamation results in a unique hybrid structure that possesses a small physical size, excellent flexibility, and improved impedance bandwidth. The antenna is at 2.45 GHz. The constructed prototype of the antenna has demonstrated a satisfactory level of concordance between the numerical and experimental findings. The achieved performance includes a peak gain of 2.06 decibels relative to an isotropic radiator (dBi) and provides 75 % of efficiency.

The antenna design aims to achieve mechanical robustness as proposed in [12]. The proposed methodology entails the process of etching Moore's curve onto a standard rectangular patch, followed by enclosing it within a rectangle loop that incorporates a defective ground structure. The antenna under consideration demonstrates dual-band characteristics, with a bandwidth that spans from 1.38 to 1.8 GHz and 2.25 to 4.88 GHz, encompassing the WMTS, ISM, and Personal Communication bands. Additionally, it partially covers a section of the UWB band. The antenna possesses two operational bands, each providing a fractional bandwidth of 38.5% and 73.7% correspondingly.

The present study introduces a newly developed monopole antenna element [13]. The compact nature of this device allows for its utilisation in portable devices with twin antennas that operate within the 2.4 GHz frequency range. This enables the achievement of antenna diversity across various combining schemes. The computing demands of the finite element simulations used to estimate the concert of the verified shapes were found to be higher compared to the moment method counterparts.

The present study introduces a novel design for a microstrip antenna with circular polarisation (CP), as described in reference [14]. The parameter studies of the Iterative Fourier Domain Gradient Sensing (IFDGS) method are presented in order to demonstrate the methodology for achieving coherent population (CP) radiation. The CP antenna is constructed and assessed using the third iterative IFDGS. The obtained experimental data confirms the accuracy of the simulated results and demonstrates a high level of agreement.

979-8-3503-6052-3/24/$31.00 ©2024 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:58:48 UTC from IEEE Xplore. Restrictions apply.

Energy harvesting systems encompass the process of converting pre-existing energy sources found in the environment, such as solar energy, wind energy, thermal energy, and electromagnetic energy, into a useful form [15]. The electromagnetic energy density is observed to be progressively growing as a result of the ongoing advancements in communication technologies [16]. RF energy harvesting (RFEH) has emerged as a viable approach for the purpose of energy recycling. The utilisation of this acquired energy will prove advantageous in the charging of diverse electronic devices in scenarios where a conventional connected power source is challenging to establish, such as in rural regions, submerged environments, and aerial settings [17]-[19]. Energy harvesting technology has the potential to serve as a viable alternative to batteries, primarily due to its limited lifespan, high cost, requirement for frequent replacement, and environmental impact [20]. Recycled energy possesses environmentally friendly attributes, exhibits an extended availability period, and obviates the need for system replacement.

The antenna under consideration has been designed and simulated utilising CST MW Studio 2020. The antenna under consideration is designed on a substrate made of FR-4 material, which has a dielectric constant of 4.4 and a substrate height of 1.6 mm. The subsequent part introduces a proposed multiband high gain antenna and presents a comprehensive analysis of its outcomes. In Section 3, the conclusion is provided and thereafter addressed.

II. Antenna Design

The subsequent section introduces the sunflower-shaped multiband fractal antenna. The proposed antenna underwent a design and simulation process at multiple stages. The resulting output, including return loss, gain, and 3D radiation patterns, were examined and compared. Figure 1 depicts the initial stage (Stage 1) of the projected antenna, wherein the dimensions of the larger and smaller sections are 3 mm and 1.8 mm, respectively, measured from the centre. The thickness of the strips in this stage is 0.5 mm. Figure 2 displays Stage 2, when nine entities of identical kind are replicated, each possessing a radius of 5 mm. Figure 3 illustrates Stage 3, wherein nine entities from Stage 2 are replicated at a radius of 10 mm, resulting in a configuration resembling that of a sunflower. Consequently, this particular antenna is referred to as a sunflower-shaped antenna. Figure 4 illustrates Stage 4, whereby nine entities are arranged in a repetitive pattern. One entity is positioned at the centre, while the remaining eight entities are evenly distributed over a radius of 15 mm. Notably, the central entity is scaled to be 1.5 times larger than the outside entities. The ultimate antenna is seen in Figure 5.

The antenna has been built and simulated for stages 2, 3, and 4. The comparison of return loss for various stages is illustrated in Figure 6. The minimum required RL for stage 2 is -6 dB. The stage 3 fractal antenna exhibits resonance throughout the frequency range of 6.78 GHz to 6.97 GHz, with a minimum return loss (RL) of -19 dB occurring at 6.87 GHz. The stage 4 fractal antenna, seen in Figure 5, has operational capabilities throughout the 3.9 GHz to 4.1 GHz, 4.55 GHz to 4.66 GHz, 4.8 GHz to 5.05 GHz, and 7.3 GHz to 7.85 GHz. This antenna offers a bandwidth of 120 MHz, 110 GHz, 250 MHz, and 550 MHz, respectively. Additionally, it achieves a minimum return loss (RL) of -34.4 dB at 7.44 GHz. The suggested antenna exhibits multiband features that make it well-suited for RF energy harvesting applications. Specifically, it is designed to efficiently capture RF energy in the mid 5G band, WLAN, and 5G extended mid band, taking advantage of the available power in the surrounding environment.



Fig. 1. Stage 1 of the planned antenna

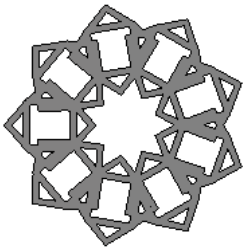


Fig. 2. Stage 2 of the planned antenna

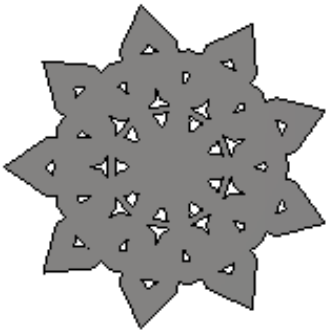


Fig. 3. Stage 3 of the planned antenna

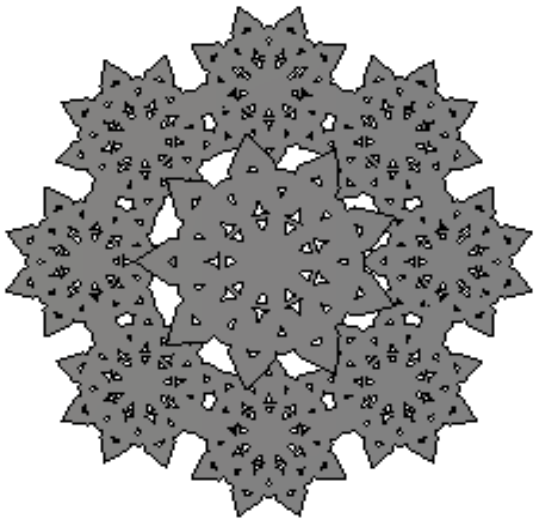


Fig. 4. Stage 4 of the planned antenna

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:58:48 UTC from IEEE Xplore. Restrictions apply.

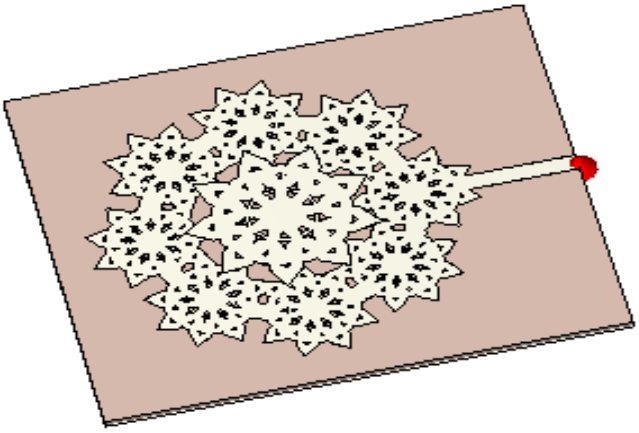


Fig. 5. Planned antenna

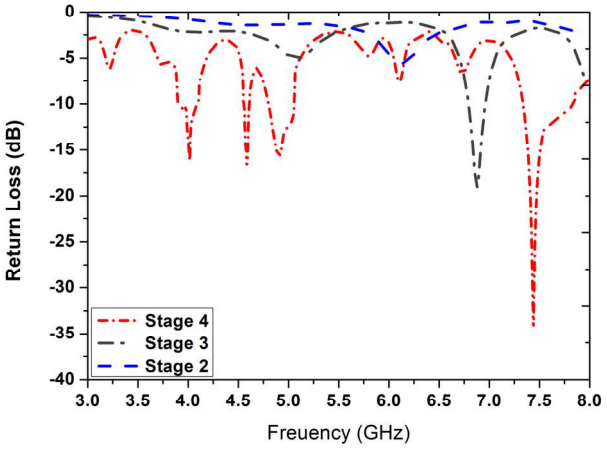


Fig. 6. RL Comparison of various stages of fractal antenna

Figure 6, 7, 8, and 9 depict the gain variation of the antenna under consideration at 4.1 GHz, 4.6 GHz, 4.9 GHz, and 7.45 GHz correspondingly. The antenna under consideration exhibits a peak gain of 5.1 dBi, 2.3 dBi, 2.2 dBi, and 4.9 dBi at 4.1 GHz, 4.6 GHz, 4.9 GHz, and 7.45 GHz correspondingly. The three-dimensional radiation pattern (RP) of the antenna under consideration is depicted in Figures 11, 12, 13, and 14. The greatest directivity achieved at 4.1 GHz, 4.6 GHz, 4.9 GHz, and 7.45 GHz is measured to be 6.1 dBi, 6.45 dBi, 8.34 dBi, and 10.1 dBi, respectively. The antenna that has been proposed exhibits multiband and high gain properties, making it suitable for many wireless applications. The proposed antenna exhibits potential for utilisation in RF energy harvesting applications, as it is capable of efficiently harvesting power from several frequency bands.

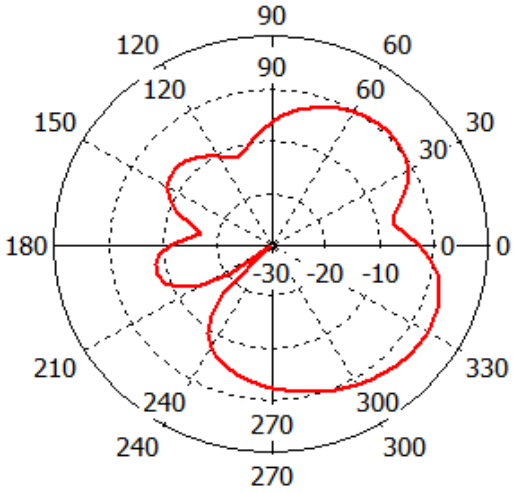


Fig. 7. Gain pattern at 4.1 GHz

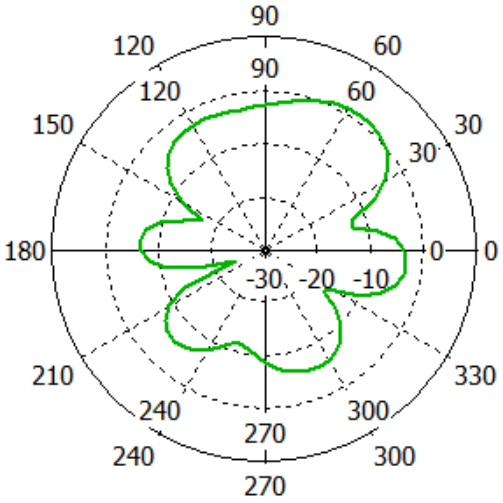


Fig. 8. Gain pattern at 4.6 GHz

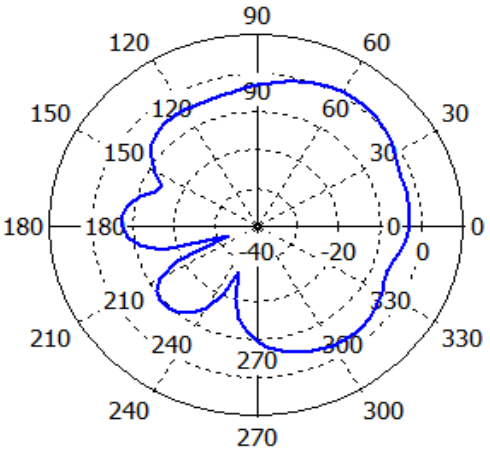


Fig. 9. Gain pattern at 4.9 GHz

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:58:48 UTC from IEEE Xplore. Restrictions apply.

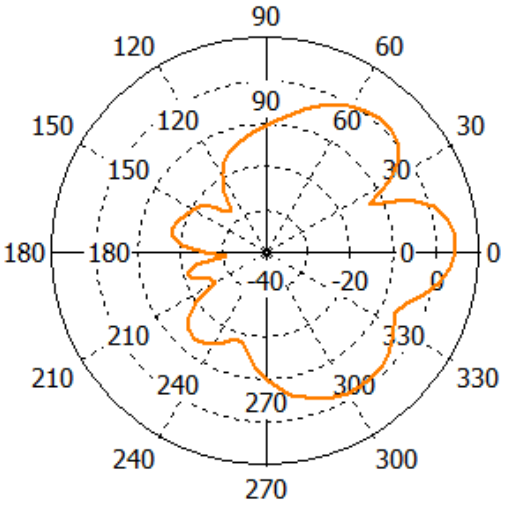


Fig. 10. Gain pattern at 7.45 GHz

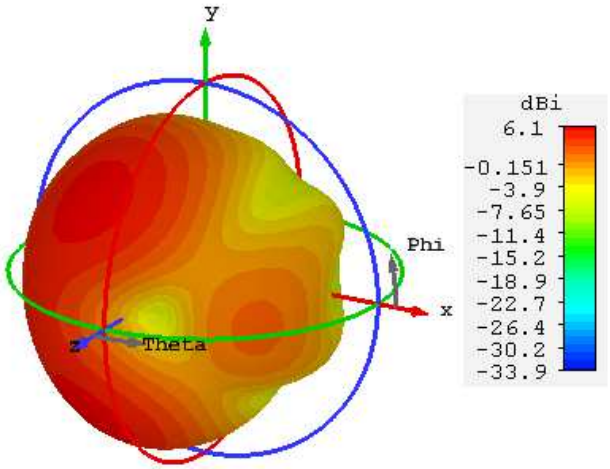


Fig. 11. 3D RP at 4.1 GHz

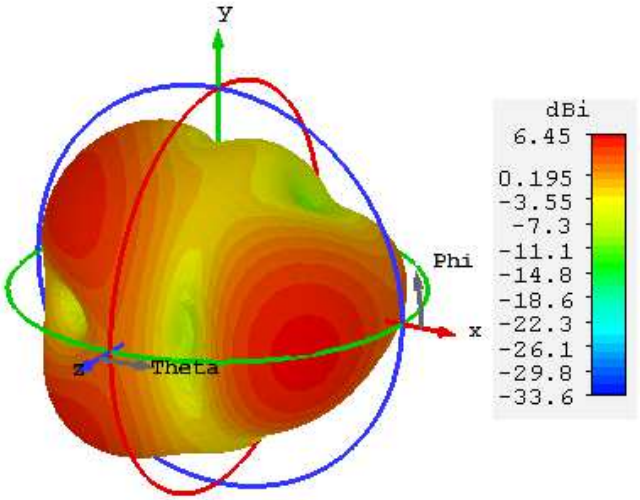


Fig. 12. 3D RP at 4.6 GHz

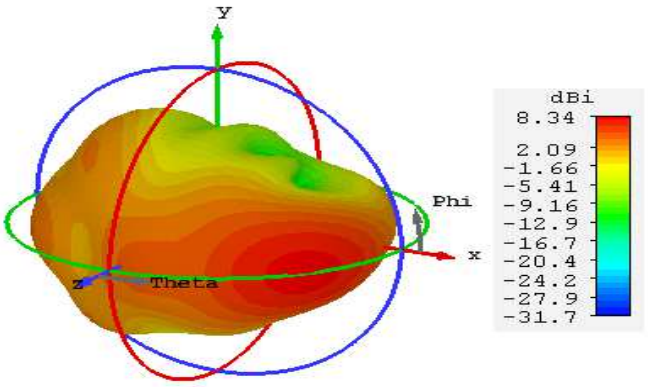


Fig. 13. 3D RP at 4.9 GHz

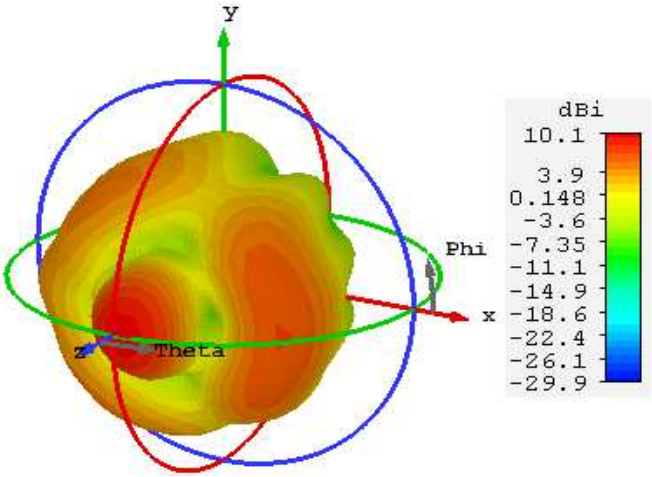


Fig. 14. 3D RP at 7.45 GHz

TABLE I. COMPARISON WITH PREVIOUS WORK

S. No.	Frequency (GHz)	Bandwidth (MHz)	Gain (dBi)
1.	2.45	500	3
2.	2.63	144	4.07
15.	1.9	50	--
	1.8	60	--
20	2.45	--	4.7
Proposed Work	4.1	120	5.1
	4.6	110	2.3
	4.9	250	2.2
	7.45	550	4.9

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:58:48 UTC from IEEE Xplore. Restrictions apply.

III. Conclusion

A multiband fractal antenna in the form of a sunflower, with a specific focus on its suitability for wireless energy harvesting purposes is presented. The antenna under consideration has the capability to operate on four discrete frequency ranges: 3.9 GHz to 4.1 GHz, 4.55 GHz to 4.66 GHz, 4.8 GHz to 5.05 GHz, and 7.3 GHz to 7.85 GHz. The bandwidths of each frequency band are 120 MHz, 110 GHz, 250 MHz, and 550 MHz, respectively. The antenna being examined demonstrates a maximum gain of 5.1 decibels isotropic (dBi) at a frequency of 4.1 gigahertz (GHz), 2.3 dBi at 4.6 GHz, 2.2 dBi at 4.9 GHz, and 4.9 dBi at 7.45 GHz, respectively. The antenna being presented demonstrates a high gain and the ability to operate over multiple frequency bands, making it highly suitable for applications involving the collection of RF energy. In particular, it possesses the capability to efficiently focus on the mid 5G band, WLAN, and 5G extended mid band, while harnessing the existing power within the surrounding environment. Following this, the obtained energy could be utilised for the objective of providing power to electronic sensors with low power requirements.

References

[1] Gupta S., Kanaujia B. K., Dalela C., & Patil S., "Design of circularly polarized antenna using inclined fractal defected ground structure for S-band applications," Electromagnetics, vol. 40, no. 7, pp. 526-540, 2020.

[2] Cannon, J. W., "The fractal geometry of nature. by Benoit B. Mandelbrot," The American Mathematical Monthly, vol. 91, no. 9, pp. 594-598, 1984.

[3] Anguera J., Andujar A., Jayasinghe J., Chakravarthy V.S., Chowdary P.S.R., Pijoan J.L., Ali T. and Cattani C., "Fractal antennas: An historical perspective," Fractal and Fractional, vol. 4, no. 1, pp. 1-26, 2020.

[4] Werner D.H., Haupt R.L., Werner P.L., "Fractal antenna engineering: The theory and design of fractal antenna arrays," IEEE Antennas Propag. Mag., vol. 41, pp. 37-58, 1999.

[5] Gianvittorio J.P., Rahmat-Samii, Y. Fractal antennas "A novel antenna miniaturization technique and applications," IEEE Antennas Propag. Mag., vol. 44, pp. 20-36, 2002.

[6] Anguera, J., C. Puente, and C. Borja, "Fractal shaped Antennas: A review," Encyclopedia of RF and Microwave Engineering, vol. 2, pp. 1620–1635, 2005.

[7] Puente C., Pous R., "Fractal design of multiband and low side-lobe arrays," IEEE Trans. Antennas Propag, vol. 44, pp. 730-739, 1996.

[8] Cohen N., "Fractal antenna applications in wireless telecommunications," Professional Program Proceedings. Electronic Industries Forum of New England. IEEE, pp. 43–49, 1997.

[9] Werner D.H., Ganguly S., "An overview of fractal antenna engineering research," IEEE Antennas Propag. Mag., vol. 45, pp. 38-57, 2003.

[10] Gaonkar Abhijeet, M. Ayyappan, and Pragati Patel, "A Novel Fractal RDRA for C-Band Applications," IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 13, no. 7, pp. 995-1002, 2023.

[11] Arif A., Zubair M., Ali M., Khan M. U., & Mehmood M. Q., "A compact, low-profile fractal antenna for wearable on-body WBAN applications," IEEE Antennas and Wireless Propagation Letters, vol. 18, no. 5, pp. 981-985, 2019.

[12] Khan U. R., Sheikh J. A., Junaid A., Amin R., Ashraf S., & Ahmed S, "Design of a compact hybrid Moore's fractal inspired wearable antenna for IoT enabled bio-telemetry in diagnostic health monitoring system," IEEE Access, vol. 10, pp. 116129-116140, 2022.

[13] Gupta S, Kanaujia B. K., Dalela C., & Patil S. "Design of circularly polarized antenna using inclined fractal defected ground structure for S-band applications," Electromagnetics, vol. 40, no.7, pp. 526-540, 2020.

[14] Wang L., Yu J., Xie T., & Bi K, "A novel multiband fractal antenna for wireless application," International Journal of Antennas and Propagation, vol. 2021, pp. 1-9, 2021.

[15] Ojha Shailendra Singh, P. K. Singhal, and Vandana Vikas Thakare, "Dual-band rectenna system for biomedical wireless applications," Measurement: Sensors, vol. 24, pp. 1-7, 2022.

[16] Ojha Shailendra Singh, P. K. Singhal, Anshul Agarwal, and Akhilesh Kumar Gupta, "2-GHz dual diode dipole rectenna for wireless power transmission," Int J Microw Opt Technol, vol. 8, no. 2, pp. 86-92, 2013.

[17] Huang W., Du J., Yang X. X., Che W., & Gao S, "A novel 24 GHz circularly polarised metasurface rectenna," IET Microwaves, Antennas & Propagation, vol. 17, no. 6, pp. 419-426, 2023.

[18] Ojha Shailendra Singh, P. K. Singhal, and Vandana Vikas Thakare. "Triple-Wideband Antenna for RF Energy Harvesting," International Conference on Communication, Networks and Computing. Cham: Springer Nature Switzerland, vol. 1983, pp. 226-237, 2022.

[19] Ojha S.S., Tomar R.S., Akashe S., Dhakad B., Mishra S. and Sharma M., "Dual-Band Antenna and Low Pass Filter Design for Wireless Energy Harvesting," International Conference on Communication, Networks and Computing. Cham: Springer Nature Switzerland, vol. 1893, pp. 238-248, 2022.

[20] Ojha Shailendra Singh, Pramod Kumar Singhal, and Vandana Vikas Thakare, "Highly Efficient Dual Diode Rectenna with an Array for RF Energy Harvesting," Wireless Personal Communications, vol. 131, no. 4, pp. 1-22, 2023.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:58:48 UTC from IEEE Xplore. Restrictions apply.

2024 3rd International Conference for Innovation in Technology (INOCON)
Karnataka, India. Mar 1-3, 2024

Multilingual Depression Detection in Online Social Media Across Eight Indian Languages

1st Viraj Rajderkar
Department of Computer Science
Delhi Technological University
New Delhi, India
virajraj022@gmail.com

2nd Dr. Aruna Bhat
Department of Computer Science
Delhi Technological University
New Delhi, India
aruna.bhat@dtu.ac.in

2024 3rd International Conference for Innovation in Technology (INOCON) | 979-8-3503-8193-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/INOCON60754.2024.10511869

***Abstract*—Depression, a prevalent mental health concern, manifests itself on social media, providing opportunities for early detection. This study aims to fill the gap in research for low-resource Indian languages by proposing a comprehensive approach to depression detection across eight languages. By utilizing the Long Short-Term Memory and the Gated Recurrent Unit models, we provide an ensemble model that improves overall performance. Our translation-based dataset solves data restrictions related to language. Drawing insights from various studies, the literature review delves into the complexities of analyzing mental health on social media. We translate these insights into practical applications through model implementation and preprocessing techniques. The LSTM and GRU models, enriched with pre-trained FastText embeddings, demonstrate competitive accuracy. The ensemble model, amalgamating their outputs, exhibits robustness in depression detection. Nevertheless, certain limitations persist, including concerns related to dataset representativeness and the nuances of informal social media language. Our study concludes with a roadmap for future research, emphasizing the need to expand datasets, develop language-specific models, and explore advanced architectures. This research contributes to a nuanced understanding of depression detection in linguistically and culturally diverse contexts, building upon insights gained from the literature review and model implementations across various Indian languages.**

***Index Terms*—Depression Detection, Social Media Analysis, Multilingual Sentiment Analysis, Indian Languages, LSTM and GRU Models, Ensemble Learning, Translation Approach, Text Preprocessing, Mental Health Detection, Low-Resource Languages, Social Media Data Mining, Cross-Lingual Sentiment Analysis, Word Embeddings, Natural Language Processing**

I. Introduction

Understanding human psychology is a complex yet crucial process, especially when extracting insights from text-based data. Analyzing emotions, opinions, and judgment formation is key, with social media platforms like Facebook, Twitter, and Reddit playing a pivotal role in contemporary expression [1]. Psychology's influence spans mental health, product reviews, customer satisfaction, and business decisions. Online reviews impact product perception, offering insights for improvement [2]. This paper focuses on major depressive disorder, prevalent globally, and explores Twitter as a valuable source for mental health data [4] [9]. Social media, including Twitter, Facebook, Instagram, and Reddit, provides rich datasets for psychological classification tasks, fostering concise expression and discussions on various topics [9]. The study investigates the connection between artificial intelligence (AI) and psychology, using machine learning classifiers to detect depression on Twitter. Preprocessed datasets are subjected to four different analyses: logistic regression, decision trees, K-nearest neighbor, support vector machines, and long short-term memory. The enhanced performance of LSTM with imbalanced datasets is highlighted.

The methodology, results, and a discussion follow, contributing insights to the psychology-AI nexus. Social media platforms, especially Twitter, offer crucial psychological insights due to character limits and diverse discussion groups on topics, including mental health. Psychological analysis of social media data extracts meaningful insights from platforms like Twitter, Facebook, Instagram, and Reddit. Twitter's character limit encourages concise expression, making it valuable for understanding user psychology. Focusing on major depressive disorder in mental health research, this paper highlights symptoms, emphasizing Twitter's role as a space for users to share thoughts and feelings concisely [4] [9]. The study's core is applying machine learning classifiers to analyze social media data for depression detection. The comparative analysis showcases LSTM's effectiveness, particularly with imbalanced datasets, contributing to the intersection of psychology and AI. The structure of this work is as follows: The literature is reviewed in Section 2, the criteria for identifying depression are covered in Section 3, the approach and five machine learning classifiers are covered in Section 4, and the experimental results are presented in Section 5. Concluding remarks and research directions are provided in Section 6.

II. Literature review

The exploration of mental health through social media has been a subject of substantial research, employing various approaches across different languages and

platforms.

Reputable studies, such as [1] have examined real-time detection on Thai Twitter. The PHQ-9 questionnaire, demographic information, and Twitter IDs were utilized in the study to categorize participants as Moderately Depressed, Slightly Depressed, and Normal. F-measure, accuracy, precision, and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:58:28 UTC from IEEE Xplore. Restrictions apply.

recall were used to gauge the job's complexity using models including Decision Tree, Random Forest, Naïve Bayes, and Deep Learning. PySpark and other statistical tools improved the research.

[2] examined datasets from Reddit and Twitter in research done in the English language. With the use of sophisticated models like LSTM, Bi-LSTM, and BERT, this study was able to get remarkable accuracy rates of 94%, 96%, and 98%, respectively. It was emphasized how crucial word embedding methods are to accurately expressing textual data. Multimodal analysis had a role in another aspect of Twitter mental health detection in [3] This study relied on self-declared tweets for classification, highlighting the difficulties of varied user expressions, with over 11 million tweets from depressed users and 16 million from non-depressed users. [4] concentrated on Weibo in Chinese, extending beyond Western platforms. In order to get insight into language-specific difficulties, users were categorized according to the presence or lack of explicit terms connected to depression for depressed users and these keywords for non-depressed users.

[5] examined sentiment in 1.6 million tweets using well-established datasets, highlighting the need of nuanced sentiment interpretation in the context of sentiment analysis methodologies.

[6] examined Twitter and Reddit for cross-language comparisons, using explicit self-declaration as the classification criterion. The two language-specific approaches, [9] and [8]examined Bengali sentiment analysis on social media platforms such as Facebook and Twitter.

[10] showcased the difficulties of multilingual analysis by utilizing the source of articles for labeling, while also examining Japanese mood. [11]examined Arabic tweets and labeled them based on depressive and melancholic phrases. To sum up, the literature review highlights the complexity of comprehending mental health through social media by presenting a diverse range of approaches, languages, and platforms.The diverse strategies employed contribute to a nuanced understanding of sentiment and user expressions across different linguistic and cultural contexts. Comparative Analysis of Depression Detection Models Recent studies have explored the effectiveness of various depression detection models across different languages and social media platforms. Diverse machine learning techniques were employed to analyze user-generated content and identify depressive sentiments. Here, we provide an overview of several studies and their outcomes: Bengali Depression Detection Models

Finding the Severity of Depression Through Bengali Social Media Posts on Mental Health: 86% accuracy was attained with a BiGRU model. Depression Analysis using Gated Recurrent Neural Network (GRU) on Bangla Social Media Data: 75.7% accuracy was attained using GRU. Multilingual Tweet Analysis Multilabel Classification for Emotion Analysis of Multilingual Tweets: Gujarati: Accuracy of 84% using Linear SVC. Hindi: Accuracy of 78% using Linear SVC. Hindi Depression Recognition Models

CNN-BiLSTM was utilized to identify emotions in code-mixed social media in Hinglish (Hindi+English), having an accuracy of 83.21%.

BERT was utilized to recognize emotions in data containing mixed English and Hindi codes, using a six transformer technique.

Sentiment Analysis in Malayalam Deep Learning-Mixed Malayalam-English Text Analytics on Social Media: 76.33% accuracy was attained using GRU. Sentiment Analysis of Dravidian Languages

Benchmarking multi-task learning reveals that Tamil has a precision of 74.4% when using BERT regarding Offensive Language Identification.

Malayalam: 64.7% precision with BERT. Sentiment Analysis in Multiple South Indian Languages (Malayalam, Tamil, Kannada) Deep Learning Based Sentiment Analysis: Malayalam: Accuracy of 84.28% using LSTM. Tamil: Accuracy of 52.93% using LSTM. Kannada: Accuracy of 57.45% using LSTM. These comparative insights highlight the diversity in model performances across languages, emphasizing the need for language-specific and culturally sensitive approaches in social media-based mental health research.

III. Model Architecture

A. Long Short-Term Memory (LSTM) Model

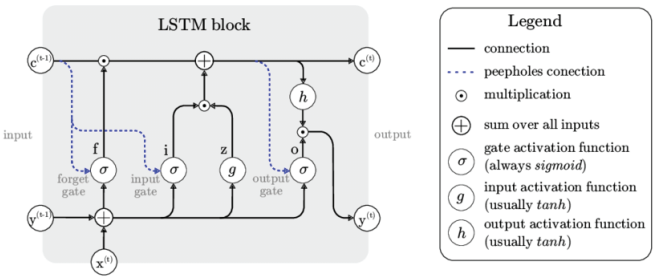


Fig. 1. LSTM Architecture [21]

The LSTM model is a type of recurrent neural network (RNN) designed to address the vanishing gradient problem, enabling the capture of long-range dependencies in sequential data. The architecture consists of memory cells, input gates, output gates, and forget gates. Given an input sequence $X = (x_1, x_2, ..., x_t)$, the LSTM computations at time t are defined in equations 1-6:

B. Gated Recurrent Unit (GRU) Model

The GRU is another type of RNN designed for simpler computations and faster training. It comprises a reset gate, an update gate, and a candidate hidden state. Given an input sequence $X = (x_1, x_2, ..., x_t)$, the GRU computations at time t are defined in equations 7-10:

C. Ensemble Model

The ensemble model combines the predictions of the LSTM and GRU models to enhance overall performance. Given individual predictions P_{LSTM} and P_{GRU} from the LSTM and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:58:28 UTC from IEEE Xplore. Restrictions apply.

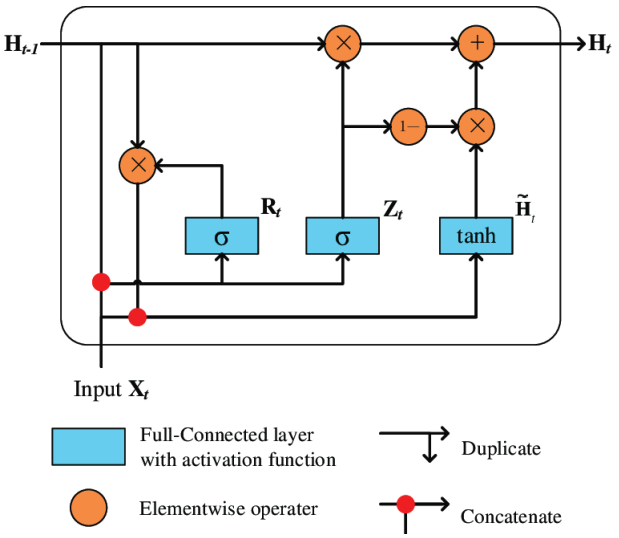


Fig. 2. GRU Architecture [22]

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{4}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{6}$$

Fig. 3. Equations describing LSTM model

GRU models, respectively, the ensemble prediction P_{Ensemble} is computed as shown in the equation 11:

The ensemble model benefits from the diverse representations learned by LSTM and GRU, providing a more robust and accurate prediction for the task of depression detection in social media data.

IV. Methodology

A. Dataset

1. Collection of Depressive and Non-Depressive Tweets The dataset used in this study came from Kaggle and is made up of tweets that Google Translate has translated into a number of different languages. Two subsets were created: depressive tweets, containing content indicative of depression, and non-depressive tweets, representing normal Twitter discourse. The creation of a balanced dataset ensures a representative training environment for sentiment analysis models. 2. Data Preprocessing Steps To enhance the dataset's quality and facilitate

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \tag{7}$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \tag{8}$$

$$\tilde{h}_t = \tanh(W \cdot [r_t \cdot h_{t-1}, x_t] + b) \tag{9}$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \tag{10}$$

Fig. 4. Equations describing GRU model

$$P_{\text{Ensemble}} = \frac{P_{\text{LSTM}} + P_{\text{GRU}}}{2} \tag{11}$$

Fig. 5. Equation describing ensemble model

effective model training, several preprocessing steps were employed. These steps included the removal of noise and irrelevant information, text normalization, and the creation of labeled datasets for depressive and non-depressive tweets.

B. Preprocessing Techniques

1. Text Cleaning and Normalization:

Text cleaning involved the removal of special characters, URLs, and unnecessary symbols. Additionally, text normalization techniques, such as converting text to lowercase, were applied to maintain uniformity in the dataset. The preprocessing aimed to ensure that the input data for the models is consistent and free from irrelevant artifacts.

2. Tokenization and Word Analysis:

Tokenization, performed using the Keras Tokenizer, involved breaking down the cleaned text into individual tokens. This step aids in creating a vocabulary for the models. Furthermore, an analysis of words and their frequencies was conducted to understand the composition of both depressive and non-depressive tweets.

3. Word Cloud Visualization: The most common terms in the depressing tweets were shown visually in a word cloud created with the WordCloud software. This visualization contributes to a qualitative comprehension of the dataset by helping to discover important phrases linked to depression.

C. LSTM and GRU Models

1. Word Embedding with Pre-trained FastText Vectors:

Word embedding for the LSTM and GRU models was done with pre-trained FastText vectors. These vectors improve the input text's representation by capturing the semantic connections between words. Using FastText makes it easier to comprehend linguistic subtleties by making one more aware of the context.

2. Training and Model Architecture:

Both LSTM and GRU models were constructed using specific architectures that included embedding layers, recurrent layers (LSTM/GRU), and a dense output layer. The models

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:58:28 UTC from IEEE Xplore. Restrictions apply.

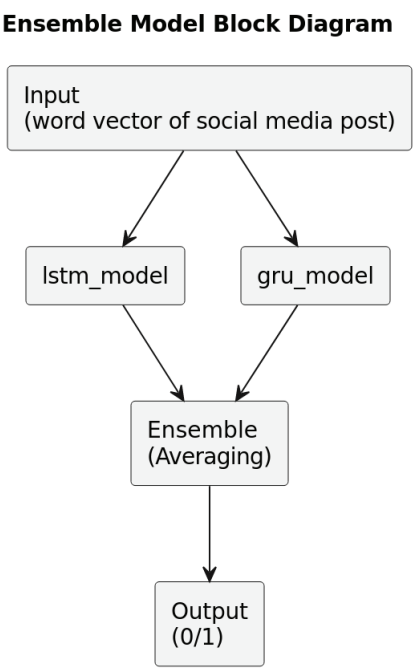


Fig. 6. Ensemble Architecture

have been trained using the preprocessed data to find hidden trends in the data, enabling sentiment analysis.

3. Early Stopping and Optimization Strategies:

To prevent overfitting and enhance training efficiency, early stopping mechanisms were implemented. Additionally, optimization strategies, such as dropout and regularization, were applied to fine-tune the models. These measures aimed to improve the generalization ability of the LSTM and GRU models. The technique described here offers a thorough rundown of the procedures used for gathering datasets, preparing them, and applying baseline and deep learning methods to evaluate sentiment on multilingual Twitter data.

V. Results and Discussion

In our comprehensive research of sentiment analysis across several Indian languages, we give an in-depth examination of the model's performance metrics, including precision, recall, accuracy, and F1 score. Our approach, leveraging a translation methodology for the creation of language-specific databases, adds a layer of complexity and sophistication to the study. The subsequent table encapsulates nuanced performance metrics for each language.

Gujarati: With an F1 score of 0.90280, recall of 0.89615, accuracy of 0.91669, precision of 0.91058, and robust performance, our sentiment analysis model performed well in Gujarati. This shows how effective our translation-based dataset construction is, since it outperforms the accuracy standard established by [15]

Kannada: Our model showed excellent recall (0.89360), accuracy (0.92232), precision (0.92664), and F1 score (0.90753) in Kannada. The accuracy and recall measures demonstrated the flexibility of our technique to the subtleties of Kannada sentiment expression, outperforming previous models.

Hindi: We obtained an accuracy of 0.92513, precision of 0.92843, recall of 0.89936, and an F1 score of 0.91181 for our Hindi sentiment analysis model. These outcomes demonstrated the model's efficacy in identifying emotion subtleties in Hindi social media content, surpassing the F1 score published in [16]

Bengali: With an accuracy of 0.91922, precision of 0.91842, recall of 0.89437, and an F1 score of 0.90492, our model performed exceptionally well in Bengali. These data show how effective our sentiment analysis model is when applied to Bengali social media material, surpassing the performance of current standards.

Telugu: With an F1 score of 0.90416, accuracy of 0.91866, precision of 0.93089, recall of 0.88704, and other notable outcomes, our sentiment analysis model performed quite well in Telugu. The model's overall usefulness is bolstered by its high accuracy, which demonstrates its ability to reliably recognize both positive and negative attitudes in Telugu text.

Marathi: With an accuracy of 0.91247, precision of 0.91511, recall of 0.88310, and an F1 score of 0.89649, our model performed admirably in Marathi. The model's 30 ability to keep recall and accuracy in check indicates its dependability when it comes to sentiment analysis of Marathi social media material.

Malayalam: We obtained an accuracy of 0.91388, precision of 0.90700, recall of 0.89309, and F1 score of 0.89952 using our sentiment analysis model for Malayalam. These findings demonstrate the efficacy of our technique in sentiment analysis of code-mixed Malayalam-English text, surpassing the F1 scores published in [18]

Tamil: With an accuracy of 0.91416, precision of 0.91904, recall of 0.88777, and an F1 score of 0.90065, our model performed competitively in Tamil. As demonstrated by the F1 scores obtained in [19] these outcomes demonstrate the adaptability of our sentiment analysis model in the context of the Tamil language. These language-specific discussions underscore the adaptability and efficacy of our sentiment analysis model across diverse linguistic landscapes, providing valuable insights into the nuances of sentiment expression in Indian languages on social media. The integration of a translation approach for dataset creation contributes to the model's success in capturing sentiment variations across different languages.

VI. Limitations

Despite the promising results obtained, our study has certain limitations. First off, the caliber and representativeness of the training data have a major impact on how well the models function. While we expanded the dataset to include many Indian languages using a translation technique, the effectiveness of this strategy could differ depending on the language, which might affect the generality of the model. Moreover, the ensemble model, while designed to enhance overall performance, is not immune to the individual limitations of the LSTM and GRU modelsThe quality of the data and the subtle differences

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:58:28 UTC from IEEE Xplore. Restrictions apply.

TABLE I
OUR RESULTS

Language	Accuracy	Precision	Recall	F1 Score
Gujarati	0.91669	0.91058	0.89615	0.90280
Kannada	0.92232	0.92664	0.89360	0.90753
Hindi	0.92513	0.92843	0.89936	0.91181
Bengali	0.91922	0.91842	0.89437	0.90492
Telugu	0.91866	0.93089	0.88704	0.90416
Marathi	0.91247	0.91511	0.88310	0.89649
Malayalam	0.91388	0.90700	0.89309	0.89952
Tamil	0.91416	0.91904	0.88777	0.90065

TABLE II
COMPARATIVE ANALYSIS

Paper	Language	Accuracy	Precision	Recall	F1 Score	Model
[12]	Bengali	0.86	0.85	0.86	0.86	BiGRU
[13]	Bengali	0.757	-	-	-	GRU
[14]	Bengali	90.3	0.694	0.547	0.562	BiGRU
[15]	Gujarati	-	-	0.84	-	Linear SVC
[16]	Hindi	-	-	0.78	-	Linear SVC
[16]	Hindi	0.8321	0.8323	0.8321	0.8319	CNN-BiLSTM
[17]	Hindi	0.7143	-	-	-	BERT
[18]	Malayalam	0.7633	0.7607	0.7633	0.7617	GRU
[19]	Tamil	-	0.744	0.645	0.625	BERT
[20]	Malayalam	-	0.647	0.69	0.668	BERT
[20]	Malayalam	0.8428	0.6303	0.6304	0.627	LSTM
[20]	Tamil	0.5293	0.4232	0.5272	0.441	LSTM
[20]	Kannada	0.5745	0.5062	0.5455	0.5193	LSTM

in vocabulary between the languages determine how effective each model is. There are inherent difficulties with using social media data to identify mental health issues. These platforms' texts are frequently informal, context-dependent, and noisy, all of which affect how accurate the models are. Additionally, the diverse linguistic landscape of Indian languages introduces complexities that may not be fully addressed in our current approach.

VII. CONCLUSION

In conclusion, our research explores the detection of depression in social media across eight Indian languages, addressing the paucity of work in low-resource Indian languages. We have implemented LSTM and GRU models, leveraging their ability to capture sequential dependencies, and proposed an ensemble model to amalgamate their strengths. The translation approach has been employed to overcome data limitations in each language. Our findings indicate promising results, with competitive accuracy and precision scores across languages. The ensemble model, combining the strengths of LSTM and GRU, demonstrates robustness in depression detection.

VIII. FUTURE SCOPE

The current study opens avenues for future research in several directions. Firstly, continuous efforts should be directed towards expanding the dataset, encompassing a more diverse range of social media posts to enhance model generalization. Moreover, fine-tuning models on language-specific linguistic characteristics could further improve performance. To address the challenge of limited work in low-resource Indian languages, future research should focus on developing dedicated models for these languages, considering the linguistic intricacies unique to each. Contextual data, user demographics, and temporal features incorporated into the models may offer a more comprehensive knowledge of how mental health is expressed on social media. Furthermore, investigating sophisticated deep learning architectures and utilizing multimodal data sources might potentially enhance the precision and subtlety of depression detection algorithms.

REFERENCES

[1] J. Angskun, S. Tipprasert, T. Angskun, "Big data analytics on social networks for real-time depression detection," Journal of Big Data, vol. 9, no. 1, p. 69, 2022, Springer.

[2] K. Zeberga, M. Attique, B. Shah, F. Ali, Y. Jembre, T.-S. Chung, "A novel text mining approach for mental health prediction using Bi-LSTM and BERT model," Computational Intelligence and Neuroscience, vol. 2022, 2022, Hindawi.

[3] R. Safa, P. Bayat, L. Moghtader, "Automatic detection of depression symptoms in twitter using multimodal analysis," The Journal of Supercomputing, vol. 78, no. 4, p. 4709–4744, 2022, Springer.

[4] G. Li, B. Li, L. Huang, S. Hou, et al., "Automatic construction of a depression-domain lexicon based on microblogs: text mining study," JMIR medical informatics, vol. 8, no. 6, p. e17650, 2020, JMIR Publications Inc., Toronto, Canada.

[5] S. Gupta, L. Goel, A. Singh, A. Prasad, M.A. Ullah, "Psychological analysis for depression detection from social networking sites," Computational Intelligence and Neuroscience, vol. 2022, 2022, Hindawi.

[6] J. de Jesús Titla-Tlatelpa, R.M. Ortega-Mendoza, M. Montes-y-Gómez, L. Villaseñor-Pineda, "A profile-based sentiment-aware approach for depression detection in social media," EPJ data science, vol. 10, no. 1, p. 54, 2021, Springer Berlin Heidelberg.

[7] H. Zogan, I. Razzak, X. Wang, S. Jameel, G. Xu, "Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media," World Wide Web, vol. 25, no. 1, p. 281–304, 2022, Springer.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:58:28 UTC from IEEE Xplore. Restrictions apply.

[8] M.R.H. Khan, U.S. Afroz, A.K.M. Masum, S. Abujar, S.A. Hossain, "A deep learning approach to detect depression from Bengali text," Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 2, p. 777–785, 2021, Springer.
[9] D.B. Victor, J. Kawsher, M.S. Labib, S. Latif, "Machine learning techniques for depression analysis on social media-case study on bengali community," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), p. 1118–1126, 2020, IEEE.
[10] Y.M. Niimi, Y. Miyaji Yuka, "Machine learning approach for depression detection in japanese," Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation, p. 350–357, 2021.
[11] A.A. Alaskar, "Depression Detection from Arabic Tweets Using Machine Learning Techniques," J Comput Sci Software Dev, vol. 2, p. 1–10, 2022.
[12] D.B. Victor, J. Kawsher, M.S. Labib, S. Latif, "Machine learning techniques for depression analysis on social media-case study on Bengali community," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), p. 1118–1126, 2020, IEEE.
[13] M.K. Kabir, M. Islam, A.N.B. Kabir, A. Haque, M.K. Rhaman, "Detection of Depression Severity Using Bengali Social Media Posts on Mental Health: Study Using Natural Language Processing Techniques," JMIR Formative Research, vol. 6, no. 9, p. e36118, 2022, JMIR Publications Inc., Toronto, Canada.
[14] A.H. Uddin, D. Bapery, A.S.M. Arif, "Depression analysis of bangla social media data using gated recurrent neural network," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), p. 1–6, 2019, IEEE.
[15] L. Gohil, D. Patel, "Multilabel classification for emotion analysis of multilingual tweets," Int J Innov Technol Explor Eng, vol. 9, no. 1, p. 4453–4457, 2019.
[16] T.T. Sasidhar, B. Premjith, K.P. Soman, "Emotion detection in hinglish (hindi+ english) code-mixed social media text," Procedia Computer Science, vol. 171, p. 1346–1352, 2020, Elsevier.
[17] A. Wadhawan, A. Aggarwal, "Towards emotion recognition in Hindi-English code-mixed data: A transformer based approach," arXiv preprint arXiv:2102.09943, 2021.
[18] S. Thara, P. Poornachandran, "Social media text analytics of Malayalam–English code-mixed using deep learning," Journal of big Data, vol. 9, no. 1, p. 45, 2022, Springer.
[19] A. Hande, S.U. Hegde, R. Priyadharshini, R. Ponnusamy, P.K. Kumaresan, S. Thavareesan, B.R. Chakravarthi, "Benchmarking multi-task learning for sentiment analysis and offensive language identification in under-resourced dravidian languages," arXiv preprint arXiv:2108.03867, 2021.
[20] P.H.V. Pavan Kumar, B. Premjith, J.P. Sanjanasri, K.P. Soman, "Deep Learning Based Sentiment Analysis for Malayalam, Tamil and Kannada Languages," 2021.
[21] G. Van Houdt, C. Mosquera, G. Nápoles, "A Review on the Long Short-Term Memory Model," Artificial Intelligence Review, vol. 53, p. , Dec. 2020, doi: 10.1007/s10462-020-09838-1.
[22] Y. Fang, S. Yang, B. Zhao, C. Huang, "Cyberbullying Detection in Social Networks Using Bi-GRU with Self-Attention Mechanism," Information, vol. 12, p. 171, Apr. 2021, doi: 10.3390/info12040171.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:58:28 UTC from IEEE Xplore. Restrictions apply.

International Journal of Information Security
https://doi.org/10.1007/s10207-024-00861-9

REGULAR CONTRIBUTION

Non-Alpha-Num: a novel architecture for generating adversarial examples for bypassing NLP-based clickbait detection mechanisms

Ashish Bajaj[1] · **Dinesh Kumar Vishwakarma**[1]

Accepted: 23 April 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
The vast majority of online media rely heavily on the revenues generated by their readers' views, and due to the abundance of such outlets, they must compete for reader attention. It is a common practise for publishers to employ attention-grabbing headlines as a means to entice users to visit their websites. These headlines, commonly referred to as clickbaits, strategically leverage the curiosity gap experienced by users, enticing them to click on hyperlinks that frequently fail to meet their expectations. Therefore, the identification of clickbaits is a significant NLP application. Previous studies have demonstrated that language models can effectively detect clickbaits. Deep learning models have attained great success in text-based assignments, but these are vulnerable to adversarial modifications. These attacks involve making undetectable alterations to a small number of words or characters in order to create a deceptive text that misleads the machine into making incorrect predictions. The present work introduces "*Non-Alpha-Num*", a newly proposed textual adversarial assault that functions in a black box setting, operating at the character level. The primary goal is to manipulate a certain NLP model in a manner that the alterations made to the input data are undetectable by human observers. A series of comprehensive tests were conducted to evaluate the efficacy of the suggested attack approach on several widely-used models, including Word-CNN, BERT, DistilBERT, ALBERTA, RoBERTa, and XLNet. These models were fine-tuned using the clickbait dataset, which is commonly employed for clickbait detection purposes. The empirical evidence suggests that the attack model being offered routinely achieves much higher attack success rates (ASR) and produces high-quality adversarial instances in comparison to traditional adversarial manipulations. The findings suggest that the clickbait detection system has the potential to be circumvented, which might have significant implications for current policy efforts.

Keywords Clickbait · Convolutional Neural Network · Transformer models · Adversarial Attacks · Deep Learning · Vulnerability

1 Introduction

Throughout the preceding decennium, Machine Learning models exhibited remarkable efficacy across a range of tasks, including regression, classification, and decision-making. The susceptibility of these models to adversarial instances, which are genuine inputs that have been modified by subtle perturbations, has been noted. Recent studies have made significant advancements in the creation of hostile pictures, which have the capability to undermine the effectiveness of computer vision algorithms [1]. There exists a scarcity of research on adversarial scenarios within several NLP applications, including but not limited to sentiment analysis [2], topic categorization, machine translation, hate content identification, and false news detection [3]. Due to its efficacy in the realm of visual perception, adversarial learning has emerged as a relatively new subject of heightened interest, warranting more investigation.

The focus of previous research on adversarial examples has predominantly been on the domain of vision, mainly due to the discrete nature of text data and the challenges associated with optimisation. Moreover, it is worth noting that the perturbations introduced in the image domain often possess the ability to appear almost imperceptible to the

✉ Dinesh Kumar Vishwakarma
dvishwakarma@gmail.com

Ashish Bajaj
bajaj.ashish25@gmail.com

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

Published online: 13 May 2024

human visual system. As a result, this can give rise to disparities between the way humans perceive visual stimuli and the performance of state-of-the-art computational models [1]. The manipulation of a single word within the context of a text can significantly impact the meaning of a statement, with even minor changes being readily identifiable. The field of natural language processing (NLP) encompasses a diverse array of applications, spanning various domains such as machine translation, sentiment analysis, and question answering. However, despite its broad scope, there has been a limited amount of research conducted on adversarial cases within NLP. Adversarial learning has emerged as a promising technique in the field of image analysis. Its effectiveness has garnered significant interest and attention in recent years. Consequently, there is a growing need to delve deeper into this relatively new and valuable research area. Multiple studies have examined the security of current machine learning models and have proposed different attack strategies, including *causative* and *exploratory* attacks, as suggested by various research findings [4]. Exploratory attacks involve the manipulation of testing instances, known as adversarial examples, with the intention of deceiving a targeted classifier. On the other hand, causative attacks focus on modifying the training data with the intention of misleading the classification algorithm [5].

The generation of adversarial instances involve the application of two distinct adversarial environments. The concept of a "*black-box setup*" refers to a specific design approach for creating adversarial examples. In this setup, the malignant is intentionally kept unconscious of both the model being targeted & the specific training data used to train the classifier. This lack of knowledge about the inner workings of the classifier and the data it has been trained on adds an additional layer of complexity to the adversarial example creation process. In contrast, the adversary possesses complete knowledge about the classifier and training data within a "*white-box*" setting [6]. To be considered a natural language attacking system, the system must meet certain criteria. These criteria include successfully deceiving the target models and adhering to the 3 key conditions which maintain practicality. (1) Semantic similarity refers to requirement that the generated instance should possess the identical meaning as the genuine example, as assessed by human evaluators. (2) The inclusion of adversary examples should prioritise their naturalness and grammatical correctness. (3) The Integrity of Individual Predictions: Examining the Stability of Predictions Made by Humans.

The investigation into the resilience of transformer-based "clickbait-detection" is a significant area of study, particularly due to the increasing use of adversarial techniques. These techniques involve malicious users making slight modifications to deceive the classifiers and generate inaccurate outcomes. The concept of perceptual distinctiveness refers to the ability to differentiate between what should be perceived in the surrounding environment and what is actually perceived. The architecture of deceiving deep learning-based clickbait detection mechanisms is shown in Fig. 1. In addition, it is imperative to investigate and evaluate the susceptibilities of these classifiers when subjected to various adversarial attack strategies, with the aim of providing potential expedient that can be implemented to mitigate these assaults. The research studies indicate that the identified susceptibilities have the potential to manipulate the perception of classifiers, resulting in a reversal or cancellation of their intended functionality. This poses a significant risk when these vulnerabilities are exploited in applications that rely on the same classifier.

1.1 Importance of the study

The advent of social media platforms, such as Facebook and Twitter, has facilitated global connectivity and the exchange of ideas among individuals worldwide. By engaging in such actions, individuals have inadvertently opened avenues for cyber assailants and online harassers to disseminate false information, engage in offensive behaviour, cyberbully users, and incite conflicts through social media platforms and public online forums. Clickbaiting may be seen as a marketing tactic employed to attract readers by exploiting their curiosity or interest, particularly when there is a lack of compelling or captivating content. Clickbaiting is often perceived as lacking integrity. Digital marketing organisations are implementing methods to mitigate the prevalence of clickbait postings in users' timelines through the utilisation of clickbait detecting techniques. Despite the notable advancements made by deep learning systems in recognising clickbait language, they still exhibit susceptibility to adversarial settings. Adversarial machine learning refers to the intentional manipulation of input data by introducing tiny perturbations, with the objective of causing intended paradigm may have a glitch, resulting in the generation of inaccurate results. The diagram illustrating the architecture for adversarial machine learning may be observed in Fig. 5. Therefore, the effectiveness of the model may be undermined. The present work investigates the adversarial limitations inherent in sophisticated deep learning algorithms. This study is of potential significance to those who often utilise existing and advanced deep learning frameworks for analytical & commercial activities.

This finding will provide motivation for the next generation of scholars to not just focus on the improvement of model accuracy scores, nevertheless, to focus on developing resilient architectures that can survive malicious manipulations, which are already seen as a substantial security concern.

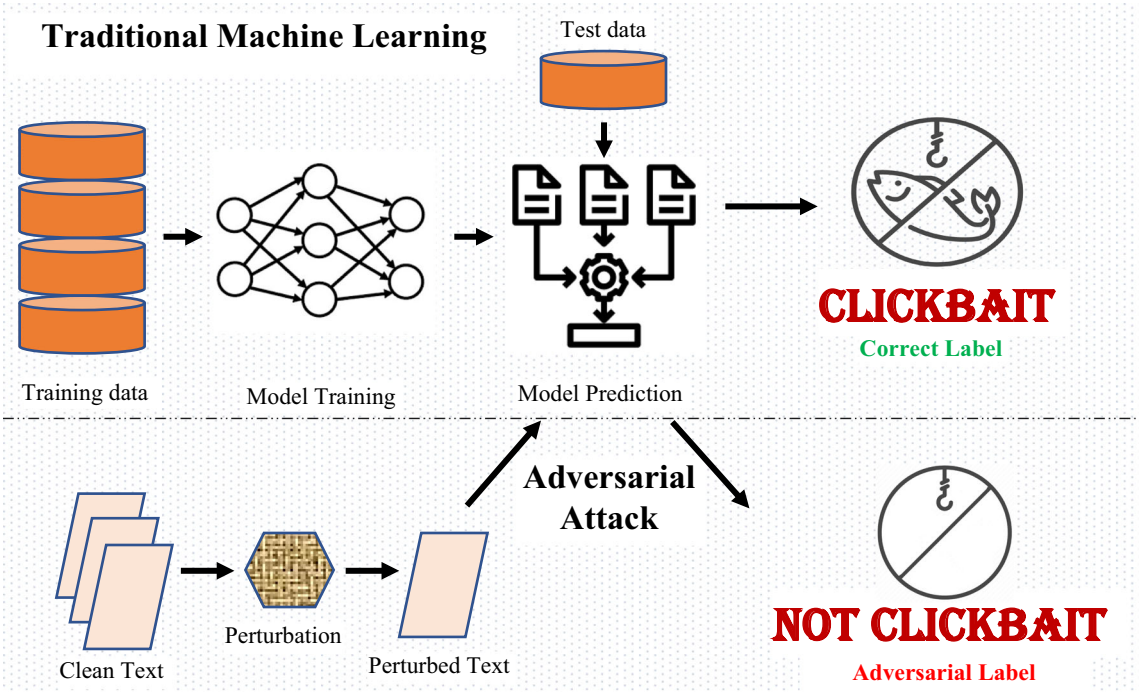


Fig. 1 Adversarial attack framework in clickbait detection mechanism

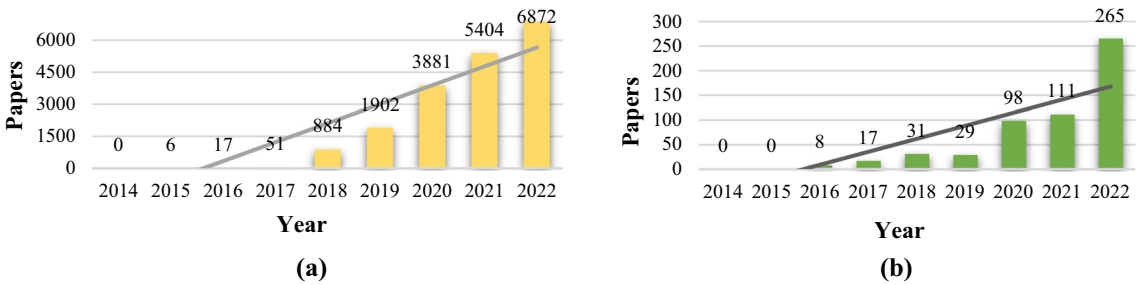


Fig. 2 **a** Represents the number of publications in the field of adversarial example, as compiled by Carlini, including image, audio, and text across a broad spectrum. **b** Depicts the number of publications in the adversarial text-domain

1.2 Motive of the investigation

The domain of adversarial machine learning has witnessed significant growth in recent years. The current body exists of a substantial body of scholarly work about the adversary interpretation of textual modelling but is still limited, indicating a scarcity of research in this area. In contemporary literature, Fig. 2 provides a visual representation of the prevalence of adversarial examples.

Threats are being extensively investigated in numerous studies, specifically focusing on the field of computer vision. Nevertheless, there's a shortage of scholarly articles about the domain of textual analysis. Adversarial attacks have garnered significant attention in the realm of research, particularly in the domain of image manipulation. Therefore, in this study, we analyse prominent research papers in the field of NLP to ascertain the datasets employed within the textual context. Table 1 showcases the specifications of the extensively employed datasets that have been utilised in various studies pertaining to adversarial attacks [7].

Figure 3 illustrates the distribution of percentages for the NLP assignments. A significant majority, over 50%, of databases are specifically allocated for the purpose of categorising textual data. This activity has considerable importance within the realm of NLP [7]. Clickbait intended to deceive readers by promoting misleading claims. However, it can also serve more nefarious purposes, such as engaging in phishing activities to get personal information or, in more severe cases, acting as a platform for hosting malware. Due to the aforementioned factors, the matter of clickbait assumes a significant magnitude and needs prompt attention and resolution. The initial stage involves the identification of bait and non-clickbait content through the utilisation of

Table 1 The prior research utilizes twelve widely used text datasets in the examination of adversarial assaults, with a focus on the domains of Neural Machine Translation (NMT), Question and Answer (QA), and Natural Language Inference (NLI)

Problem	Title	Size	Specification	Dataset link
Classification	SST	240 T	The standard sentiment dataset from Stanford	https://github.com/stanfordnlp/sentiment-treebank
	MR	10 T	Information extracted from movie review	http://www.cs.cornell.edu/people/pabo/movie-review-data/
	Yelp	140 T	Customer testimonials of business reviews	https://huggingface.co/docs/datasets/index
	Amazon	2 M	Amazon merchandise evaluation	
	Yahoo	1.4 M	Yahoo! responds to detailed questions	
	IMDB	50 T	different opinions about films	
	AG news	144 T	More than 2,000 sources of news	
	DBPedia	45 T	The organised information of Wikimedia projects	
NMT	WMT14	–	Texts that are next to each other (for example, translation models) German/English)	http://www.statmt.org/wmt14/translation-task.html
NLI	MultiNLI	4.33T	Crowdsourced collection of sentence pairs	https://cims.nyu.edu/~sbowman/multinli/
	SNLI	570T	Authors compose sentence pairs in the English language	https://nlp.stanford.edu/projects/snli/
QA	SQuAD	100T	Data collection from Wikipedia for question answering and reading comprehension	https://datarepository.wolframcloud.com/resources/SQuAD-v1.1

T Thousand, *M* Million

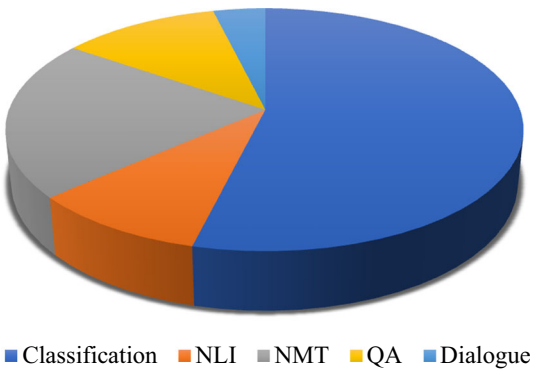


Fig. 3 Statistics pertaining to datasets utilized in the study of adversarial attacks

deep learning algorithms. Because of a pressing need, this study used a clickbait dataset to show how fragile clickbait identification is when there are attackers around. As a result, the authors decided to do a thorough testing study of how deep learning models for clickbait detection can be attacked. Currently, there's a dearth of empirical research about the resilience of various deep neural network algorithms against adversarial assaults. This study adds something new to the field because it questions how stable existing cutting-edge deep learning models are that are often used in NLP tasks. Individuals who use these models and want to better comprehend their flaws will find this information useful.

This study employs cutting-edge techniques to attack and investigate vulnerabilities in Clickbait detection. The objective is to conduct experiments with adversarial attacks to recognize and compute the weaknesses of clickbait detectors. Therefore, the purpose of this paper is to address the following research question: How susceptible is the detection of clickbait to adversarial attacks?

1.2.1 The key *findings* of the intended research are outlined below

- "'*Non-Alpha-Num*" a novel textual adversarial attack architecture is developed that generates undetectable adversarial texts against cutting-edge text categorization systems in black-box settings. Extensive experimentation demonstrates the ease with which an adversary can deceive Deep Learning-based clickbait detection mechanisms. Such adversarial instances raise numerous concerns and significantly diminish the utility of classification systems for detecting clickbaits.
- To find the perturbation technique (attack method) that has the most significant effect on deep learning models that were trained on the clickbait dataset. Among the models are the well-known Word-CNN and five potent transformer models: BERT, DistilBERT, ALBERT, RoBERTa, and XLNet.
- This study also aims to compare and analyze the performance of deep learning-based clickbait detectors in order

to ascertain the framework that demonstrates the highest sensitivity & resilience under all adversarial scenarios.

1.2.2 Organization

The present study is structured in the following manner: Sect. 2 of this paper provides an overview of the background information and past publications that are relevant to the topic being discussed. Section 3 presents the recommended methodology for evaluating clickbait detection mechanisms in hostile circumstances. The design of our experimental settings is outlined in Sect. 4, while Sect. 5 presents the experimental outcomes. In Sect. 6, an analysis of the experimental data is presented. Following that, Sect. 7 offers a comprehensive analysis of the findings. Section 8 of the paper provides a comprehensive summary of the study's findings, along with suggestions for future investigations and a discussion of the limitations inherent in the present research.

2 Related work

This section provides an overview of the problem of clickbait detection and its use in many contexts. Subsequently, the paper delves into the contextual framework of adversarial machine learning in NLP, specifically focusing on the manipulation of text through elementary perturbations in order to deceive machine algorithms. In addition to the aforementioned points, this section also presents a compilation of previously established advanced attack tactics in the field of NLP.

2.1 Clickbait detection

Due to the widespread dissemination of information on the internet, individuals are currently exposed to a constant influx of advertising and news headlines that appear on nearly every webpage they visit. Given the widespread availability of the internet, websites and news providers are engaged in a perpetual competition to attract and retain a significant viewership [8]. Consequently, there exists a significant burden on individuals to generate more captivating, attention-grabbing, and thought-provoking titles for articles, irrespective of their veracity. As a consequence, there has been an observable increase in sensationalist "news" headings in recent years, which lack substantive information about the subject but are designed to captivate readers. The word "clickbait" pertains to deceptive hyperlinks accompanied with exaggerated headlines, designed to captivate the audience's interest and encourage them to click on the hyperlink [9]. In a general context, clickbait headlines may be categorised based on two primary factors:

(1) their tendency to deceive readers regarding the actual substance of the article.
(2) their use of the "curiosity gap" by withholding complete information about the article's contents [10]. In a more concise manner, the material contained in these headlines either generates curiosity in the reader regarding the remaining content of the article or explores subjects that are not covered inside the article's main body.
(3) Also, clickbaits are utilised by those involved in fraudulent operations, including phishing and smishing, with the intention of acquiring personal information or facilitating the dissemination of malware [11].

The initial step in resolving this issue involves differentiating clickbait headlines from authentic headline links. On the surface, clickbait headlines can be difficult to identify because they are designed to deceive the user. However, there are key semantic characteristics that can be used to identify clickbait headlines. AI tools based on machine learning algorithms can systematically detect and block clickbaits. Clickbait identification has been a subject of significant accomplishment through the use of classical machine learning methods [12]. The utilization of deep learning methods has led to advancements in model sophistication and intelligence, as demonstrated by the notable example of transformer models [13]. The models have attracted considerable attention as a result of their outstanding confidence ratings in the field of text categorization [14]. The present work utilized highly efficient classification models that are rooted in deep learning techniques. This study highlights the need to assess clickbait detectors based on deep learning techniques before their use in decision support systems, through the utilization of adversarial assaults.

2.2 Preliminaries of adversarial machine learning in NLP

This section provides an introductory understanding of adversarial instances, encompassing formulaic descriptions, interpretations, and the categorization of such cases.

Deep neural networks (DNNs) are a type of artificial neural network that are composed of the function $\boldsymbol{F}$, which represents a typical Deep Neural Network (DNN), can be defined as $\boldsymbol{F}$:$\mathbf{X} \rightarrow \mathbf{Y}$. This function maps elements from the input set $\mathbf{X}$ to corresponding labels in the label set $\mathbf{Y}$. The set $\mathbf{Y}$ consists of $\boldsymbol{k}$ classes, denoted as $\{1, 2, \ldots, k\}$. In the case of a given sample $\boldsymbol{x}$ belonging to the set $\mathbf{X}$, it is observed that the classification function $\boldsymbol{F}$ accurately assigns the true label $\boldsymbol{y}$ to $\boldsymbol{x}$, denoted as $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{y}$.

Adversarial examples refer to a phenomenon in machine learning where slight modifications or perturbations are made to input data. The objective of the attacker is to introduce a slight perturbation $\boldsymbol{\varepsilon}$ into the input variable $\boldsymbol{x}$ in order to generate an adversarial example $\boldsymbol{x}'$. This adversarial example is designed to have a different output label $\boldsymbol{y}'$ (where $\boldsymbol{y}$ is not equal to $\boldsymbol{y}'$) when evaluated by the target model $\boldsymbol{F}$. Simultaneously, it is imperative for $\boldsymbol{x}'$ to deceive $\boldsymbol{F}$ while also maintaining imperceptibility to human observers [15]. In order to ensure that the generated $\boldsymbol{x}'$ remains imperceptible, various metrics such as semantic similarity are utilised to accomplish this objective, specifically by ensuring that the norm of $\boldsymbol{\varepsilon}$ is less than $\boldsymbol{\delta}$. Here, $\boldsymbol{\delta}$ serves as a threshold to restrict the magnitude of perturbations.

Robustness The concept of robustness refers to the ability of a system or process to maintain its functionality and when it comes to defending against adversarial examples, it is crucial for a robust deep neural network (DNN) model to possess the ability to withstand adversarial attacks [16]. This means that the model should be able to maintain accurate predictions even when faced with imperceptible additive malicious manipulations [3]. Therefore, it is hypothesised that in a robust deep neural network (DNN) model, the anticipated outcome of the adversarial example $\boldsymbol{x}'$ should be $\boldsymbol{y}$ instead of $\boldsymbol{y}'$, denoted as $\boldsymbol{F}(\mathbf{x}') = \boldsymbol{y}$. In order to improve the resilience of models, it is crucial to implement defence mechanisms that effectively address a broad spectrum of $\boldsymbol{\varepsilon}$.

2.2.1 Categorization of adversarial examples

This section provides a detailed explanation of the hierarchy of adversarial examples in Natural Language Processing, as depicted in Fig. 4.

Attacker's range A wide range of attack techniques have been proposed, including *causative* and *exploratory* attacks, in order to explore potential vulnerabilities in these models. Exploratory attacks, also referred to as test time evasion attacks, involve the generation of adversarial examples with the intention of evading a specific classifier. These malicious testing instances are designed to exploit vulnerabilities in the classifier's decision-making process. On the other hand, causative attacks focus on manipulating the training data to deceive the classifier. By altering the training data, these attacks aim to mislead the classifier during its learning process [17]. The focus of this article is solely on the topic of test time evasion, specifically examining attacks that exploit classification models as a potential security vulnerability.

Attack precision Adversarial assaults can be categorised into *Targeted* and *Non-targeted* attacks based on the intentions of adversaries. In the context of the targeted assault, the adversarial example $\boldsymbol{x}'$ is intentionally classified into a predetermined precise class $\boldsymbol{t}$, which represents the desired target for the adversary [18]. The primary mechanism of this approach is centred around augmenting the confidence score associated with class $\boldsymbol{t}$. In the context of non-targeted attack the primary objective of the adversary is to deceive the model without precisely aiming for a particular desired result. The outcome $\boldsymbol{y}'$ has the potential to belong to any class, with the exception of $\boldsymbol{y}$. In contrast to a targeted assault, a non-targeted attack functions by diminishing the confidence score associated with the right class $\boldsymbol{y}$.

Attacker's expertise Adversarial assaults can occur with expertise of attacker which includes *white-box*, *black-box*, and *grey-box* circumstances [4]. The white-box scenario gives adversaries complete target model access. Adversaries may create ideal adversarial instances by using target model knowledge, such as architectures, parameters, and training data. In the black-box scenario, adversaries cannot determine the target models [19]. Black-box attacks use transferability of hostile instances or recurrent inquiries for optimization. Since attackers can only access model parameters, adversaries in the grey-box only partially comprehend the model. Grey-box assaults assume the target model is available throughout training, unlike the other two kinds.

Perturbation granularity Text-based adversarial assaults can be categorised as *char-level*, *word-level*, *sentence-level*, or *hybrid-level* based on the perturbation granularity used to generate samples. Char-level assaults include altering several characters in words to create instances that deceive detectors. The majority of updates are misspellings and simple operations such as insertion, swap, and deletion. Attacks at the word level disturb words. Attackers create hostile instances by altering words through substitution, replacement, or deletion. Sentence-level assaults insert or rewrite sentences while retaining their meaning [20]. Hybrid-level assaults use many perturbation attacks to obtain high success rates and imperceptibility [4].

2.2.2 Constituents of textual adversarial attack

A schematic representation of generating an adversarial assault within the textual domain is depicted in Fig. 5.

Morris et al. [21] outline four components of the procedure for generating adversarial natural language instances. (1) A *sequential search* technique performs an exhaustive investigation to determine the optimal perturbations [22]. (2) A *transformation module* is used to transform a given text input, denoted $\boldsymbol{x}$, into a modified version, $\boldsymbol{x}'$. Numerous techniques, such as synonym substitutions, random character insertion, etc. are utilized to accomplish this disruption which are imperceptible to human observation. (3) A set of *constraints* is used to exclude unwanted changes to $\boldsymbol{x}'$, ensuring that the modified $\boldsymbol{x}'$ retains the semantics and fluency of the original $\boldsymbol{x}$. (4) The *objective function* is to identify an adversarial example that generates a distinct predicted label from the actual label.

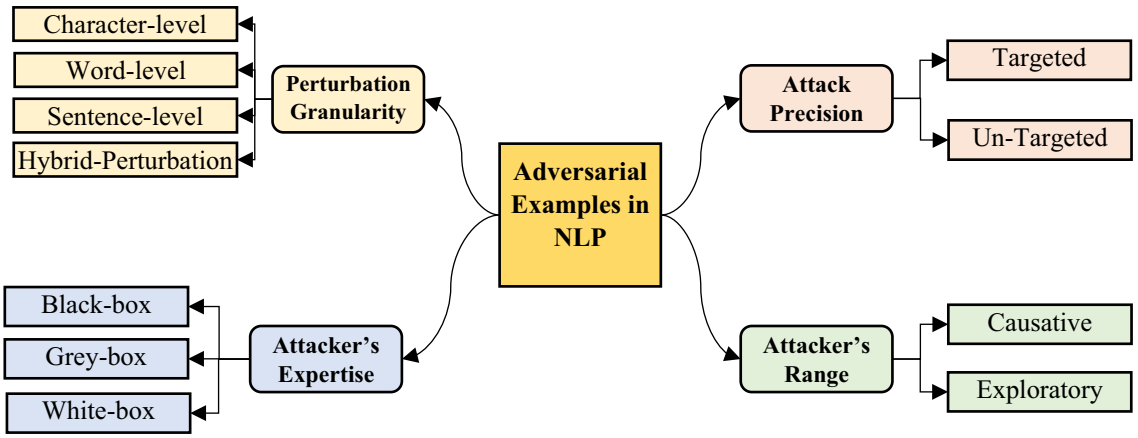


Fig. 4 Taxonomy of adversarial attacks in NLP

Fig. 5 Adversarial example generation

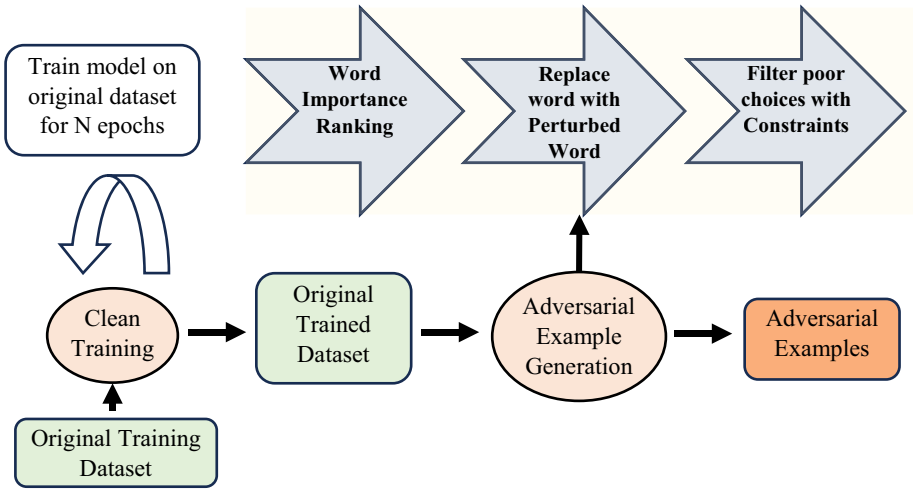


Table 2 complies with the parameters outlined in Sects. 2.2.1 and 2.2.2 by presenting notable textual adversarial assault methods across six disciplines. It comprises of the Attacker's expertise, perturbation granularity, Attack precision, searching technique, transformation used & set of constraints chosen for their attack approach, which demonstrates the uniqueness of every assault methodology outlined in the current body of academic work.

Based on current attack constraints, this study introduces "*Non-Alpha-Num*", a powerful adversarial NLP assault. This attack method creates adversarial instances by altering textual input at the character level. The assault approach aims to deceive a target NLP model when the disruption meets specific language requirements. (e.g., grammatical and semantic similarity constraints). Black-box attacks are more realistic as attackers lack knowledge of trained models or input feature representations. This analysis focuses on a black-box situation. The new strategy for clickbait identification in NLP tasks beats baseline attack techniques by improving attack success rates. The introduced approach will be explained in the subsequent section.

3 Proposed methodology

This section presents a comprehensive analysis of the proposed *'Non-Alpha-Num'* adversarial attack framework. The first segment of this section provides an overview of the problem. The subsequent section provides an overview of the attack's architecture, which comprises four distinct modules. These modules will be further elucidated. Furthermore, as an integral component of the second approach, the present investigation undertook a comparative analysis and evaluation of various neural clickbait classifiers to determine their vulnerability to adversarial perturbations. Additionally, the study aimed to identify the classifier that exhibits greater resistance to such perturbations.

3.1 Problem definition

In the context of text classification, an adversary seeks to create an adversarial document, denoted as $\boldsymbol{x}_{adv}$, from a legitimate document, denoted as $\boldsymbol{x}$, which belongs to the feature space $\mathbf{X}$. The pre-trained text classification model, denoted

Table 2 Comparing standard adversarial attack strategies with the proposed approach

Attack	Transformation	Search method	Constraints	Attack precision	Attacker's expertise	Perturbation granularity	Limitation of the work
TextFooler [23]	Word replaced with counter-fitted GLOVE Word embedding	Greedy—WIR (random)	word embedding similarity, BERT Score, POS consistency	UT	Black-box	Word-level	This type of attack is easily defended against by robust adversarial training that employs lexical alternatives to the words in a text input
TextBugger [24]	Swapping, deleting, substituting, inserting for characters	Greedy—WIR	USE similarity, POS consistency	T, UT	Black-box	Char-level	At times the modifications to the input sequence become visible
PWWS [25]	WordNet-based similar word swapping	Greedy—WIR (saliency)	USE similarity, POS consistency	UT	White-box	Word-level	An adversary must have access to the model's parameters in a white box technique
PSO [26]	How-Net word swapping	PSO	Leven-shtein edit-distance	UT	White-box	Word-level	The search technique utilises a method that requires a significant amount of time to find possible disruptions
Pruthi et al. [27]	Inserting, keyboard-char swapping, deleting, adjacent char swapping	Greedy search	USE similarity, word embedding similarity, POS consistency	T, UT	Black-box	Char-level	By utilising an orthography and grammar checker, it becomes very achievable to prevent any modifications produced to legitimate input

Table 2 (continued)

Attack	Transformation	Search method	Constraints	Attack precision	Attacker's expertise	Perturbation granularity	Limitation of the work
Kuleshov et al. [28]	Word replaced with counter-fitted GLOVE word embedding	Greedy—WIR (random)	USE similarity, word embedding similarity, POS consistency	UT	White-box	Word-level	In order to employ a white box approach, the adversary is required to possess access to the model, the feature set of inputs, and the model parameters
IGA [29]	Word replaced with counter-fitted GLOVE word embedding	Genetic algorithm	BERT Score, POS consistency	T, UT	White-box	Word-level	The modifications applied to the input may often be easily discerned by humans
DWB [30]	Deleting, substituting, inserting, adjacent char swapping	Greedy—WIR (random)	Leven-shtein edit-distance	T, UT	Black-box	Char-level	The adjustments made to genuine input may be readily avoided by employing a spelling and grammar checker
Checklist [31]	Word swap changing location of word, swapping contracting, changing name word swap extending, changing number	Greedy search	Repeat word Modification, POS consistency	UT	Black-box	Word-level	The assault strategy is less successful, and the number of alterations to input sequences is significantly more substantial
BAE [32]	BERT masked token prediction	Greedy—WIR (random)	USE similarity, POS consistency	UT	Black-box	Word-level	Resilience is lacking since word-level perturbations may be easily discovered by utilizing numerous detection procedures

Table 2 (continued)

Attack	Transformation	Search method	Constraints	Attack precision	Attacker's expertise	Perturbation granularity	Limitation of the work
A2T [33]	Word replaced with counter-fitted GLOVE word embedding	Greedy—WIR (gradient-based)	word embedding similarity, BERT Score, POS consistency	UT	White-box	Word-level	Token replacements can be complex and out-of-context at certain times, but are nevertheless readily identifiable by humans
Iyyer et al. [34]	Produce alternative versions of the given text that meet the desired criteria without compromising the overall standard	–	Part of speech checker, grammatical constraints	UT	Black-box	Sentence-level	Paraphrasing the entire input phrase to generate antagonistic input can sometimes impact the semantic meaning
Liang et al. [35]	Modification approaches: transformation deletion and addition	Greedy search	Maximum number of words altered, minimum word length	UT	White-box, black-box	Mixed-perturbations character and word level	The concept of imperceptibility is compromised when the alterations applied to the input sequence occur at both the word and character levels
NonAlphaNum (proposed approach)	Examines the use of punctuation insertions as adversarial assaults on NLP systems	Beam search	Part of speech consistency, minimum cosine similarity	UT	Black-box	Char-level	–

as $\boldsymbol{F}$, is responsible for mapping the features in $\mathbf{X}$ to a set of classes, denoted as $\mathbf{Y}$. The ground truth label of the legitimate document $\boldsymbol{x}$ is represented as $\boldsymbol{y}$, where $\boldsymbol{y}$ belongs to the set of classes $\mathbf{Y}$. The objective of the adversary is to generate an adversarial document $\boldsymbol{x}_{adv}$ such that when it is passed through the text classification model $\boldsymbol{F}$, the predicted label, denoted as $\boldsymbol{y'}$, is different from the ground truth label $\boldsymbol{y}$. In order to ensure utility-preserving alteration, the adversary must satisfy the condition $\boldsymbol{Sim(x, x_{adv}) \geq \varepsilon}$ for a domain-specific similarity function $\boldsymbol{S : X \times X \rightarrow R_+}$, where $\boldsymbol{\varepsilon \in R}$ represents the bound that captures the notion of utility-preserving alteration. In the context of text classification tasks, researchers often utilize $\boldsymbol{Sim}$ as a means to capture the semantic similarity between $\boldsymbol{x}$ and $\boldsymbol{x}_{adv}$.

3.2 Attack design

The process of generating textual adversarial examples is structured as a system consisting of four key components: an objective function, a set of restraints, a modification mechanism, and a searching technique which are discussed in depth in this section [21]. The aim of this system is to search for a perturbation from $\boldsymbol{x}$ to $\boldsymbol{x}_{adv}$ that can deceive a predictive NLP model. This perturbation should be able to achieve a specific objective, such as causing the model to predict an incorrect classification label. Additionally, it must adhere to a predefined set of limitations. The searching technique aims to identify a series of transformations that lead to a successful perturbation. The Proposed adversarial attack architecture is depicted in Fig. 6.

3.2.1 Modification function

The input undergoes a transformation process, resulting in the generation of several potential perturbations. If $\mathbf{X}$ is represented as a vector ($\boldsymbol{x}_1$, …,$\boldsymbol{x_i}$, …, $\boldsymbol{x_n}$), substituting $\boldsymbol{x_i}$ with a modified version of $\boldsymbol{x_i'}$ will lead to a perturbed text. The focus of this work is to examine the substitution of significant words with character-level perturbed terms. The Non-Alpha-Num adversarial attack involves the random insertion of non-alphanumeric characters from a list that contain (!"#$%&'() * + , − . /:; < = > ? @ [\] ^‘ {|}) into the most significant words. Subsequently, the words that have been perturbed are substituted with their original counterparts. The aforementioned selections are made based on their ability to maintain the semantic closeness of the phrase as seen by human observers, despite the fact that neural text classifiers tokenize them in a different manner. Non-Alpha-numeric characters are also referred to as punctuation. Punctuation insertion might be a viable attack vector since grammar checkers struggle to identify punctuation while also not significantly impacting the content of the statement. Deep learning models perform poorly when punctuation is removed because punctuation includes crucial information that models require to function properly. Furthermore, punctuation can include antagonistic downstream information that undesirable users might use.

One crucial point in these perturbations is that words are symbolic entities, and Deep Learning frameworks that depends on learning-based approaches often employ a dictionary to represent a finite collection of potential words. The size of the average word dictionary is considerably less compared to the potential combinations of characters of a comparable length. In the context of English words, it may be seen that the total number of potential combinations is around $26^{\boldsymbol{n}}$, where n signifies the word length. This implies that premeditatedly perturbed important tokens allow for their effortless conversion into "unknown" words, which are not recognized by the dictionary. In deep learning modelling, any words that are not recognized or known will be allocated to the “unknown” embedding vector [24]. The investigation from the study provides compelling evidence that the use of random punctuation insertion is a straightforward method can significantly manipulate the behaviour of text categorization models, leading to erroneous outcomes.

3.2.2 Searching technique

The search mechanism is responsible for identifying the most optimal perturbations based on the modification function. The score is allocated to the optimal collection of altered words. The task entails the utilization of the beam search technique [22]. This algorithm employs a heuristic scoring function, as described in Eq. (1). In this function, for a given text $\boldsymbol{x}$, all potential perturbed texts $\boldsymbol{x'}$ are formed by substituting each word $\boldsymbol{x_i}$, and then scored.

$$Score(\boldsymbol{x'}) = 1 - \boldsymbol{F}_{\boldsymbol{y}}(\boldsymbol{x'}) \tag{1}$$

The function $F_y(\boldsymbol{x})$ represents the projected probability of class $\boldsymbol{y}$ by the model, whereas $\boldsymbol{y}$ represents the actual output of the original text $\boldsymbol{x}$. The highest-ranking $\boldsymbol{b}$ texts are retained, with $\boldsymbol{b}$ being referred to as the “beam width”. The iterative process continues by applying additional perturbations to each of the top $\boldsymbol{b} = 8$ perturbed texts, resulting in the generation of the subsequent set of candidate texts. The computational complexity of this operation is $\boldsymbol{O(b*W^2*T)}$, where $\boldsymbol{W}$ is the number of words in the input. The variable $\boldsymbol{T}$ denotes the maximum number of modification options that are accessible for a given input.

3.2.3 Set of restraints

A collection of linguistic restrictions is utilized to ensure that $\boldsymbol{x}$ and perturbed $\boldsymbol{x'}$ exhibit similarity in terms of both

meaning and fluency, rendering x' a viable prospective adversarial example. The search space should be designed in a way that ensures the proximity of x and x' in the semantic embedding space. We employed the Universal Sentence Encoder (USE) in this work to assess the semantic similarity of textual occurrences by utilizing cosine similarity [36]. The cosine similarity between two n-dimensional vectors, represented as m and n, is defined by Eq. (2). The word embedding vectors e_{x_i} and $e_{x_{i'}}$ must satisfy a specified minimal threshold, in Non-Alpha-Num attack, the value of the threshold is taken as $\tau = 0.6$.

$$\tau = Cosine\ similarity(m, n) = \frac{m \cdot n}{||m| \cdot |n||} = \frac{\sum_{i=1}^{k} m_i \times n_i}{\sqrt{\sum_{i=1}^{k} m_i^2} \times \sqrt{\sum_{i=1}^{k} n_i^2}} \quad (2)$$

The notation $T(x) = x'$ is employed to represent transformations that perturb x to x'. Additionally, it is assumed that the j**-th** constraints are represented as Boolean functions $C_j(x, x')$, which indicate whether x' satisfies the constraint C_j. Next, the search space S can be formally defined using mathematical notation as shown in Eq. (3).

$$Sim(x) = \{T(x) | C_j(x, T(x)) \forall j \in [m]\} \quad (3)$$

The objective of the algorithm for searching is to identify an element x' that belongs to the set $Sim(x)$ and is capable of deceiving the victim model. Additionally, in the set of restraints, the Language Tool [37] is employed to minimize the occurrence of grammatical mistakes while ensuring uniformity in part-of-speech usage. Specifically, the alternative term selected should possess the identical part of speech as the original word. The support taggers offered by SpaCy, NLTK and flair aim to maintain semantic consistency between x and x'.

3.2.4 Objective function

The effectiveness of an assault is assessed by considering the model outcomes and employing an objective function. The search strategy is explored, coupled with modifications and a specific set of restrictions, until it results in the misinterpretation of the true output [21]. The adversarial examples created by the proposed technique are illustrated in Fig. 6. Table 3 presents the algorithm employed in the suggested architecture.

3.3 Comparison of clickbait detector

In order to get a deeper comprehension of the challenges associated with clickbait detection approaches, the experiment has been bifurcated into two distinct phases. The first training of cutting-edge deep learning-based linguistic models is conducted utilizing the clickbait dataset. The baseline textual assaults and Non-Alpha-Num adversarial attacks are employed to alter the trained models in order to misclassify clickbait as non-clickbait and vice versa. Figure 7 illustrates the methodology for conducting an assault on the clickbait detector. The objective of this approach is to ascertain the relative susceptibility of different models to adversarial manipulations. The authors assert that this study represents the inaugural endeavour within the literature to juxtapose clickbait detectors in order to assess their susceptibility to hostile scenarios.

4 Experimental settings

This section provides an introduction to the dataset, victim models, attack tactics, assessment metrics, and execution details. Subsequently, in the subsequent part, we will evaluate the data and examine numerous probable factors contributing to the measured outcome.

4.1 Dataset description

This study aims to examine the impact of malicious textual examples on a widely used benchmark dataset within the domain of clickbait detection. The test set is employed for the production & assessment of the adversarial instances. Table 4 offers a concise summary of the dataset.

Clickbait [38]: The current investigation employs the clickbait dataset in order to conduct a clickbait identification task. The clickbait dataset originates from Chakraborty et al.'s paper [38]. The corpus comprises news titles extracted from a corpus of news articles annotated with two labels, namely: **0**: "*Not-Clickbait*" & **1**: "*Clickbait*". For Non-Clickbait, the headlines were taken from a repository of 18,513 Wikinews articles compiled by NewsReader, whereas for clickbait, the headlines were retrieved from 8,069 online articles of 'BuzzFeed', 'Upworthy', 'ViralNova', 'Scoopwhoop', and 'ViralStories' news portals. The average sample length of the dataset is 8.50 words, while the average length of clickbait headlines is 10 and the average length of non-clickbait headlines is 7.

Figure 8 illustrates the percentage of distribution of both clickbait and non-clickbait headlines, also it depicts the percentage of word contractions, hyperbolic words, determiners, and stop words in both clickbait and non-clickbait headlines. The data has already been pre-processed based on the methodology outlined in their paper, and it is accessible to the public via the Hugging face library. The hugging face datasets library offers an API to facilitate the acquisition of public datasets. There are a total of 16,000 samples in the dataset. In our experimental configuration, we divided the

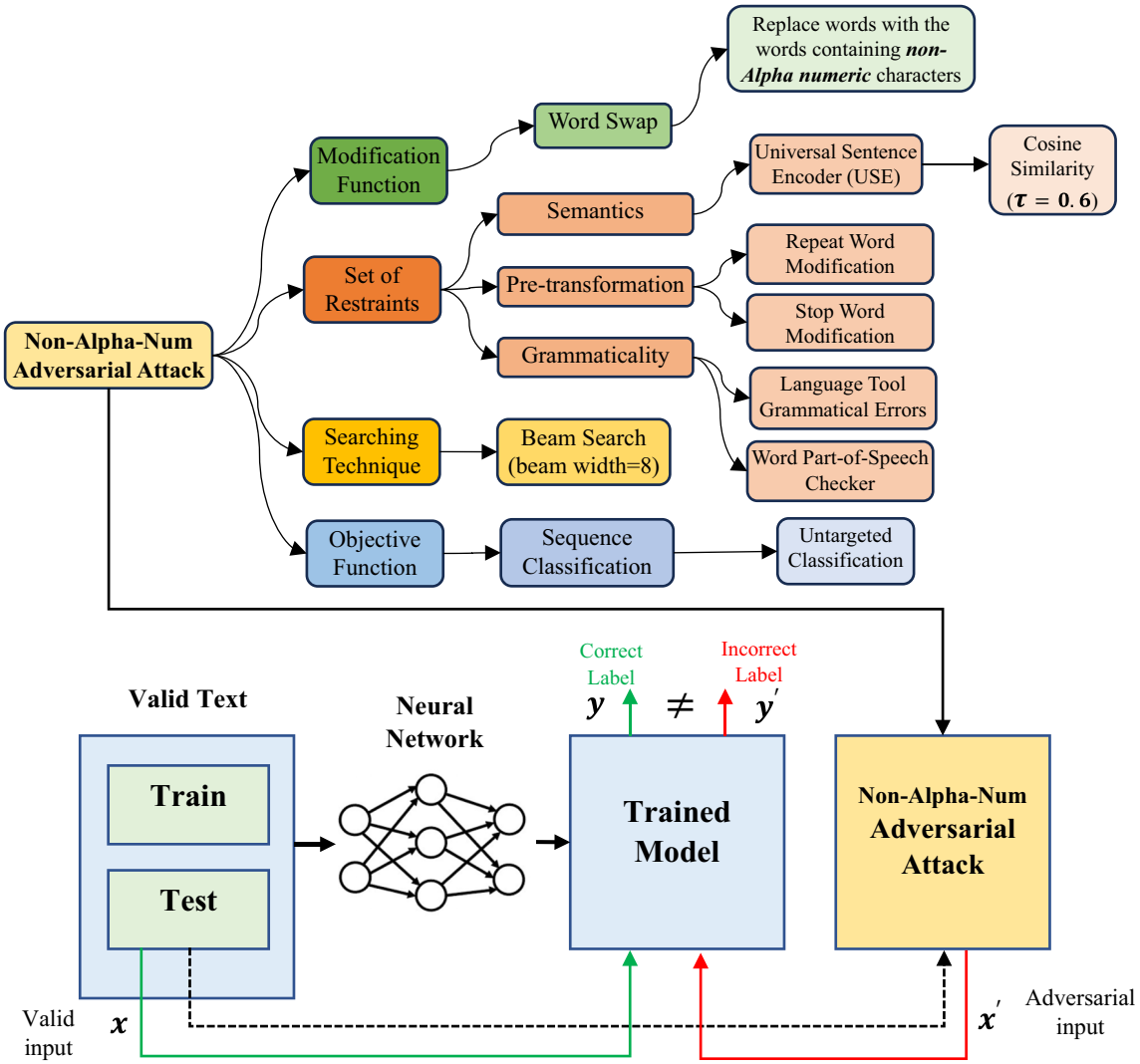


Fig. 6 Proposed adversarial attack architecture

dataset into three, namely 0.80, 0.10, and 0.10, for training, testing, and validation purposes, respectively.

4.2 Victim models

This part encompasses the depiction of the models that underwent training on the clickbait dataset, along with their respective parameter sets. The metric employed for assessing models in the context of binary classification for clickbait detection is the accuracy score. Table 5 presents the comprehensive range & corresponding description for the accuracy formula. The next section contains empirical evidence that showcases the vulnerability of each clickbait model to adversarial situations. This vulnerability is measured by a decrease in accuracy resulting from various adversarial assault strategies.

4.2.1 Description of the models

Deep neural network models demonstrate the capacity to independently acquire knowledge and identify relevant characteristics, resulting in enhanced precision and efficacy. The study employed a variety of well-known deep-learning classification methods, such as basic CNN & various transformer networks. Figure 9 presents a comprehensive representation of the models employed in this study.

Convolutional neural network (CNN) In Convolutional Neural Networks, Layers of convolution extract features through the process of screening the input information, wherein the combination of numerous filters results in the generation of outputs. subsampling or pooling is an approach utilized in CNNs to decrease the granularity of feature maps in various tiers. This process aims to improve the network's robustness against distortions and disturbances. Pooling is a

Table 3 Algorithm of the proposed attack framework

Algorithm 1: Non-Alpha-Num Adversarial Attack

Aim: Adversarial attack framework to fool neural clickbait classifier
Input: legitimate input x and its ground truth label y, classifier $F(.)$, threshold τ, semantic similarity $Sim(.)$.
Output: Generated an adversarial sequence x'

1. Initialization: $x' \Leftarrow x$
2. **for** x_i in X **do**
3. Compute score (x_i') according to **Equation (1)** — Searching Technique
4. **end for**
5. $W_{Ranked} \Leftarrow Ranking(x_1, x_2, x_3 \ldots\ldots x_n)$ in descending order
6. Remove the stop words in W_{Ranked}
7. **for** x_i in W_{Ranked} **do**
8. $x' \Leftarrow$ *Substituting x with (words with Punctuations)* — Modification
9. **if** $Sim(x, x') <= \tau$ **then** — Restraints
10. **return None**:
11. **else if** $F(x') \neq y_{true}$ **then** — Objective Function
12. Solution found **return** x'.
13. **end if**
14. **end for**
15. **return None**

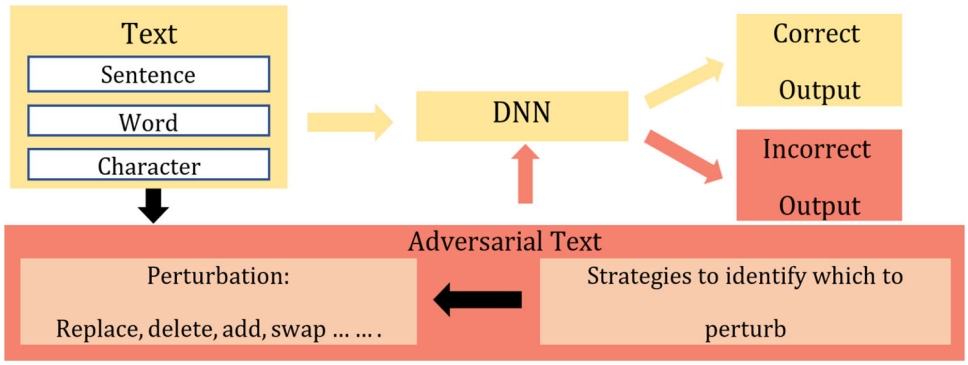


Fig. 7 Procedure of generating adversarial text

Table 4 Concise description of the dataset

Task	Application domain	Granularity	Classification	Dataset	Labels	Train	Validation	Test	Average length
Clickbait detection	News media	News headlines	Binary	Clickbait dataset	2	12.8 K	1.6 K	1.6 K	08.50

technique used to decrease the dimensionality of the output produced by a certain layer in order to pass it on to the succeeding layer. The execution of categorization operations is carried out by the use of completely linked layers [39]. Convolutional Neural Networks (CNNs) have a notable capability in effectively identifying local patterns as well as patterns that are invariant to changes in position. The utilization of Convolutional Neural Networks (CNNs) has been found to be highly beneficial in the domain of text classification.

Transformer models Deep neural networks that employ transformers incorporate a self-attention mechanism to allocate various levels of importance to distinct regions of the input data. Table 6 displays a comprehensive descriptive study of many pre-trained classifiers.

4.2.2 Parameter configurations

Table 7 displays the settings of the parameters for every framework that underwent training on the clickbait dataset,

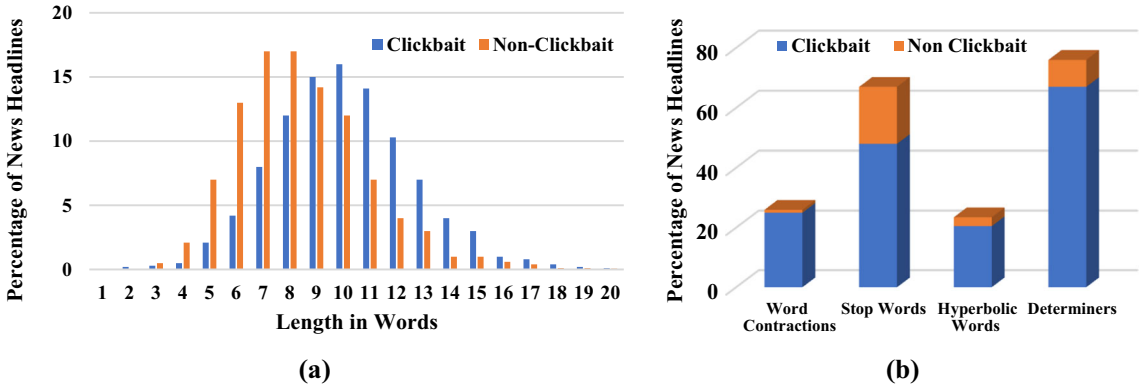


Fig. 8 **a** Percentage distribution of clickbait & non-clickbait news headlines, **b** Percentage distribution of word contractions, hyperbolic words, determiners, and stop words in News titles

Table 5 Classification metric for evaluating clickbait detectors

Metric	Formulae	Range
Accuracy (ACC)	$\frac{TP+TN}{FN+FP+TN+TP}$	[0,1]

aiming to evaluate the effectiveness of adversarial attack approaches.

Table 8 displays the testing scores for the accuracy of all of the models that completed training using the clickbait dataset.

4.3 Baseline adversarial attacks

The assault strategies were employed on the testing samples of the clickbait dataset in order to produce adversarial examples. The adversarial instances are subsequently employed to incorrectly identify the authentic output of the input sentence when assigned to the pre-trained algorithms. The scope of the testing was restricted to attack approaches that have been published in reputable conferences & publications in the domains of AI and NLP. The aforementioned papers encompass a range of esteemed publications which includes ICLR, ACL, AAAI, EMNLP, IJCAI, NAACL, COLING TACL, TKDE & JMLR.

The following portion provides a succinct summary of the baselines that have been chosen to demonstrate the efficacy of the proposed approach in deceiving clickbait detection algorithms.

- A2T [33]: Yoo and Qi utilized a gradient-based method to ascertain the level of significance of words. Subsequently, a process was undertaken wherein every single word was sequentially replaced with synonyms obtained from counterfeit word embeddings.
- BAE [45]: Garg and Ramakrishnan generate samples that are adversarial with the purpose of misleading sentiment classification. The researchers proposed utilizing the BERT model for the purpose of inserting and replacing tokens inside a phrase, taking into consideration their individual levels of importance. The substitution possibilities generated by the model occasionally display antonymous or unrelated associations with the original terms, leading to a decrease in semantic consistency.
- Checklist [31]: Drawing upon the foundational principles of behavioral testing. The use of perturbations in terminology, numerical values, and places, together with contractions and expansions of the significant elements inside the sentence.

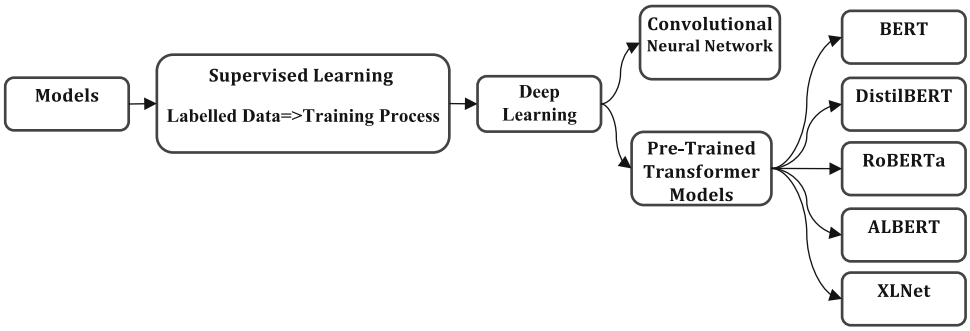


Fig. 9 Synopsis of the models employed in the investigation

Table 6 A thorough examination of the transformer variants

Model	Trained on datasets	Tuning	Description	Advantages	challenges
BERT [40]	English wikipedia & the BookCorpus	Fine-tuned on target dataset, pre-trained on specific parameters	The model under consideration is a bidirectional transformer architecture that incorporates both Masked Language Modelling (MLM) methods and Next Sentence Prediction (NSP)	1. The ability to efficiently handle and analyse context-specific data 2. Training at a Faster Pace	1. The process of categorizing is limited to a certain vocabulary 2. The total length of the input phrases is predetermined and remains constant 3. Exhibits issues pertaining to reasoning that is logical 4. The processing expense is substantial
DistilBERT [41]	English wikipedia & the BookCorpus	Fine-tuned on target dataset, pre-trained on specific parameters	The approach adopted in this study utilizes an early iteration of BERT, with a reduction in the total amount of tiers by a factor of 2. Furthermore, the implementation of adaptive masking has been carried out	1. The implementation of preliminary training has been employed to enhance the effectiveness of linguistic simulation proficiency 2. When comparing the freshly designed model to BERT, it demonstrates improved acceleration and decreased weight	1. The set-length restriction is a limitation imposed on a system or process that requires a specific length to be adhered to. This constraint ensures that the system or process operates within
ALBERT [42]	English wikipedia & the BookCorpus	Fine-tuned on target dataset, pre-trained on specific parameters	A modified version of BERT has undergone feature reduction, leading to a more lightweight version. The diminution of factors is easily accomplished by employing factorized embedding layer parameterization & cross-layer parameter allocation	1. Reduced memory utilization 2. One potential area of improvement for BERT is the enhancement of its training pace	1. Framework exhibits incompatibility for problems that involve the production of textual content 2. The text demonstrates deficiencies in logical reasoning

Table 6 (continued)

Model	Trained on datasets	Tuning	Description	Advantages	challenges
RoBERTa [43]	English wikipedia & the BookCorpus, CC-news, stories, OpenWebText	Fine-tuned on target dataset, pre-trained on specific parameters	The present study aims to replicate the BERT model by employing an enlarged training dataset and fine-tuning the hyper-parameters. Additionally, the utilization of dynamic masking is incorporated in this replication effort	1. The utilization of a larger volume of preliminary training data has been shown to enhance effectiveness 2. The performance of the model surpasses that of both XLNet and BERT	1. The attribute requires a significant number of resources 2. The current undertaking necessitates substantial computer resources and involves a lengthy processing time
XLNet [44]	English wikipedia & the BookCorpus, Giga5, ClueWeb 2012-B, Common Crawl	Fine-tuned on target dataset, pre-trained on specific parameters	It blends auto-regressive models with bi-directional context modelling to overcome BERT's drawbacks and surpass BERT on 20 tasks, frequently by a wide margin in sentiment analysis, question answering, document rating, natural language inference	Using greater amounts of initial training data improves efficacy	The absence of fictitious symbols, such as [MASK], utilized by BERT during the pretraining phase, in actual data during the finetuning phase leads to a disparity between the training and finetuning processes

- HotFlip [46]: The method involves the replacement of one token with another, using the gradients of the one-hot input vectors. The disrupted input exhibits imperceptibility to human inspection while inducing a reversal in the output, leading to inaccurate predictions.
- DeepWordBug [30]: The assault methodology described in this recipe involves the combination of four distinct forms of character disruptions, namely insertion, deletion, modification, and swapping. The authors utilize the concept of edit distance to assess the degree of resemblance between the unaltered and edited samples. This study introduces two unique scoring functions, specifically temporal and temporal tail scores, to evaluate the relevance of each word in the final model prediction. Following this, the words are sorted based on their individual scores, and a perturbation approach is then employed on the words of high relevance.
- Pruthi [27]: The authors proposed the use of a blend of character-level perturbations, such as swapping nearby characters, eliminating characters, exchanging a character with its neighboring keys on the QWERTY keyboard, and introducing new characters, as a means to produce adversarial cases.
- PWWS [25]: Ren et al. suggested defining the order of substitution candidates based on not just the classification probability but also word saliency. This enabled modeling not just the quantitative but also the qualitative aspect of the change produced by the substituted word. The substitution candidates are derived using WordNet
- PSO [26]: In their study, Zang et al. introduced a word-level adversarial assault strategy that relies on sememe substitution. The authors proposed that the challenge of choosing the best appropriate adversarial example from a set of sememe-based candidates might be seen as a multivariate combinatorial optimization problem.
- TextBugger [24]: Li et al. introduced a series of composite transformations, which encompass character deletion, swapping, substitution, and synonym substitution. In the context of the white-box setup, the identification of the most crucial words in a given sample was achieved by utilizing the Jacobian matrix of the victim classifier.
- TextFooler [23]: An attack technique with the objective of replacing the most crucial words in a given sample with a set of N synonyms. The process of selecting synonym candidates was determined by calculating the cosine similarity

Table 7 Configuration of parameters for the intended classifiers

Models	Parameter configurations
Word-CNN	For the aim of this investigation, the CNN model developed by Kim et al. [39] was used. The Word-CNN model utilizes a total of 100 filters and incorporates three different window sizes, namely 3, 4, and 5. The selected dropout rate of the framework is 0.3, and it utilizes a basis of 200-dimensional GLoVE embeddings. Afterward, a fully linked layer is utilized, which is then followed by a time-dependent max-pooling layer in order to facilitate the process of categorization. The algorithm used has achieved a test set for accuracy of 89.47% on the clickbait dataset
BERT	The "bert-base-uncased" model is subjected to a training process consisting of 10 iterations. Each iteration involves a batch size of 64, a learning rate of 2e-05, and a maximum sequence length of 128. The objective of this training is to enhance the model's performance in sequence classification specifically for the clickbait dataset. The framework underwent training by employing a cross-entropy loss mechanism. The framework attained its maximum efficacy in this assignment, as determined by the accuracy of the test set, which reached 91.55% after eight epochs
DistilBERT	The "distilbert-base-uncased" network was trained for 10 epochs using a batch size of 64, an average rate of learning of 2e-05, and an optimal sequence length of 128. This optimization was performed specifically for sequence classification on the clickbait dataset. The cross-entropy loss function was employed during the training of the model. The accuracy of the evaluation set, determined after 5 epochs, revealed that the model achieved a maximum accuracy of 90.83% on this particular job
ALBERT	We enhanced the performance of the "albert-base-v2" model for text categorization on the clickbait dataset. The model was trained for 10 epochs using a batch size of 64, a learning rate of 2e-05, and a maximum sequence length of 128 bits. The model was trained using a cross-entropy loss function. The greatest accuracy score achieved by the model in this task, as evaluated by the test set accuracy, was 91.72%
RoBERTa	The performance of the "Roberta-base" model has demonstrated enhancement in sequence classification when applied to the specific dataset utilized in our experiment. The model was executed using a maximum sequence length of 128 and a batch size of 64 for a total of 10 epochs, with a learning rate of 2e-05. The model was trained using a cross-entropy loss function, as it was a problem with categorization. The assessment of the model's performance on this particular task resulted in the highest score of 92.12% for accuracy, which was attained after 8 epochs
XLNet	The efficiency of the "xlnet-base-cased" classifier on the clickbait dataset was improved. The training process involved training the model for a total of 10 epochs. During each epoch, a batch size of 64 was used to process the data. The learning rate, which determines the step size at each iteration of the training process, was set at 2e-05. Additionally, a maximum sequence that measured 128 bits was specified to handle the input data. The training of the model was conducted with a cross-entropy loss function. The model earned a maximum accuracy score of 91.96% in this task, as determined by evaluating the test set accuracy

between word embeddings. Before finishing adversarial examples, a limitation based on part-of-speech was also enforced.

4.4 Attack assessment criteria

The effectiveness of strategies for attack has been empirically demonstrated by employing three evaluation parameters: Post Attack Accuracy, Attack Success Rate, and Average Perturbed Rate. The approach employed to evaluate and clarify the three indicators is outlined below.

- *Post attack accuracy (PAA)* The primary objective of adversarial assaults is to undermine the effectiveness of the classifiers. Performance measures, such as accuracy, are commonly employed for the evaluation of classification work. The accuracy scores have been presented both before and after the attack. The utilization of potent adversarial attacks has been found to result in a notable reduction in accuracy scores due to the efficacy of their attack techniques.
- *Attack success rate (ASR)* To analyze the effectiveness of the attack techniques, a random sample of five hundred correctly classified instances is chosen from the test set. This ensures that the evaluation is not influenced by the classifiers' classification accuracy. The source texts are subsequently subjected to attack algorithms in order to produce adversarial examples. Subsequently, the adversarial instances are forwarded to neural clickbait classifiers in order to provide the ultimate prediction. The success rate of the attack algorithm is determined by utilizing the percentage of erroneous predictions generated by these classifiers. A higher success rate indicates that the assault algorithm possesses the ability to generate more powerful adversaries, perhaps leading to the malfunctioning of

Table 8 Testing accuracy of the targeted models

	Word-CNN	BERT	DistilBERT	ALBERT	RoBERTa	XLNet
ACC	89.47%	91.55%	90.83%	91.72%	92.12%	91.96%

these clickbait classifiers. We use attack success rate ASR (ratio of successful attack samples to the total of successful and failed samples $\frac{(successful\ samples)}{successful+failed\ samples}$) to determine the effectiveness of an attack technique against the victim classifier. The successful samples refer to those that can misclassify the true prediction, whereas failed samples are unable to incorrectly categorize the genuine outcome. In an analytical context, an assault is considered effective when the algorithm $\boldsymbol{F}$ accurately classifies the initial legal input $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{y}$, but predicts incorrectly the attacked input $\boldsymbol{F}(\boldsymbol{x} + \boldsymbol{\Delta x}) = \boldsymbol{y}'$. Therefore, the ASR may be mathematically represented as depicted in Eq. (4).

$$\text{ASR} = \frac{\boldsymbol{F}(\boldsymbol{x} + \boldsymbol{\Delta x}) = \boldsymbol{y}'}{(\boldsymbol{F}(\boldsymbol{x} + \boldsymbol{\Delta x}) = \boldsymbol{y}') + (\boldsymbol{F}(\boldsymbol{x} + \boldsymbol{\Delta x}) = \boldsymbol{y})} \tag{4}$$

In the setting of untargeted attacks, the symbol $\boldsymbol{y}'$ represents any label that differs from $\boldsymbol{y}$. The sign $\boldsymbol{\Delta x}$ is used to represent modifications made to the initial specimen. In the present context, a successful assault is defined as the ability of an adversarial sample to make inaccurate predictions with a significantly high level of confidence. In the event of an unsuccessful assault, the adversarial sample lacks the ability to cause misclassification of the true prediction. The missing statements refer to those that the model wrongly categorized throughout the training process. Our research focuses on analysing the success rates of assaults and evaluating the effectiveness of those assaults in incorrectly identifying outcomes.

- *Average perturbed rate (APR)* The average perturbation rate (APR) is another core metric employed to assess an adversarial text attack. It is defined as the ratio of the number of perturbed words to the text length as shown in Eq. (5). Less perturbation implies retention of higher levels of semantic similarity.

$$\text{APR} = \frac{Number\ of\ words\ pertured}{Total\ number\ of\ words\ in\ input\ sequence} \tag{5}$$

5 Experimental results and analysis

To comprehend the susceptibility of clickbait classifiers. The initial stage of the investigation involves employing the offered clickbait dataset to train sophisticated deep-learning models. The settings for the parameters of each model are shown in Sect. 4.2. The models that have been trained can potentially be manipulated by employing the *Non-Alpha-Num* adversarial attack strategy. After subjecting the test samples to perturbation using the technique described, the resulting drop in accuracy scores is presented in Table 9. The first measurement and recording of the accuracy of the intended models on the original test specimens is referred to as the Before-Attack Accuracy (BAA). Following this, the effectiveness of the target models is assessed by exposing them to adversarial samples created using the offered attack technique. Post-attack accuracy (PAA) refers to the score of accuracy obtained after the proposed attack has been executed. In addition, the study presents findings about the proportion of altered words in relation to the initial sentence length, denoted as Perturbation Rate (PR).

In order to assess the efficacy of the suggested attack technique in comparison to traditional attack methods against clickbait classifiers, the (ASR) metric is employed. For this, a set of 500 samples, which were accurately classified were taken from the test set. Following this, the generation of adversarial cases is achieved by the utilization of different attack methods. The current investigation entailed exposing a collection of adversarial instances to six state-of-the-art clickbait detectors. In this study, the performance of several adversarial approaches was evaluated and compared with the proposed attack, with the metric of ASR being used. This statistic can demonstrate the level of effectiveness of the assault approach. A greater ASR value signifies that a particular assault type has a higher level of effectiveness in misleading the model. Table 10 provides a concise overview of the main results obtained from the Non-Alpha-Num attack method when applied to the clickbait dataset. Additionally, it includes a comparative analysis of the effectiveness of this attack strategy with earlier attack methods.

The Non-Alpha-Num technique exhibited a notable impact on a restricted set of words, resulting in a significantly elevated rate of success in attacks. The aforementioned technique demonstrates superior performance compared to the benchmark perturbation techniques throughout all of the models. The manipulation of a restricted set of words within the data yielded a success rate of 95.8% when tested against a Word-CNN method, with an alteration rate of just under 11%. In contrast, none of the baselines were able to exceed this level of success. The mean length of the clickbait dataset sample ranges from 8 to 9 words. To execute effective assaults, the Non-Alpha-Num adversarial attack method

Table 9 Comparison of the accuracy of each model before and after the proposed adversarial attack algorithm is conducted

	Word-CNN (%)	BERT (%)	DistlBERT (%)	ALBERT (%)	RoBERTa (%)	XLNet (%)
BAA	89.47	91.55	90.83	91.72	92.12	91.96
PAA	08.48	19.30	13.70	15.80	24.80	21.90
PR	12.52	13.81	14.07	14.75	12.09	15.44

BAA Before attack accuracy, *PAA* Post attack accuracy, *PR* Perturbed rate

Table 10 Attack outcomes on different models

Attacks	Word-CNN		BERT		DistilBERT		ALBERT		RoBERTa		XLNet	
	ASR%	APR%	ASR%	APR%	ASR%	APR%	ASR%	APR%	ASR%	APR%	ASR%	APR%
A2T [33]	15.8	14.7	12.3	16.3	13.8	15.5	14.7	14.4	07.2	13.6	09.2	06.7
BAE [45]	64.4	18.6	56.5	20.2	58.4	14.6	56.2	11.8	49.6	14.9	51.7	10.6
Checklist [31]	16.8	13.4	13.4	19.7	18.8	20.9	13.8	18.1	09.2	17.1	19.9	08.7
DWB [30]	68.7	11.8	53.9	19.2	59.6	10.3	55.2	13.1	48.5	11.4	50.6	12.8
HotFlip [46]	76.6	14.9	68.8	11.6	70.6	13.7	69.1	14.1	60.2	12.1	63.7	11.1
IGA [29]	69.4	15.2	59.0	12.9	63.7	15.5	60.1	14.9	50.6	13.9	54.4	12.8
InputReduction [47]	34.9	11.0	21.6	10.8	29.1	13.8	25.4	11.6	15.3	14.7	19.2	13.6
Kuleshov et al. [28]	76.5	17.5	65.4	18.3	69.9	12.6	68.4	13.7	49.6	14.5	55.5	14.2
Pruthi et al. [27]	32.8	08.7	21.3	09.2	28.4	08.2	26.6	09.9	19.7	09.2	26.6	16.3
PSO [26]	75.3	16.8	63.9	14.7	69.7	13.2	63.0	15.8	52.9	15.1	56.7	13.6
PWWS [25]	55.8	18.2	47.1	17.4	55.8	11.7	50.6	16.5	42.5	15.6	45.8	14.3
Textbugger [24]	95.6	13.4	90.1	12.8	93.8	13.6	90.6	13.8	86.2	14.8	89.6	13.8
TextFooler [23]	72.7	15.9	65.3	13.4	68.9	14.7	66.7	12.6	59.7	13.7	61.6	14.5
NonAlphaNum	**95.8**	10.6	**91.6**	12.4	**94.0**	11.1	**92.2**	11.2	**86.5**	12.6	**91.4**	12.5

The findings of our proposed methodology are displayed in bold
ASR Attack success rate, *APR* Average perturbed rate

disturbed a mere percentage, namely less than 14%, of the words inside a singular sample. Research investigations have indicated that transformer representations can be deceived by the proposed attack approach, as long as the interference ratio remains under 20%. The attack approach under consideration introduced little perturbations, affecting just a small fraction of words, typically about 2–3, inside a single sample. The deception of transformer models through attacks has been seen, as long as the tampering rate stays under 20%. The ALBERT model has the highest (ASR) of 92.2% when compared to alternative attack strategies on the given dataset. The assault mechanism utilized in this work effectively deceives the DistilBERT model's predictions to a minimal extent, accomplished by introducing an average word perturbation rate of just 13.10%. Whilst being widely regarded as the leading model for a range of tasks relating to natural language processing, BERT, a sophisticated model with 110 M parameters, is vulnerable to Non-Alpha-Num adversarial attack. The results suggest that by maintaining a perturbance rate below 14.4%, the suggested methodology may successfully get an ASR of 91.6% on the clickbait dataset. Moreover, in the case of RoBERTa and XLNet, the suggested methodology demonstrates superior performance compared to previous state-of-the-art attack methodologies. This observation suggests that the suggested attack system possesses the ability to manipulate classifiers in order to provide inaccurate predictions.

In addition to the aforementioned decline in accuracy scores, we propose the concept of attack Success Rates (ASR) as a method for evaluating the efficacy of each assault, as outlined in Sect. 4.4. The assessment technique measures the proportion of successful attack samples to the total number of successful and failed samples in order to ascertain the efficacy of an attack technique against the victim classifier.

Successful samples have the potential to modify the accurate forecast, whereas failed samples do not, as outlined in Sect. 4.4. Furthermore, the Average perturbed Rates (APR) are provided, whereby the calculation entails dividing the number of disturbed words by the whole length of the text.

The application of these 2 criteria enables the assessment of different adversarial assault strategies across many models. The objective of this research is to illustrate the relative level of danger that certain attack methods pose to distinct models. Furthermore, the employment of average ASR scores assists in illustrating the classifier that has the most susceptibility to adversarial manipulations. The work also aims to analyse the vulnerability of various clickbait classifiers to various forms of adversarial interference. This study aims to assess the relative susceptibility & robustness of different models when subjected to adversarial perturbations. The success rate of each assault is assessed on each targeted paradigm in order to ascertain the relative vulnerability of each of the models as shown in Fig. 10. The mean attack success rate for each model is calculated using Eq. (6).

$$\boldsymbol{Success_{avg}} = \frac{\sum_{i=1}^{n} \frac{\boldsymbol{Successful_i}}{\boldsymbol{Successful_i + Failed_i}}}{\boldsymbol{n_{attacks}}} \tag{6}$$

The Attack Success Rate will be denoted as $\boldsymbol{Success_{avg}}$, where $\boldsymbol{Successful_i}$ represents the number of successful attacks, $\boldsymbol{Failed_i}$ represents the number of unsuccessful assaults, and $\boldsymbol{n_{attacks}}$ represents the number of attack recipes. The skipped sentences are the assertions that the machine originally mis predicted during its training. They were excluded from the computation.

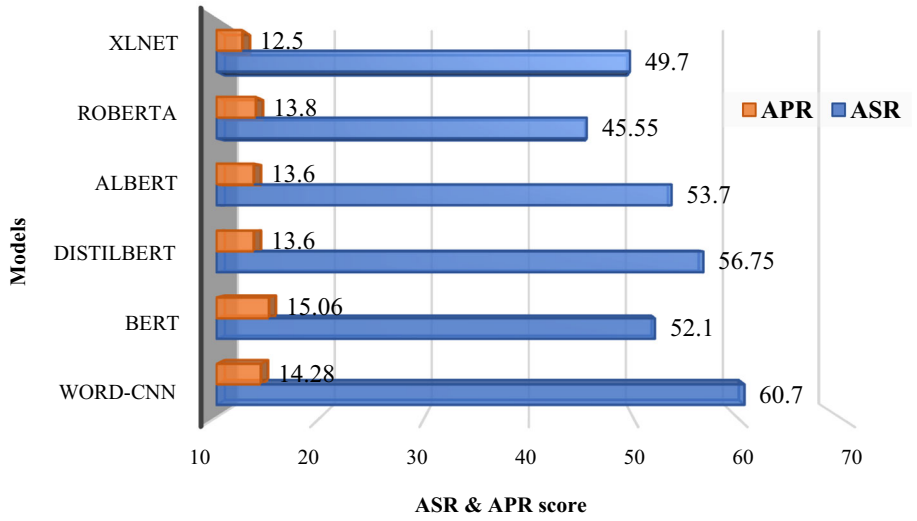
As seen in Fig. 10, subsequent examination has revealed that the Word-CNN model demonstrates the greatest vulnerability to adversarial assaults. Among transformer models, it has been observed that the RoBERTa model exhibits the least vulnerability, while the DistilBERT model is found to be the most susceptible. This observation leads to the conclusion that lighter models, such as DistilBERT [41] and ALBERT models are more prone to these attacks. This vulnerability can be attributed to the fact that the ALBERT [42] model possesses an architecture, consisting of 128 embedding layers, 768 hidden layers, and with only 12 million parameters, whereas the DistilBERT model is also a condensed variant of the BERT model. In the preliminary training stage, the BERT model experienced a process known as knowledge distillation, which led to a decrease in its overall size by 40%. Significantly, the aforementioned reduction in size was accomplished while maintaining 97% of the model's language understanding skills. However, it should be noted that the BERT base model consists of 110 million parameters, making it a computationally intensive model with a significant level of complexity. In contrast, the RoBERTa model underwent training using a significantly larger dataset and implemented a more optimized training methodology [43]. The RoBERTa model was trained using a dataset including 160GB of textual data, which is more than 10 times larger than the dataset used to train BERT. Furthermore, RoBERTa employs a dynamic masking strategy throughout its training process, which enhances the model's ability to acquire more resilient and versatile word representations. This characteristic renders RoBERTa the least susceptible among all the models. The results might be important for those who frequently utilize well-established, cutting-edge algorithms in their efforts to spot clickbait. The reader will possess the ability to determine the most appropriate model that corresponds to their particular problem. Moreover, this phenomenon acts as a motivating factor for academics to develop models that exhibit adversarial robust generalizations rather than traditional generalizations. To ascertain the comparative efficacy of assault approaches in deceiving the model with a reduced average perturbance rate. The average rate of success and alteration rate for each assault type across all models have been computed and are presented in Fig. 11. Next, we proceed to assign rankings to different assault techniques, as presented in Table 12.

The results depicted in Fig. 11 clearly demonstrate that the Non-Alpha-Num attack algorithm, which operates at the character level, exhibits the highest level of effectiveness in terms of perturbation. Following closely behind is TextBugger, another character-level perturbation attack. On the other hand, Hot-Flip emerges as the most effective word-level perturbation technique. Conversely, the attack method A2T, which employs gradient-based synonym word swap within the white-box adversarial setting, is found to be the least effective across all models, as indicated by a comprehensive evaluation of assault tactics.

The most effective attacks shown in Table 11 are char-level perturbation assaults, specifically the Non-Alpha-Num suggested attack technique and TextBugger, which is a prominent conventional attack tactic. While the majority of attacks operate at the granularity of word-level perturbations, it is worth noting that the top two assaults specifically target character-level perturbations.

6 Further investigation and analysis

A further investigation is undertaken to assess the efficacy of the Non-Alpha-Num assault methodology across several circumstances, encompassing its entire execution duration in generating an adversarial instance. The evaluation also includes an assessment of the transferability property of adversarial instances created by the proposed technique. To assess the property of transferability of our suggested approach across different classifiers, we utilize the attack success rate (ASR). ASR is calculated as the ratio of successful



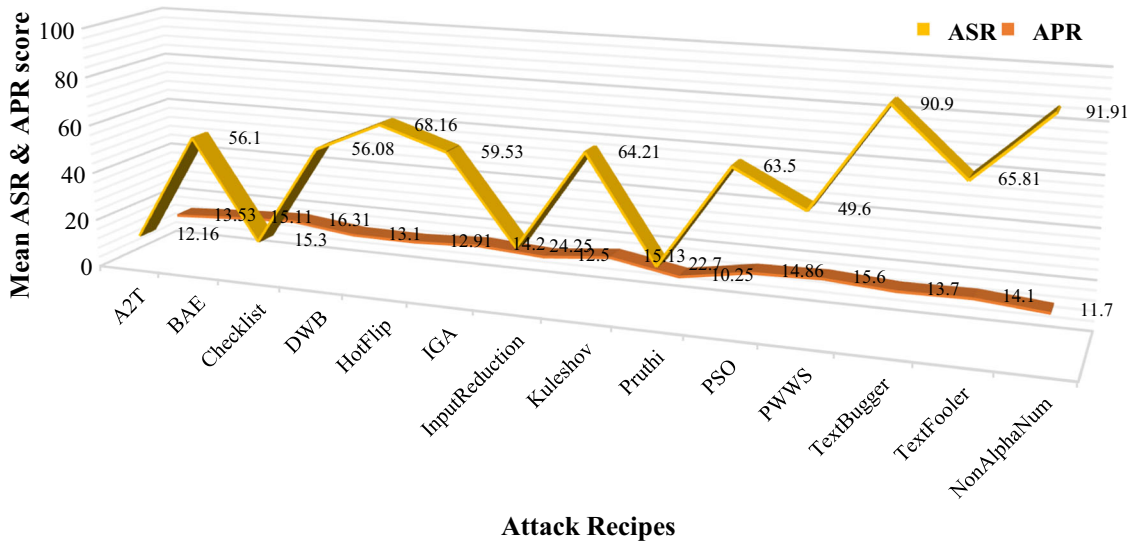


Fig. 10 Average ASR and APR scores of all classifiers

Fig. 11 Mean ASR & APR score of each attack type on all models

attack samples to the total of successful and failed samples, i.e., successful samples divided by the sum of successful and failed samples. The successful samples refer to those that can misclassify the true prediction, whereas failed samples are unable to incorrectly categorize the genuine outcome. The metric is explained in detail in Sect. 4.4. Moreover, the utility evaluation of the adversarial words generated by Non-Alpha-Num demonstrates a certain level of resemblance to the genuine text.

6.1 Execution time

A study was conducted to evaluate the computing time implications of the proposed framework. The fundamental aim of an attacker is to manipulate a model by implementing the intended attack. Figure 12 displays the probable outcomes of the average runtime required to produce a solitary adversarial example utilizing the Non-Alpha-Num assault architecture for each specific model. Analysing the research findings, it is evident from the Figure that BERT requires the longest duration in developing an adversarial example. Conversely, lighter models like ALBERT and DistilBERT exhibit less time. Furthermore, the non-transformer-based model, namely Word-CNN, demonstrates the least amount of time needed for producing an adversarial sequence.

6.2 Accountability and applicability

The attack method presented in this study generates hostile texts that exhibit a greater degree of resemblance to the source texts. Based on the findings shown in Fig. 13,

Table 11 The average ASR for each assault recipe on each classifier

Attack methodology	Mean ASR%	Mean APR%	Perturbation level
NonAlphaNum	91.91	11.7	Char-level
Textbugger [24]	90.95	13.7	Char-level
HotFlip [46]	68.16	12.91	Word-level
TextFooler [23]	65.81	14.1	Word-level
Kuleshov et al. [28]	64.21	15.13	Word-level
PSO [26]	63.5	14.86	Word-level
IGA [29]	59.53	14.2	Word-level
BAE [45]	56.10	15.11	Word-level
DWB [30]	56.08	13.1	Char-level
PWWS [25]	49.6	15.6	Word-level
Input Reduction [47]	24.25	12.5	Word-level
Pruthi et al. [27]	22.7	10.25	Char-level
Checklist [31]	15.3	16.31	Char-level
A2T [33]	12.16	13.53	Word-level

it can be inferred that the viability preservation of adversarial instances created by Non-Alpha-Num is comparatively higher in terms of its accountability and applicability. Char-level attacks refer to a range of language blunders, such as misspellings, transpositions, and random character swaps [24]. Both human individuals and spell-checking algorithms possess the ability to easily identify and recognize assaults that occur at the level of individual characters. Considering the proposed solution, the adversary can create changes through the insertion of punctuation marks that cannot be concealed by detection techniques. These transformations exhibit improved levels of imperceptibility and legibility.

6.3 Adversarial transferability

The present study aimed to investigate the property of adversarial transferability in the text by assessing the effectiveness of adversarial instances developed by one classifier in deceiving a various classifier [48, 49]. This research study involved the collection of 100 adversarial examples generated through the utilization of Non-Alpha-Num Adversarial Attack. These examples were specifically chosen as they were incorrectly classified by a designated target model. The ASR scores of these examples were subsequently evaluated against alternative target models. The results displayed in Table 12.

Table 12 demonstrate that the Word-CNN model exhibits a moderate level of transferability. In contrast, the transferability of the transformer models is comparatively lower. According to the investigation, the BERT model has been found to demonstrate superior transferability compared to other transformer models.

6.4 Explainability and interpretability

The LIME approach, as introduced by Ribeiro et al. [50], is utilized to provide localized interpretations regarding our algorithms. The LIME methodology employs a linear framework to estimate the local decision boundary for each example by fitting it to the associated data. The example was perturbed in order to get the acquired information. To evaluate the accuracy of the regional interpretations obtained from LIME, the area over perturbation curve (AOPC) is utilized as a quantitative measure. The mathematical expressions for this metric are provided in Eq. (7).

$$\mathrm{AOPC} = \frac{1}{M+1}\sum_{m=1}^{M}\frac{1}{N}\sum_{1=1}^{N}\left(X_{(0)}^{(j)}\right) - F\left(X_{(m)}^{(j)}\right) \tag{7}$$

Within the framework of this research, the notation $\boldsymbol{X}_{(0)}^{(\boldsymbol{i})}$ refers to a specific occurrence where $\boldsymbol{X}^{(j)}$ is denoted as

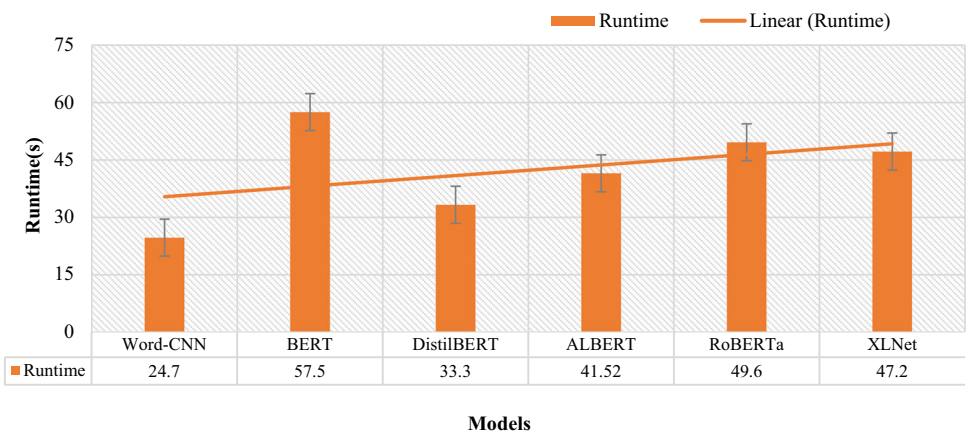


Fig. 12 Mean time to produce an adversarial sample for each model

(0) without any words being removed. On the other hand, $X^{(j)}_{(m)}$ represents an instance where the m most significant words are excluded and denoted as (m). The function $F(X)$ is employed to represent the degree of confidence of the model in relation to the predicted target label $Y^{(j)}$. The function $\mathbf{F}(\mathbf{X})$ is employed to represent the degree of confidence of the model in relation to the predicted target label $\mathbf{Y^{(j)}}$. The AOPC idea refers to the average change in the model's confidence towards the target label when the top-m most significant words, as recognized by LIME, are removed. This concept is drawn from intuitive understanding. For assessment, an arbitrary number of 500 occurrences was picked from the test set. During the explanation generation process utilizing LIME, a set of 500 altered samples is created for each instance, using the proposed attack algorithm to evaluate the explanations. The selection of $m = 10$ was made for the AOPC metric. Based on the data shown in Table 13, it

Table 12 The transferability of adversarial examples on a clickbait dataset

	Word-CNN (%)	BERT (%)	DistilBERT (%)	ALBERT (%)	RoBERTa (%)	XLNet (%)
Word-CNN	–	65.57	63.21	68.82	59.44	58.91
BERT	42.14	–	39.63	34.44	32.16	40.86
DistilBERT	48.76	45.54	–	49.75	48.91	45.09
ALBERT	52.12	49.96	51.16	–	47.72	41.66
RoBERTa	43.37	39.28	50.63	38.56	–	40.63
XLNet	42.90	42.00	48.87	49.27	48.84	–

The ASR of adversaries developed for model "p", when assessed on model "q", is represented by row "p" and column "q"



Fig. 13 Adversarial example generation using punctuation marks (non-alpha numeric characters) to evade clickbait detection mechanisms. The words enclosed in double square brackets **[[]]** have undergone alterations

Table 13 The AOPC ratings were calculated for the LIME explanations of each model. A model that has a higher AOPC score possesses greater interpretability

Models	AOPC scores
Word-CNN	0.68
BERT	0.36
DistilBERT	0.48
ALBERT	0.47
RoBERTa	0.23

is evident that the Word-CNN model achieves the greatest (AOPC) score. The analysis reveals that the AOPC scores of RoBERTa models are notably lower than those of other models, suggesting that RoBERTa may exhibit a comparatively decreased degree of explainability in comparison to all models.

7 Discussion and future work

Natural Language Processing systems that depend on textual data are susceptible to many nuanced perturbations that can alter the outcome of the neural network and extend the duration required for inference, all while keeping the visual representation of the input unaltered. The aforementioned assaults involve the utilization of random character swaps, as well as the addition, elimination, & replacement of crucial words with synonymous alternatives. This article introduces an innovative approach, employing punctuation, to deceive neural clickbait detectors. It appears that the creators of the several natural language processing algorithms currently deployed on a significant magnitude have completely overlooked these difficulties. The present work investigates the occurrence of adversarial assaults on text-based neural clickbait detectors by employing punctuation marks. The empirical results demonstrate that the Non-Alpha-Num assault outperforms traditional approaches in terms of both attack success rates as well as mean perturbations.

7.1 Future work

- *Extension to wider NLP applications* In future endeavors, we intend to broaden the reach of Non-Alpha-Num attack to encompass a diverse range of natural language processing (NLP) applications [51]. These include sentiment analysis, toxic content identification, rumor detection, as well as smishing, and phishing detection, in addition to its current focus on clickbait detection.
- *Adversarial training (defensive algorithm)* Incorporating an adversarial training methodology [52], we want to develop a framework that demonstrates resilience to adversarial disturbances specifically in the context of Non-Alpha-Num attacks. The present study provides a more comprehensive exploration of adversarial instances in the context of text categorization. Our future research endeavors to bolster these frameworks and improve the precision of classification algorithms by incorporating adversarial training methods [33]. To enhance the robustness of applications against malicious adjustments, it is recommended that all organizations engaged in the development and dissemination of natural language processing technologies integrate these precautions.
- *Extension to targeted attack* This study only concentrates on the execution of untargeted assaults, that encompass the alteration of the model's results. Nevertheless, it is important to acknowledge that Non-Alpa-Num demonstrates the capability for easy adaptation in the setting of specific assaults, whereby the framework may be coerced into generating a pre-determined classification [21]. The intentional alteration of a model with the aim of producing a specific result, sometimes known as a targeted attack, entails adjusting the objective functionality component inside the suggested approach.

7.2 Limitation of the work

The results of this investigation suggest the existence of adversarial manipulations inside the textual domain. Nevertheless, the efficacy of the modifications might be amplified by employing a more sophisticated technique, such as particle swarm optimization [26], for the purpose of identifying and altering pivotal terms. This methodology possesses the capacity to enhance the results of the suggested assault.

8 Conclusion

The primary objective of the research is to examine the susceptibility of clickbait detection algorithms to adversarial assaults. The findings unequivocally demonstrate that the identification of clickbait may be impeded by modifying the lexicon and sentence structure in algorithms used for machine learning yet retaining semantic correspondence for human assessors. The current investigation focuses on a limitation seen in deep learning models employed for the purpose of clickbait detection tasks. This study introduces Non-Alpha-Num, a unique framework that exploits a vulnerability to produce hostile text sequences with the intention of deceiving deep learning networks. Additionally, the research also conducted a comparative examination of several neural clickbait detection systems in order to ascertain the relative vulnerability of each model to adversarial perturbations, as well

as their respective levels of resilience against such perturbations. Empirical data, in a broad sense, has substantiated the viability of perturbing strategies designed for clickbait detection through the implementation of adversarial alterations. Hence, it is crucial to give precedence to the advancement of antagonistic strong generalizations to foster societal growth.

Acknowledgements The present study was not given any explicit financial support from governmental, corporate, or philanthropic organizations.

Author contributions Ashish Bajaj: Software, Validation, Investigation, Data Curation, Writing – Original Draft, Visualization, Conceptualization, Methodology Dinesh Kumar Vishwakarma: Formal Analysis, Resources, Writing – Review & Editing, Supervision, Project Administration, Funding Acquisition.

Data availability The data utilized to substantiate the conclusions of this study may be obtained by contacting the relevant author.

Declarations

Conflict of interest The contributors affirm that they do not own any identifiable conflicting financial interests or personal affiliations that may potentially be interpreted as exerting an impact on the research conducted in this investigation.

References

1. Bajaj, A., Vishwakarma, D.K.: A state-of-the-art review on adversarial machine learning in image classification. Multimed Tools Appl (2023). https://doi.org/10.1007/s11042-023-15883-z
2. Bajaj, A., Vishwakarma, D.K.: Bypassing deep learning based sentiment analysis from business reviews. In: 2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN), IEEE, May 2023, pp. 1–6. https://doi.org/10.1109/ViTECoN58111.2023.10157098
3. Bajaj, A., Kumar Vishwakarma, D.: Evading text based emotion detection mechanism via adversarial attacks. Neurocomputing **558**, 126787 (2023). https://doi.org/10.1016/j.neucom.2023.126787
4. Goyal, S., Doddapaneni, S., Khapra, M.M., Ravindran, B.: A survey of adversarial defences and robustness in NLP. ACM Comput. Surv. (2023). https://doi.org/10.1145/3593042
5. Yerlikaya, F.A., Bahtiyar, Ş: Data poisoning attacks against machine learning algorithms. Expert Syst. Appl. (2022). https://doi.org/10.1016/j.eswa.2022.118101
6. Machado, G.R., Silva, E., Goldschmidt, R.R.: Adversarial machine learning in image classification: a survey toward the defender's perspective. ACM Comput. Surv. **55**(1), 1–38 (2023). https://doi.org/10.1145/3485133
7. Wang, W., Wang, R., Wang, L., Wang, Z., Ye, A.: Towards a robust deep neural network against adversarial texts: a survey. IEEE Trans. Knowl. Data Eng. (2023). https://doi.org/10.1109/TKDE.2021.3117608
8. Potthast, M., et al.: Crowdsourcing a large corpus of clickbait on Twitter. In: COLING 2018 - 27th International Conference on Computational Linguistics, Proceedings (2018)
9. Agrawal, A.: Clickbait detection using deep learning. In: Proceedings on 2016 2nd International Conference on Next Generation Computing Technologies, NGCT 2016 (2017). https://doi.org/10.1109/NGCT.2016.7877426
10. Fakhruzzaman, M.N., Jannah, S.Z., Ningrum, R.A., Fahmiyah, I.: Flagging clickbait in Indonesian online news websites using fine-tuned transformers. Int. J. Electr. Comput. Eng. (2023). https://doi.org/10.11591/ijece.v13i3.pp2921-2930
11. Al-Sarem, M., et al.: An improved multiple features and machine learning-based approach for detecting clickbait news on social networks. Appl. Sci. (Switzerland) (2021). https://doi.org/10.3390/app11209487
12. Pujahari, A., Sisodia, D.S.: Clickbait detection using multiple categorisation techniques. J. Inf. Sci. (2021). https://doi.org/10.1177/0165551519871822
13. Uddin Rony, M.M., Hassan, N., Yousuf, M.: Diving deep into clickbaits: Who use them to what extents in which topics with what effects? In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2017 (2017). https://doi.org/10.1145/3110025.3110054
14. Kaur, S., Kumar, P., Kumaraguru, P.: Detecting clickbaits using two-phase hybrid CNN-LSTM biterm model. Expert Syst. Appl. (2020). https://doi.org/10.1016/j.eswa.2020.113350
15. Bajaj, A., Vishwakarma, D.K.: Exposing the vulnerabilities of deep learning models in news classification. In: 2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT), IEEE, pp. 1–5 (2023). https://doi.org/10.1109/ICITIIT57246.2023.10068577
16. Qiu, S., Liu, Q., Zhou, S., Huang, W.: Adversarial attack and defense technologies in natural language processing: a survey. Neurocomputing (2022)
17. Ha, T., Dang, T.K., Le, H., Truong, T.A.: Security and privacy issues in deep learning: a brief review. SN Comput. Sci. (2020). https://doi.org/10.1007/s42979-020-00254-4
18. Miller, D.J., Xiang, Z., Kesidis, G.: Adversarial learning targeting deep neural network classification: a comprehensive review of defenses against attacks. Proc. IEEE (2020). https://doi.org/10.1109/JPROC.2020.2970615
19. Liu, J., et al.: Aliasing black box adversarial attack with joint self-attention distribution and confidence probability. Expert Syst. Appl. (2023). https://doi.org/10.1016/j.eswa.2022.119110
20. Li, A., Zhang, F., Li, S., Chen, T., Su, P., Wang, H.: Efficiently generating sentence-level textual adversarial examples with Seq2seq Stacked Auto-Encoder. Expert Syst. Appl. (2023). https://doi.org/10.1016/j.eswa.2022.119170
21. Morris, J.X., Lifland, E., Yoo, J.Y., Qi, Y.: TextAttack: a framework for adversarial attacks in natural language processing. ArXiv, pp. 119–126 (2020)
22. Yoo, J.Y., Morris, J.X., Lifland, E., Qi, Y.: Searching for a search method: benchmarking search algorithms for generating NLP adversarial examples. In: Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, pp. 323–332. [Online] (2020). Available: https://github.com/QData/TextAttack
23. Jin, D., Jin, Z., Zhou, J.T., Szolovits, P.: Is BERT really robust? A strong baseline for natural language attack on text classification and entailment. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 8018–8025. [Online] (2019). Available: http://arxiv.org/abs/1907.11932
24. Li, J., Ji, S., Du, T., Li, B., Wang, T.: TextBugger: generating adversarial text against real-world applications. In: 26th Annual Network and Distributed System Security Symposium, pp. 1–15 (2019). https://doi.org/10.14722/ndss.2019.23138
25. Ren, S., Deng, Y., He, K., Che, W.: Generating natural language adversarial examples through probability weighted word saliency. In: ACL 2019—57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (2020). https://doi.org/10.18653/v1/p19-1103

26. Zang, Y., et al.: Word-level textual adversarial attacking as combinatorial optimization. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 6067–6080 (2020). https://doi.org/10.18653/v1/2020.acl-main.540
27. Pruthi, D., Dhingra, B., Lipton, Z.C.: Combating adversarial misspellings with robust word recognition. In: ACL 2019—57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (2020). https://doi.org/10.18653/v1/p19-1561
28. Kuleshov, V., Thakoor, S., Lau, T., Ermon, S.: Adversarial examples for natural language classification problems. In: ICLR 2018: International Conference on Learning Representations (2018)
29. Wang, X., Jin, H., He, K.: Natural language adversarial attacks and defenses in word level. ArXiv (2019)
30. Gao, J., Lanchantin, J., Lou Soffa, M., Qi, Y.: Black-box generation of adversarial text sequences to evade deep learning classifiers. In: Proceedings—2018 IEEE Symposium on Security and Privacy Workshops, SPW 2018, pp. 1–21 (2018). https://doi.org/10.1109/SPW.2018.00016
31. Ribeiro, M.T., Wu, T., Guestrin, C., Singh, S., Beyond accuracy: behavioral testing of NLP models with CheckList. In: ACL 2020—58th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), pp. 4902–4912 (2020)
32. Garg, S., Ramakrishnan, G.: BAE: BERT-based adversarial examples for text classification. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 6174–6181 (2020)
33. Yoo, J.Y., Qi, Y.: Towards improving adversarial training of NLP models. In: Findings of the Association for Computational Linguistics, Findings of ACL: EMNLP 2021 (2021). https://doi.org/10.18653/v1/2021.findings-emnlp.81
34. Iyyer, M., Wieting, J., Gimpel, K., Zettlemoyer, L.: Adversarial example generation with syntactically controlled paraphrase networks. In: NAACL HLT 2018—2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies—Proceedings of the Conference (2018). https://doi.org/10.18653/v1/n18-1170
35. Liang, B., Li, H., Su, M., Bian, P., Li, X., Shi, W.: Deep text classification can be fooled. In: IJCAI International Joint Conference on Artificial Intelligence, pp. 4208–4215 (2018). https://doi.org/10.24963/ijcai.2018/585
36. Cer, D., et al.: Universal sentence encoder (2018), [Online]. Available: http://arxiv.org/abs/1803.11175
37. Naber, D., Kummert, P.F., Fakultät, T., Witt, A.: A rule-based style and grammar checker, Technische Fakultät, Universität Bielefeld. (2003), [online]. https://www.danielnaber.de/languagetool/download/style_and_grammar_checker.pdf. Accessed 10 May 2024
38. Chakraborty, A., Paranjape, B., Kakarla, S., Ganguly, N.: Stop clickbait: detecting and preventing clickbaits in online news media. In: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016 (2016). https://doi.org/10.1109/ASONAM.2016.7752207
39. Kim, Y.: Convolutional neural networks for sentence classification. In: EMNLP 2014—2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference (2014). https://doi.org/10.3115/v1/d14-1181
40. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: NAACL HLT 2019—2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies—Proceedings of the Conference (2019)
41. Sanh, V., Debut, L., Chaumond, J., Wolf, T.: DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, pp. 2–6 (2019)
42. Lan, Z., et al.: Albert: a lite bert for self-supervised learning of language representations. In: International Conference on Learning Representations (ICLR), pp. 1–17 (2020). [Online]. Available: https://github.com/google-research/ALBERT
43. Liu, Y., et al.: RoBERTa: a robustly optimized BERT pretraining approach. In: International Conference on Learning Representations (ICLR), pp. 1–15 (2019). [Online]. Available: http://arxiv.org/abs/1907.11692
44. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., Le, Q.V.: XLNet: Generalized autoregressive pretraining for language understanding. In: 33rd conference on neural information processing systems (NeurIPS 2019). Vancouver, Canada, pp 1–11 (2019)
45. Garg, S., Ramakrishnan, G.: BAE: BERT-based adversarial examples for text classification. In: EMNLP 2020—2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference (2020). https://doi.org/10.18653/v1/2020.emnlp-main.498
46. Ebrahimi, J., Rao, A., Lowd, D., Dou, D.: HotFlip: white-box adversarial examples for NLP. In: Proceedings of the 56th annual meeting of the association for computational linguistics (Short Papers), pp 31–36 (2018)
47. Feng, S., Wallace, E., Grissom, A., Iyyer, M., Rodriguez, P., Boyd-Graber, J.: Pathologies of neural models make interpretations difficult. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018 (2018). https://doi.org/10.18653/v1/d18-1407
48. Xie, C., et al.: Improving transferability of adversarial examples with input diversity. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2019). https://doi.org/10.1109/CVPR.2019.00284
49. Zhang, J., Peng, W., Wang, R., Lin, Y., Zhou, W., Lan, G.: Enhance domain-invariant transferability of adversarial examples via distance metric attack. Mathematics (2022). https://doi.org/10.3390/math10081249
50. Ribeiro, M.T., Singh, S., Guestrin, C.: 'Why should I trust you?' Explaining the predictions of any classifier. In: NAACL-HLT 2016—2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Demonstrations Session (2016). https://doi.org/10.18653/v1/n16-3020
51. Patwardhan, N., Marrone, S., Sansone, C.: Transformers in the real world: a survey on NLP applications. Information (Switzerland) (2023). https://doi.org/10.3390/info14040242
52. He, L., Ai, Q., Yang, X., Ren, Y., Wang, Q., Xu, Z.: Boosting adversarial robustness via self-paced adversarial training. Neural Netw. (2023). https://doi.org/10.1016/j.neunet.2023.08.063

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

Received: 9 November 2023 | Revised: 17 April 2024 | Accepted: 24 April 2024

DOI: 10.1111/exsy.13619

ORIGINAL ARTICLE

Expert Systems WILEY

Nonlinear dynamical system approximation and adaptive control based on hybrid-feed-forward recurrent neural network: Simulation and stability analysis

R. Shobana[1] | Rajesh Kumar[2] | Bhavnesh Jaint[1]

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

[2]Department of Electrical Engineering, National Institute of Technology Kurukshetra, Kurukshetra, India

Correspondence
Rajesh Kumar, Department of Electrical Engineering, National Institute of Technology Kurukshetra, Kurukshetra 136119, India.
Email: rajeshmahindru23@gmail.com; rajeshmahindru23@nitkkr.ac.in

Abstract

We proposed an online identification and adaptive control framework for the nonlinear dynamical systems using a novel hybrid-feed-forward recurrent neural network (HFRNN) model. The HFRNN is a combination of a feed-forward neural network (FFNN) and a local recurrent neural network (LRNN). We aim to leverage the simplicity of FFNN and the effectiveness of RNN to capture changing dynamics accurately and design an indirect adaptive control scheme. To derive the weights update equations, we have applied the gradient-descent-based Back-Propagation (BP) technique, and the stability of the proposed learning strategy is proven using the Lyapunov stability principles. We also compared the proposed method's results with those of the Jordan network-based controller (JNC) and the local recurrent network-based controller (LRNC) in the simulation examples. The results demonstrate that our approach performs satisfactorily, even in the presence of disturbance signals.

KEYWORDS
disturbance rejection, hybrid structures, identification based control, recurrent neural networks, robustness

1 | INTRODUCTION

Nonlinear control theory has gained popularity in recent times for designing nonlinear control systems. While linear control theory approaches, such as PID controllers, state space control approaches, and gain scheduling, have been used to identify and handle nonlinear dynamic systems since the late 1930s, they have many limitations (Oaks & Cook, 1976; Zhao & Guo, 2021). The linear controller design pre assumes that both the plant and the controller are linearized, which is not always the case in realtime. Second, because the linear control approaches do not follow the superposition principle, their solutions are invalid for nonlinear systems. They are stable on one range of inputs and unstable on another. Third, they exhibit chaotic behaviour at their output, and hard nonlinearities are difficult to linearized. All physical systems are subject to parameter variations, and controllers developed using linear control theory fail to account for this. This resulted in the development of nonlinear control theory methods including sliding mode control (Alanis, 2022), robust control (Qin et al., 2022), adaptive control (Swarnkar et al., 2014), and fuzzy-based control (Wang et al., 2017). Adaptive controllers, among these, have received a lot of interest from researchers. Unlike other control approaches, adaptive control generates control inputs by taking the system's slowly varying dynamics into account and driving it to follow the desired control law (Li et al., 2018; Zheng et al., 2019). As a result, even in the presence of disturbance, the controller performance is faster and better. There are two types of adaptive control approaches: direct adaptive control and indirect adaptive control. In direct adaptive control, the error between the plant and reference output is used to adjust the controller parameters. In indirect adaptive control, the controller parameters are adjusted by estimating the plant parameters. The indirect adaptive control is more prevalent than direct adaptive control. The direct adaptive control approach does not guarantee stability as the controller parameters are adjusted by calculating the error between the plant and the

 © 2024 John Wiley & Sons Ltd.

reference. On the other hand, the indirect adaptive control estimates the plant's behaviour and designs a controller based on that estimated behaviour. This approach is more stable and can effectively handle time-varying systems and disturbance (Narendra & Valavani, 1979). The working of indirect adaptive control involves both identification and controller blocks. The identification block gathers knowledge about the unknown plant, while the control block generates an appropriate control signal to make the plant track the reference model. Artificial neural networks (ANNs) are increasingly used in identification-based adaptive control due to their ability to handle complex nonlinearities and self-adaptation. ANNs are suitable for modelling and control of any nonlinear system as they can approximate any nonlinear function (Akpan & Hassapis, 2011; Polycarpou & Ioannou, 1991). Additionally, they have fast processing capabilities, allowing them for real-time control. A self-recurrent wavelet neural network-based control is proposed by Yoo et al. (2007) for the control of nonlinear dynamical systems. The modified structure is trained using the gradient descent algorithm. Adaptive learning rates are used for faster convergence. The simulation result shows the effectiveness of the suggested structure. In the study by Slama et al. (2019), an indirect adaptive switch controller that switches online between PID and IAC is proposed. The adaptation laws are developed using the gradient descent method and the stability is analyzed using the Lyapunov function. The results show that the proposed controller performs better than an indirect adaptive controller, and adaptive switch controller. In the study by Ciliz (2009), an adaptive control approach combining both the direct and indirect adaptive scheme is proposed. The proposed controller convergence is proved by using the Lyapunov function and the proposed controller is found to improve the tracking accuracy of the controller. Anavatti et al. (2015) have surveyed various state-of-the-art methods in adaptive control system theory. The paper briefly discusses the various pros and cons of various techniques employed in adaptive control strategy. In the study by Cheng et al. (2021), adaptive neural control is proposed for the control of dynamic systems. Input-output linearization technique is used to convert the object's model into a multi-integrator system and for online training, the state observation technique is used. Further, in the study by Chen et al. (2016), a modified adaptive controller for the MIMO system is developed. The controller used is the backstepping control and an adaptive algorithm was developed to find the states and deal with non-symmetrical input. In the study by He et al. (2019), online adaptive control is developed with a policy iteration (PI) algorithm. The optimal control laws are developed using the Ricatti equation. In the study by Bonassi and Scattolini (2022), a gated RNN structure such as GRU and LSTM for control of a stable nonlinear system is proposed. Identification-based control was carried out and the stability was proven by using incremental input and output state stability. The results show that the proposed controller has better convergence over other controllers. Zhou, Tong, et al. (2021) have designed a neural controller with a dynamic surface control scheme to overcome the effect of uncertainties on the system. A control law is proposed and is approximated using a radial basis function neural network (RBF-NN). The stability is also analyzed using the Lyapunov function. In the study by Kumar, Srivastava, Gupta, and Mohindru (2018), a recurrent wavelet neural structure is used to identify and control nonlinear systems. The parameters are updated using the back-propagation algorithm and adaptive learning rates are designed to ensure fast convergence. In the study by Kumar et al. (2017), a diagonal recurrent neural network (DRNN) based adaptive control is proposed for identification-based control of nonlinear systems, and the Lyapunov function is used for training the algorithm. The results show better response than feed-forward networks. In the study by Kumar, Srivastava, and Gupta (2018), a comparative analysis of various FFNN and RNN-based controllers for adaptive control of nonlinear dynamic systems is carried out. Dynamic back propagation is used for updating the weights of the controller and the convergence of the structure is also analyzed. Simulation results prove that the RBFN-based controller performs better than NARX and FFNN-based controllers even in the presence of uncertainties. Perrusquía and Yu (2021) have designed an optimal controller using RNN, and learning of the structure is done using reinforcement training. The controller is designed for both continuous and discrete-time systems. The performance is analyzed for robustness. Though the results look promising, the optimal controllers require the entire knowledge of the system to be known to compute the Jacobian matrix. In the study by Truong et al. (2023), an adaptive back-stepping-based controller for a higher-order nonlinear system is proposed. An RBFNN-based time-delayed estimator (TDE) architecture is designed and controlled using a back-stepping controller. The controller model is designed off-line and tested for efficiency by comparing the model with PID based controller and the original back-stepping controller model-free approach. Kapnopoulos et al. (2024) have designed an RBF neural network-based back-stepping controller for solving the tracking problem for quad-rotor systems. Though RBFN is a memory-less structure in combination with the fuzzy controller, it is found to increase the accuracy by taking into account the unmodelled dynamics. The performance is compared with an MLP-based back-stepping controller. It is found to provide better tracking performance. In another work (Bukhari et al., 2023), the authors have analyzed the dynamic behaviour of fractional order chaotic systems with RBFN. The parameters of the system are computed with a Runge-Kutta solver and designed with RBFN for various initial conditions of the Lorentz system. Average mutual information (AMI) techniques are used to evaluate the time delay pattern and the performance of the proposed RBFN is found to be exceptionally high. In the study by Hesamian et al. (2023), an RBFN model is developed based on fuzzy responses and predictions. A squared error-based method is used to estimate the fuzzy parameters and tuning constants. The performance of the proposed Fuzzy RBFN is found to be effective in comparison to conventional fuzzy controllers. A fuzzy RBFN-based sliding mode control is used for the control of nonlinear systems of varying degrees in this work (Tavoosi, 2020). The proposed method does not require prior information about the system as the controller is simulated online. The method was found to improve the final threshold stability and yielded better performance than RBFN-based SMC. Further, the based adaptive controllers have found numerous applications in real-time. In the study by Zhou and Tan (2023), the characteristics and progress of NN-based adaptive control are tested on Nuclear power plants (NPP).The performance is compared with NNPID, self-tuning PID, and tested for efficiency with respect to disturbance. A car driving-based simulator model based on adaptive and robust control with online learning is designed by Manh et al. (2023). This RBFNN structure-based adaptive control does not require

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

system information. In another work (Patan & Patan, 2020), an iterative learning control of a Nonlinear system based on ANN is designed. The controller update rule is updated by iterative training and the updated rules were incorporated in the next trial process to achieve a zero error at convergence even in terms of uncertainties. The update rules are further tested in servomechanism and magnetic levitation applications. Di Ferdinando et al., 2020 have studied the Lyapunov-kravoskii theorem for global asymptotic and local exponential stability (GALES) for a nonlinear delay-free system. Point-wise dissipation rates are used. The proposed theorem is found to replace the use of the Lipschitz property of the function describing the system, thus maintaining local and global asymptotic stability. Xu et al. (2023) have studied a Lyapunov-Kravovskii function for semi-globally practical fixed-time stable systems. L'Hopital's rule is used to determine the boundness of the lyapunov-kravoskii function and the signals are bounded within the integral function range. This makes the tracking error converge into a small region around zero. Further, it is extended to adaptive control of higher-order stochastic time-delay nonlinear systems. Lale et al. (2022) have studied and proposed a Kravoskii-based stability theorem for a model-based Reinforcement training framework. The changing dynamics of the system are captured using feature extraction methods. A prime dual approach based Kravoskii theorem is used to derive the Lyapunov function. The proposed approach is found to have applications in online control of safety-critical systems such as robotic manipulators and locomotive tasks. Further in the study by Dashkovskiy and Pavlichkov (2020), a novel I_{∞} input to state stability for infinite networks is proposed. Feedback for each agent is established and an ISS-based Lyapunov function is developed for infinite networks. The Lyapunov function is further extended for the constructive design of decentralized controllers to stabilize infinitely larger networks. Zhou, Hu, et al. (2021) have discussed the various techniques that could effectively optimize the gain in the Lyapunov function. The Hamilton energy as a suitable Lyapunov function is proved against Lorentz and Chua oscillation in a chaotic state. The paper claims that the control of energy flow when done completely can control chaos in a Nonlinear system and improve stability. After reviewing the literature, we discovered significant shortcomings in the existing approaches to controlling nonlinear dynamic systems. Firstly, the feed-forward models such as MLP and RBFN are not much suited for real-time implementation as they lack dynamic memory structures. The adaptive training requires samples derived at every instant to properly control the plant. These structures cannot capture the changing dynamics of the system and must be supplied with system order in advance. Secondly, since the control applications are generally implemented in real-time, a simple control algorithm is necessary to derive the control inputs and hence the algorithm needs to be kept simple and efficient. Many of the algorithms used in the literature involve a large number of parameters, which leads to poor convergence. Lastly, while many models and approaches have been cited in the literature for the identification-based control of nonlinear dynamic systems, a benchmark approach for the control of dynamic systems is not yet available. To the best of our knowledge, there is no hybrid fusion of FFNN and RNN available in the literature. These shortcomings have motivated us to try to enhance the control approach for effective identification-based control of nonlinear dynamic systems.

The rest of the paper is organized as follows: The introduction section briefly describes the need for more controller structures and the motivation behind the work. Section 3 elaborates on the problem statement and Section 4 discusses the various architectures of the neural structure used in this work to design the controller. Section 4 discusses briefly the proposed control scheme, learning algorithm, and its convergence in the sense of Lyapunov stability. Section 5 discusses the simulation result obtained. The proposed HFRNN-based controller is tested on two benchmark nonlinear plant equations and their performance and efficiency are discussed. In section 6, the conclusion and future scope of the work are discussed.

2 | PROBLEM STATEMENT

Let's consider a plant $y_p(k)$ represented by the following difference equation as follows:

$$y_p(k+1) = f\left[y_p(k), y_p(k-1), .., y_p(k-n), u_c(k), u_c(k-1), .., u_c(k-m)\right] \tag{1}$$

where $y_p(k+1)$ denotes the one-step ahead output of plant model, n refers to the order of the plant and and m refers to the order of the controller input $u_c(k)$, f refers to the unknown differentiable nonlinear function. Let's consider a reference model with the following difference equation:

$$y_m(k+1) = \ g\left[y_m(k), y_m(k-1), .., y_m(k-n), r(k), r(k-1), .., r(k-m)\right] \tag{2}$$

where $y_m(k+1)$ refers to the one-step ahead output of the reference model, n denotes the order of the reference model and m denotes the order of external input $r(k)$, g denotes the unknown nonlinear function that is assumed differentiable. The main objective of our work is as follows: 1. To identify the unknown plant model by minimizing the error between the plant and the identified neural model. This is done by updating the identifier weight equations based on the error between the plant and the identified neural model such that

$$\lim_{k\to\infty} |\, y_p(k+1) - y_{hni}(k+1) \,| \leq \epsilon \tag{3}$$

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where $\epsilon \to 0$. 2. To make the plant follow the reference model by determining the control input $u_c(k)$. This is achieved by continuously training the ANN-based controller update equations to generate a control input $u_c(k)$ such that

$$\lim_{k\to\infty} |\, y_m(k+1) - y_p(k+1) \,| \le \epsilon \tag{4}$$

where $\epsilon \to 0$. The error between the plant and the reference model updates the controller equations.

3 | THEORETICAL FOUNDATION OF HFRNN

A brief introduction and applications of the Jordan neural network and Local recurrent neural network are discussed by many authors in the literature. For more details on the Jordan neural network, readers can refer to these references Jordan (1986), Wysocki and Lawryńczuk (2015), Huang et al. (2007) Şen et al. (2020) and for the local recurrent neural network readers can refer to these references Kumar et al. (2017), Campolucci et al. (1999), and Coban (2013). In this section, we have outlined briefly the structures of HFRNN.

3.1 | HFRNN structure

Throughout the paper, the neural architecture is referred to as HFRNN. The structure of HFRNN is depicted in Figure 1. It is the combination of both the FFNN and LRNN models. The structure is explained in brief below:

1. **Input layer:** The input layer comprises the input samples supplied to the network. Here, the present and past external input is denoted as $r(k)$ and $r(k-1)$ respectively, and the present and the delayed output of the plant are denoted as $y_p(k)$ and $y_p(k-1)$ respectively. They are considered as inputs. The input vector is denoted using the vector $x = [r(k), r(k-1), y_p(k), y_p(k-1)]$. These inputs are further passed to the feed-forward network (indicated by green lines in the figure) and recurrent network paths (indicated by orange lines in the figure) of the hidden

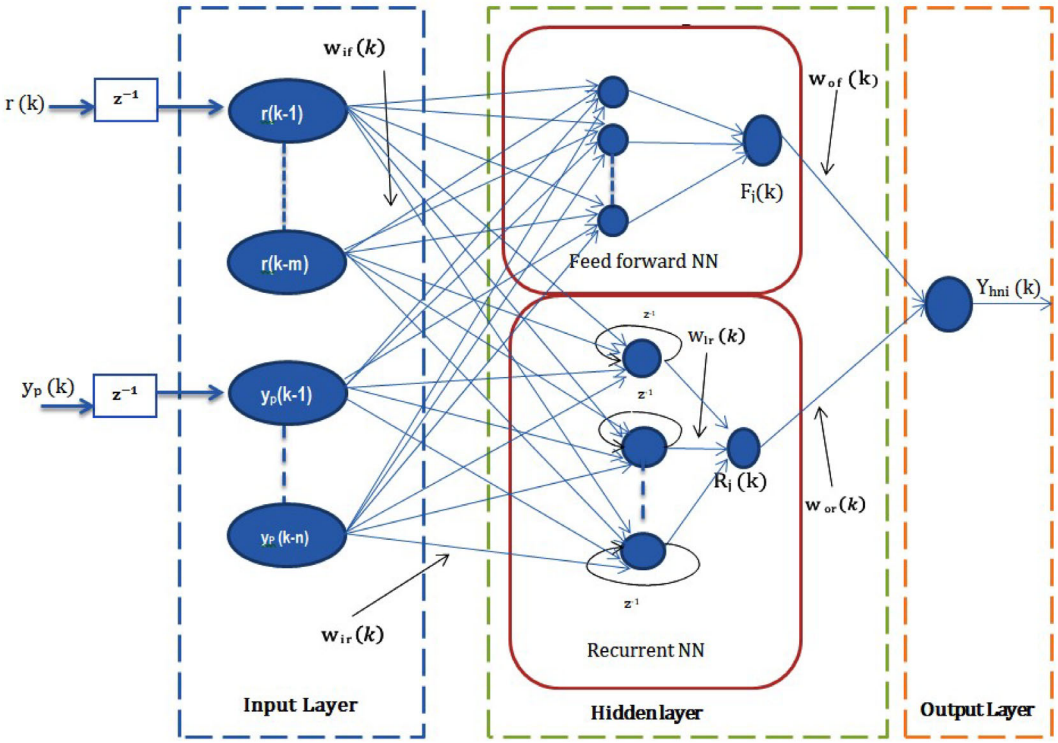


FIGURE 1 HFRNN structure.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

layer. The weight connections between the feedforward network and the input layers are denoted as $w_{if}(k)$ and that of the recurrent network and the input layer is denoted as $w_{ir}(k)$ in Figure 1.

2. **Hidden layer:** The weighted samples from the input layer enter the hidden layer and the output of the hidden layer is computed in two paths. The first path generates the induced field output of the feed-forward network and is denoted as F_j and the second path generates the induced field output of the recurrent network and is denoted as R_j. The induced field output of the feed-forward network is obtained by passing the weighted samples through a nonlinear activation function. The induced field output of the recurrent network is obtained by multiplying the weighted samples received from the input layer with the local self-hidden recurrent layer input samples. The recurrent structure being an LRNN, the past outputs of the hidden layers are fed as input to the hidden layer. Both the induced output of the hidden layers are sent to the output layer.
3. **Output layer:** The output layer computes the final hybrid network output with static and recurrent network characteristics.

The output of the induced field at any kth instant is computed below:

$$F_j(k) = f_1\left(\sum_{j=1}^{n} x(k-j)w_{if_j}(k) + b_{fx_j}w_{b_j}(k)\right), \tag{5}$$

$$R_j(k) = g_1\left(\sum_{j=1}^{m} x(k-j)w_{if_j}(k) + L(k)w_{lr}(k) + b_{fx_j}w_{b_j}(k)\right), \tag{6}$$

where f_1 and g_1 indicate the non-linear tangent hyperbolic activation function at the hidden node and b_{fx_j} denotes the associated weight of the input bias vector. The HFRNN output at any k^{th} instant is given by:

$$Y_{hni}(k) = f\left(\sum_{i=1}^{m} F_j(k)w_{of}(k) + b_{oh}w_{bh}(k) + R_j(k)w_{or}(k)\right), \tag{7}$$

where $w_{of}(k)$ and $w_{or}(k)$ denote the weights of the feed-forward and recurrent path respectively proceeding to the output layer. $w_{bh}(k)$ is the weight of the bias at the output node and b_{oh} denotes the associated weight of output bias vector. f denotes the linear activation function of the output layer.

4 | HFRNN BASED INDIRECT ADAPTIVE CONTROL SCHEME FOR NONLINEAR DYNAMICAL SYSTEMS

We will discuss the proposed HFRNN-based indirect adaptive controller scheme in this section. Figure 2 shows the block diagram of the proposed scheme. The main objective of our work is to predict the unknown controller parameters and obtain a control input to make the plant follow the reference model. The controller scheme consists of two major blocks: the identifier and the controller. We use the HFRNN-based identifier to identify the unknown plant and the HFRNN-based controller to produce the control input $u_c(k)$. We train both of these block's parameters using the backpropagation algorithm. The HFRNN-based controller takes the reference input $r(k)$, previous control input $u_c(k-1)$, and the past output of the plant $y_p(k-1)$ as inputs. The weights are updated by detecting the error between the plant and the reference model output and are applied to the HFRNN-based controller until the error is minimized. Although the plant remains generally unknown, the information about the plant is crucial for properly tuning the controller. In this case, we use an HFRNN-based identifier, to estimate the plant sensitivity for the controller. The HFRNN-based identifier takes the present control input $u_c(k)$ and previous plant output $y_p(k-1)$ as inputs. The identifier is trained using the error generated between the output of the plant and the neural identifier. The training is done online and the plant generates efficient weights and better control through simultaneous identification and control.

5 | LEARNING ALGORITHM FOR HFRNN-BASED CONTROLLER STRUCTURE

The gradient descent algorithm is used for updating the controller and identifier weight equations. $E_c(k)$ is the instantaneous error calculated between the model output and the plant output. $E_c(k)$ is given by:

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

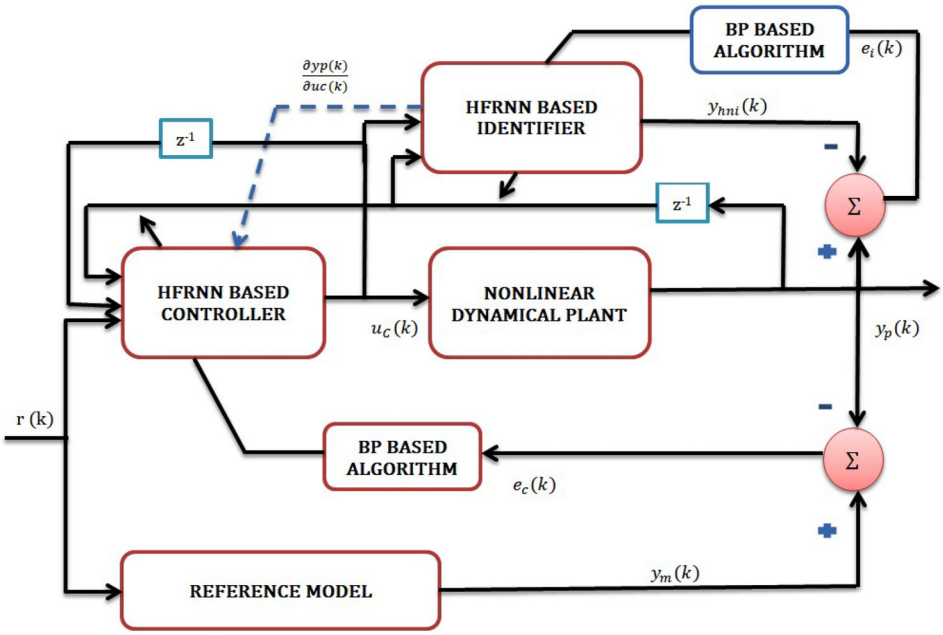


FIGURE 2 Block diagram of the proposed control scheme.

$$E_c(k) = \frac{1}{2}[e_c(k)]^2, \tag{8}$$

where, $e_c(k) = y_m(k) - y_p(k)$. Here, $y_m(k)$ and $y_p(k)$ denote the reference model output and the plant response respectively at kth instant of time. The gradient chain rule is applied by calculating the change of $E_c(k)$ with respect to every element present in the link. Starting with the calculation for updating the output recurrent weights $w_{orc}(k)$, it is obtained by differentiating $E_c(k)$ with respect to $w_{orc}(k)$ as below:

$$\frac{\partial E_c(k)}{\partial w_{orc}(k)} = \frac{\partial E_c(k)}{\partial y_p(k)} \times \frac{\partial y_p(k)}{\partial u_c(k)} \times \frac{\partial u_c(k)}{\partial w_{orc}(k)}. \tag{9}$$

The right hand side of the above equation can be written as:

$$\frac{\partial E_c(k)}{\partial w_{orc}(k)} = -e(k) \times R_j(k) \times J(k). \tag{10}$$

The weight update equation of $w_{orc}(k)$ is given by:

$$w_{orc}(k+1) = w_{orc}(k) + \eta\left(-\frac{\partial E_c(k)}{\partial w_{orc}(k)}\right). \tag{11}$$

The gradient chain rule is again applied for the calculation of output feed-forward weights $w_{ofc}(k)$. The change of $E_c(k)$ with respect to $w_{ofc}(k)$ is calculated as follows:

$$\frac{\partial E_c(k)}{\partial w_{ofc}(k)} = \frac{\partial E_c(k)}{\partial y_p(k)} \times \frac{\partial y_p(k)}{\partial u_c(k)} \times \frac{\partial u_c(k)}{\partial w_{ofc}(k)}, \tag{12}$$

where $\frac{\partial E_c(k)}{\partial y_p(k)} = -e(k)$, $\frac{\partial y_p(k)}{\partial u_c(k)} = F_j(k)$ and $\frac{\partial u_c(k)}{\partial w_{ofc}(k)} = J(k)$ so above equation can be written as follows:

$$\frac{\partial E_c(k)}{\partial w_{ofc}(k)} = -e(k) \times F_j(k) \times J(k). \tag{13}$$

The weight update equation of $w_{ofc}(k)$ is given by:

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

$$w_{ofc}(k+1) = w_{ofc}(k) + \eta\left(-\frac{\partial E_c(k)}{\partial w_{ofc}(k)}\right). \tag{14}$$

Every calculation of the controller requires knowledge of the changing dynamics of the plant, and $J(k)$ computes the sensitivity of the controller. If the plant is completely unknown, it becomes very necessary to design an online identifier in parallel to the controller to automatically adapt to the changing dynamics. The proposed structure is trained online, hence requiring the weights and errors of the identifier to be updated simultaneously with the controller parameters. $E_i(k)$ is the instantaneous error calculated between the plant, $y_p(k)$ and the neural identifier, $y_{hni}(k)$. This is calculated by:

$$E_i(k) = \frac{1}{2}[e_i(k)]^2, \tag{15}$$

where, $e_i(k) = y_p(k) - y_{hni}(k)$. As the structure is trained online, the output of the controller $u_c(k)$ becomes one of the inputs of the identifier. This results in calculation of $\frac{\partial y_p(k)}{\partial u_c(k)}$. This indicates the Jacobian matrix, J(k). The sensitivity of the plant is determined through the HFRNN-based identifier block, as the plant is completely unknown to the HFRNN-based controller. This computation updates both the input recurrent $w_{irc}(k)$ and feed-forward identifier weights $w_{ifc}(k)$ with respect to $E_c(k)$. This starts with updating the hidden layer local weights of the controller $w_{hc}(k)$ with respect to $E_c(k)$.

$$\frac{\partial E_c(k)}{\partial w_{hc}(k)} = \frac{\partial E_c(k)}{\partial y_p(k)} \times \frac{\partial y_p(k)}{\partial y_{rc}(k)} \times \frac{\partial y_{rc}(k)}{\partial u_c(k)} \times \frac{\partial u_c(k)}{\partial w_{hc}(k)}. \tag{16}$$

The above equation can be written as follows:

$$\frac{\partial E_c(k)}{\partial w_{hc}(k)} = -e(k) \times \left(I - R_j^2(k)\right) \times w_{orc}(k) \times T(k), \tag{17}$$

where $T(k)$ denotes the memory of the past hidden context layer. Further, the input feed-forward layer weights of the controller $w_{ifc}(k)$ and input recurrent weights of the controller $w_{irc}(k)$ are updated as follows:

$$\frac{\partial E_c(k)}{\partial w_{ifc}(k)} = \frac{\partial E_c(k)}{\partial y_p(k)} \times \frac{\partial y_p(k)}{\partial y_{fc}(k)} \times \frac{\partial y_{fc}(k)}{\partial u_c(k)} \times \frac{\partial u_c(k)}{\partial w_{ifc}(k)}. \tag{18}$$

On substituting the values for the terms appearing on the right hand side of the above equation:

$$\frac{\partial E_c(k)}{\partial w_{ifc}(k)} = -e(k) \times w_{ofc}(k) \times \left(I - F_j^2(k)\right) \times X(k), \tag{19}$$

$$\frac{\partial E_c(k)}{\partial w_{irc}(k)} = \frac{\partial E_c(k)}{\partial y_p(k)} \times \frac{\partial y_p(k)}{\partial y_{rc}(k)} \times \frac{\partial y_{rc}(k)}{\partial u_c(k)} \times \frac{\partial u_c(k)}{\partial w_{irc}(k)}, \tag{20}$$

or

$$\frac{\partial E_c(k)}{\partial w_{irc}(k)} = -e(k) \times w_{orc}(k) \times \left(I - R_j^2(k)\right) \times X(k). \tag{21}$$

The weights $w_{ifc}(k)$ and $w_{irc}(k)$ are updated at every instant of time k using the weight update rule applied as below:

$$w_{ifc}(k+1) = w_{ifc}(k) + \eta\left(-\frac{\partial_c(k)}{\partial w_{ifc}(k)}\right), \tag{22}$$

$$w_{irc}(k+1) = w_{irc}(k) + \eta\left(-\frac{\partial E_c(k)}{\partial w_{irc}(k)}\right). \tag{23}$$

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

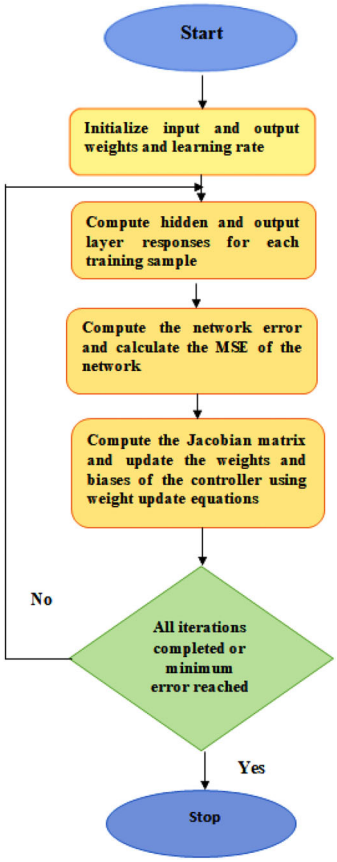


FIGURE 3 Steps followed in training HFRNN-based controller structure.

Here, η denotes the static learning rate and is taken between 0 and 1. As the training progresses, the neural network starts following the plant model, that is, $\frac{\partial y_{hni}(k)}{\partial u_c(k)} \approx \frac{\partial y_p(k)}{\partial u_c(k)}$. Once this is achieved, the identified model can accurately represent the changing dynamics of the plant. With the unknown plant identified, the HFRNN-based controller starts to learn the dynamics and hence the plant model starts to follow the reference model, that is, $\frac{\partial y_p(k)}{\partial u_c(k)} \approx \frac{\partial y_m(k)}{\partial u_c(k)}$. Figure 3 illustrates the various iterative steps followed in learning the controller algorithm.

6 | STABILITY AND CONVERGENCE

Stability is a main characteristic of a system's behaviour. Lyapunov-based stability is a stability concept that states that the system's equilibrium is maintained as long as its states remain close to the initial condition. The Lyapunov stability can be used in two ways to analyze the stability of the system. First, it is used to derive the weight update equation, which ensures that the derived equations are always stable. Second, we can declare the weight adjustment equation and then use Lyapunov stability analysis to check whether the declared equations satisfy the Lyapunov stability conditions. If they do, they will always be stable; otherwise, the stability cannot be guaranteed. In this work, we use the second approach of stability analysis. The weight update equations derived are used to check whether stability has been achieved. For this, let us consider a Lyapunov-based error function such that

$$J = min(E), \tag{24}$$

where E is the Lyapunov-based function. The system is considered to be stable if the Lyapunov-based error remains positive and minimal regardless of the system's condition.

$$\dot{V}(x) > 0 \text{ for } x > 0 \text{ and } \dot{V}(x) = 0 \text{ for } x = 0. \tag{25}$$

The rate of change of error is given by

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

$$\dot{V}(k) = \left(\frac{\partial V(k)}{\partial w_{ofc}(k)} + \frac{\partial V(k)}{\partial w_{orc}(k)} + \frac{\partial V(k)}{\partial w_{hc}(k)} + \frac{\partial V(k)}{\partial w_{irc}(k)} + \frac{\partial V(k)}{\partial w_{ifc}(k)}\right) \tag{26}$$

or

$$\dot{V}(k) = \left(\frac{\partial V(k)}{\partial w_{ofc}(k)} \times \frac{dw_{ofc}(k)}{dt} + \frac{\partial V(k)}{\partial w_{orc}(k)} \times \frac{dorc(k)}{dt} + \frac{\partial V(k)}{\partial w_{hc}(k)} \times \frac{dhc(k)}{dt} + \frac{\partial E(k)}{\partial w_{irc}(k)} \times \frac{dw_{irc}(k)}{dt} + \frac{\partial E(k)}{\partial w_{ifc}(k)} \times \frac{dw_{ifc}(k)}{dt}\right). \tag{27}$$

When $\frac{dw_{ofc}(k)}{dt}$, $\frac{dw_{orc}(k)}{dt}$ $\frac{dw_{hc}(k)}{dt}$, $\frac{dw_{ifc}(k)}{dt}$, and $\frac{dw_{irc}(k)}{dt}$, is taken as $(x-y)^2$ and the weight update equations from E_q. (9) to (20), the rate of change of error becomes

$$\dot{V}(k) = \eta \times (x-y)^2 \times \left((R_j(k)J(k)) + F_j(k)J(k) + w_{orc}(k)T(k) + w_{ofc}(k)\left(1 - F_j^2(k)X(k)\right) + w_{orc}(k)\left(1 - R_j^2(k)X(k)\right)\right) \tag{28}$$

or

$$\dot{V}(k) = \eta \times (x-y)^2 \times \left(\frac{\partial V(k)}{\partial w_{orc}(k)} + \frac{\partial V(k)}{\partial w_{ofc}(k)} + \frac{\partial V(k)}{\partial w_{hc}(k)} + \frac{\partial V}{\partial w_{ifc}(k)} + \frac{\partial V}{\partial w_{irc}(k)}\right). \tag{29}$$

Now to make the system ensure stability, the term $(x-y)^2$ will always be positive and $(x-y)^2 \neq 0$. Along with this $\eta > 1$. As long as these conditions are ensured, the weight update equations ensure stability and the controller remains BIBO stable.

7 | SIMULATIONS AND DISCUSSIONS

To efficiency of the proposed HFRNN-based controller is compared with two other neural-based controllers such as the Jordan-based neural network controller (JNC), and the local recurrent-based neural network controller (LRNC) on two different nonlinear plant equations of varying degrees. All of the controllers are designed with a single hidden layer with 20 neurons and a static learning rate of 0.001. The initial weights and biases are taken randomly in the interval (0,1).

7.1 | Example 1: A nonlinear dynamic plant equation with degree 3

In this example, a plant with the following differential equation is considered as in the study by Narendra and Parthasarathy (1992).

$$y_p(k+1) = \frac{5y_p(k)y_p(k-1)}{\left(1 + y_p(k)^2 + y_p(k-1)^2 + y_p(k-2)^2\right)} + u_c(k) + 0.8u_c(k-1) + v(k). \tag{30}$$

The nonlinear plant takes the identification structure as below:
$y_p(k+1) = f\left[y_p(k), y_p(k-1), y_p(k-2), u_c(k), u_c(k-1)\right]$ Consider a reference model with the following differential equation:

$$y_m(k+1) = 0.72y_m(k) + 0.64y_m(k-1) - 0.5y_m(k-2) + r(k), \tag{31}$$

where f is the nonlinear function, $r(k)$ denotes the external reference input, $u_c(k)$ is the control input generated by the controller and $v(k)$ denotes the disturbance signal. The reference signal applied to the system is, $r(k) = sin\left(\frac{2\pi k}{25}\right)$. The disturbance applied to the system initially is $v(k) = 0$. The main objective is to obtain the control input $u_c(k)$ and make the plant follow the reference model. The inputs of the controller are $\left[y_p(k), y_p(k-1), y_p(k-2), r(k), u_c(k-1), u_c(k-2)\right]$ and that of identifier are $\left[u_c(k), y_p(k-1), y_p(k)\right]$. Figure 4 shows the open-loop response of the plant. It can be seen that, without controller action, the plant does not track the reference model. Figure 5 and Figure 6 show the response of the plant during the initial and final stages of training. It is observed initially, that the plant does not follow the reference model. But, as training progresses, the plant starts to follow the reference model and after 15000 epochs, the error minimizes and converges to a minimum value. The instantaneous MSE obtained is also shown in Figure 7.

It is evident from Figure 7 and Table 1 that the proposed controller gives good control performance even with a minimum number of parameters. On the other hand, both JNC and LRNC take time to converge to a minimum MSE.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

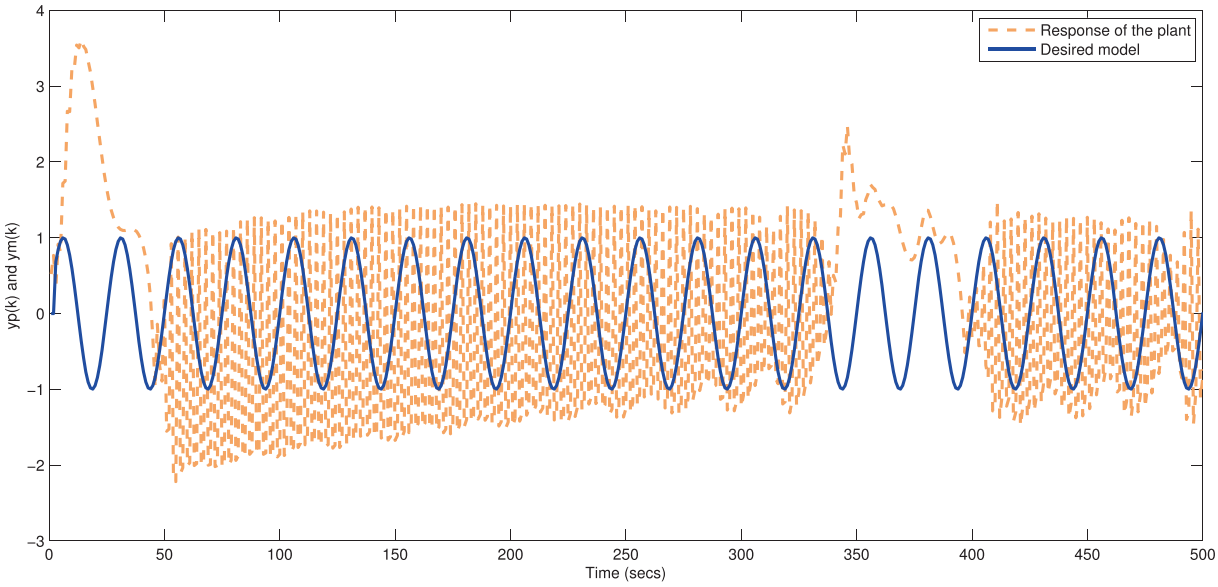


FIGURE 4 Plant's response without controller action (Example 1).

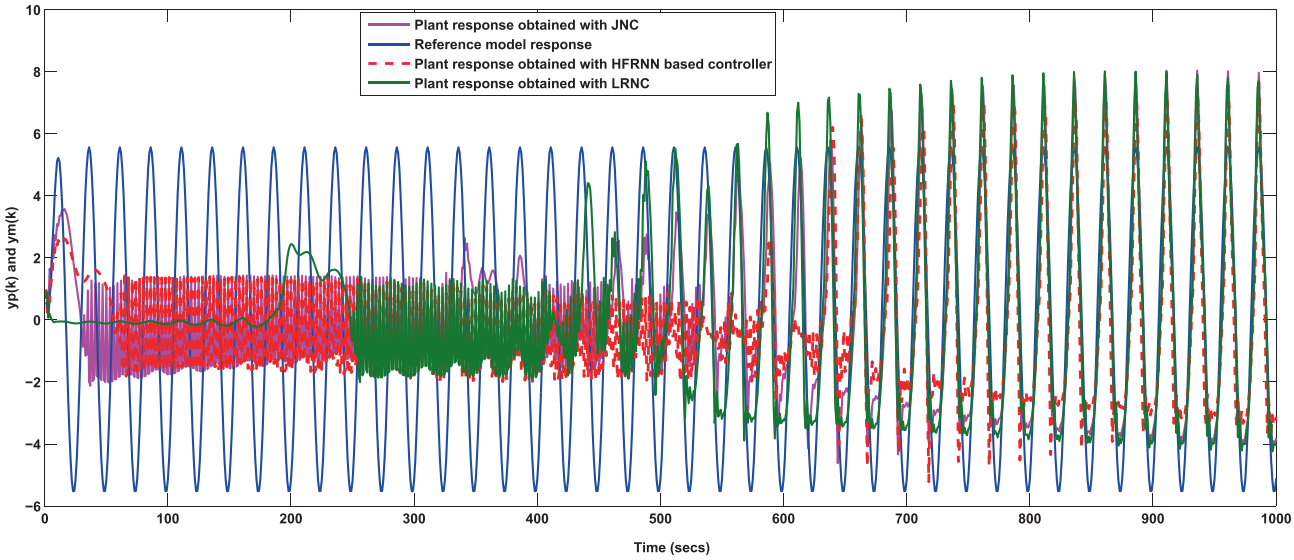


FIGURE 5 Plant response obtained during the initial training phase (Example 1).

7.2 | Disturbance testing

To examine the recovering ability of the proposed controller, a sudden disturbance signal $v(k) = sin(\frac{2\pi k}{40})$ is introduced between the time interval $9000 < k \leq 11000$. A static learning rate of 0.001 is considered, and the effect of external noise on the HFRNN-based controller, JNC, and LRNC was analyzed. Initially, the system deviated from the reference model due to the disturbance signal. But with training, the control parameters were adjusted to the changing environment and the system recovered. However, it took a little longer time. The instantaneous MSE during the disturbance signal of a based controller is found to be very small as compared to JNC and LRNC. The MSE obtained for the based controller was around 0.0653. Refer to Figure 8 for the comparison of the effect of disturbance signal over HFRNN-based controller, JNC, and LRNC.

7.3 | Recovering ability of the HFRNN based controller

The recovering ability of the proposed controller was also tested by changing the reference input. When the reference input was changed from the previous one to the new one, the network was able to capture and adapt to the new reference input. For this, a new reference input $r(k)$ was considered as below between the interval $5000 < k \leq 15000$:

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

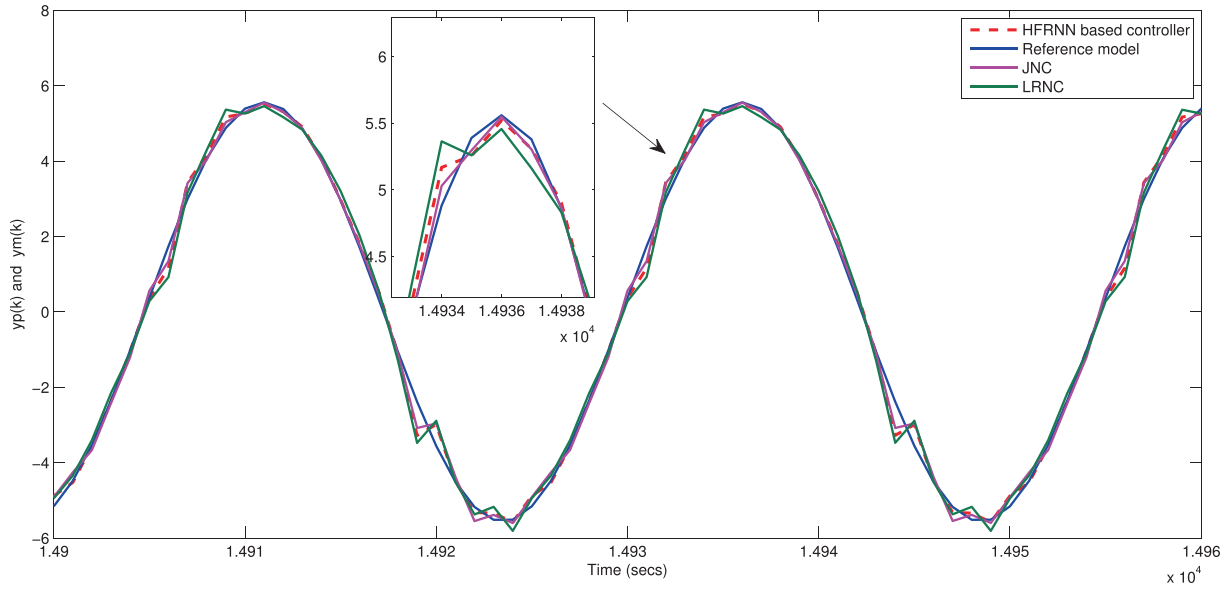


FIGURE 6 Response of controllers during the final phase of training (Example 1).

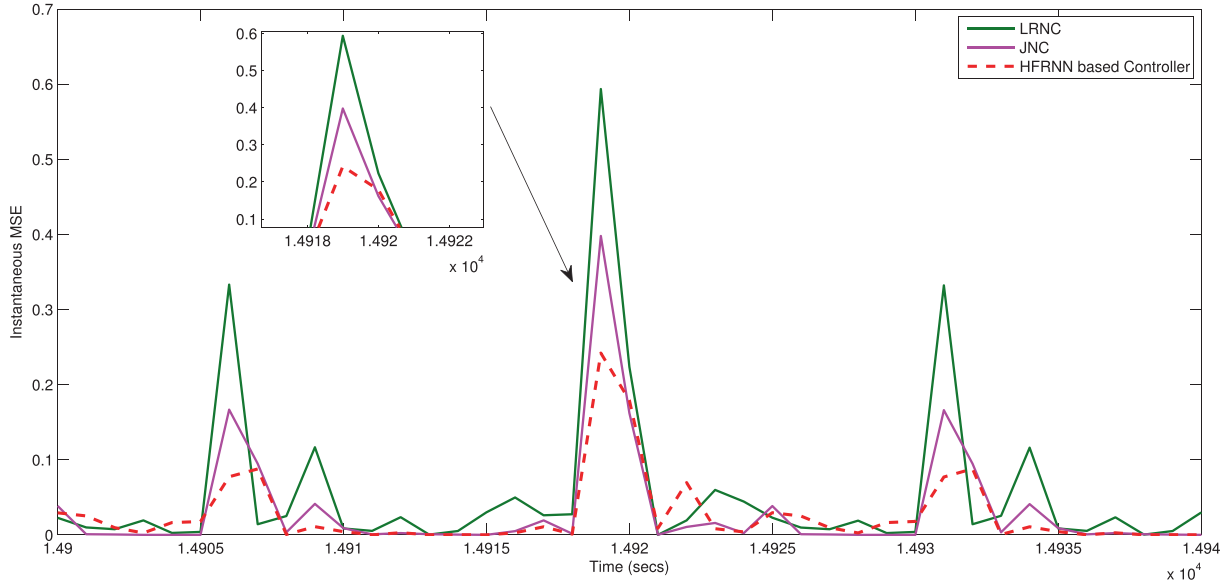


FIGURE 7 Instantaneous MSE (Example 1).

TABLE 1 Comparison of controller parameters (Example 1).

S. No	Structure	Instantaneous MSE	Simulation time (s)
1.	HFRNN based controller	0.0010	1.39
2.	JNC	0.0244	1.66
3.	LRNC	0.0095	2.35

$$r(k) = \begin{cases} \sin\left(\frac{2\pi k}{25}\right), & \text{for } 0 < k \le 5000 \\ \sin\left(\frac{2\pi k}{25}\right) + \sin\left(\frac{2\pi k}{10}\right), & \text{for } 5000 < k \le 15000 \end{cases} \quad (32)$$

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

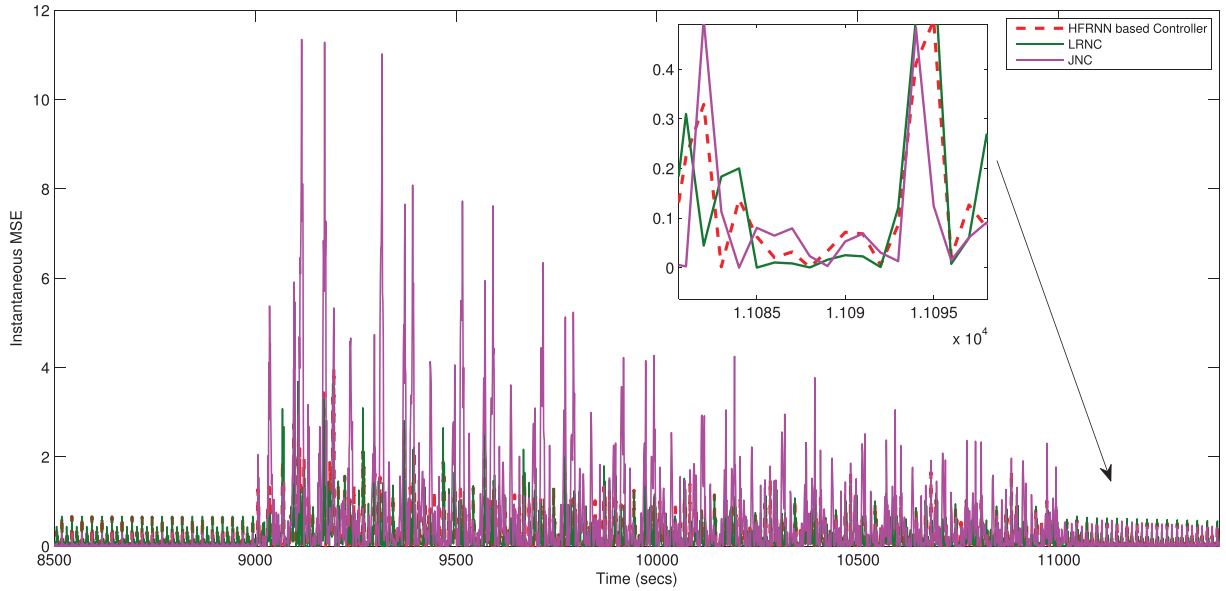


FIGURE 8 Comparison of the effect of external noise on controllers (Example 1).

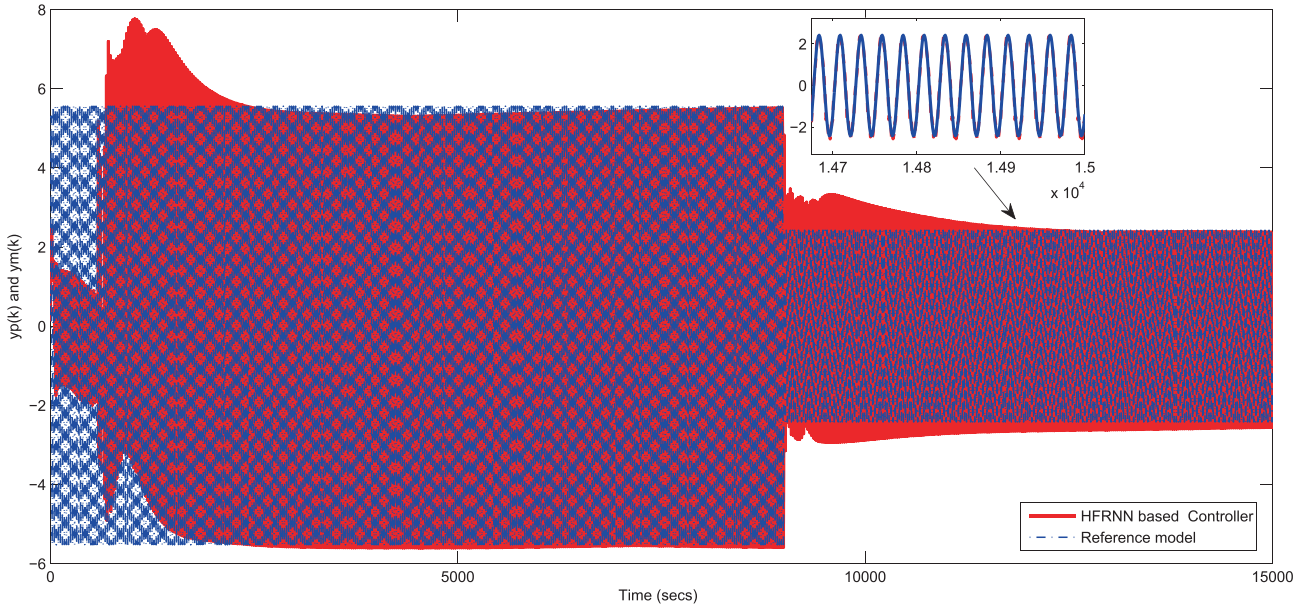


FIGURE 9 Recovering ability of HFRNN-based controller (Example 1).

When the reference input was changed after 5000 training epochs, the controller was found to show good recovering ability. Figure 9 shows the online adaptability of the controller.

7.3.1 | Testing of models with square wave as a reference signal

The controller was also supplied with a square waveform as a reference signal to test the recovery ability. The network was found to capture and adapt to a new reference input. For this, the new reference input $r(k) = square\left(\frac{\pi k}{40}\right)$ was considered. The controller showed good adaptability, with the plant slowly following the desired model. During the end of the training, the plant response resembled the desired model. The instantaneous MSE of the proposed HFRNN controller is found to be 4.7×10^{-7} and that of JNC and LRNC are found to be 4.88×10^{-8} and 6.17×10^{-7}, respectively. Figure 10 shows the responses of the controller during the final stages of training with a square waveform as a reference signal.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

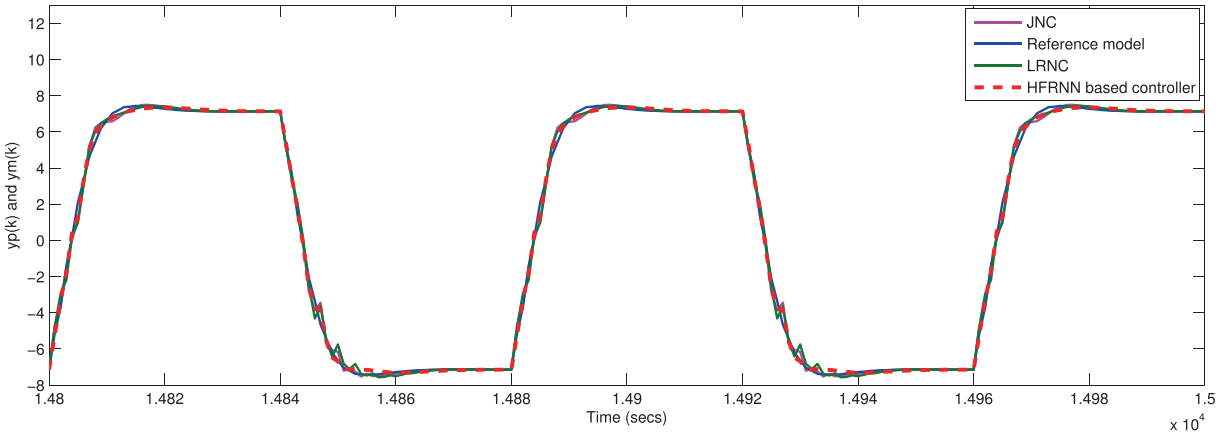


FIGURE 10 Response of controllers with a square waveform as a reference signal (Example 1).

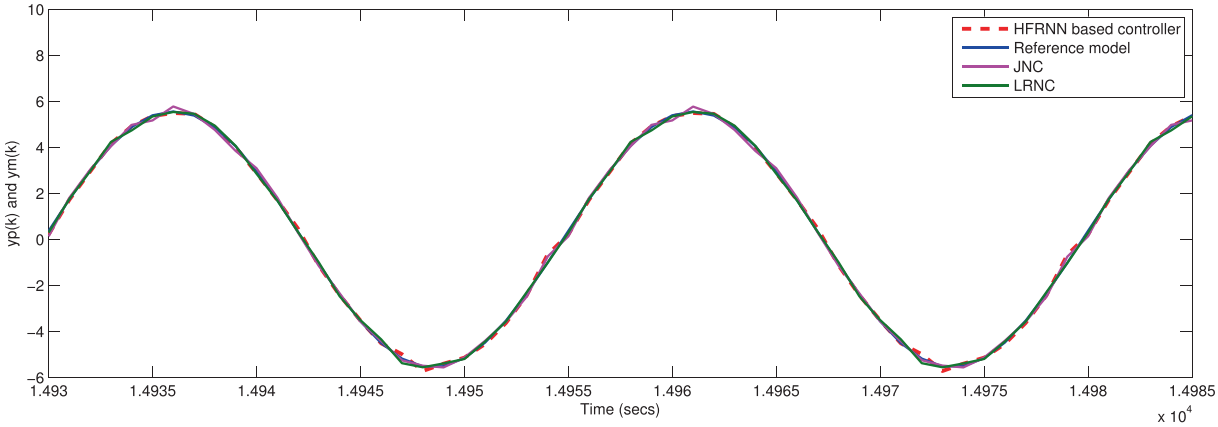


FIGURE 11 Response of controllers for different initial conditions (Example 1).

7.3.2 | Testing of models with different initial conditions

The controller was also tested by assigning a different initial condition of interval (−1, 1) for weights and biases. Initially, the plant failed to track the reference model, but with training and updates to network weights through the BP algorithm, the controller started to learn the dynamics of the plant. Towards the end of training, it is found to resemble the desired response of the plant. Figure 11 shows the response of the controller during the final stages of training. The instantaneous MSE is also shown in Figure 12. The proposed HFRNN controller is seen to provide better MSE compared to other selected JNC and LRNC controllers.

7.4 | Example 2: A nonlinear dynamic plant having degree of 2

To test the efficiency of the proposed controller, another BIBO plant with degree 2 as in the study by Narendra and Parthasarathy (1992) is considered. The plant takes the following differential equation.

$$y_p(k+1) = \frac{y_p(k)}{\left(1+y_p(k-1)^2\right)+u_c(k)^3} + v(k). \tag{33}$$

The nonlinear plant has the identification structure as below: $y_p(k+1) = f\left[y_p(k), y_p(k-1), u_c(k)\right]$ The reference model is given by the following differential equation:

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

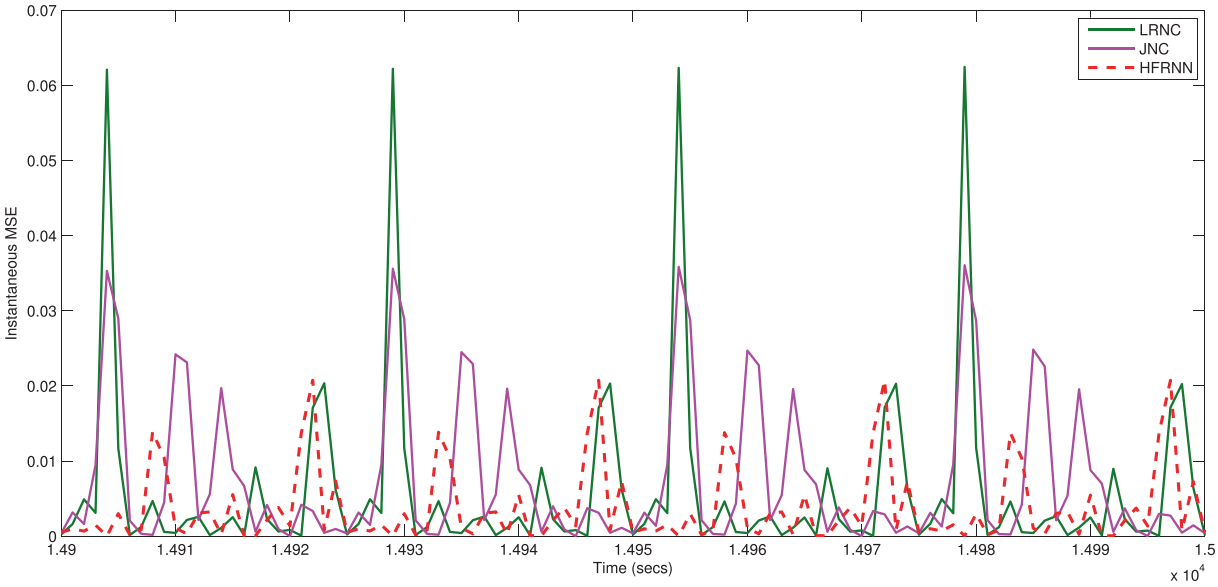


FIGURE 12 Instantaneous MSE of controllers for different initial conditions (Example 1).

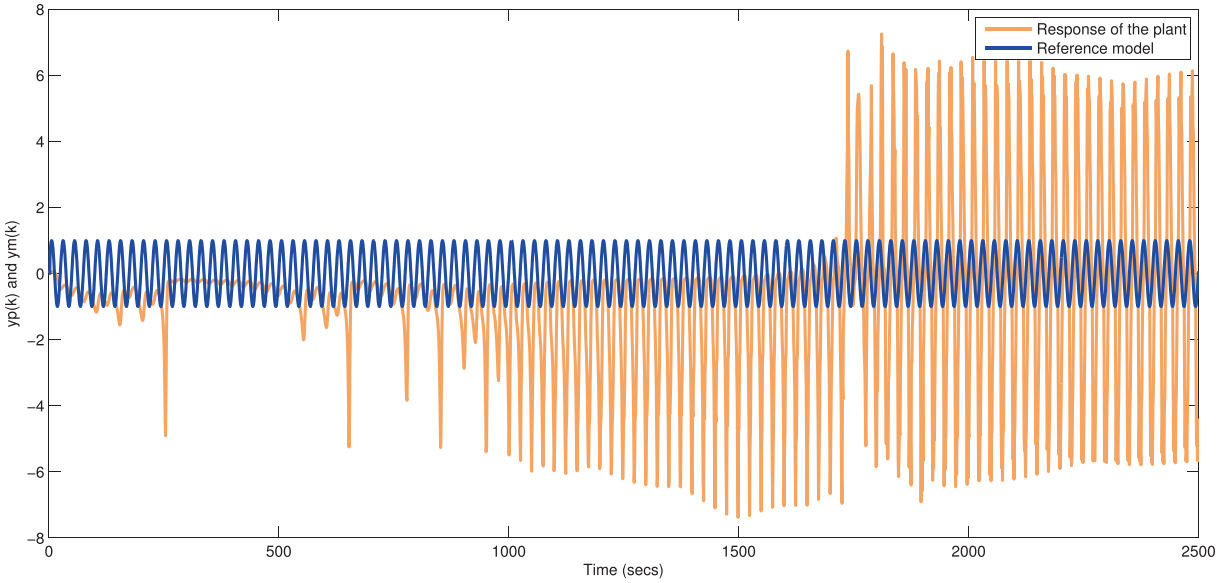


FIGURE 13 Plant's response without controller action (Example 2).

$$y_m(k+1) = 0.72y_m(k) + 0.64y_m(k-1) - 0.5y_m(k-2) + r(k), \tag{34}$$

where f is the nonlinear function, $r(k)$ denotes the external reference input, $u_c(k)$ is the control input generated by the controller and $v(k)$ denotes the disturbance signal. The reference signal applied to the system is, $r(k) = sin\left(\frac{2\pi k}{25}\right)$. The inputs of the controller are $\left[y_p(k), y_p(k-1), y_p(k-2), r(k), u_c(k-1), u_c(k-2)\right]$ and that of identifier are $\left[u_c(k), y_p(k-1), y_p(k)\right]$. The response of the plant without controller action is shown in Figure 13. The plant does not track the desired model response without controller action. With the application of the controller, the plant slowly starts to follow the desired model and after 15000 epochs the plant resembles the desired model response. Figure 14 shows the responses of the controller during the initial training phase and Figure 15 shows the responses of the controller during the final stages of training. The plant is trained online and the instantaneous MSE is also shown in Figure 16. The MSE is found to increase to the maximum during the initial

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

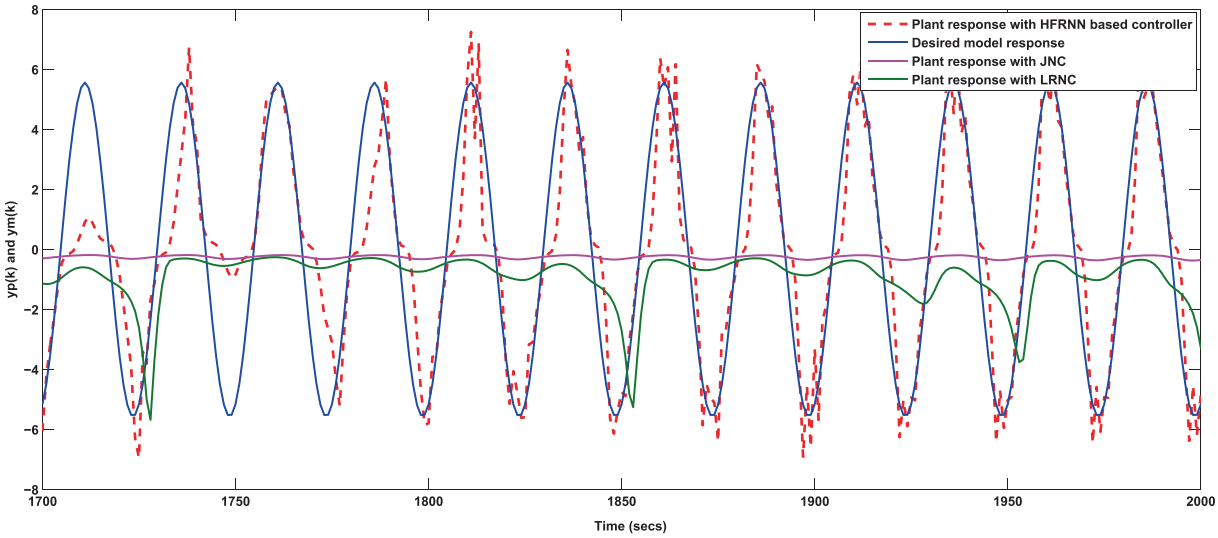


FIGURE 14 Response obtained from the plant during the initial training phase (Example 2).

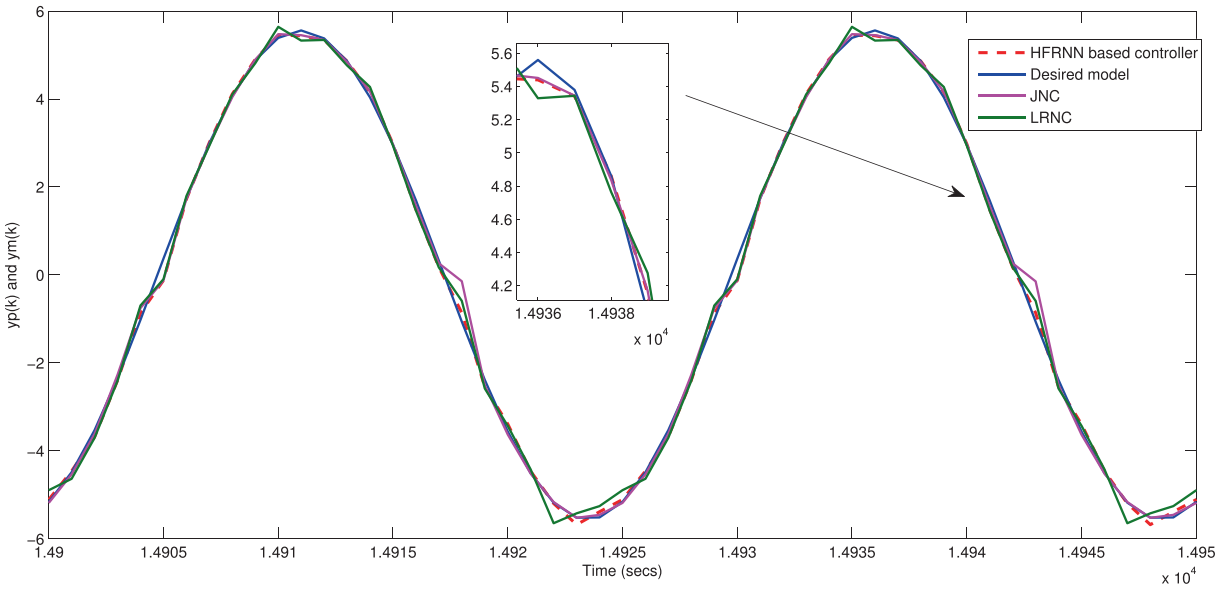


FIGURE 15 Response of controllers during the final phase of training (Example 2).

stages of training and has reached a minimum value after 15000 epochs. Figures 14 and 15 and Table 2 show that the HFRNN based controller performs better than the Jordan and the LRNN based controller scheme.

7.5 | Disturbance testing

To examine the recovering ability of the proposed controller, a sudden disturbance signal, $v(k) = sin\left(\frac{2\pi k}{40}\right)$ was added between the time interval $9000 < k \leq 11000$. A static learning rate of 0.001 was considered and the effect of external noise on HRC, JNC, and LRNC was observed. The MSE of all controllers is found to increase with the addition of external noise. This is due to the model deviating from the reference model. After a few epochs of training, the system is seen to adapt to its new changes with MSE decreasing. The instantaneous MSE of the proposed controller is found to be very small around 0.0080 as compared to JNC and LRNC. Figure 17 shows the effect of disturbance on the proposed controller, JNC, and LRNC.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

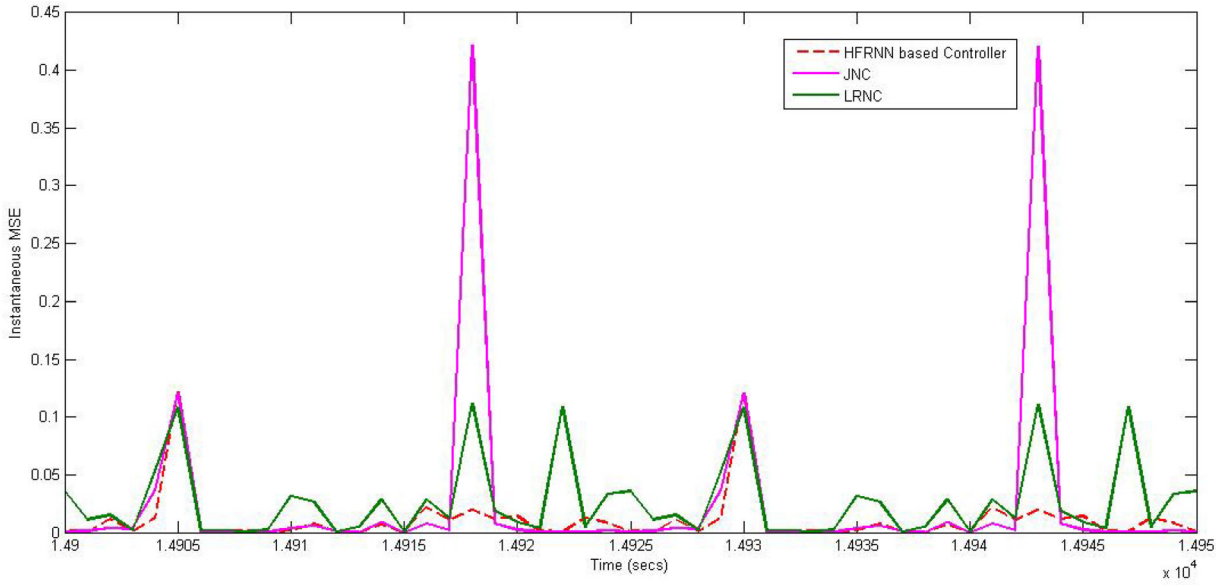


FIGURE 16 Instantaneous MSE (Example 2).

TABLE 2 Comparison of controller parameters (Example 2).

S. No	Structure	Instantaneous MSE	Simulation time (s)
1.	HFRNN based controller	3.160×10^{-4}	1.52
2.	JNC	0.0012	1.43
3.	LRNC	0.0111	2.43

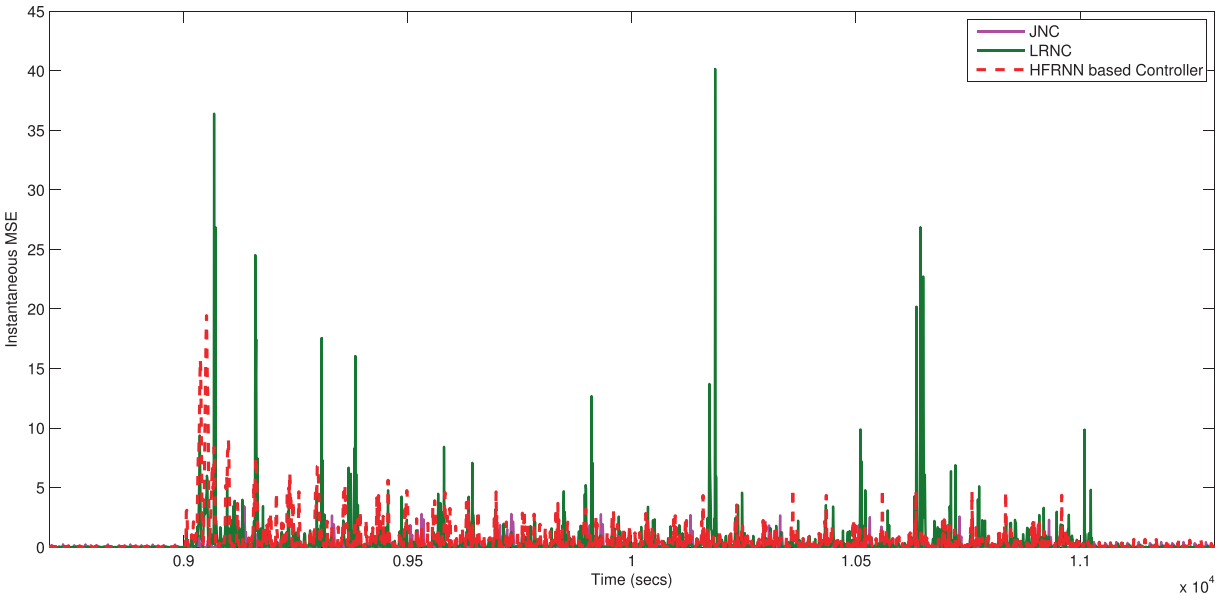


FIGURE 17 Comparison of the effect of external noise on controllers (Example 2).

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

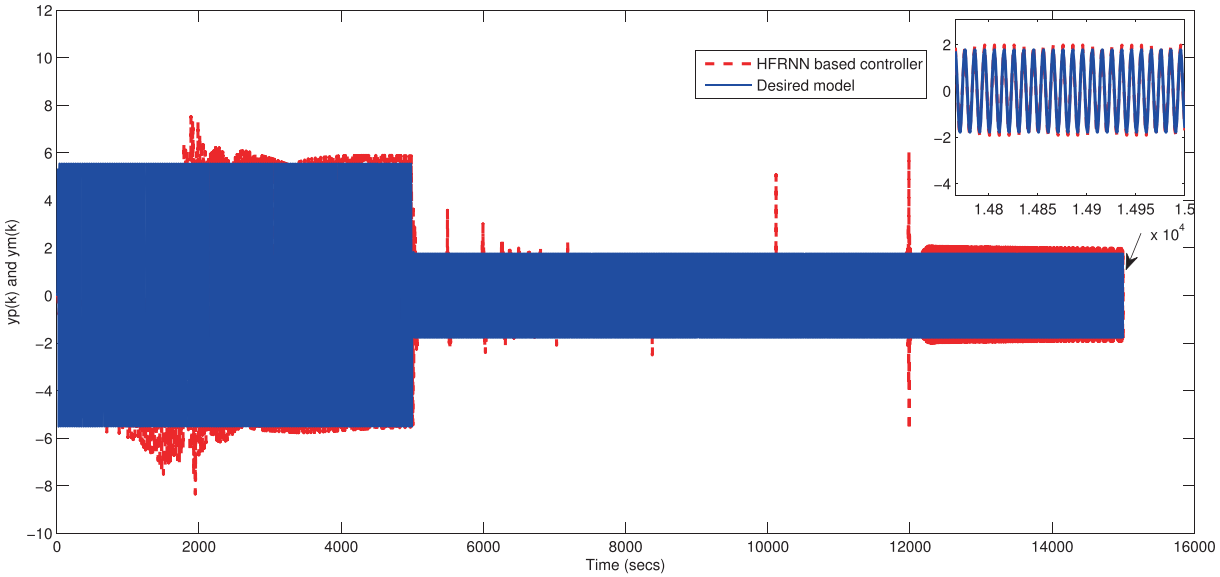


FIGURE 18 Recovering ability of HFRNN based controller (Example 2).

7.6 | Recovering ability of the HFRNN based controller

The recovering ability of the proposed controller was also tested by changing the reference input. When the reference input was changed from the previous one to the new one, the network was able to capture and adapt to the new reference input. For this, the new reference input $r(k)$ is considered as below between the interval $5000 < k \leq 15000$:

$$r(k) = \begin{cases} sin\left(\frac{2\pi k}{25}\right), & \text{for } 0 < k \leq 5000 \\ sin\left(\frac{2\pi k}{10}\right), & \text{for } 5000 < k \leq 15000 \end{cases}. \tag{35}$$

Figure 18 shows the online adaptability of the controller.

7.6.1 | Testing of models with square wave as a reference signal

The controller was also supplied with a square waveform as a reference signal to test the recovery ability. The network was found to capture and adapt to a new reference input. For this, the new reference input $r(k) = square\left(\frac{\pi k}{40}\right)$ was considered. The controller was found to show good adaptability. The plant slowly starts to follow the desired model and after 13000 epochs the plant resembles the desired model response. The instantaneous MSE of the proposed HFRNN controller is found to be 1.660×10^{-7} and that of JNC and LRNC is found to be 9.880×10^{-7} and 6.000×10^{-7}, respectively. Figure 19 shows the responses of the controller during the final stages of training with a square waveform as a reference signal.

7.6.2 | Testing of models with different initial conditions

The controller was also tested by assigning a different initial condition of interval (-1,1) for weights and biases. Initially, the plant does not track the reference model, but with training and an update of network weights through the BP algorithm, the controller starts to learn the dynamics of the plant, and slowly towards the end of training it is found to resemble the desired response of the plant. Figure 20 shows the response of the controller during the initial and final stages of training. The instantaneous MSE is also shown in Figure 21. The proposed HFRNN controller is seen to provide a better MSE as compared to other selected JNC and LRNC controllers.

From the results, it is very evident that the structure could capture the changing dynamics of the plant very efficiently. The proposed approach is order-independent and the structure is also simple as compared to other recurrent structures in the literature. All the selected

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

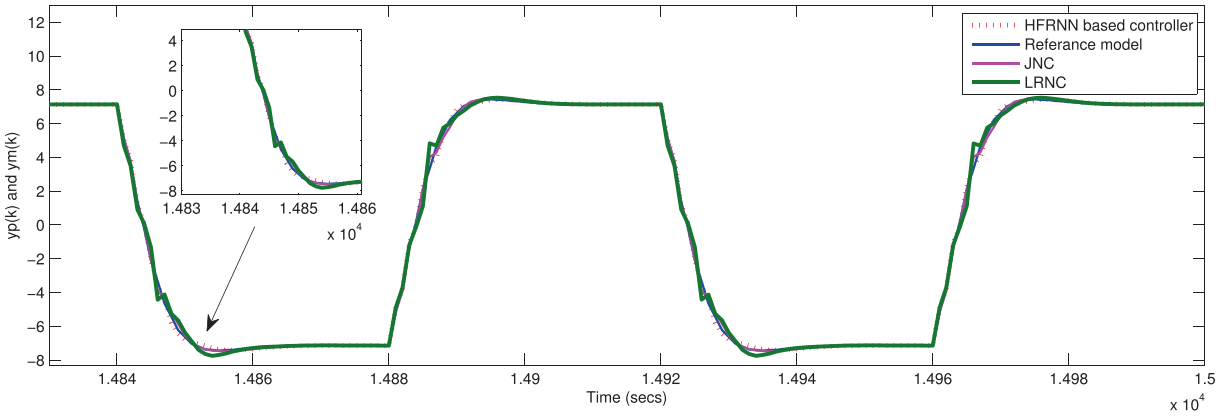


FIGURE 19 Response of controllers with a square waveform as a reference signal (Example 2).

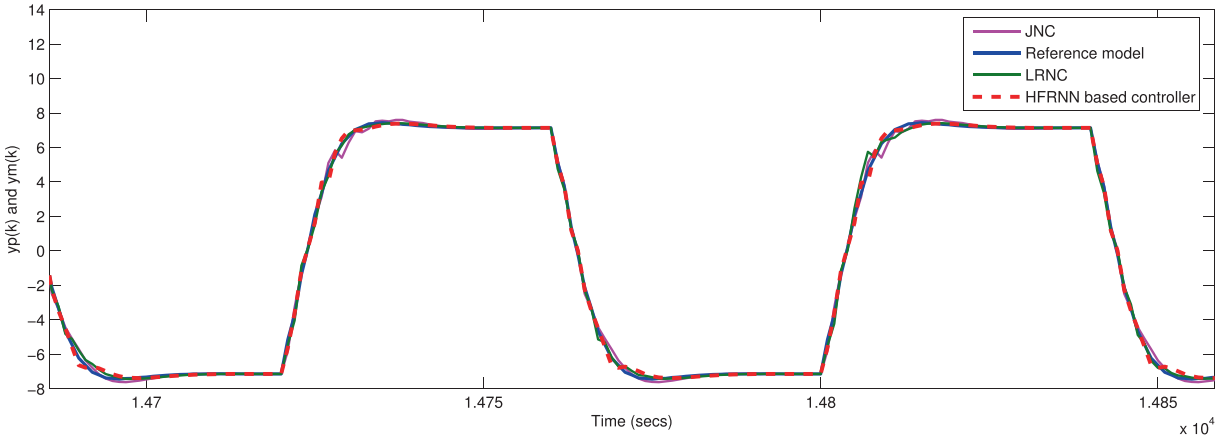


FIGURE 20 Response of controllers for different initial conditions (Example 2).

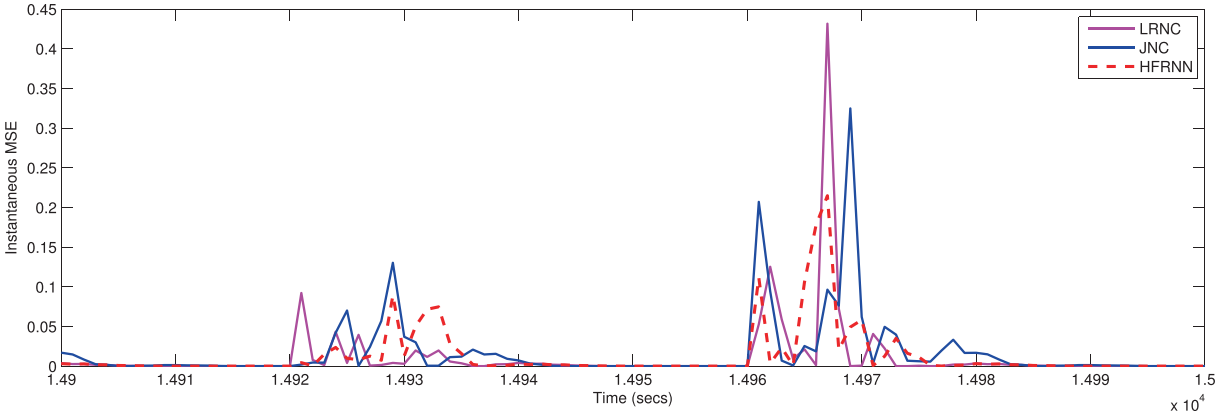


FIGURE 21 Instantaneous MSE of controllers for different initial conditions (Example 2).

controller structures were provided with the same inputs and hidden neurons. The proposed HFRNN though initially took time to adapt to changes, but with iterative learning, it can track the reference model. The MSE obtained for HFRNN also shows the efficiency of the model over JNC and LRNC.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

In this paper, an HFRNN-based indirect adaptive control scheme for the identification-based control of nonlinear dynamic systems. The controller scheme has both identifier and controller blocks. The controller's performance is compared with the Jordan neural controller (JNC) and local recurrent neural controller (LRNC) models. The proposed model gives better MSE than JNC and LRNC for any degree of BIBO system. The proposed model can be used to control unknown plant applications as the sensitivity of the plant is identified online through the HFRNN-based identifier block. The stability and convergence of the updated equations are guaranteed by Lyapunov stability principles. The proposed model is also tested for its recovering ability and its ability to recover from disturbance. The results suggest that the proposed model can be used for real-time nonlinear applications in the future.

8.1 | Limitations and future works

The learning rate is the most important hyperparameter for every neural network. It specifies the extent of correction applied to the learning step. Our future works will focus on replacing the static learning rate with the adaptive one. This paper is also limited to the use of the BP algorithm for network learning due to its simplicity. To improve the overall performance of the learning algorithm, BP can be combined with any other new optimization technique. Our future works will also focus on creating new optimization techniques with the BP method.

CONFLICT OF INTEREST STATEMENT

All three authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Rajesh Kumar https://orcid.org/0000-0001-7172-1081

REFERENCES

Akpan, V. A., & Hassapis, G. D. (2011). Nonlinear model identification and adaptive model predictive control using neural networks. *ISA Transactions*, *50*(2), 177–194.

Alanis, A. Y. (2022). Adaptive neural sensor and actuator fault-tolerant control for discrete-time unknown nonlinear systems. *Franklin Open*, *1*, 9–16.

Anavatti, S. G., Santoso, F., & Garratt, M. A. (2015). Progress in adaptive control systems: past, present, and future. In *2015 international conference on advanced mechatronics, intelligent manufacture, and industrial automation (ICAMIMIA)* (pp. 1–8). IEEE.

Bonassi, F., & Scattolini, R. (2022). Recurrent neural network-based internal model control design for stable nonlinear systems. *European Journal of Control*, *65*, 100632.

Bukhari, A. H., Shoaib, M., Kiani, A. K., Chaudhary, N. I., Raja, M. A. Z., & Shu, C.-M. (2023). Dynamical analysis of nonlinear fractional order Lorenz system with a novel design of intelligent solution predictive radial base networks. *Mathematics and Computers in Simulation*, *213*, 324–347.

Campolucci, P., Uncini, A., Piazza, F., & Rao, B. D. (1999). On-line learning algorithms for locally recurrent neural networks. *IEEE Transactions on Neural Networks*, *10*(2), 253–271.

Chen, Z., Li, Z., & Chen, C. P. (2016). Adaptive neural control of uncertain mimo nonlinear systems with state and input constraints. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(6), 1318–1330.

Cheng, L., Wang, Z., Jiang, F., & Li, J. (2021). Adaptive neural network control of nonlinear systems with unknown dynamics. *Advances in Space Research*, *67*(3), 1114–1123.

Ciliz, M. K. (2009). Combined direct and indirect adaptive control for a class of nonlinear systems. *IET Control Theory and Applications*, *3*(1), 151–159.

Coban, R. (2013). A context layered locally recurrent neural network for dynamic system identification. *Engineering Applications of Artificial Intelligence*, *26*(1), 241–250.

Dashkovskiy, S., & Pavlichkov, S. (2020). Stability conditions for infinite networks of nonlinear systems and their application for stabilization. *Automatica*, *112*, 108643.

Di Ferdinando, M., Pepe, P., & Di Gennaro, S. (2020). A converse lyapunov–krasovskii theorem for the global asymptotic local exponential stability of nonlinear time–delay systems. *IEEE Control Systems Letters*, *5*(1), 7–12.

He, S., Fang, H., Zhang, M., Liu, F., & Ding, Z. (2019). Adaptive optimal control for a class of nonlinear systems: The online policy iteration approach. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(2), 549–558.

Hesamian, G., Johannssen, A., & Chukhrova, N. (2023). Fuzzy nonlinear regression modeling with radial basis function networks. *IEEE Transactions on Fuzzy Systems*, *13*, 1–10. https://doi.org/10.1109/TFUZZ.2023.3332918

Huang, T.-Y., Li, C. J., & Hsu, T.-W. (2007). Structure and parameter learning algorithm of jordan type recurrent neural networks. In *2007 international joint conference on neural networks* (pp. 1819–1824). IEEE.

Jordan, M. (1986). Attractor dynamics and parallelism in a connectionist sequential machine. In *Eighth annual conference of the cognitive science society, 1986* (pp. 513–546). Erlbaum.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

Kapnopoulos, A., Kazakidis, C., & Alexandridis, A. (2024). Quadrotor trajectory tracking based on backstepping control and radial basis function neural networks. *Results in Control and Optimization*, *14*, 100335.

Kumar, R., Srivastava, S., & Gupta, J. (2017). Diagonal recurrent neural network based adaptive control of nonlinear dynamical systems using Lyapunov stability criterion. *ISA Transactions*, *67*, 407–427.

Kumar, R., Srivastava, S., & Gupta, J. (2018). Comparative study of neural networks for control of nonlinear dynamical systems with Lyapunov stability-based adaptive learning rates. *Arabian Journal for Science and Engineering*, *43*, 2971–2993.

Kumar, R., Srivastava, S., Gupta, J., & Mohindru, A. (2018). Self-recurrent wavelet neural network–based identification and adaptive predictive control of nonlinear dynamical systems. *International Journal of Adaptive Control and Signal Processing*, *32*(9), 1326–1358.

Lale, S., Shi, Y., Qu, G., Azizzadenesheli, K., Wierman, A., & Anandkumar, A. (2022). KCRL: Krasovskii-constrained reinforcement learning with guaranteed stability in nonlinear dynamical systems, *arXiv preprint* arXiv:2206.01704.

Li, D.-P., Liu, Y.-J., Tong, S., Chen, C. P., & Li, D.-J. (2018). Neural networks-based adaptive control for nonlinear state constrained systems with input delay. *IEEE Transactions on Cybernetics*, *49*(4), 1249–1258.

Manh, C. N., Manh, T. N., Thi Kim, D. H., Van, Q. N., & Nguyen, T. L. (2023). An adaptive neural network-based controller for car driving simulators. *International Journal of Control*, *96*(1), 82–93.

Narendra, K. S., & Parthasarathy, K. (1992). Neural networks and dynamical systems. *International Journal of Approximate Reasoning*, *6*(2), 109–131.

Narendra, K. S., & Valavani, L. S. (1979). Direct and indirect model reference adaptive control. *Automatica*, *15*(6), 653–664.

Oaks, O. J., & Cook, G. (1976). Piecewise linear control of nonlinear systems. *IEEE Transactions on Industrial Electronics and Control Instrumentation*, *IECI-23* (1), 56–63.

Patan, K., & Patan, M. (2020). Neural-network-based iterative learning control of nonlinear systems. *ISA Transactions*, *98*, 445–453.

Perrusquía, A., & Yu, W. (2021). Identification and optimal control of nonlinear systems using recurrent neural networks and reinforcement learning: An overview. *Neurocomputing*, *438*, 145–154.

Polycarpou, M. M., & Ioannou, P. A. (1991). *Identification and control of nonlinear systems using neural network models: Design and stability analysis*. Citeseer.

Qin, C., Wang, J., Zhu, H., Zhang, J., Hu, S., & Zhang, D. (2022). Neural network-based safe optimal robust control for affine nonlinear systems with unmatched disturbances. *Neurocomputing*, *506*, 228–239.

Şen, G. D., Günel, G. Ö., & Güzelkaya, M. (2020). Extended Kalman filter based modified Elman-Jordan neural network for control and identification of nonlinear systems. In *2020 Innovations in intelligent systems and applications conference (ASYU)* (pp. 1–6). IEEE.

Slama, S., Errachdi, A., & Benrejeb, M. (2019). Neural adaptive PID and neural indirect adaptive control switch controller for nonlinear mimo systems. *Mathematical Problems in Engineering*, *2019*, 7340392.

Swarnkar, P., Jain, S. K., & Nema, R. (2014). Adaptive control schemes for improving the control system dynamics: a review. *IETE Technical Review*, *31*(1), 17–33.

Tavoosi, J. (2020). Sliding mode control of a class of nonlinear systems based on recurrent type-2 fuzzy RBFN. *International Journal of Mechatronics and Automation*, *7*(2), 72–80.

Truong, H. V. A., Nguyen, M. H., Tran, D. T., & Ahn, K. K. (2023). A novel adaptive neural network-based time-delayed estimation control for nonlinear systems subject to disturbances and unknown dynamics. *ISA Transactions*, *142*, 214–227.

Wang, J., Gao, Y., Qiu, J., & Ki Ahn, C. (2017). Sliding mode control for non-linear systems by Takagi–Sugeno fuzzy model and delta operator approaches. *IET Control Theory and Applications*, *11*(8), 1205–1213.

Wysocki, A., & Lawryńczuk, M. (2015). Jordan neural network for modelling and predictive control of dynamic systems. In *2015 20th international conference on methods and models in automation and robotics (MMAR)* (pp. 145–150). IEEE.

Xu, K., Wang, H., & Liu, P. X. (2023). Adaptive fixed-time control for high-order stochastic nonlinear time-delay systems: An improved lyapunov–krasovskii function. *IEEE Transactions on Cybernetics*, *54*, 776–786.

Yoo, S. J., Park, J. B., & Choi, Y. H. (2007). Indirect adaptive control of nonlinear dynamic systems using self recurrent wavelet neural networks via adaptive learning rates. *Information Sciences*, *177*(15), 3074–3098.

Zhao, C., & Guo, L. (2021). Control of nonlinear uncertain systems by extended PID. *IEEE Transactions on Automatic Control*, *66*(8), 3840–3847.

Zheng, D.-D., Pan, Y., Guo, K., & Yu, H. (2019). Identification and control of nonlinear systems using neural networks: A singularity-free approach. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(9), 2696–2706.

Zhou, G., & Tan, D. (2023). Review of nuclear power plant control research: Neural network-based methods. *Annals of Nuclear Energy*, *181*, 109513.

Zhou, P., Hu, X., Zhu, Z., & Ma, J. (2021). What is the most suitable Lyapunov function? *Chaos, Solitons & Fractals*, *150*, 111154.

Zhou, Z., Tong, D., Chen, Q., Zhou, W., & Xu, Y. (2021). Adaptive NN control for nonlinear systems with uncertainty based on dynamic surface control. *Neurocomputing*, *421*, 161–172.

AUTHOR BIOGRAPHIES

R. Shobana graduated in Electronics and Instrumentation Engineering from Anna University, Chennai, India. She has a Master's degree in control and Instrumentation from Anna University, Chennai, India. She is pursuing her Ph.D. in the electrical engineering department at Delhi Technological University, New Delhi. She is currently working as an Assistant Professor in GCET, Greater Noida, and has experience of 9 years in the field of teaching. Her research interest includes the application of Artificial Neural networks in modelling and control.

Rajesh Kumar is currently working as an Assistant Professor [Grade-I] in the Department of Electrical Engineering at the National Institute of Technology Kurukshetra, Kurukshetra-136119, India. Before this, he worked as an Assistant Professor in the Department of Electrical Department at Delhi Technological University (DTU), New Delhi. Before joining DTU, he worked as an Assistant Professor in the Electrical & Instrumentation department of Thapar Institute of Engineering and Technology (deemed to be a university), in Patiala, India. He received a B. Tech. degree in Applied Electronics & Instrumentation Engineering from Manav Rachna College of Engineering, Faridabad (now known as Manav

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

Rachna University, Faridabad), India (M.D.U Rohak) in 2011, M. Tech degree in Electrical Engineering (Control Systems) from National Institute of Technology, Kurukshetra, India in 2013, and a Ph.D. degree in the area of Intelligent modelling and control of nonlinear dynamical systems using artificial neural networks from Netaji Subhas Institute of Technology (now known as Netaji Subhas University of Technology, New Delhi), University of Delhi, India in 2019. His current research interests include Soft Computing, Intelligent modeling, and control, Artificial Neural Networks, and Optimization. He is a senior member of IEEE and a fellow of IETE India. He is currently serving as the editor for various reputed international journals such as: 1. Editorial Board Member (Associate Editor) in Complex Intelligent Systems Journal – Springer (SCIE). 2. Associate Editor in Applied Intelligence-Springer (SCIE). 3. Editorial Board Member in Intelligent Automaton & soft computing – Tech Science Press (SCIE). 4. Associate Editor in Results in Control and Optimization Journal- Elsevier (Scopus). 5. Academic Editor in PLOS ONE Journal (Scopus). 6. Editorial Board Member in Automatic Control and Computer Sciences- Springer-Emerging Source Citation Index (ESCI). 7. Academic Editor in Computational and Mathematical Methods-Wiley (ESCI) 8. Associate Editor in Journal of Electrical Engineering & Technology – Springer (SCIE) 9. Associate Editor (Editorial Board Member) in Ain Shams Engineering Journal – Elsevier (SCIE) He is also a reviewer for many reputed international Journals and transactions. He has authored so far 26 SCIE transactions/journal research papers, many book chapters, and international conferences.

Bhavnesh Jaint did her B.E. in Electronics & Telecommunication Engineering in 2000 & M.E. in Control Systems in 2003 from Government Engineering College, Jabalpur, India. She did her Ph.D. in control systems from Delhi Technological University (D.T.U), New Delhi, India in 2020. She is currently working as an assistant professor in the Department of Electrical Engineering, DTU, New Delhi. Her areas of interest include control systems, wireless sensor networks, and soft computing.

How to cite this article: Shobana, R., Kumar, R., & Jaint, B. (2024). Nonlinear dynamical system approximation and adaptive control based on hybrid-feed-forward recurrent neural network: Simulation and stability analysis. *Expert Systems*, e13619. https://doi.org/10.1111/exsy.13619

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13619 by Delhi Technological University, Wiley Online Library on [21/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

arXiv:2405.07995v1 [math.CV] 22 Apr 2024

ON ESTIMATION OF HANKEL DETERMINANTS FOR CERTAIN CLASS OF STARLIKE FUNCTIONS

S. SIVAPRASAD KUMAR AND NEHA VERMA

Abstract. In the present study, we consider two subclasses starlike and convex functions, denoted by $\mathcal{S}^*_{\mathcal{B}}$ and $\mathcal{C}_{\mathcal{B}}$ respectively, associated with a bean-shaped domain. Further, we estimate certain sharp initial coefficients, as well as second, third and fourth-order Hankel determinants for functions belonging to the class $\mathcal{S}^*_{\mathcal{B}}$. Additionally, we compute sharp second and third-order Hankel determinants for functions belonging to the $\mathcal{C}_{\mathcal{B}}$ class.

1. Introduction

Let $\mathcal{A}$ denote the class of normalized analytic functions defined on the open unit disk $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$, having the form

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n \tag{1.1}$$

and suppose $\mathcal{S}$ be a subclass of $\mathcal{A}$ comprising univalent functions. Consider $\mathcal{P}$ to be the class of analytic functions defined on $\mathbb{D}$ with a positive real part, expressed as $p(z) = 1 + \sum_{n=1}^{\infty} p_n z^n$. Suppose h and g are two analytic functions, we say h is subordinate to g, symbolically $h \prec g$, if there exists a Schwarz function w with $w(0) = 0$ and $|w(z)| \leq |z|$ such that $h(z) = g(w(z))$. A substantial body of literature exists on coefficient problems, ranging from the seminal Bieberbach's conjecture of 1916 to contemporary research (see [5]). There are two prominent subclasses of $\mathcal{S}$ consisting of starlike and convex functions, respectively denoted by $\mathcal{S}^*$ and $\mathcal{C}$. Further, in 1992, Ma and Minda [14] unfified various subclasses of $\mathcal{S}^*$ and $\mathcal{C}$ by introducing the following two classes:

$$\mathcal{S}^*(\varphi) = \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec \varphi(z) \right\} \tag{1.2}$$

and

$$\mathcal{C}(\varphi) = \left\{ f \in \mathcal{A} : 1 + \frac{zf''(z)}{f'(z)} \prec \varphi(z) \right\} \tag{1.3}$$

where φ is an analytic univalent function satisfying the conditions $\operatorname{Re}\varphi(z) > 0$, $\varphi(\mathbb{D})$ symmetric about the real axis and starlike with respect to $\varphi(0) = 1$ with $\varphi'(0) > 0$. Recently, many Ma-Minda classes are introduced and studied by several authors by appropriately choosing $\varphi(z)$ in (1.2). See the various Ma-Minda subclasses of starlike functions listed in the first column with the corresponding choice of $\varphi(z)$ in second column of Table 1.

The concept of Hankel determinants, introduced in 1966 (see [16]), continues to be a topic of significant interest for researchers today. The definition of the qth Hankel determinant $H_q(n)$ of

2010 *Mathematics Subject Classification.* 30C45, 30C50.

Key words and phrases. Coefficient problems, Bean-shaped domain, Starlike functions, Hankel determinants.

TABLE 1. List of sharp third order Hankel determinants

Class	$\varphi(z)$	Sharp $\lvert H_3(1)\rvert$	Reference
$\mathcal{S}^*$	$(1+z)/(1-z)$	$4/9$	[3, 7]
$\mathcal{S}^*_\varrho$	$1+ze^z$	$1/9$	[19]
$\mathcal{SL}^*$	$\sqrt{1+z}$	$1/36$	[4]
$\mathcal{S}^*_e$	e^z	$1/9$	[20]
$\mathcal{S}^*_\rho$	$1+\sinh^{-1}(z)$	$1/9$	[10]
$\mathcal{S}^*_\tau$	$1+\arctan z$	$1/9$	[9]

analytic functions $f \in \mathcal{A}$, under the assumption that $a_1 := 1$, is as follows:

$$H_q(n) = \begin{vmatrix} a_n & a_{n+1} & \dots & a_{n+q-1} \\ a_{n+1} & a_{n+2} & \dots & a_{n+q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n+q-1} & a_{n+q} & \dots & a_{n+2q-2} \end{vmatrix}, \quad n, q \in \mathbb{N}. \tag{1.4}$$

The expressions for the second and third-order Hankel determinants for specific values of q and n, are denoted by $H_2(3)$ and $H_3(1)$, respectively, given by

$$H_2(3) := a_3a_5 - a_4^2, \tag{1.5}$$

and

$$H_3(1) := 2a_2a_3a_4 - a_3^3 - a_4^2 - a_2^2a_5 + a_3a_5. \tag{1.6}$$

Deriving a sharp bound for Hankel determinants is a formidable challenge, prompting numerous researchers to endeavor to do so for various subclasses of starlike functions, see [1, 2, 4, 6, 15] and some are listed in third column of Table 1.

Recently, Kumar and Yadav [11], by choosing $\varphi(z) = \sqrt{1+\tanh z}$, introduced and studied the Ma-Minda subclass of starlike functions $\mathcal{S}^*_{\mathcal{B}}$ associated with a bean-shaped domain, given by

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

Motivated by it, we introduce the following convex counterpart of the above class:

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

The authors have investigated the geometric properties of the univalent function $\sqrt{1+\tanh z}$, along with some inclusion and sharp radius results involving $\mathcal{S}^*_{\mathcal{B}}$, as well as implications of first-order differential subordination. This class can be further studied to know more about the behavior of the coefficients of functions belonging to this class, second and third order differential subordination, etc. Thus, studying such subclasses of starlike and convex functions opens new avenues in the field of research. Taking this aspect in account, in our current investigation, we focus on coefficient-related problems concerning the aforementioned classes $\mathcal{S}^*_{\mathcal{B}}$ and $\mathcal{C}_{\mathcal{B}}$, which are yet not addressed in the literature. In fact, we are finding the sharp bounds of the initial coefficients, second and third-order Hankel determinants, as well as the bound of the fourth-order Hankel determinant for functions belonging to the class $\mathcal{S}^*_{\mathcal{B}}$. Additionally, we establish sharp bounds of the second and third-order Hankel determinants for functions belonging to the class $\mathcal{C}_{\mathcal{B}}$.

2. Coefficient related problems for $\mathcal{S}^*_{\mathcal{B}}$

In this section, we start by determining the sharp bounds of the initial coefficients a_i for $(i = 2, 3, 4, 5)$ followed by establishing the sharp bounds of the second and third-order Hankel determinants for functions $f \in S^*_{\mathcal{B}}$. Subsequently, we derive the bounds for a_6 and a_7 to deduce a possible bound of the fourth-order Hankel determinant for functions $f \in S^*_{\mathcal{B}}$.

2.1. **Sharp initial coefficient bounds.** Let $f \in \mathcal{S}^*_{\mathcal{B}}$, then there exists a Schwarz function $w(z)$ such that

$$\frac{zf'(z)}{f(z)} = \sqrt{1 + \tanh w(z)}. \tag{2.1}$$

Suppose that $p(z) = 1 + p_1 z + p_2 z^2 + \cdots \in \mathcal{P}$ and consider $w(z) = (p(z) - 1)/(p(z) + 1)$. Further, by substituting the expansions of $w(z)$, $p(z)$ and $f(z)$ in equation (2.1) and then comparing the coefficients, we obtain the expressions of $a_i (i = 2, 3, ..., 7)$ in terms of $p_j (j = 1, 2, ..., 5)$, given as

$$a_2 = \frac{p_1}{4}, \quad a_3 = \frac{1}{64}\Big(8p_2 - 3p_1^2\Big), \quad a_4 = \frac{1}{2304}\Big(23p_1^3 - 168p_1p_2 + 192p_3\Big), \tag{2.2}$$

$$a_5 = \frac{1}{18432}\Big(- 11p_1^4 + 528p_1^2p_2 - 576p_2^2 - 1056p_1p_3 + 1152p_4\Big), \tag{2.3}$$

$$a_6 = \frac{1}{1843200}\Big(50880p_1p_2^2 - 2367p_1^5 - 8560p_1^3p_2 + 46080p_1^2p_3 - 96000p_2p_3 - 86400p_1p_4 + 92160p_5\Big) \tag{2.4}$$

and

$$a_7 = \frac{1}{530841600}\Big(601421p_1^6 - 2365320p_1^4p_2 - 4818240p_1^2p_2^2 + 4723200p_2^3 + 26150400p_1p_2p_3 \\ - 2648640p_1^3p_3 - 11980800p_3^2 + 11577600p_1^2p_4 - 23500800p_2p_4 - 21012480p_1p_5\Big). \tag{2.5}$$

The results stated below are necessary for proving our main result.

Lemma 2.1. *[13] Let $p \in \mathcal{P}$ be of the form $1 + \sum_{n=1}^{\infty} p_n z^n$. Then*

$$|p_1^4 - 3p_1^2p_2 + p_2^2 + 2p_1p_3 - p_4| \leq 2 \tag{2.6}$$

and

$$|p_3 - 2p_1p_2 + p_1^3| \leq 2. \tag{2.7}$$

Lemma 2.2. *[14] Let $p \in \mathcal{P}$ be of the form $1 + \sum_{n=1}^{\infty} p_n z^n$. Then*

$$|p_2 - \beta p_1^2| \leq \begin{cases} 2 - 4\beta, & \beta \leq 0; \\ 2, & 0 \leq \beta \leq 1; \\ 4\beta - 2, & \beta \geq 1 \end{cases}$$

when $\beta < 0$ or $\beta > 1$, the equality holds if and only if $p(z) = (1+z)/(1-z)$ or one of its rotations. If $0 < \beta < 1$, then the inequality holds if and only if $p(z) = (1+z^2)/(1-z^2)$ or one of its rotations. If $\beta = 0$, the equality holds if and only if $p(z) = (1+\eta)(1+z)/(2(1-z))+(1-\eta)(1-z)/(2(1+z))(0 \leq \eta \leq 1)$ or one of its rotations. If $\beta = 1$, the equality holds if and only if p is the reciprocal of one of the functions such that the equality holds in case of $\beta = 0$. Though the above upper bound is sharp for $0 < \beta < 1$, still it can be improved as follows:

$$|p_2 - \beta p_1^2| + \beta|p_1|^2 \leq 2 \quad (0 < \beta \leq 1/2) \tag{2.8}$$

and

$$|p_2 - \beta p_1^2| + (1 - \beta)|p_1|^2 \leq 2 \quad (1/2 < \beta \leq 1).$$

Also, we recall that

$$\max_{0\le t\le 4}(At^2+Bt+C)=\begin{cases} C, & B\le 0, A\le \frac{-B}{4};\\ 16A+4B+C, & B\ge 0, A\ge \frac{-B}{8} \quad \text{or} \quad B\le 0, A\ge \frac{-B}{4};\\ \dfrac{4AC-B^2}{4A}, & B>0, A\le \frac{-B}{8}. \end{cases} \tag{2.9}$$

Theorem 2.3. *If $f\in\mathcal{S}^*_{\mathcal{B}}$, then (i) $|a_2|\le 1/2$, (ii) $|a_3|\le 1/4$, (iii) $|a_4|\le 1/6$ and (iv) $|a_5|\le 847/3216\simeq 0.263371\cdots$. These bounds are sharp.*

Proof. (*i*) Since $|p_n|\le 2$ for $n\ge 1$, therefore, from (2.2), $|a_2|\le 1/2$.
(*ii*) For a_3, we use the inequality $|p_2-\mu p_1^2|\le 2\max\{1,|2\mu-1|\}$ given by Ma and Minda [14], which yields $|a_3|\le 1/4$.
(*iii*) For a_4, (2.1) is re-written as:

$$zf'(z)=\sqrt{1+\tanh(w(z))}f(z). \tag{2.10}$$

On substituting $f(z)=z+\sum_{n=2}^{n=\infty}a_nz^n$ and $w(z)=\sum_{k=1}^{\infty}w_kz^k$ in (2.10) and comparing the coefficients of z^4, we get

$$6a_4=\left(w_3+\frac{1}{4}w_1w_2-\frac{13}{48}w_1^3\right).$$

Now using [17, Lemma 2, p.128], we deduce that $|6a_4|\le 1$ and hence the result follows.
(*iv*) From (2.3), we get the expression of a_5 as

$$\begin{aligned} a_5&=\frac{1}{16}\bigg(-\frac{11}{1152}p_1^4+\frac{11}{24}p_1^2p_2-\frac{p_2^2}{2}-\frac{11}{12}p_1p_3+p_4\bigg)\\ &=\frac{1}{16}\bigg(-\frac{1}{2}P+\frac{1}{12}p_1Q-\frac{7}{8}p_1^2R+\frac{1}{2}p_4\bigg), \end{aligned}$$

which further gives

$$|a_5|\le\frac{1}{16}\left(\frac{1}{2}|P|+\frac{1}{12}|p_1||Q|+\frac{7}{8}|p_1|^2|R|+\frac{1}{2}|p_4|\right),$$

where $P=p_1^4-3p_1^2p_2+p_2^2+2p_1p_3-p_4$, $Q=p_3-2p_1p_2+p_1^3$ and $R=p_2-(67/144)p_1^2$. Moreover, using the bounds of $|P|\le 2$ from (2.6), $|Q|\le 2$ from (2.7) and $|R|\le 2$ from (2.8), respectively, we obtain

$$|a_5|\le\frac{1}{16}\left(\frac{7}{3}+\frac{7}{4}|p_1|^2-\frac{469}{1152}|p_1|^4\right).$$

Now, we obtain $|7|p_1|^2/4-469|p_1|^4/1152|\le 126/67$ using (2.9) by taking $A=-469/1152$, $B=7/4$ and $C=0$, which leads to the desired estimate for $|a_5|$. The sharpness of the result can be witnessed when $p_1=p_4=2$, $p_2=(737+\sqrt{963326})/402$ and $p_3=-2$. The function

$$f_n(z)=z\exp\left(\int_0^z\frac{\sqrt{1+\tanh(t^{n-1})}-1}{t}dt\right)$$

acts as the extremal function for the initial coefficients a_n for $n=2,3$ and 4. ∎

The formula for p_j $(j=2,3,4)$, which is included in the Lemma 2.4 below, plays a vital role in establishing the sharp bounds for Hankel determinants and forms the foundation for our main results.

Lemma 2.4. *[12, 13] Let $p\in\mathcal{P}$ has the form $1+\sum_{n=1}^{\infty}p_nz^n$. Then for some γ, η and ρ such that $|\gamma|\le 1$, $|\eta|\le 1$ and $|\rho|\le 1$, we have*

$$2p_2=p_1^2+\gamma(4-p_1^2), \tag{2.11}$$

$$4p_3 = p_1^3 + 2p_1(4-p_1^2)\gamma - p_1(4-p_1^2)\gamma^2 + 2(4-p_1^2)(1-|\gamma|^2)\eta, \tag{2.12}$$

and

$$8p_4 = p_1^4 + (4-p_1^2)\gamma(p_1^2(\gamma^2-3\gamma+3)+4\gamma) - 4(4-p_1^2)(1-|\gamma|^2)(p_1(\gamma-1)\eta + \bar{\gamma}\eta^2 - (1-|\eta|^2)\rho). \tag{2.13}$$

2.2. **Sharp Hankel determinants for $\mathcal{S}^*_{\mathcal{B}}$.** The following theorem presents the sharp bound for $|H_3(1)|$ for functions belonging to the class $\mathcal{S}^*_{\mathcal{B}}$.

Theorem 2.5. *Let $f \in \mathcal{S}^*_{\mathcal{B}}$, then*

$$|H_3(1)| \leq 1/36. \tag{2.14}$$

This result is sharp.

Proof. Since the class $\mathcal{P}$ is invariant under rotation, we can take $p_1 =: p$ belongs to the interval [0,2]. Substitute the values of $a_i (i = 2, 3, 4, 5)$ in (1.6) from (2.2) and (2.3). We get

$$H_3(1) = \frac{1}{21233664}\bigg(-3511p^6 - 5160p^4p_2 - 14400p^2p_2^2 - 124416p_2^3 + 56256p^3p_3 + 216576pp_2p_3 - 147456p_3^2 - 145152p^2p_4 + 165888p_2p_4\bigg).$$

After simplifying the calculations using (2.11)-(2.13), we obtain

$$H_3(1) = \frac{1}{21233664}\bigg(\beta_1(p,\gamma) + \beta_2(p,\gamma)\eta + \beta_3(p,\gamma)\eta^2 + \beta_4(p,\gamma,\eta)\rho\bigg),$$

for $\gamma, \eta, \rho \in \mathbb{D}$. Here

$$\begin{aligned}
\beta_1(p,\gamma) :&= -1099p^6 - 1872\gamma^2p^2(4-p^2)^2 - 20736\gamma^3(4-p^2)^2 - 5760\gamma^3p^2(4-p^2)^2 + 1152\gamma^4p^2(4-p^2)^2 \\
&\quad + 3084\gamma p^4(4-p^2) + 624p^4\gamma^2(4-p^2) - 7776p^4\gamma^3(4-p^2) - 31104\gamma^2p^2(4-p^2),\\
\beta_2(p,\gamma) :&= 96(1-|\gamma|^2)(4-p^2)(149p^3 + 324\gamma p^3 + 228p\gamma(4-p^2) - 48p\gamma^2(4-p^2)),\\
\beta_3(p,\gamma) :&= 1152(1-|\gamma|^2)(4-p^2)(-32(4-p^2) - 4|\gamma|^2(4-p^2) + 27p^2\bar{\gamma}),\\
\beta_4(p,\gamma,\eta) :&= 10368(1-|\gamma|^2)(4-p^2)(1-|\eta|^2)(4(4-p^2)\gamma - 3p^2).
\end{aligned}$$

By choosing $x = |\gamma|$, $y = |\eta|$ and utilizing the fact that $|\rho| \leq 1$, the above expression reduces to the following:

$$|H_3(1)| \leq \frac{1}{21233664}\bigg(|\beta_1(p,\gamma)| + |\beta_2(p,\gamma)|y + |\beta_3(p,\gamma)|y^2 + |\beta_4(p,\gamma,\eta)|\bigg) \leq A(p,x,y),$$

where

$$A(p,x,y) = \frac{1}{21233664}\bigg(a_1(p,x) + a_2(p,x)y + a_3(p,x)y^2 + a_4(p,x)(1-y^2)\bigg), \tag{2.15}$$

with

$$\begin{aligned}
a_1(p,x) :&= 1099p^6 + 1872x^2p^2(4-p^2)^2 + 20736x^3(4-p^2)^2 + 5760x^3p^2(4-p^2)^2 + 1152x^4p^2(4-p^2)^2 \\
&\quad + 3084xp^4(4-p^2) + 624p^4x^2(4-p^2) + 7776p^4x^3(4-p^2) + 31104x^2p^2(4-p^2),\\
a_2(p,x) :&= 96(1-x^2)(4-p^2)(149p^3 + 324xp^3 + 228px(4-p^2) + 48px^2(4-p^2)),\\
a_3(p,x) :&= 1152(1-x^2)(4-p^2)(32(4-p^2) + 4x^2(4-p^2) + 27p^2x),\\
a_4(p,x) :&= 10368(1-x^2)(4-p^2)(4x(4-p^2) + 3p^2).
\end{aligned}$$

In the closed cuboid $Q : [0,2] \times [0,1] \times [0,1]$, we now maximise $A(p,x,y)$, by locating the maximum values in the interior of the six faces, on the twelve edges, and in the interior of Q.

(1) We start by taking into account every internal point of Q. Assume that $(p,x,y) \in (0,2)\times(0,1)\times(0,1)$. We calculate $\partial A/\partial y$, partially differentiate (2.15) with respect to y to identify the points of maxima in the interior of Q. We get

$$\frac{\partial A}{\partial y} = \frac{(4-p^2)(1-x^2)}{21233664}\bigg(24px(5+2x)+p^3(17+24x-12x^2)+96(8-9x+x^2)y - 12p^2(25-27x+2x^2)y\bigg).$$

Now $\partial A/\partial y = 0$ gives

$$y = y_0 := \frac{48px(19+4x)+p^3(149+96x-48x^2)}{24(1-x)(16(x-8)+p^2(59-4x))}.$$

The existence of critical points requires that y_0 belong to $(0,1)$, which is only possible when

$$300p^2+864x+24p^2x^2 > 17p^3+120px+24p^3x+48px^2-12p^3x^2 + 768+864x+24p^2x^2. \tag{2.16}$$

Now, we find the solution satisfying the inequality (2.16) for the existence of critical points using the hit and trial method. If we assume p tends to 0, then there does not exist any $x \in (0,1)$ satisfying (2.16). But, when p tends to 2, (2.16) holds for all $x < 175/648$. We also observe that there does not exist any $p \in (0,2)$ when $x \in (174/648, 1)$. Similarly, if we assume x tends to 0, then for all $p > 1.61687$, (2.16) holds. After calculations, we observe that there does not exist any $x \in (0,1)$ when $p \in (0, 1.61687)$. Thus, the domain for the solution of the equation is $(1.61687, 2)\times(0, 175/648)$. Now, we examine that $\frac{\partial A}{\partial y}|_{y=y_0} \neq 0$ in $(1.61687,2)\times(0,175/648)$. So, we conclude that the function A has no critical point in $(0,2)\times(0,1)\times(0,1)$.

(2) The interior of each of the cuboid Q's six faces is now being considered.

<u>On $p = 0$</u>

$$k_1(x,y) := \frac{16y^2-14x^2y^2-2x^4y^2-9x^3(1-2y^2)+18x(1-y^2)}{576}, \quad x,y \in (0,1). \tag{2.17}$$

Since

$$\frac{\partial k_1}{\partial y} = \frac{(1-x^2)(x+1)(8-x)y}{144} \neq 0, \quad x,y \in (0,1),$$

indicates that k_1 has no critical points in $(0,1)\times(0,1)$.

<u>On $p = 2$</u>

$$A(2,x,y) := \frac{1099}{331776} = 0.00331248, \quad x,y \in (0,1). \tag{2.18}$$

<u>On $x = 0$</u>

$$k_2(p,y) := \frac{1099p^6+(4-p^2)(14304p^3y+36864y^2(4-p^2)+31104p^2(1-y^2)}{21233664} \tag{2.19}$$

with $p \in (0,2)$ and $y \in (0,1)$. To determine the points of maxima, we solve $\partial k_2/\partial p = 0$ and $\partial k_2/\partial y = 0$. After solving $\partial k_2/\partial y = 0$, we get

$$y = \frac{149p^3}{24(59p^2-128)} (=: y_p). \tag{2.20}$$

In order to have $y_p \in (0,1)$ for the given range of y, $p_0 := p > \approx 1.61687$ is required. Based on calculations, $\partial k_2/\partial p = 0$ gives

$$p(41472-20736p^2+1099p^4+28608py-11920p^3y-139776y^2+45312p^2y^2) = 0. \tag{2.21}$$

After substituting (2.20) in (2.21), we have

$$75497472p - 107347968p^3 + 50314752p^5 - 8246384p^7 + 133989p^9 = 0. \tag{2.22}$$

A numerical calculation suggests that $p \approx 1.24748 \in (0,2)$ is the solution of (2.22). So, we conclude that k_2 does not have any critical point in $(0,2)\times(0,1)$.

On $x = 1$

$$A(p,1,y) = k_3(p,y) := \frac{331776 + 99072p^2 - 34704p^4 - 1601p^6}{21233664}, \quad p \in (0,2). \tag{2.23}$$

While computing $\partial k_3/\partial p = 0$, $p_0 := p \approx 1.14405$ comes out to be the critical point. Undergoing simple calculations, k_3 achieves its maximum value ≈ 0.0187629 at p_0.

On $y = 0$

$$k_4(p,x) := \frac{1}{21233664}\bigg(-1601p^6 - 331776(-1-2x+2x^3) + 2304p^2(97 - 144x - 54x^2 + 144x^3) - 144p^4(457 - 288x - 216x^2 + 288x^3)\bigg).$$

After further calculations such as,

$$\frac{\partial k_4}{\partial x} = \frac{(4-p^2)}{1024}\bigg(8 - 24x^2 + p^2(-2-3x+6x^2)\bigg)$$

and

$$\frac{\partial k_4}{\partial p} = \frac{1}{3538944}\bigg(-1601p^5 + 768p(97 - 144x - 54x^2 + 144x^3) - 96p^3(457 - 288x - 216x^2 + 288x^3)\bigg),$$

we observe that only real solutions (p,x) of the system of equations $\partial k_4/\partial x = 0$ and $\partial k_4/\partial p = 0$ are $(2, 1.74724)$, $(2, -1.74724)$, $(-1.36584, -0.835809)$, $(1.36584, -0.835809)$, $(2, -1.74724)$ and $(-0.854598, 0.524203)$. Thus, no solution exists in $(0,2)\times(0,1)$, resulting in no critical points.

On $y = 1$

$$k_5(p,x) := \frac{1}{21233664}\bigg\{1099p^6 + 31104p^2(4-p^2) + 11484p^4(4-p^2) + 20736(4-p^2)^2 + 8784p^2(4-p^2)^2 - 1152(4-p^2)(1-x^2)\bigg(-16(8+x^2) + p^2(32 - 27x + 4x^2)\bigg) + 96(4-p^2)(1-x^2)\bigg(48px(19+4x) + p^3(149 + 96x - 48x^2)\bigg)\bigg\}.$$

Simple calculations leads to

$$\frac{\partial k_5}{\partial x} = \frac{(4-p^2)}{110592}\bigg(-192x(7+2x^2) + 6p^2(27 + 56x - 81x^2 + 16x^3) + 24p(19 + 8x - 57x^2 - 16x^3) + p^3(48 - 197x - 144x^2 + 96x^3)\bigg)$$

and

$$\frac{\partial k_5}{\partial p} = \frac{1}{3538944}\Big(-1601p^5 - 3072x(-19 - 4x + 19x^2 + 4x^3) + 768p(-85 + 54x + 112x^2 - 54x^3 + 16x^4) - 96p^3(-15 + 216x + 224x^2 - 216x^3 + 32x^4) - 80p^4(149 + 96x - 197x^2 - 96x^3 + 48x^4) + 192p^2(149 - 132x - 245x^2 + 132x^3 + 96x^4)\Big).$$

We note that the only real solutions (p, x) of the system of equations $\partial k_5/\partial x = 0$ and $\partial k_5/\partial p = 0$ are $(-5.5858, 2.7083)$, $(2, -2.2645)$, $(-2, 0.357662)$, $(-1.98983, 0.350993)$, $(0.932759, -1.56488)$ and $(-1.03049, -0.27789)$. Thus, no solution exists in $(0, 2) \times (0, 1)$, resulting in no critical points.

(3) We next examine the maxima attained by $A(p, x, y)$ on the edges of the cuboid Q. From (2.19), we have $A(p, 0, 0) = e_1(p) := (124416p^2 - 31104p^4 + 1099p^6)/21233664$. It is easy to observe that $e_1'(p) = 0$ whenever $p = \delta_0 := 0$ and $p = \delta_1 := 1.50801 \in [0, 2]$ as its points of minima and maxima respectively. Hence,

$$A(p, 0, 0) \leq 0.00635802, \quad p \in [0, 2].$$

Now considering (2.19) at $y = 1$, we get $A(p, 0, 1) = e_2(p) := (589824 - 294912p^2 + 57216p^3 + 36864p^4 - 14304p^5 + 1099p^6)/21233664$. It is easy to observe that $e_2'(p) < 0$ in $[0, 2]$ and hence $p = 0$ serves as the point of maxima. So,

$$A(p, 0, 1) \leq \frac{1}{36} \approx 0.0277778, \quad p \in [0, 2].$$

Through computations, (2.19) shows that $A(0, 0, y)$ attains its maxima at $y = 1$. This implies that

$$A(0, 0, y) \leq \frac{1}{36}, \quad y \in [0, 1].$$

Since, (2.23) does not involve x, we have $A(p, 1, 1) = A(p, 1, 0) = e_3(p) := (331776 + 99072p^2 - 34704p^4 - 1601p^6)/21233664$. Now, $e_3'(p) = 33024p - 23136p^3 - 1601p^5 = 0$ when $p = \delta_2 := 0$ and $p = \delta_3 := 1.14405$ in the interval $[0, 2]$ with δ_2 and δ_3 as points of minima and maxima, respectively. Hence

$$A(p, 1, 1) = A(p, 1, 0) \leq 0.0187629, \quad p \in [0, 2].$$

After considering $p = 0$ in (2.23), we get, $A(0, 1, y) = 1/64$. The equation (2.18) has no variables. So, on the edges, the maximum value of $M(p, x, y)$ is

$$A(2, 1, y) = A(2, 0, y) = A(2, x, 0) = A(2, x, 1) = \frac{1099}{331776}, \quad x, y \in [0, 1].$$

Using (2.17), we obtain $A(0, x, 1) = e_4(x) := (16 - 14x^2 + 9x^3 - 2x^4)/576$. Upon calculations, we see that $e_4(x)$ is a decreasing function in $[0, 1]$ and attains its maxima at $x = 0$. Thus

$$A(0, x, 1) \leq \frac{1}{36}, \quad x \in [0, 1].$$

Again utilizing (2.17), we get $A(0, x, 0) = e_5(x) := x(2 - x^2)/64$. On further calculations, we get $e_5'(x) = 0$ for $x = \delta_4 := \sqrt{2/3}$. Also, $e_5(x)$ is an increases in $[0, \delta_4)$ and decreases in $(\delta_4, 1]$. So, it reaches its maximum value at δ_4. Thus

$$A(0, x, 0) \leq 0.0170103, \quad x \in [0, 1].$$

Given all the cases, the inequality (2.14) holds. Let the function $f_0(z) \in \mathcal{S}^*_{\mathcal{B}}$, be defined as

$$f_0(z) = z\exp\left(\int_0^z \frac{\sqrt{1+\tanh t^3}-1}{t}dt\right) = z + \frac{z^4}{6} - \frac{z^7}{144} + \cdots, \tag{2.24}$$

with $f_0(0) = 0$ and $f_0'(0) = 1$, acts as an extremal function for the bound of $|H_3(1)|$ for $a_2 = a_3 = a_5 = 0$ and $a_4 = 1/6$. ∎

Next, we find the sharp bound of $|H_2(3)|$ for functions belonging to the class $S^*_{\mathcal{B}}$, given by

Theorem 2.6. *Let $f \in \mathcal{S}^*_{\mathcal{B}}$, then*

$$|H_2(3)| \leq \frac{1}{36}. \tag{2.25}$$

This bound is sharp.

Proof. We proceed on the similar lines as in the proof of Theorem 2.5. Assuming $p_1 =: p \in [0,2]$, we substitute the values of $a_i'(i = 3,4,5)$ from (2.2) and (2.3) into (1.5), we obtain

$$H_2(3) = \frac{1}{10616832}\bigg(-761p^6 + 408p^4p_2 - 2880p^2p_2^2 - 41472p_2^3 + 10848p^3p_3 + 52992pp_2p_3 - 73728p_3^2 - 31104p^2p_4 + 82944p_2p_4\bigg).$$

Using (2.11)-(2.13), we arrive at

$$H_2(3) = \frac{1}{5308416}\bigg(\beta_5(p,\gamma) + \beta_6(p,\gamma)\eta + \beta_7(p,\gamma)\eta^2 + \beta_8(p,\gamma,\eta)\rho\bigg),$$

where $\gamma, \eta, \rho \in \mathbb{D}$,

$$\begin{aligned}
\beta_5(p,\gamma) :&= -437p^6 - 5904\gamma^2p^2(4-p^2)^2 + 1440\gamma^3p^2(4-p^2)^2 + 576\gamma^4p^2(4-p^2)^2 \\
&\quad + 5184\gamma^2p^2(4-p^2) - 852\gamma p^4(4-p^2) - 4008p^4\gamma^2(4-p^2) + 1296p^4\gamma^3(4-p^2),\\
\beta_6(p,\gamma) :&= (1-|\gamma|^2)(4-p^2)(5424p^3 - 5184p^3\gamma - 2880p\gamma(4-p^2) - 2304p\gamma^2(4-p^2)),\\
\beta_7(p,\gamma) :&= 576(1-|\gamma|^2)(4-p^2)(-32(4-p^2) - 9p^2\bar{\gamma} - 4|\gamma|^2(4-p^2))\\
\beta_8(p,\gamma,\eta) :&= 5184(1-|\gamma|^2)(4-p^2)(1-|\eta|^2)(p^2 + 4\gamma(4-p^2)).
\end{aligned}$$

Additionally, by using the fact that $|\rho| \leq 1$ and taking $x = |\gamma|$, $y = |\eta|$, we obtain

$$|H_2(3)| \leq \frac{1}{10616832}\bigg(|\beta_5(p,\gamma)| + |\beta_6(p,\gamma)|y + |\beta_7(p,\gamma)|y^2 + |\beta_8(p,\gamma,\eta)|\bigg) \leq B(p,x,y),$$

where

$$B(p,x,y) = \frac{1}{10616832}\bigg(b_1(p,x) + b_2(p,x)y + b_3(p,x)y^2 + b_4(p,x)(1-y^2)\bigg), \tag{2.26}$$

with

$$\begin{aligned}
b_1(p,x) :&= 437p^6 + 5904p^2x^2(4-p^2)^2 + 1440p^2x^3(4-p^2)^2 + 576p^2x^4(4-p^2)^2 \\
&\quad + 5184p^2x^2(4-p^2) + 852p^4x(4-p^2) + 4008p^4x^2(4-p^2) + 1296p^4x^3(4-p^2),\\
b_2(p,x) :&= (4-p^2)(1-x^2)(5424p^3 + 5184p^3x + 2880px(4-p^2) + 2304px^2(4-p^2)),\\
b_3(p,x) :&= 576(4-p^2)(1-x^2)(32(4-p^2) + 9p^2x + 4x^2(4-p^2)),\\
b_4(p,x) :&= 5184(4-p^2)(1-x^2)(p^2 + 4x(4-p^2)).
\end{aligned}$$

At this point, we must maximise $B(p,x,y)$ in the closed cuboid $R : [0,2] \times [0,1] \times [0,1]$. By identifying the maximum values on the twelve edges, the interior of R, and the interiors of the six faces, we can prove this.

(1) We start by taking into account, every interior point of R. Assume that $(p,x,y) \in (0,2)\times(0,1)\times(0,1)$. On differentiating (2.26) with respect to y, to locate the points of maxima in the interior of R, we obtain

$$\frac{\partial B}{\partial y} = \frac{(1-x^2)}{221184}(4-p^2)\Big(48px(5+4x)+p^3(113+48x-48x^2) + 384(8-9x+x^2)y - 24p^2(41-45x+4x^2)y\Big).$$

Now $\partial B/\partial y = 0$ gives

$$y = y_1 := \frac{48px(5+4x)+p^3(113+48x-48x^2)}{24(1-x)(16(x-8)+p^2(41-4x))}.$$

Since y_1 must be a member of $(0,1)$ for critical points to exist, this is only possible if

$$48px(5+4x)+p^3(113+48x-48x^2) < 24(1-x)(16(x-8)+p^2(41-4x)). \tag{2.27}$$

Now, we find the solutions satisfying the inequality (2.27) for the existence of critical points using hit and trial method. If p tends to 0 and 2, then no $x \in (0,1)$ exists satisfying (2.27). Similarly, if take x tending to 0 and 1, then there does not exist any $p \in (0,2)$ such that (2.27) holds. So, we conclude that the function B has no critical point in $(0,2)\times(0,1)\times(0,1)$.

(2) Now, we study the interior of six faces of the cuboid R.

On $p = 0$

$$l_1(x,y) := \frac{(1-x^2)(8y^2+x^2y^2+9x(1-y^2))}{288}, \tag{2.28}$$

with $x,y \in (0,1)$. We note that, in $(0,1)\times(0,1)$, l_1 does not have any critical point. As

$$\frac{\partial l_1}{\partial y} = \frac{y(1-x^2)(x-1)(x-8)}{144} \neq 0 \quad x,y \in (0,1).$$

On $p = 2$

$$B(2,x,y) := \frac{437}{165888} = 0.00263431, \quad x,y \in (0,1). \tag{2.29}$$

On $x = 0$

$$l_2(p,y) := \frac{437p^6+(4-p^2)(5424p^3y+18432(4-p^2)y^2+5184p^2(1-y^2))}{10616832} \tag{2.30}$$

with $p \in (0,2)$ and $y \in (0,1)$. We solve $\partial l_2/\partial p$ and $\partial l_2/\partial y$ to locate the points of maxima. On solving $\partial l_2/\partial y = 0$, we obtain

$$y = \frac{113p^3}{24(41p^2-128)} =: y_p. \tag{2.31}$$

For the given range of y, we should have $y_p \in (0,1)$ but no such $p \in (0,2)$ exists.

On $x = 1$

$$B(p,1,y) = l_3(p,y) := \frac{(147456p^2-43920p^4+2201p^6)}{10616832}, \quad p \in (0,2). \tag{2.32}$$

And when computing $\partial l_3/\partial p = 0$, $p = p_0 \approx 1.40378$ turns out to be the critical point. According to elementary calculations, l_3 reaches its maximum value ≈ 0.0128915 at p_0.

On $y = 0$

$$l_4(p,x) := \frac{1}{10616832}\Big(331776x(1-x^2) + 2304p^2(9 - 72x + 41x^2 + 82x^3 + 4x^4) - 48p^4(108 - 503x + 650x^2 + 564x^3 + 96x^4) + p^6(437 - 852x + 1896x^2 + 144x^3 + 576x^4)\Big) = B(p,x,0).$$

On computations, we obtain

$$\frac{\partial l_4}{\partial x} = \frac{(4-p^2)}{884736}\Big(6912(1-3x^2) + 96p^2(-18 + 41x + 69x^2 + 8x^3) - p^4(-71 + 316x + 36x^2 + 192x^3)\Big)$$

and

$$\frac{\partial l_4}{\partial p} = \frac{1}{1769472}\Big(768p(9 - 72x + 41x^2 + 82x^3 + 4x^4) - 32p^3(108 - 503x + 650x^2 + 564x^3 + 96x^4) + p^5(437 - 852x + 1896x^2 + 144x^3 + 576x^4)\Big).$$

The common real solutions (p,x) of the system of equations, $\partial l_4/\partial x = 0$ and $\partial l_4/\partial p = 0$ are $(-2,-2.86143)$, $(-2.6516,-0.571214)$, $(-2,-0.247504)$, $(-2,-2.86143)$, $(2,0.0163426)$, $(-2,0.0163426)$, $(2.6516,-0.571214)$, $(2,-0.247504)$, $(0,0.57735)$, $(0,-0.57735)$ and $(-0.914024,0.721238)$. Thus, no solution exists in $(0,2)\times(0,1)$, resulting in no critical points.

On $y = 1$

$$l_5(p,x) := \frac{1}{10616832}\Big(9216px(5 + 4x - 5x^2 - 4x^3) + 36864(8 - 7x^2 - x^4) - 2304p^2(64 - 9x - 106x^2 - x^3 - 12x^4) - 48p^5(113 + 48x - 161x^2 - 48x^3 + 48x^4) + 192p^3(113 - 12x - 209x^2 + 12x^3 + 96x^4) + 48p^4(384 - 37x - 1094x^2 - 24x^3 - 144x^4) + p^6(437 - 852x + 1896x^2 + 144x^3 + 576x^4)\Big).$$

On computations, we get

$$\frac{\partial l_5}{\partial x} = \frac{(4-p^2)}{884736}\Big(-1536x(7 + 2x^2) + 192p(5 + 8x - 15x^2 - 16x^3) + 48p^2(9 + 156x + 3x^2 + 32x^3) + 8p^3(24 - 161x - 72x^2 + 96x^3) + p^4(71 - 316x - 36x^2 - 192x^3)\Big)$$

and

$$\frac{\partial l_5}{\partial p} = \frac{1}{1769472}\Big(1536x(5 + 4x - 5x^2 - 4x^3) - 768p(64 - 9x - 106x^2 - x^3 - 12x^4) - 40p^4(113 + 48x - 161x^2 - 48x^3 + 48x^4) + 96p^2(113 - 12x - 209x^2 + 12x^3 + 96x^4) + 32p^3(384 - 37x - 1094x^2 - 24x^3 - 144x^4) + p^5(437 - 852x + 1896x^2 + 144x^3 + 576x^4)\Big).$$

The common real solutions (p, x) of the system of equations, $\partial l_5/\partial x = 0$ and $\partial l_5/\partial p = 0$ are $(10.3578, 0.237179)$, $(2.61706, 3.83978)$, $(-2, -2.97479)$, $(2, 2.03127)$, $(-2.52708, -1.62061)$, $(-2, -0.783621)$, $(-1.37805, -1.02453)$, $(-1.77448, 0.0488452)$, $(2.00019, -0.515602)$, $(0, 0)$ and $(1.35192, -1.00909)$. Thus, no solution exists in $(0, 2) \times (0, 1)$, resulting in no critical points.

(3) Now, we calculate the maximum values achieved by $B(p, x, y)$ on the edges of the cuboid R. From (2.30), we have $B(p, 0, 0) = f_1(p) := (20736p^2 - 5184p^4 + 437p^6)/10616832$. It is easy to observe that $f_1'(p) = 0$ for $p = \delta_0 := 0$ in the interval $[0, 2]$. The maximum value of $f_1(p)$ is 0.
Now considering (2.30) at $y = 1$, we get $B(p, 0, 1) = f_2(p) := (294912 - 147456p^2 + 21696p^3 + 18432p^4 - 5424p^5 + 437p^6)/10616832$. It is easy to observe that $f_2'(p)$ is a decreasing function in $[0, 2]$ and hence $p = 0$ acts as its point of maxima. Thus

$$B(p, 0, 1) \leq \frac{1}{36} = 0.0277778, \quad p \in [0, 2].$$

Through computations, (2.30) shows that $B(0, 0, y) = y^2/36$, attains its maximum value at $y = 1$. This implies that

$$B(0, 0, y) \leq \frac{1}{36}, \quad y \in [0, 1].$$

Since, (2.32) is independent of x, we have $B(p, 1, 1) = B(p, 1, 0) = f_3(p) := (147456p^2 - 43920p^4 + 2201p^6)/10616832$. Now, $f_3'(p) = 294912p - 175680p^3 + 13206p^5 = 0$ when $p = \delta_2 := 0$ and $p = \delta_3 := 1.40378$ in the interval $[0, 2]$ with δ_2 and δ_3 as points of minima and maxima respectively. Hence

$$B(p, 1, 1) = B(p, 1, 0) \leq 0.0128915, \quad p \in [0, 2].$$

On substituting $p = 0$ in (2.32), we get, $B(0, 1, y) = 0$. The equation (2.29) is independent of the all the variables namely p, x and y. Thus the maximum value of $B(p, x, y)$ on the edges $p = 2, x = 1; p = 2, x = 0; p = 2, y = 0$ and $p = 2, y = 1$, respectively, is given by

$$B(2, 1, y) = B(2, 0, y) = B(2, x, 0) = B(2, x, 1) = \frac{437}{165888}, \quad x, y \in [0, 1].$$

From (2.30), we obtain $B(0, 0, y) = y^2/36$. A simple calculation shows that

$$B(0, 0, y) \leq \frac{1}{36}, \quad y \in [0, 1].$$

Using (2.28), we obtain $B(0, x, 1) = f_4(x) := (8 - 7x^2 - x^4)/288$. Upon calculations, we see that f_4 is a decreasing function in $[0, 1]$ and hence attains its maximum value at $x = 0$. Thus

$$B(0, x, 1) \leq \frac{1}{36}, \quad x \in [0, 1].$$

On again using (2.28), we get $B(0, x, 0) = f_5(x) := 9x(1 - x^2)/288$. On further calculations, we get $f_5'(x) = 0$ for $x = \delta_4 := 1/\sqrt{3}$. Also, $f_5(x)$ is an increasing function in $[0, \delta_4)$ and decreasing in $(\delta_4, 1]$. So, it attains its maximum value at δ_4. Thus

$$B(0, x, 0) \leq 0.0120281, \quad x \in [0, 1].$$

In view of all the cases, the inequality (2.25) holds. The function specified in (2.24) acts as an extremal function for the bounds of $|H_2(3)|$ having values $a_3 = a_5 = 0$ and $a_4 = 1/6$. ∎

2.3. **Fourth Hankel determinant.** Given that sharp bounds for third-order Hankel determinants have been attained for various subclasses of stalike functions, as shown in Table 1, determining bounds for fourth-order Hankel determinants proves to be considerably challenging, necessitating extensive computations. Consequently, there have been relatively few efforts in this direction within the existing literature, for recent advancements, we refer [8, 10, 19, 20]. Subsequently, we introduce a lemma which is required in forthcoming results.

Lemma 2.7. *[8, 18] Let* $p = 1 + \sum_{n=1}^{\infty} p_n z^n \in \mathcal{P}$. *Then*

$$|p_n| \leq 2, \quad n \geq 1,$$

$$|p_{n+k} - \nu p_n p_k| \leq \begin{cases} 2, & 0 \leq \nu \leq 1; \\ 2|2\nu - 1|, & otherwise, \end{cases}$$

and

$$|p_1^3 - \nu p_3| \leq \begin{cases} 2|\nu - 4|, & \nu \leq 4/3; \\ 2\nu\sqrt{\dfrac{\nu}{\nu - 1}}, & 4/3 < \nu. \end{cases}$$

Now, we try to estimate the bound of sixth and seventh coefficient bounds of function $f \in \mathcal{S}_{\mathcal{B}}^*$ as follows:

Lemma 2.8. *Let* $f \in \mathcal{S}_{\mathcal{B}}^*$, *then* $|a_6| \leq 0.611233$ *and* $|a_7| \leq 0.690994$.

Proof. From (2.4), we have

$$\begin{aligned} 1843200a_6 = &-2367p_1^5 - 8560p_1^3p_2 + 50880p_1p_2^2 - 86400p_1p_4 + 92160p_5 - 96000p_2p_3 \\ &+ 46080p_1^2p_3 \end{aligned}$$

or

$$\begin{aligned} 1843200|a_6| \leq &|p_1^3(-2367p_1^2 - 8560p_2)| + |p_1(50880p_2^2 - 86400p_4)| + |92160p_5 - 96000p_2p_3| \\ &+ |46080p_1^2p_3|. \end{aligned}$$

Using Lemma 2.7 and the triangle inequality, we arrive at

$$|a_6| \leq \frac{35207}{57600} \approx 0.611233. \tag{2.33}$$

Similarly,

$$\begin{aligned} 530841600a_7 = &601421p_1^6 - 2365320p_1^4p_2 + 4723200p_2^3 - 4818240p_1^2p_2^2 - 2648640p_1^3p_3 - 11980800p_3^2 \\ &+ 26150400p_1p_2p_3 - 23500800p_2p_4 + 11577600p_1^2p_4 - 21012480p_1p_5 \end{aligned}$$

or

$$\begin{aligned} 530841600|a_7| \leq &|p_1^4(601421p_1^2 - 2365320p_2)| + |p_2^2(4723200p_2 - 4818240p_1^2)| \\ &+ |p_3(2648640p_1^3 - 11980800p_3)| + |p_2(26150400p_1p_3 - 23500800p_4)| \\ &+ |p_1(11577600p_1^2p_4 - 21012480p_5)|. \end{aligned}$$

By employing Lemma 2.7 and the triangle inequality, we obtain

$$|a_7| \leq \frac{31841}{46080} \approx 0.690994. \tag{2.34}$$

∎

We derive the expression of the fourth Hankel determinant, upon substituting $q = 4$ and $n = 1$ in (1.4) as follows:

$$H_4(1) = a_7H_3(1) - a_6B_1 + a_5B_2 - a_4B_3, \tag{2.35}$$

where

$$B_1 := a_6(a_3 - a_2^2) + a_3(a_2a_5 - a_3a_4) - a_4(a_5 - a_2a_4), \tag{2.36}$$

$$B_2 := a_3(a_3a_5 - a_4^2) - a_5(a_5 - a_2a_4) + a_6(a_4 - a_2a_3), \tag{2.37}$$

and

$$B_3 := a_4(a_3a_5 - a_4^2) - a_5(a_2a_5 - a_3a_4) + a_6(a_4 - a_2a_3). \tag{2.38}$$

Next, we determine the bound of these B_i for $i = 1, 2, 3$. By substituting the values from (2.2)-(2.4) in (2.36), we get

$$\begin{aligned} 707788800B_1 = {} & 110689p_1^7 - 299496p_1^5p_2 - 855360p_1^4p3 + 269760p_1^3p_2^2 + 944640p_1p_2^3 \\ & - 1128960p_1^2p_2p_3 - 3686400p_2^2p_3 + 4608000p_1p_3^2 + 2668800p_1^3p_4 \\ & + 460800p_1p_2p_4 - 3686400p_3p_4 - 3870720p_1^2p_5 + 4423680p_2p_5 \end{aligned}$$

or

$$\begin{aligned} 707788800|B_1| \leq {} & |p_1^4(110689p_1^7 - 855360p_3)| + |p_2^2(269760p_1^3 - 3686400p_3)| + |4608000p_1p_3^2| \\ & + |p_1p_2(944640p_2^2 + 460800p_4)| + |p_1^2p_2(-299496p_1^3 - 1128960p_3)| \\ & + |p_4(2668800p_1^3 - 3686400p_3)| + |p_5(4423680p_2 - 3870720p_1^2)| \end{aligned}$$

Using Lemma 2.7 and the triangle inequality, we get

$$\begin{aligned} |B_1| &\leq \frac{1}{707788800}\left(110590464 + 255027456\sqrt{\frac{165}{744671}} + 471859200\sqrt{\frac{15}{3559}} + 117964800\sqrt{\frac{3}{53}}\right) \\ &\approx 0.244544. \end{aligned} \tag{2.39}$$

Similarly,

$$\begin{aligned} 5662310400B_2 = {} & 927648p_1^5p_3 - 149225p_1^8 + 785840p_1^6p_2 + 3939840p_1^3p_2p_3 - 2073600p_1^2p_2^3 - 8294400p_2^4 \\ & + 23316480p_1p_2^2p_3 - 2025600p_1^4p_2^2 - 11704320p_1^2p_3^2 - 29491200p_2p_3^2 + 25804800p_1p_3p_4 \\ & - 3676800p_1^4p_4 - 3225600p_1^2p_2p_4 - 29491200p_1p_2p_5 + 27648000p_2^2p_4 - 22118400p_4^2 \\ & + 6144000p_1^3p_5 + 23592960p_3p_5 \end{aligned}$$

or

$$\begin{aligned} 5662310400|B_2| \leq {} & |p_1^5(927648p_3 - 149225p_1^3)| + |p_1^3p_2(785840p_1^3 + 3939840p_3)| \\ & + |p_2^3(-2073600p_1^2 - 8294400p_2)| + |p_1p_2^2(23316480p_3 - 2025600p_1^3)| \\ & + |p_3^2(-11704320p_1^2 - 29491200p_2)| + |p_1p_4(25804800p_3 - 3676800p_1^3)| \\ & + |p_1p_2(-3225600p_1p_4 - 29491200p_5)| + |p_4(27648000p_2^2 - 22118400p_4)| \\ & + |p_5(6144000p_1^3 + 23592960p_3|. \end{aligned}$$

With the aid of Lemma 2.7 and the triangle inequality, we obtain

$$\begin{aligned} |B_2| &\leq \frac{1}{5662310400}\left(1982371840 + \frac{22263552}{149225}\sqrt{\frac{6442}{778423}} + \frac{43008}{383}\sqrt{\frac{42}{2305}} + \frac{97152}{1055}\sqrt{\frac{759}{11089}}\right) \\ &\approx 0.350099. \end{aligned} \tag{2.40}$$

and

$$\begin{aligned}305764761600B_3 &= 133166592p_1^5p_3 - 8521200p_1^8 + 126489600p_1^4p_2p_3 - 621000p_1^7p_2 + 331084800p_1^4p_2^2\\ &+ 2362245120p_1p_2^2p_3 + 6566400p_1^3p_2^3 - 199065600p_2^3p_3 + 99532800p_1p_2^4\\ &- 49766400p_1p_2^2p_4 - 1260195840p_1^3p_2p_3 - 1327104000p_2p_3^2 - 103680000p_1^2p_2^2p_3\\ &- 20260800p_1^5p_2^2 - 177638400p_1^3p_3^2 - 176947200p_3^3 + 398131200p_1^2p_3p_4\\ &- 12182400p_1^5p_4 + 331776000p_1^3p_5 + 1274019840p_3p_5 + 398131200p_2p_3p_4\\ &- 311040000p_1^4p_4 - 1194393600p_1p_3p_4 + 1492992000p_1^2p_2p_4 - 1592524800p_1p_2p_5\\ &+ 637009920p_1^2p_3^2 - 160625p_1^9 - 879206400p_1^2p_2^3 + 99532800p_1p_2p_3^2 - 298598400p_1p_4^2\end{aligned}$$

or

$$\begin{aligned}305764761600|B_3| &\le |p_1^5(133166592p_3 - 8521200p_1^3)| + |p_1^4p_2(126489600p_3 - 621000p_1^3)|\\ &+ |p_1p_2^2(331084800p_1^3 + 2362245120p_3)| + |p_2^3(6566400p_1^3 - 199065600p_3)|\\ &+ |p_1p_2^2(99532800p_2^2 - 49766400p_4)| + |p_2p_3(-1260195840p_1^3 - 1327104000p_3)|\\ &+ |p_1^2p_2^2(-103680000p_3 - 20260800p_1^3)| + |p_3^2(-177638400p_1^3 - 176947200p_3)|\\ &+ |p_1^2p_4(398131200p_3 - 12182400p_1^3)| + |p_5(331776000p_1^3 + 1274019840p_3)|\\ &+ |p_2p_4(398131200p_3 - 95385600p_1^3)| + |p_1p_4(-311040000p_1^3 - 1194393600p_3)|\\ &+ |p_1p_2(1492992000p_1p_4 - 1592524800p_5)| + |637009920p_1^2p_3^2| + |160625p_1^9|\\ &+ |879206400p_1^2p_2^3| + |99532800p_1p_2p_3^2| + |298598400p_1p_4^2|.\end{aligned}$$

Through Lemma 2.7 and the triangle inequality, we have

$$\begin{aligned}|B_3| &\le \frac{1}{305764761600}\left(210371822080 + 147726139392\sqrt{\frac{114}{32059}} + 101921587200\sqrt{\frac{6}{1489}}\right.\\ &\left. + 64762675200\sqrt{\frac{366}{23309}} + 12740198400\sqrt{\frac{6}{73}} + \frac{76441190400}{\sqrt{557}}\right)\\ &\approx 0.787068. \qquad (2.41)\end{aligned}$$

Upon substituting values from (2.33), (2.34), (2.5), (2.39)-(2.41), and Theorem 2.3 in (2.35), we obtain the following result, given by

Theorem 2.9. *Let $f \in \mathcal{S}^*_{\mathcal{B}}$, then $|H_4(1)| \le 0.169251$.*

3. Sharp Hankel Determinants for $\mathcal{C}_{\mathcal{B}}$

In this section, we determine the sharp bounds for the second and third-order Hankel determinants for functions $f \in \mathcal{C}_{\mathcal{B}}$. Below, we provide the expressions for the initial coefficients of functions $f \in \mathcal{C}_{\mathcal{B}}$ in terms of Carathéodory coefficients, serving as a foundation for subsequent calculations.

Let $f \in \mathcal{C}_{\mathcal{B}}$, then there exists a Schwarz function $w(z)$ such that

$$1 + \frac{zf''(z)}{f'(z)} = \sqrt{1 + \tanh w(z)}. \qquad (3.1)$$

Suppose that $p(z) = 1 + p_1z + p_2z^2 + \cdots \in \mathcal{P}$ and consider $w(z) = (p(z) - 1)/(p(z) + 1)$. Further, by substituting the expansions of $w(z)$, $p(z)$ and $f(z)$ in (3.1) and then comparing the coefficients,

we obtain the expressions of $a_i (i = 2, 3, ..., 7)$ in terms of $p_j (j = 1, 2, ..., 5)$, given by

$$a_2 = \frac{1}{8}p_1, \quad a_3 = \frac{1}{192}\Big(8p_2 - 3p_1^2\Big), \quad a_4 = \frac{1}{9216}\Big(23p_1^3 - 168p_1p_2 + 192p_3\Big) \tag{3.2}$$

and

$$a_5 = \frac{1}{92160}\Big(- 11p_1^4 + 528p_1^2p_2 - 576p_2^2 - 1056p_1p_3 + 1152p_4\Big). \tag{3.3}$$

The following theorem presents the sharp bound for $|H_3(1)|$ for functions belonging to the class $\mathcal{C}_{\mathcal{B}}$.

Theorem 3.1. *Let $f \in \mathcal{C}_{\mathcal{B}}$, then*

$$|H_3(1)| \leq 1/576. \tag{3.4}$$

This result is sharp.

Proof. Since the class $\mathcal{P}$ is invariant under rotation, the value of p_1 belongs to the interval [0,2]. Let $p := p_1$ and then substitute the values of $a_i (i = 2, 3, 4, 5)$ in (1.6) from (3.2) and (3.3). We get

$$H_3(1) = \frac{1}{424673280}\Big(- 3581p^6 - 11184p^4p_2 - 2880p^2p_2^2 - 141312p_2^3 + 73344p^3p_3 + 211968pp_2p_3$$
$$- 184320p_3^2 - 165888p^2p_4 + 221184p_2p_4\Big).$$

After simplifying the calculations through (2.11)-(2.13), we obtain

$$H_3(1) = \frac{1}{424673280}\Big(\beta_9(p,\gamma) + \beta_{10}(p,\gamma)\eta + \beta_{11}(p,\gamma)\eta^2 + \beta_{12}(p,\gamma,\eta)\rho\Big),$$

for $\gamma, \eta, \rho \in \mathbb{D}$. Here

$$\begin{aligned}
\beta_9(p,\gamma) :&= -1157p^6 - 5328\gamma^2p^2(4-p^2)^2 - 15360\gamma^3(4-p^2)^2 - 4224\gamma^3p^2(4-p^2)^2 + 2304\gamma^4p^2(4-p^2)^2\\
&\quad + 3144\gamma p^4(4-p^2) - 1056p^4\gamma^2(4-p^2) - 6912p^4\gamma^3(4-p^2) - 27648\gamma^2p^2(4-p^2),\\
\beta_{10}(p,\gamma) :&= 192(1-|\gamma|^2)(4-p^2)(83p^3 + 144\gamma p^3 + 84p\gamma(4-p^2) - 48p\gamma^2(4-p^2)),\\
\beta_{11}(p,\gamma) :&= 9216(1-|\gamma|^2)(4-p^2)(-5(4-p^2) - |\gamma|^2(4-p^2) + 3p^2\bar{\gamma}),\\
\beta_{12}(p,\gamma,\eta) :&= 27648(1-|\gamma|^2)(4-p^2)(1-|\eta|^2)(2(4-p^2)\gamma - p^2).
\end{aligned}$$

By choosing $x = |\gamma|$, $y = |\eta|$ and utilizing the fact that $|\rho| \leq 1$, the above expression reduces to the following:

$$|H_3(1)| \leq \frac{1}{424673280}\Big(|\beta_9(p,\gamma)| + |\beta_{10}(p,\gamma)|y + |\beta_{11}(p,\gamma)|y^2 + |\beta_{12}(p,\gamma,\eta)|\Big) \leq C(p,x,y),$$

where

$$C(p,x,y) = \frac{1}{424673280}\Big(c_1(p,x) + c_2(p,x)y + c_3(p,x)y^2 + c_4(p,x)(1-y^2)\Big), \tag{3.5}$$

with

$$\begin{aligned}
c_1(p,x) :&= 1157p^6 + 5328x^2p^2(4-p^2)^2 + 15360x^3(4-p^2)^2 + 4224x^3p^2(4-p^2)^2 + 2304x^4p^2(4-p^2)^2\\
&\quad + 3144xp^4(4-p^2) + 1056p^4x^2(4-p^2) + 6912p^4x^3(4-p^2) + 27648x^2p^2(4-p^2),\\
c_2(p,x) :&= 192(1-x^2)(4-p^2)(83p^3 + 144xp^3 + 84px(4-p^2) + 48px^2(4-p^2)),\\
c_3(p,x) :&= 9216(1-x^2)(4-p^2)(5(4-p^2) + x^2(4-p^2) + 3p^2x),\\
c_4(p,x) :&= 27648(1-x^2)(4-p^2)(2x(4-p^2) + p^2).
\end{aligned}$$

In the closed cuboid $S : [0,2] \times [0,1] \times [0,1]$, we now maximise $C(p,x,y)$, by locating the maximum values in the interior of the six faces, on the twelve edges, and in the interior of S.

(1) We start by taking into account every internal point of S. Assume that $(p,x,y) \in (0,2) \times (0,1) \times (0,1)$. We calculate $\partial C/\partial y$, partially differentiate (3.5) with respect to y to identify the points of maxima in the interior of U. We get

$$\frac{\partial C}{\partial y} = \frac{(4-p^2)(1-x^2)}{2211840}\Big(48px(7+4x) + p^3(83+60x-48x^2) - 96p^2(8-9x+x^2)y + 384(5-6x+x^2)y\Big).$$

Now $\partial C/\partial y = 0$ gives

$$y = y_0 := \frac{48px(7+4x) + p^3(83+60x-48x^2)}{96(1-x)(p^2(8-x)-4(5-x))}.$$

The existence of critical points require that y_0 belong to $(0,1)$, which is only possible when

$$83p^3 + 336px + 60p^3x + 192px^2 - 48p^3x^2 < 96(-1+x)(20-8p^2-4x+p^2x). \tag{3.6}$$

Now, we find the solution satisfying the inequality (3.6) for the existence of critical points using the hit and trial method. If we assume p tends to 0, then there does not exist any $x \in (0,1)$ satisfying (3.6). But, when p tends to 2, (3.6) holds only when $x < 61/288$. Similarly, if we assume x tends to 0, then for all $p > 1.75665$, (3.6) holds. After calculations, we observe that there does not exist any $x \in (0,1)$ when $p \in (0, 1.75665)$. Thus, the domain for the solution is $(1.75665, 2) \times (0, 61/288)$. Now, we examine that $\frac{\partial C}{\partial y}|_{y=y_0} \neq 0$ in $(1.75665, 2) \times (0, 61/288)$. So, we conclude that the function M has no critical point in $(0,2) \times (0,1) \times (0,1)$.

(2) The interior of each of the cuboid S's six faces is now being considered.

On $p = 0$

$$m_1(x,y) := \frac{20x^3 + 12(1-x^2)(5+x^2)y^2 + 72x(1-x^2)(1-y^2)}{34560}, \quad x,y \in (0,1). \tag{3.7}$$

Since

$$\frac{\partial m_1}{\partial y} = \frac{(1-x)^2(x+1)(5-x)y}{1440} \neq 0, \quad x,y \in (0,1),$$

indicates that m_1 has no critical points in $(0,1) \times (0,1)$.

On $p = 2$

$$C(2,x,y) := \frac{1157}{6635520} \approx 0.000174365, \quad x,y \in (0,1). \tag{3.8}$$

On $x = 0$

$$m_2(p,y) := \frac{1157p^6 + (4-p^2)(15936p^3y + 46080(4-p^2)y^2 + 27648p^2(1-y^2))}{424673280} \tag{3.9}$$

with $p \in (0,2)$ and $y \in (0,1)$. To determine the points of maxima, we solve $\partial m_2/\partial p = 0$ and $\partial m_2/\partial y = 0$. After solving $\partial m_2/\partial y = 0$, we get

$$y = \frac{83p^3}{384(2p^2-5)} =: y_p. \tag{3.10}$$

In order to have $y_p \in (0,1)$ for the given range of y, $p_0 := p > \approx 1.75665$ is required. Based on calculations, $\partial m_2/\partial p = 0$ gives

$$36864p - 18432p^3 + 1157p^5 + 31872p^2y - 13280p^4y - 159744py^2 + 49152p^3y^2 = 0. \tag{3.11}$$

On substituting equation (3.10) into equation (3.11), we have

$$1843200p - 2396160p^3 + 1021152p^5 - 152402p^7 + 2367p^9 = 0. \tag{3.12}$$

A numerical calculation suggests that $p \approx 1.28894 \in (0,2)$ is the solution of (3.12). So, we conclude that m_2 does not have any critical point in $(0,2)\times(0,1)$.
On $x=1$

$$C(p,1,y) = m_3(p,y) := \frac{245760 + 177408p^2 - 62688p^4 + 1901p^6}{424673280}, \quad p \in (0,2). \tag{3.13}$$

While computing $\partial m_3/\partial p = 0$, $p_0 := p \approx 1.23293$ comes out to be the critical point. After simple calculations, m_3 achieves its maximum value ≈ 0.000888358 at p_0.
On $y=0$

$$\begin{aligned} m_4(p,x) :=& \frac{1}{424673280}\Big(49152x(18-13x^2) + 2304p^2(48 - 192x + 37x^2 + 168x^3 \\ &+ 16x^4) - 96p^4(288 - 707x + 400x^2 + 480x^3 + 192x^4) \\ &+ p^6(1157 - 3144x + 4272x^2 - 2688x^3 + 2304x^4)\Big). \end{aligned}$$

After further calculations such as

$$\begin{aligned} \frac{\partial m_4}{\partial x} =& \frac{(4-p^2)}{17694720}\Big(1536(6-13x^2) - 48p^2(48 - 37x - 148x^2 - 32x^3) \\ &+ p^4(131 - 356x + 336x^2 - 384x^3)\Big) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial m_4}{\partial p} =& \frac{1}{70778880}\Big(768p(48 - 192x + 37x^2 + 168x^3 + 16x^4) - 64p^3(288 - 707x \\ &+ 400x^2 + 480x^3 + 192x^4) + p^5(1157 - 3144x + 4272x^2 - 2688x^3 \\ &+ 2304x^4)\Big), \end{aligned}$$

we observe that only real solutions (p,x) of the system of equations $\partial m_4/\partial x = 0$ and $\partial m_4/\partial p = 0$ are $(3.85748, 0.257377)$, $(3.76933, 0.0851082)$, $(-3.76933, 0.0851082)$, $(2, -0.644779)$, $(-3.85748, 0.257377)$, $(-1.57205, -1.03976)$, $(-2, -0.644779)$, $(0, -0.679366)$, $(0, 0.679366)$, $(-1.12904, 0.941715)$ and $(1.57205, -1.03976)$. Thus, no solution exists in $(0,2)\times(0,1)$, resulting in no critical points.
On $y=1$

$$\begin{aligned} m_5(p,x) :=& \frac{1}{424673280}\Big\{1157p^6 + 3144p^4(4-p^2)x + 27648p^2(4-p^2)x^2 \\ &+ 1056p^4(4-p^2)x^2 + 5328p^2(4-p^2)^2x^2 + 6912p^4(4-p^2)x^3 \\ &+ 15360(4-p^2)^2x^3 + 4224p^2(4-p^2)^2x^3 + 2304p^2(4-p^2)^2x^4 \\ &- 9216(4-p^2)(1-x^2)(-4(5+x^2) + p^2(5-3x+x^2)) \\ &+ 192(4-p^2)(1-x^2)(48px(7+4x) + p^3(83 + 60x - 48x^2))\Big\}. \end{aligned}$$

After calculations,

$$\frac{\partial m_5}{\partial p} = \frac{1}{70778880}\Big(6144x(7+4x-7x^2-4x^3) - 768p(160-48x-213x^2+72x^3-48x^4) - 160p^4(83+60x-131x^2-60x^3+48x^4) + 384p^2(83-24x-179x^2 + 24x^3+96x^4) + 64p^3(480-157x-1072x^2+384x^3-288x^4) + p^5(1157-3144x+4272x^2-2688x^3+2304x^4)\Big)$$

and

$$\frac{\partial m_5}{\partial x} = \frac{(4-p^2)}{17694720}\Big(1536x(5x-8-4x^2) + 384p(7+8x-21x^2-16x^3) + 48p^2(24 + 149x-68x^2+64x^3) + 16p^3(30-131x-90x^2+96x^3) + p^4(131-356x+336x^2-384x^3)\Big),$$

we observe that only real solutions (p, x) of the system of equations $\partial m_5/\partial x = 0$ and $\partial m_5/\partial p = 0$ are $(27.1136, 0.413453)$, $(-1.50007, -6.38485)$, $(2.68178, 5.28944)$, $(2, 1.39993)$, $(-1.80124, -0.775344)$, $(-4.98296, 1.51367)$, $(-2.89665, 3.0885)$, $(2, -0.192442)$, $(0, 0)$, $(-1.74974, 0.159528)$, $(-2, 0.737618)$, $(-0.776187, 0.905316)$, $(1.6981, -0.807967)$, $(2, -1.20749)$ and $(1.5135, -0.953552)$. Thus, no solution exists in $(0, 2) \times (0, 1)$, resulting in no critical points.

(3) We next examine the maxima attained by $C(p, x, y)$ on the edges of the cuboid S.
From (3.9), we have $M(p, 0, 0) = g_1(p) := (110592p^2 - 27648p^4 + 1157p^6)/424673280$. It is easy to observe that $g_1'(p) = 0$ whenever $p = \delta_0 := 0$ and $p = \delta_1 := 1.53142 \in [0, 2]$ as its points of minima and maxima respectively. Hence,

$$C(p, 0, 0) \leq 0.0002878, \quad p \in [0, 2].$$

Now consider (3.9) at $y = 1$, we get $C(p, 0, 1) = g_2(p) := (737280 - 368640p^2 + 63744p^3 + 46080p^4 - 15936p^5 + 1157p^6)/424673280$. It is easy to observe that $g_2'(p) < 0$ in $[0, 2]$ and hence $p = 0$ serves as the point of maxima. So,

$$C(p, 0, 1) \leq \frac{1}{576} \approx 0.00173611, \quad p \in [0, 2].$$

Through computations, (3.9) shows that $C(0, 0, y) = y^2/576$ attains its maxima at $y = 1$. This implies that

$$C(0, 0, y) \leq \frac{1}{36}, \quad y \in [0, 1].$$

Since, (3.13) does not involve x, we have $C(p, 1, 1) = C(p, 1, 0) = g_3(p) := (245760 + 177408p^2 - 62688p^4 + 1901p^6)/424673280$. Now, $g_3'(p) = 354816p - 250752p^3 + 11406p^5 = 0$ when $p = \delta_2 := 0$ and $p = \delta_3 := 1.23293$ in the interval $[0, 2]$ with δ_2 and δ_3 as points of minima and maxima, respectively. Hence

$$C(p, 1, 1) = C(p, 1, 0) \leq 0.000888358, \quad p \in [0, 2].$$

After considering $p = 0$ in (3.13), we get, $C(0, 1, y) = 1/1728 \approx 0.000578704$.
The equation (3.8) has no variables. So, on the edges, the maximum value of $C(p, x, y)$ is

$$C(2, 1, y) = C(2, 0, y) = C(2, x, 0) = C(2, x, 1) = \frac{1157}{6635520}, \quad x, y \in [0, 1].$$

Using (3.7), we obtain $C(0,x,1) = g_4(x) := (15-12x^2+5x^3-3x^4)/8640$. Upon calculations, we see that $g_4(x)$ is a decreasing function in $[0,1]$ and attains its maxima at $x=0$. Hence

$$C(0,x,1) \leq \frac{1}{576}, \quad x \in [0,1].$$

Again utilizing (3.7), we get $C(0,x,0) = g_5(x) := x(18-13x^2)/8640$. On further calculations, we get $g_5'(x) = 0$ for $x = \delta_4 := \sqrt{6/13}$. Also, $g_5(x)$ increases in $[0,\delta_4)$ and decreases in $(\delta_4,1]$. So, it reaches its maximum value at δ_4. Thus

$$C(0,x,0) \leq 0.000943564, \quad x \in [0,1].$$

Given all the cases, the inequality (3.4) holds. Let the function $f_1(z) \in \mathcal{C}_{\mathcal{B}}$, be defined as

$$f_1(z) = \int_0^z \left(\exp\left(\int_0^t \frac{\sqrt{1+\tanh u^3}-1}{u} du \right) \right) dt = z + \frac{z^4}{24} - \frac{z^7}{1008} - \cdots, \tag{3.14}$$

with $f_1(0) = 0$ and $f_1'(0) = 1$, acts as an extremal function for the bound of $|H_3(1)|$ for $a_2 = a_3 = a_5 = 0$ and $a_4 = 1/24$. ∎

Theorem 3.2. *Let $f \in \mathcal{C}_{\mathcal{B}}$, then*

$$|H_2(3)| \leq \frac{1}{576}. \tag{3.15}$$

This bound is sharp.

Proof. We proceed on the similar lines as in the proof of Theorem 2.5. Assuming $p_1 =: p \in [0,2]$, we substitute the values of $a_i's (i = 3,4,5)$ from (3.2) and (3.3) in (1.5), we obtain

$$H_2(3) = \frac{1}{424673280}\Big(-1853p^6 - 1488p^4p_2 + 1728p^2p_2^2 - 110592p_2^3 + 31872p^3p_3 + 119808pp_2p_3 - 184320p_3^2 - 82944p^2p_4 + 221184p_2p_4 \Big).$$

Using (2.11)-(2.13) for simplification, we arrive at

$$H_2(3) = \frac{1}{424673280}\Big(\beta_{13}(p,\gamma) + \beta_{14}(p,\gamma)\eta + \beta_{15}(p,\gamma)\eta^2 + \beta_{16}(p,\gamma,\eta)\rho \Big),$$

where $\gamma, \eta, \rho \in \mathbb{D}$,

$$\begin{aligned}
\beta_{13}(p,\gamma) :&= -1109p^6 - 15696\gamma^2p^2(4-p^2)^2 + 3456\gamma^3p^2(4-p^2)^2 + 2304\gamma^4p^2(4-p^2)^2 \\
&\quad + 13824\gamma^2p^2(4-p^2) - 2376\gamma p^4(4-p^2) - 10272p^4\gamma^2(4-p^2) + 3456p^4\gamma^3(4-p^2), \\
\beta_{14}(p,\gamma) :&= 192(1-|\gamma|^2)(4-p^2)(71p^3 - 72p^3\gamma - 36p\gamma(4-p^2) - 48p\gamma^2(4-p^2)), \\
\beta_{15}(p,\gamma) :&= 4608(1-|\gamma|^2)(4-p^2)(-10(4-p^2) - 3p^2\bar{\gamma} - 2|\gamma|^2(4-p^2)) \\
\beta_{16}(p,\gamma,\eta) :&= 13824(1-|\gamma|^2)(4-p^2)(1-|\eta|^2)(p^2 + 4\gamma(4-p^2)).
\end{aligned}$$

Additionally, by using the fact that $|\rho| \leq 1$, and taking $x = |\gamma|$, $y = |\eta|$, we have

$$|H_2(3)| \leq \frac{1}{424673280}\Big(|\beta_{13}(p,\gamma)| + |\beta_{14}(p,\gamma)|y + |\beta_{15}(p,\gamma)|y^2 + |\beta_{16}(p,\gamma,\eta)| \Big) \leq D(p,x,y),$$

where

$$D(p,x,y) = \frac{1}{424673280}\Big(d_1(p,x) + d_2(p,x)y + d_3(p,x)y^2 + d_4(p,x)(1-y^2) \Big), \tag{3.16}$$

with

$$\begin{aligned}
d_1(p,x) :&= 1109p^6 + 15696p^2x^2(4-p^2)^2 + 3456p^2x^3(4-p^2)^2 + 2304p^2x^4(4-p^2)^2 \\
&\quad + 13824p^2x^2(4-p^2) + 2376p^4x(4-p^2) + 10272p^4x^2(4-p^2) + 3456p^4x^3(4-p^2),\\
d_2(p,x) :&= 192(4-p^2)(1-x^2)(71p^3 + 72p^3x + 36px(4-p^2) + 48px^2(4-p^2)),\\
d_3(p,x) :&= 4608(4-p^2)(1-x^2)(10(4-p^2) + 3p^2x + 2x^2(4-p^2)),\\
d_4(p,x) :&= 13824(4-p^2)(1-x^2)(p^2 + 4x(4-p^2)).
\end{aligned}$$

Now, we must maximise $D(p,x,y)$ in the closed cuboid $T : [0,2]\times[0,1]\times[0,1]$. By identifying the maximum values on the twelve edges, the interior of T, and the interiors of the six faces, we can prove this.

(1) We start by taking into account, every interior point of T. Assume that $(p,x,y) \in (0,2)\times(0,1)\times(0,1)$. We partially differentiate (3.16) with respect to y in order to locate the points of maxima in the interior of T. We obtain

$$\frac{\partial D}{\partial y} = \frac{(1-x^2)}{2211840}(4-p^2)\Big(48px(3+4x) + p^3(71+36x-48x^2) + 384(5-6x+x^2)y - 48p^2(13-15x+2x^2)y\Big).$$

Now $\partial D/\partial y = 0$ gives

$$y = y_1 := \frac{48px(3+4x) + p^3(71+36x-48x^2)}{48(1-x)(8(x-5)+p^2(13-2x))}.$$

Since y_1 must be a member of $(0,1)$ for critical points to exist, this is only possible if

$$71p^3 + px(144 + 36p^2 + 192x - 48p^2x) < 48(-40 + 13p^2 + x(48 - 15p^2 - 8x + 2p^2x)). \quad (3.17)$$

Now, we find the solutions satisfying the inequality (3.17) for the existence of critical points using hit and trial method. If p tends to 2, then (3.17) holds whenever $x < 1/144$. Also, no such $x \in (0,1)$, satisfying (3.17) when p tends to 0. Similarly, if take x tending to 0, then (3.17) holds for $p > 1.99513$ only, whereas there does not exist any $p \in (0,2)$ such that equation (3.17) holds when x tends to 1. Thus, the domain for the solution of the equation is $(1.99513, 2)\times(0, 1/144)$. Now, we examine that $\frac{\partial D}{\partial y}|_{y=y_0} \neq 0$ in $(1.99513,2)\times(0,1/144)$. So, we conclude that the function M has no critical point in $(0,2)\times(0,1)\times(0,1)$.

(2) Now, we study the interior of each of six face of the cuboid T.

On $p = 0$

$$n_1(x,y) := \frac{(1-x^2)(5y^2 + x^2y^2 + 6x(1-y^2))}{2880}, \quad (3.18)$$

with $x, y \in (0,1)$. We note that, in $(0,1)\times(0,1)$, $n_1(x,y)$ does not have any critical point. As

$$\frac{\partial n_1}{\partial y} = \frac{(1-x^2)(x-1)(x-5)y}{1440} \neq 0 \quad x,y \in (0,1).$$

On $p = 2$

$$D(2,x,y) := \frac{1109}{6635520} \approx 0.000167131, \quad x,y \in (0,1). \quad (3.19)$$

On $x = 0$

$$n_2(p,y) := \frac{1109p^6 + (4-p^2)(13632p^3y + 46080(4-p^2)y^2 + 13824p^2(1-y^2))}{42467328} \quad (3.20)$$

with $p \in (0,2)$ and $y \in (0,1)$. We solve $\partial n_2/\partial p$ and $\partial n_2/\partial y$ to locate the points of maxima. On solving $\partial n_2/\partial y = 0$, we obtain

$$y = \frac{71p^3}{48(13p^2-40)} (=: \epsilon). \tag{3.21}$$

For the given range of y, we should have $\epsilon \in (0,1)$ which is possible only when $p > p_0 :\approx 1.99513 \in (0,2)$ exists. On computations, $\partial n_2/\partial p = 0$ gives

$$18432p - 9216p^3 + 1109p^5 + 27264p^2y - 11360p^4y - 141312py^2 + 39936p^3y^2 = 0. \tag{3.22}$$

On substituting (3.21) in (3.22), we get,

$$88473600p - 101744640p^3 + 38582784p^5 - 5470944p^7 + 169065p^9 = 0. \tag{3.23}$$

The solution of (3.23) in the interval $(0,2)$ is $p \approx 1.34821$, according to a numerical calculation. In $(0,2)\times(0,1)$, n_2 does not have a critical point.

On $x = 1$

$$n_3(p,y) := D(p,1,y) = \frac{398592p^2 - 121056p^4 + 6461p^6}{424673280}, \quad p \in (0,2). \tag{3.24}$$

And when computing $\partial n_3/\partial p = 0$, $p =: p_0 \approx 1.39681$ turns out to be the critical point. According to elementary calculations, n_3 reaches its maximum value ≈ 0.000859123 at p_0.

On $y = 0$

$$\begin{aligned} n_4(p,x) :=& \frac{1}{424673280}\Big(884736x(1-x^2) + 2304p^2(24 - 192x + 109x^2 + 216x^3 + 16x^4) \\ &- 96p^4(144 - 675x + 880x^2 + 720x^3 + 192x^4) \\ &+ p^6(1109 - 2376x + 5424x^2 + 2304x^4)\Big) = D(p,x,0). \end{aligned}$$

On computations,

$$\begin{aligned} \frac{\partial n_4}{\partial x} =& \frac{(4-p^2)}{17694720}\Big(9216(1-3x^2) - 48p^2(48 - 109x - 180x^2 - 32x^3) \\ &+ p^4(99 - 452x - 384x^3)\Big) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial n_4}{\partial p} =& \frac{1}{70778880}\Big(768p(24 - 192x + 109x^2 + 216x^3 + 16x^4) \\ &- 64p^3(144 - 675x + 880x^2 + 720x^3 + 192x^4) \\ &+ p^5(1109 - 2376x + 5424x^2 + 2304x^4)\Big), \end{aligned}$$

we observe that only real solutions (p,x) of the system of equations $\partial n_4/\partial x = 0$ and $\partial n_4/\partial p = 0$ are $(-2,-2.72495)$, $(-2.64507,-0.503718)$, $(-2.23927,-0.0103472)$, $(-2,-0.163482)$, $(2.23927,-0.0103472)$, $(-2.00568,-0.120347)$, $(2,-2.72495)$, $(-2,-0.0837889)$, $(0,-0.57735)$, $(2,-0.0837889)$, $(2.00568,-0.120347)$, $(2,-0.163482)$, $(2.64507,-0.503718)$ and $(-0.91207,0.721981)$. Thus, no solution exists in $(0,2)\times(0,1)$, resulting in no critical points.

On $y = 1$

$$n_5(p,x) := \frac{1}{424673280}\Big(36864px(3+4x-3x^2-4x^3)+147456(5-4x^2-x^4) \\ -2304p^2(160-24x-261x^2-48x^4)+768p^3(71-167x^2+96x^4) \\ -192p^5(71+36x-119x^2-36x^3+48x^4)-96p^4(-480+45x \\ +1408x^2+288x^4)+p^6(1109-2376x+5424x^2+2304x^4)\Big).$$

On computations,

$$\frac{\partial n_5}{\partial x} = \frac{(4-p^2)}{17694720}\Big(-6144x(2+x^2)+384p(3+8x-9x^2-16x^3) \\ +48p^2(12+197x+64x^3)+16p^3(18-119x-54x^2+96x^3) \\ +p^4(99-452x-384x^3)\Big)$$

and

$$\frac{\partial n_5}{\partial p} = \frac{1}{70778880}\Big(6144x(3+4x-3x^2-4x^3)+768p(-160+24x+261x^2+48x^4) \\ -160p^4(71+36x-119x^2-36x^3+48x^4)+384p^2(71-167x^2+96x^4) \\ +64p^3(480-45x-1408x^2-288x^4)+p^5(1109-2376x \\ +5424x^2+2304x^4)\Big),$$

we observe that only real solutions (p,x) of the system of equations $\partial n_5/\partial x = 0$ and $\partial n_5/\partial p = 0$ are $(10.6873, 0.22889)$, $(-2,-2.84595)$, $(2.70664, 2.22228)$, $(2, 2.03047)$, $(0,0)$, $(-2,-0.783981)$, $(2,-0.385354)$, $(-1.77349, 0.0460014)$, $(2,-0.64512)$, $(1.98564,-0.516895)$, $(2.00568,-0.120347)$, $(2, 0.657704)$, $(-2.33662,-1.5235)$, $(-1.03494, 0.828224)$ and $(1.36475,-1.03084)$. Thus, no solution exists in $(0,2)\times(0,1)$, resulting in no critical points.

(3) Now, we calculate the maximum values achieved by $D(p,x,y)$ on the edges of the cuboid T. From equation (3.20), we have $D(p,0,0) = h_1(p) := (55296p^2-13824p^4+1109p^6)/424673280$. It is easy to observe that $h_1'(p) = 0$ for $p = \delta_0 := 0$ and $p = 1.83094$ in the interval $[0,2]$.

$$D(p,0,0) \le 0.000169059.$$

Now considering (3.20) at $y = 1$, we get $D(p,0,1) = h_2(p) := (737280 - 368640p^2 + 54528p^3 + 46080p^4 - 13632p^5 + 1109p^6)/424673280$. It is easy to observe that $t_2'(p)$ is a decreasing function in $[0,2]$ and hence $p = 0$ acts as its point of maxima. Thus

$$D(p,0,1) \le \frac{1}{576} = 0.00173611, \quad p \in [0,2].$$

Through computations, (3.20) shows that $D(0,0,y) = y^2/576$, attains its maximum value at $y = 1$. This implies that

$$D(0,0,y) \le \frac{1}{576}, \quad y \in [0,1].$$

Since, (3.24) is independent of x, we have $D(p,1,1) = D(p,1,0) = h_3(p) := (398592p^2 - 121056p^4 + 6461p^6)/424673280$. Now, $h_3'(p) = 294912p - 175680p^3 + 13206p^5 = 0$ when

$p = \delta_2 := 0$ and $p = \delta_3 := 1.39681$ in the interval $[0, 2]$ with δ_2 and δ_3 as points of minima and maxima, respectively. Hence

$$D(p, 1, 1) = D(p, 1, 0) \leq 0.000859123, \quad p \in [0, 2].$$

On substituting $p = 0$ in (3.24), we get, $D(0, 1, y) = 0$. The equation (3.19) is independent of the all the variables namely p, x and y. Thus the maximum value of $D(p, x, y)$ on the edges $p = 2, x = 1; p = 2, x = 0; p = 2, y = 0$ and $p = 2, y = 1$, respectively, is given by

$$D(2, 1, y) = D(2, 0, y) = D(2, x, 0) = D(2, x, 1) = \frac{1109}{6635520}, \quad x, y \in [0, 1].$$

From (3.20), we obtain $D(0, 0, y) = y^2/576$. A simple calculation shows that

$$D(0, 0, y) \leq \frac{1}{576}, \quad y \in [0, 1].$$

Using (3.18), we obtain $D(0, x, 1) = h_4(x) := (5 - 4x^2 - x^4)/2880$. Upon calculations, we see that h_4 is a decreasing function in $[0, 1]$ and hence attains its maximum value at $x = 0$. Thus

$$D(0, x, 1) \leq \frac{1}{576}, \quad x \in [0, 1].$$

On again using (3.18), we get $D(0, x, 0) = h_5(x) := x(1-x^2)/480$. On further calculations, we get $t_5'(x) = 0$ for $x = \delta_4 := 1/\sqrt{3}$. Also, $h_5(x)$ is an increasing function in $[0, \delta_4)$ and decreasing in $(\delta_4, 1]$. So, it attains its maximum value at δ_4. Thus

$$D(0, x, 0) \leq 0.000801875, \quad x \in [0, 1].$$

In view of all the cases, the inequality (3.15) holds. The function specified in (3.14) acts as an extremal function for the bounds of $|H_2(3)|$ having values $a_3 = a_5 = 0$ and $a_4 = 1/24$. ∎

Acknowledgment. Neha Verma is thankful to the Department of Applied Mathematics, Delhi Technological University, New Delhi-110042 for providing Research Fellowship.

References

[1] N. M. Alarifi, R. M. Ali and V. Ravichandran, On the second Hankel determinant for the kth-root transform of analytic functions, Filomat **31** (2017), no. 2, 227–245.
[2] D. Alimohammadi, E. Analouei Adegani, T. Bulboacă and N. E. Cho, Logarithmic coefficients for classes related to convex functions. Bull. Malays. Math. Sci. Soc., **44** (2021), no. 4, 2659–2673.
[3] S. Banga and S. S. Kumar, Sharp bounds of third Hankel determinant for a class of starlike functions and a subclass of q-starlike functions, Khayyam J. Math. **9** (2023), no. 1, 38–53.
[4] S. Banga and S. S. Kumar, The sharp bounds of the second and third Hankel determinants for the class $\mathcal{SL}^*$, Math. Slovaca **70** (2020), no. 4, 849–862.
[5] A. W. Goodman, *Univalent functions*. Vol. I, Mariner Publishing Co., Inc., Tampa, FL, 1983.
[6] S. Kanas, Coefficient estimates in subclasses of the Carathéodory class related to conical domains. Acta Mathematica Universitatis Comenianae. New Series, **74** (2005), no. 2, 149-161.
[7] B. Kowalczyk, A. Lecko and D. K. Thomas, The sharp bound of the third Hankel determinant for starlike functions, Forum Mathematicum. De Gruyter, (2022).
[8] V. Kumar, N. E. Cho, V. Ravichandran and H. M. Srivastava, Sharp coefficient bounds for starlike functions associated with the Bell numbers, Math. Slovaca **69** (2019), no. 5, 1053–1064.
[9] S. S. Kumar and N. Verma, On a Subclass of Starlike Functions Associated with a Strip Domain. arXiv e-prints, pp.arXiv:2312.15266 (2023) (accepted in Ukrainian Mathematical Journal).
[10] S. S. Kumar and N. Verma, Coefficient problems for starlike functions associated with a petal shaped domain. arXiv e-prints, pp.arXiv:2210.01435 (2022).
[11] S. S. Kumar and P. Yadav, On Starlike Functions Associated with a Bean Shaped Domain. arXiv e-prints, pp.arXiv:2403.14162 (2024).
[12] O. S. Kwon, A. Lecko and Y. J. Sim, On the fourth coefficient of functions in the Carathéodory class, Comput. Methods Funct. Theory **18** (2018), no. 2, 307–314.
[13] R. J. Libera and E. J. Złotkiewicz, Early coefficients of the inverse of a regular convex function, Proc. Amer. Math. Soc. **85** (1982), no. 2, 225–230.

[14] W. C. Ma and D. Minda, A unified treatment of some special classes of univalent functions, Proc. Confer. Complex Anal. (Tianjin, 1992), 157–169.
[15] K. Marımuthu, J. Uma and T. Bulboacă, Coefficient estimates for starlike and convex functions associated with cosine function. Hacett. J. Math. Stat., **52** (2023), no. 3, 596-618.
[16] C. Pommerenke, On the coefficients and Hankel determinants of univalent functions, J. London Math. Soc., **41** (1966), 111—122.
[17] D. V. Prokhorov and J. Szynal, Inverse coefficients for (α, β)-convex functions, Ann. Univ. Mariae Curie-Skłodowska Sect. A **35** (1981), 125–143 (1984).
[18] V. Ravichandran and S. Verma, Bound for the fifth coefficient of certain starlike functions, C. R. Math. Acad. Sci. Paris **353** (2015), no. 6, 505–510.
[19] N. Verma and S. S. Kumar, A Conjecture on $H_3(1)$ for Certain Starlike Functions. Math. Slovaca, **73** (2023), no. 5, 1197–1206.
[20] N. Verma and S. S. Kumar, Certain Coefficient Problems of $\mathcal{S}_e^*$ and $\mathcal{C}_e$. arXiv e-prints, pp.arxive:2208.14644 (2022).

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

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India
Email address: nehaverma1480@gmail.com

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Online-Learning-Based Multi-RIS-Aided Wireless Systems

Ishaan Sharma, Rohit Kumar, and Sumit J. Darak, *Senior Member, IEEE*

***Abstract*—The evolution of software-defined radios and reconfigurable intelligent surfaces (RIS) has enabled on-the-fly control and reconfigurability at the physical layer parameters and radio propagation environment. In multi-RIS-aided communication, the RIS block, comprising a certain number of RIS elements from one or more RIS, is selected to achieve high throughput reliable communication between transmitter and receiver. However, selecting an RIS block when there are multiple RIS and receivers is not trivial due to the large number of candidate blocks. In this article, a novel multiarmed bandit (MAB) framework, which can learn and select the optimal RIS block using focused exploration, is proposed. We provide the theoretical regret bound for the proposed algorithm and demonstrate the gain in performance over existing state-of-the-art statistical and MAB approaches via detailed simulation results in terms of rate, ergodic capacity, outage probability, energy efficiency, and received SNR. The proposed algorithm offers 33%–85% lower latency than existing MAB algorithms on various edge platforms. Furthermore, the gain in performance and latency improves with the increase in the number and size of the RIS in the network.**

***Index Terms*—Multiarmed bandit (MAB), multi-reconfigurable intelligent surface (RIS), RIS, upper confidence bound (UCB).**

I. Introduction

AFTER an initial breakthrough in the evolution of analog radio to digital radio, there have been exciting contributions to making digital radio software-controlled and software-defined. This is followed by intelligent radio evolution in which the radio parameters, such as carrier frequency, bandwidth, number of antennas, beam directions, and various physical layer parameters can be controlled on the fly [1], [2]. One well-known example of such intelligent radio is cognitive radio, through which dynamic spectrum access has become a reality [2]. Still, there is limited control over the radio propagation environment, and it limits the performance of cellular networks, resulting in poor throughput and a large number of call drops. In the last few years, we have seen significant interest in making the radio environment smart and controllable due to its potential to improve network performance and cellular coverage [3]. Specifically, the aim is to exploit smart radio's on-the-fly configuration capabilities by controlling the transmitted signal's reflection, scattering, and refraction to improve the network performance. One potential solution is reconfigurable intelligent surfaces (RISs) comprised of multiple passive metamaterial reflecting elements [4]. By controlling the amplitude and phase of these elements via the controller, we can redirect the transmitted signal in the direction of the intended receivers [5], [6]. Various studies showed that the RIS offers higher reliability, improved throughput, and low-cost solutions by reducing the number of RF chains in massive antenna systems and enabling full-duplex communication [7], [8]. Other RIS applications include improved security [9] and target localization [10]. The challenges related to channel estimation due to many RIS elements and robust beamforming due to imperfect channel state information (CSI) were also considered in the literature [11], [12]. In beyond 5G applications, RIS was explored for improving the coverage of millimeter wave (mmWave) systems where propagation loss is significant [13]. Recent works discussed channel modeling for RIS-based communication in the Terahertz (THz) spectrum [14]. In [9], [15], and [16], machine learning and deep learning algorithms were explored to improve the performance of RIS-based communication.

Recently, few works have considered multiple RIS to improve the throughput and reliability of communication links. In [17], cooperative beamforming design for multi-RIS-aided systems was discussed along with channel modeling. They demonstrated the superiority of distributed RIS over centralized RIS. In [18], an efficient pilot transmission method was proposed to enable the receiver to separate signals arriving from different RIS, which further helps in estimating the channel state and localization. In [19], the challenge of timing synchronization between multiple RIS was addressed while the accurate configuration of multiple RIS, especially in mmWave multiantenna system, was discussed in [20]. In [21], the application of stochastic geometry aided robust beamforming approach showed improved performance of the multi-RIS-aided system in mmWave. The accurate channel estimation in an RIS-aided system is challenging. A robust transceiver design for the multi-RIS assisted multiantenna system with erroneous channel estimation was discussed in [22]. In [23], a multiple RIS-assisted multihop communication system, in which multiple RISs work in a domino pattern, was explored along with closed-form expressions for ergodic capacity and the outage probability. In [24], the need

Manuscript received 20 July 2023; revised 21 December 2023 and 22 March 2024; accepted 17 April 2024. This work was supported by the core research grant (CRG) awarded to Dr. Sumit J. Darak from DST-SERB, GoI. The work of Rohit Kumar was supported by START-UP RESEARCH GRANT (SRG/2022/002178) from Science and Engineering Research Board (SERB), Government of India. *(Corresponding author: Rohit Kumar.)*

Ishaan Sharma and Rohit Kumar are with the Electronics and Communications Department, Delhi Technological University, Delhi 110042, India (e-mail: rohit.kumar@dtu.ac.in).

Sumit J. Darak is with the Electronics and Communications Department, IIIT-Delhi, Delhi 110020, India (e-mail: sumit@iiitd.ac.in).

Digital Object Identifier 10.1109/JSYST.2024.3391856

1937-9234 © 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 May 10,2024 at 09:44:39 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.

for backhaul support for the feasibility of a multi-RIS system was highlighted. Recent works demonstrated the usefulness of a multi-RIS-aided system for indoor THz wireless communication in the presence of mobile human blockage [25] and for vehicular communication [26]. Do et al. [27] considered a multi-RIS-aided communication system and proposed two schemes for RIS selection. In the first scheme, exhaustive RIS-aided (ERA), all RIS were used, while in the second scheme, ORA scheme, single optimal RIS was used. Yang et al. [28] considered relay-aided communication systems using multiple RIS and compared the gain in outage probability. However, most of these works assume prior knowledge of perfect CSI.

To address scalability issues, prior knowledge of CSI and RIS selection, the multiarmed bandit (MAB) framework was explored in the literature [29], [30]. MAB framework-based online learning algorithms are a type of reinforcement learning in which the player needs to learn and select the optimal arm as many times as possible. Since the arm statistics are unknown, the algorithm must optimally satisfy the exploration–exploitation tradeoff. MAB algorithms are being used extensively to develop learning algorithms for decision-making due to their analytical traceability. Furthermore, they are compute and memory efficient and do not need prior training like machine learning and deep learning algorithms [31]. Popular wireless applications of MAB-based decision-making include cognitive radio networks [32], [33], Energy harvesting networks [34], mmWave communications [29], [35], massive multiantenna system [36], [37], and 5G cellular networks [38]. We focus on the stochastic setting where classical bandit algorithms like upper confidence bound (UCB) [39], KL-UCB [40], and Thompson sampling [41] are applied to networks by suitably mapping actions (channels, modulations schemes, power level, RIS) to arms and quantity interests (success rate, throughput) to rewards to obtain learning algorithms. Recently, MAB algorithms were used for the selection of RIS in cellular Internet-of-Things, unmanned aerial vehicle, and mmWave applications in [29], [30], and [42], respectively. As discussed later in Section III-B, direct application of MAB algorithm for RIS selection in multi-RIS-aided system results in the poor performance. To the best of authors' knowledge, the design of MAB algorithms for multi-RIS-aided wireless systems has not been discussed yet in the literature. The MAB problem setup for multiple receivers is not trivial. This work aims to develop a new efficient MAB algorithm for wireless systems comprised of multiple RIS and receivers. We analyze the performance of the proposed algorithm via theoretical, simulation, and experimental results on the edge platform. In Table I, we compare various state-of-the-art works in the literature, and our contributions are summarized as follows.

1) In this work, we set up the RIS selection problem in a multi-RIS-aided system with multiple receivers as an MAB framework where each arm corresponds to one or more RIS with a certain number of passive elements. The aim is to learn the optimal combination of RIS and their elements that offer higher average SNR/throughput. Since the number of RIS blocks can be significantly large, resulting in higher exploration time of the conventional MAB algorithm, we exploit the sparsity to significantly reduce the number of candidate blocks by preidentifying the blocks whose average throughput is above a given threshold. We demonstrate the gain in performance over existing state-of-the-art approaches via theoretical and simulation results.
2) We propose two MAB algorithms: focused exploration UCB (FEUCB) and Enhanced focused exploration UCB (EFEUCB) algorithms. We compare the performance with existing learning and nonlearning-based approaches in terms of various performance metrics, such as regret, rate, outage probability, ergodic capacity, energy efficiency, transmit power, and consumed power.
3) We compare the performance of the centralized RIS and distributed RIS systems and demonstrate the superiority of the distributed multi-RIS-aided system for multiple-user cases.
4) In time-slotted communication, faster RIS selection allows more time for data communication, resulting in higher throughput. To analyze this, we design and implement the proposed and existing MAB-based RIS selection algorithms on various embedded edge platforms, such as ARM Cortex A9 and ARM Cortex A53 processors augmented with single instruction multiple data (SIMD) coprocessors. We study the effect of word length on the latency, i.e., execution time, of the algorithms. We demonstrate the lower execution time of the proposed algorithms compared to the state-of-the-art MAB algorithms.

TABLE I
COMPARISON OF STATE-OF-THE-ART WORKS

Papers	No. of RIS	No. of Recievers	Active RIS	Learning used for Selection	Hardware Results
[9]	Single	Multiple	One	Yes	No
[42]	Single	Multiple	One	Yes	No
[17]	Multiple	Multiple	Multiple	No	No
[23]	Multiple	Multiple	Multiple	No	No
[27]	Multiple	Single	All (ERA), One(ORA)	No	No
[29]	Multiple	Multiple	One	Yes	No
[30]	Multiple	Single	One	Yes	No
Proposed work	Multiple	Multiple	Multiple	Yes	Yes

The rest of this article is organized as follows. Section II discusses the network model, followed by the details of the proposed algorithms and regret analysis in Section III. Detailed simulation results are presented in Section IV followed by experimental results on the edge platform in Section V. Finally, Section VI concludes this article.

II. NETWORK MODEL

This work considers the multi-RIS-aided system comprising one transmitter, K RIS, and N receivers. Each RIS is divided into L sub-blocks, and each sub-block consists of Z sub-λ sized passive RIS elements (Fig. 1). We consider the time-slotted communication where the transmitter must select M out of KL sub-blocks in each time slot. Thus, there are $R=\binom{\mathrm{KL}}{M}$ combinations of M sub-blocks. The combination of any M sub-blocks is referred to as a block. For the chosen block, the amplitude, $A \in [0, 1]$, and phase, $\theta \in [0, 2\pi]$, of all elements are optimally configured to establish the communication with N

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 09:44:39 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.

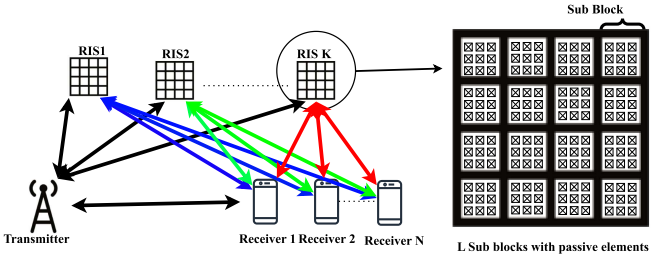


Fig. 1. Illustrations of network model for multi-RIS-aided wireless system.

receivers. We denote the block selected in the time slot, t, as $I_t \in [R] := \{1, 2, \ldots, R\}$. Note that the size of each sub-block and the number of sub-blocks per RIS are not fixed. Without loss of generality, our framework allows the transmitter to select more than one RIS in a given slot by having the sub-blocks in I_t from the same or multiple RIS.

We consider the channel model in [27], where channels associated with elements of the same RIS are assumed to be independent and identically distributed. In contrast, channels associated with different RISs are assumed to be independent but not identically distributed, and the system undergoes Nakagami-m fading. The complex channel coefficient between the transmitter and zth element of the RIS sub-block, r, is defined in the polar form as $\tilde{G}_{klz} = G_{klz}e^{j\psi_{klz}}$ where G_{klz} and ψ_{klz} denote the magnitude and phase, respectively. Similarly, the channel between the zth element of the RIS sub-block, r, and nth receiver is defined in the polar form as $\tilde{H}_{klnz} = H_{klnz}e^{j\phi_{klnz}}$ where H_{klnz} and ϕ_{klnz} denote the magnitude and phase, respectively. The channel coefficient along the direct path between the transmitter and nth receiver is $\tilde{G}_n = G_n e^{j\psi_n}$ where G_n and ψ_n denote the magnitude and phase, respectively.

For transmit symbol, x_s with power, P_s in dBm, the received signal at the nth receiver using the RIS sub-block, m, is given by [27]

$$y_n^m = \sqrt{P_s}\left(\tilde{G}_n + \sum_{z=1}^{z=Z} \tilde{G}_{klz}\alpha_{klz}e^{j\theta_{klz}}\tilde{H}_{klnz}\right)x_s + \omega_n \quad (1)$$

where α_{klnz} and θ_{klnz} are the amplitude reflection coefficient and phase-shift of the zth element of the m. ω_n denotes the additive white Gaussian noise with zero mean and variance, σ_n^2. Then, the SNR at the nth receiver, assuming optimal RIS configuration with zero phase error, is given as

$$\text{SNR}_n^m = \frac{P_s}{\sigma_n^2}\left|G_n + \sum_{z=1}^{z=Z} G_{klz}\alpha_{klz}H_{klnz}\right|^2. \quad (2)$$

For reliable communication between the transmitter and all receivers, we need to select the block, i.e., M RIS sub-blocks, which cumulatively offer an average high cumulative SNR and guarantee that the SNR at each receiver is above a certain threshold, Δ

$$I_t = \arg\max_{\substack{I_t\in[R]\\|I_t|=M}} \left[\sum_{\substack{v=i_1\\v\in[I_t]}}^{i_M}\left(\prod_{n=1}^{N} 1_{\text{SNR}_n^r>\Delta}\sum_{n=1}^{N}\text{SNR}_n^r\right)\right] \quad (3)$$

where 1_{cond} is an indicator function. Selection of all RIS results in good throughput performance but extremely high power consumption. Since SNR over each sub-block is unknown, the selection of RIS block is not trivial; hence, we need a learning framework. Due to a large number of sub-blocks and multiple receivers, one-shot learning via deep learning is not scalable [43]. The MAB algorithm is well suited for scenarios where decisions have to be made in the face of uncertainty, and the objective is to maximize accumulated rewards over time. Although such sequential learning is a promising approach, it is time-consuming, and hence, efficient algorithms are needed.

III. Proposed Work

The selection of the RIS block in a multi-RIS-aided system with multiple receivers can be set up as an MAB problem where the aim is to select the optimal block as often as possible via exploration–exploitation trade off. In this section, we discuss the selection of optimal block for multiple receiver scenarios and the drawbacks of the existing MAB algorithms. We address these challenges using the proposed algorithms.

A. MAB Framework

The standard stochastic multiplay MAB consists of a set of arms and a single player. Here, we refer to RIS sub-blocks as arms and a player as a transmitter. In each time slot, a transmitter selects a block of M RIS sub-blocks and receives a reward equal to SNR at each receiver. For each receiver, the reward is assumed to be drawn independently across time from distributions that are stationary and independent across sub-blocks. We denote the mean reward of nth receiver for block, r as SNR_n^r, and hence, total reward in time slot, t is given as

$$X^{I_t} = \sum_{\substack{v=i_1\\v\in[I_t]}}^{i_M}\left(\prod_{n=1}^{N} 1_{\text{SNR}_n^r>\Delta}\sum_{n=1}^{N}\text{SNR}_n^r\right). \quad (4)$$

The performance metric for the MAB algorithm is regret, which compares the SNR degradation due to sub-optimal block selection, and it is given as

$$R = TX^{r_*} - \mathbb{E}\left[\sum_{\substack{r=1\\r\in[R]}}^{R} X^r S^r\right] \quad (5)$$

where T is the total number of time slots, S^r is the number of times the rth block is selected by base station (BS) and r_* is the optimal block, which offers highest SNR. The expectation is with respect to the random number of selections of the block S^r. Thus, the regret can be minimized by selecting r_*, i.e., the combination of M sub-blocks with the highest SNR as often as possible in a given horizon of size T.

B. Limitations of Existing MAB Framework

In this article, we limit our discussion to the UCB-based MAB algorithm [44] and provide regret bounds. The proposed idea can be easily extended to other MAB algorithms, such as UCB variants and Thompson Sampling. In the UCB algorithm, each

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 09:44:39 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.

block is selected once in the beginning. After that, in each time slot ($t > R$), the block that offers the highest UCB factor is selected. The UCB factor of the block, r, in time slot t is given as

$$\text{UCB}^r(t) = \frac{\hat{X}^r(t)}{S^r(t)} + 2\sqrt{\frac{\log(t)}{S^r(t)}} \tag{6}$$

where $\hat{X}^r(t)$ denotes the accumulated SNR obtained over $S^r(t)$ time slots during which UCB selects the block, r. The expected regret of the UCB scales as $\mathcal{O}(\sum_{r\in[R]\setminus r_*} \frac{\log T}{\Delta_r})$ [45] where $\Delta_r = X^{r_*} - X^r$ for all $r \neq r_*$. For fixed T, the distribution-independent bounds are of the order $\sqrt{RT}$ [45]. It is evident that the UCB suffers from high exploration time, especially when R is large.

C. Proposed MAB Algorithms

In the Multi-RIS-aided system with N receivers, the number of RIS blocks that offer reasonably good SNRs for all receivers simultaneously is expected to be limited, and hence, we can safely assume that $\hat{R} \leq R$ where $\hat{R}$ are good RIS blocks. In such a case, we can reduce the distribution-independent upper bound to $\sqrt{\hat{R}T}$ and develop the algorithms that satisfy this bound. In the proposed algorithm, inspired from [46], we focus on quickly identifying and exploring these $\hat{R}$ blocks, and it is referred to as the FEUCB algorithm. As shown in Algorithm 1, all the blocks are selected once initially (Lines 4–5). In the rest of the time slots, we check whether at least $\hat{R}$ blocks have been explored sufficiently till time t (Line 9). If not, we will use the conventional UCB algorithm with R blocks (Line 12). Otherwise, we restrict the UCB only to the selected $\hat{R}$ blocks that have been explored sufficiently. The set of blocks that are sufficiently explored is given as

$$\mathcal{E}(t) := \left\{ r \in [R] \middle| \hat{X}^r(t) \geq 2\sqrt{\frac{\log(t)}{S^r(t)}} \right\}. \tag{7}$$

The FEUCB algorithm aims to avoid selecting RIS blocks with poor SNR into the $\hat{R}$ blocks. This is done by ensuring that each of the selected $\hat{R}$ blocks has been sampled a sufficient number of times, as shown in (7). Using Theorem 1, we show that the proposed FEUCB algorithm satisfies logarithm regret, and the scaling factor depends on $\hat{R}$ instead of R, thereby outperforming the UCB algorithm.

Theorem 1: The FEUCB algorithm guarantees that the expected regret is upper bounded as

$$\mathbb{E}[R(T)] \leq \log(T) \sum_{r\in[\hat{R}]\setminus r_*} \frac{1}{\Delta_r}. \tag{8}$$

Proof: We define the following three events.

1) $A_r(t)$: An event of good RIS block, r, being selected during the UCB phase when $\hat{R}$ blocks are sampled sufficiently (Line 10).
2) $B_r(t)$: An event of good RIS block, r, being selected during the UCB phase when $\hat{R}$ blocks are not sampled sufficiently (Line 12).
3) $C_r(t)$: An event of poor RIS block, r, being selected.

Algorithm 1: FEUCB.

1: **Input:** $R, \hat{R}, T$
2: **Initialize:** $\hat{X}^r(t) \leftarrow 0$ and $S^r \leftarrow 0$ for all r
3: **for** $t = 1, 2 \ldots T$, **do**
4: **if** $t \leq R$ **then**
5: Select block, $I_t = t$.
6: **else**
7: $UCB^r(t) \leftarrow 0$ for all r
8: Compute $\mathcal{E}(t)$ using (7).
9: **if** $|\mathcal{E}(t)| \geq \hat{R}$ **then**
10: $\forall r \in [\mathcal{E}(t)]$: Update $UCB^r(t)$ as given in (6)
11: **else**
12: $\forall r \in [R]$ Update $UCB^r(t)$ as given in (6)
13: **end if**
14: Select block, $I_t = \arg\max_{r\in[R]} UCB^r(t)$
15: **end if**
16: Transmitter configures RIS block, I_t and transmits a data frame.
17: Each receiver observes instantaneous normalized SNR, and communicates to the transmitter, which calculates X^{I_t} using (4).
18: $S^{I_t} \leftarrow S^{I_t} + 1$ and $\hat{X}^r(t) \leftarrow \hat{X}^r(t) + X^{I_t}$.
19: **end for**

Then,

$$\begin{aligned}\mathbb{E}[R(T)] = 1 &+ \sum_{r\in[\hat{R}]\setminus r_*} \Delta_r \,\mathbb{E}\left[\sum_{t=1}^{T} 1_{A_r(t)}\right] \\ &+ \sum_{r\in[\hat{R}]\setminus r_*} \Delta_r \,\mathbb{E}\left[\sum_{t=1}^{T} 1_{B_r(t)}\right] \\ &+ \sum_{r\notin[\hat{R}]} \Delta_r \,\mathbb{E}\left[\sum_{t=1}^{T} 1_{C_r(t)}\right].\end{aligned} \tag{9}$$

The first term is during the initialization phase when each block is selected once. Using Lemma 1–3, we can independently upper bound each of the three terms

$$\begin{aligned}\mathbb{E}[R(T)] = 1 &+ \sum_{r\in[\hat{R}]\setminus r_*} \left(\frac{16\log(T)+8}{\Delta_r} + 3\right) + \frac{R\Delta_{\hat{R}}\pi^2}{6} \\ &+ \sum_{r\notin[\hat{R}]} \frac{\pi^2\Delta_r}{6}.\end{aligned} \tag{10}$$

■

Since all except the second terms are independent of T, the second term dominates the regret.

Lemma 1: The regret due to event $A_r(t)$ is given as

$$\mathbb{E}\left[\sum_{t=1}^{T} 1_{A_r(t)}\right] \leq \sum_{r\in[\hat{R}]\setminus r_*} \left(\frac{16\log(T)+8}{\Delta_r^2} + 3\right). \tag{11}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 09:44:39 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.

Proof: The proof is directly based on the UCB algorithm regret analysis in [44]. ■

Lemma 2: The regret due to event $B_r(t)$ is given as

$$\sum_{r\in[\hat{R}]\setminus r_*} \Delta_r \,\mathbb{E}\left[\sum_{t=1}^{T} 1_{B_r(t)}\right] \leq \frac{R\Delta_{\hat{R}}\pi^2}{6}. \tag{12}$$

Proof: The FEUCB algorithm uses the UCB algorithm over all RIS blocks when good $\hat{R}$ blocks are not sampled sufficiently. Since UCB is based on an exploration–exploitation tradeoff, the probability of this happening for the entire horizon is finite. Thus, the probability that a good RIS block is selected when good $\hat{R}$ blocks are not sufficiently sampled is finite.Assuming $\Delta_{\hat{R}} = \max_{r\in[\hat{R}]\setminus r_*}\Delta_r$, we have

$$\sum_{r\in[\hat{R}]\setminus r_*} \Delta_r \,\mathbb{E}\left[\sum_{t=1}^{T} 1_{B_r(t)}\right] \leq \Delta_{\hat{R}} \sum_{t=1}^{T} \mathbb{E}\left[\sum_{r\in[\hat{R}]\setminus r_*} 1_{B_r(t)}\right].$$

Then,

$$\mathbb{E}\left[\sum_{r\in[\hat{R}]} 1_{B_r(t)}\right] \leq \sum_{r\notin\hat{R}} \mathbb{P}\left[\hat{X}^r(S^r(t)) \geq 2\sqrt{\frac{\log(t)}{S^r(t)}}\right].$$

Using the Chernoff bound, we have

$$\mathbb{E}\left[\sum_{r\in[\hat{R}]} 1_{B_r(t)}\right] \leq \sum_{r\notin\hat{R}} \mathbb{P}\left[\hat{X}^r(S^r(t)) - X^r \geq 2\sqrt{\frac{\log(t)}{S^r(t)}}\right] \leq \sum_{r\notin\hat{R}} t^{-2}.$$

Thus

$$\sum_{r\in[\hat{R}]\setminus r_*} \Delta_r \,\mathbb{E}\left[\sum_{t=1}^{T} 1_{B_r(t)}\right] \leq \Delta_{\hat{R}} \sum_{t=1}^{T}\sum_{r\notin\hat{R}} t^{-2} \leq \frac{R\Delta_{\hat{R}}\pi^2}{6}.$$

■

Lemma 3: The regret due to event $C_r(t)$ is given as

$$\mathbb{E}\left[\sum_{t=1}^{T} 1_{C_r(t)}\right] \leq \frac{\pi^2}{6}. \tag{13}$$

Proof: Since the poor RIS, $r \notin \hat{R}$, are not sufficiently explored by the algorithm, their regret is bounded. Let $r \notin \hat{R}$. Then, $1_{C_r(t)} = 0$ for $t \leq R$. Then

$$\mathbb{E}\left[\sum_{t=1}^{T} 1_{C_r(t)}\right] = \mathbb{E}\left[\sum_{t=R}^{T} 1_{C_r(t)}\right] \leq \mathbb{E}\left[\sum_{u=1}^{\infty} 1_{\hat{X}^r(S^r(u))\geq 2\sqrt{\frac{\log(u)}{S^r(u)}}}\right]. \tag{14}$$

Using the Chernoff bound, we have

$$\mathbb{E}\left[\sum_{u=1}^{\infty} 1_{\hat{X}^r(S^r(u))\geq 2\sqrt{\frac{\log(u)}{S^r(u)}}}\right] = \sum_{u=1}^{\infty} \mathbb{P}\left[\hat{X}^r(S^r(u)) \geq 2\sqrt{\frac{\log(u)}{S^r(u)}}\right] \leq \sum_{u=1}^{\infty} \mathbb{P}\left[\hat{X}^r(S^r(u)) - X^r \geq 2\sqrt{\frac{\log(u)}{S^r(u)}}\right]$$

Algorithm 2: EFEUCB.

1: **Input:** $R, \hat{R}, T$
2: **Initialize:** $\hat{X}^r(t) \leftarrow 0$ and $S^r \leftarrow 0$ for all r
3: **for** $t = 1, 2 \ldots T$, **do**
4: **if** $t \leq R$ **then**
5: Select block, $I_t = t$.
6: **else**
7: $UCB^r(t) \leftarrow 0$ for all r
8: Compute $\mathcal{E}(t)$ using (7).
9: **if** $|\mathcal{E}(t)| \geq \hat{R}$ **then**
10: $\forall r \in [\mathcal{E}(t)]$: Update $UCB^r(t)$ as given in (6)
11: **else**
12: $\forall r \in [R]$ Update $TCB^r(t)$ as given in (16)
13: $\forall r \in [R] s.t. TCB^r > SNR_d$, Update $UCB^r(t)$ as given in (6)
14: **end if**
15: Select block, $I_t = \arg\max_{r\in[R]} UCB^r(t)$
16: **end if**
17: Transmitter configures RIS block, I_t and transmits a data frame.
18: Each receiver observes instantaneous normalized SNR, and communicates to the transmitter, which calculates X^{I_t} using (4).
19: $S^{I_t} \leftarrow S^{I_t} + 1$ and $\hat{X}^r(t) \leftarrow \hat{X}^r(t) + X^{I_t}$.
20: **end for**

$$\leq \sum_{u=1}^{\infty} e^{-2\log(u) \leq \sum_{u=1}^{\infty} u^{-2} = \frac{\pi^2}{6}}. \tag{15}$$

■

D. Enhanced FEUCB Algorithm

In the FEUCB algorithm, the exploration is done over all R blocks until we have $\hat{R}$ sufficiently sampled blocks. Since R is large and few RIS blocks are good satisfying sparsity requirement, the performance of the FEUCB algorithm can be improved empirically by increasing the selection of good RIS blocks during the period where UCB is exploring all R blocks. The proposed empirical approach is based on a thresholding approach where the algorithm aims to identify all the RIS blocks that offer SNR above a certain desired threshold, SNR_d. Once such $\hat{R}$ blocks are identified, conventional UCB is used. As shown in Algorithm 2, the EFEUCB algorithm differs from the FEUCB algorithm on Lines 12–13, where we identify the blocks that offer SNR above a certain desired threshold, SNR_d. Such blocks are identified based on the threshold confidence bound (TCB), and it is given as

$$\text{TCB}^r(t) = \sqrt{S^r(t)}\left|\frac{\hat{X}^r(t)}{S^r(t)}\right|. \tag{16}$$

In the FEUCB algorithm, we must select $\hat{R}$ judiciously. When the selection of $\hat{R}$ is not trivial or $\hat{R}$ is large, then the performance of the FEUCB algorithm may degrade. The EFEUCB addresses this problem to a certain extent by identifying good RIS blocks using the thresholding approach, and threshold parameters can

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 09:44:39 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.

be selected based on prior knowledge of wireless system deployment, the number of RIS and their locations, wireless physical parameters, etc. In both algorithms, M is fixed. Future works will use an online learning approach to select the M depending on channel conditions.

Sparsity-based approaches are useful in large-scale applications where only a small fraction of options yield significant rewards. The empirical thresholding approach in EFEUCB ensures that only good RIS sub-blocks that offer SNR above a certain threshold are selected often. Due to fewer blocks, the EFEUCB algorithm is expected to converge to the optimal block faster than the FEUCB algorithm. However, the close-form expression for EFEUCB regret is challenging due to the nonlinear thresholding operation.

IV. Performance Analysis

We consider time-slotted communication, where the transmitter selects the RIS block in each time slot and transmits a certain fixed number of data packets over the selected RIS block. The aim is to select the optimal RIS block as often as possible. We compare the performance of the proposed FEUCB and EFEUCB algorithms with existing learning and nonlearning-based algorithms. In the learning-based approach, we consider UCB-based RIS selection approaches. In nonlearning-based algorithms, we consider ERA and opportunistic RIS-aided (ORA) scheme algorithms in [27]. ORA needs prior knowledge of RIS statistics since it selects the optimal RIS block in each time slot, while ERA selects all RIS simultaneously in each time slot. In addition, we consider a simple random selection approach.

We consider the number of receivers, $N = \{1, 2, 3, 4\}$, and the number of RIS, $K = \{5, 20\}$. Each RIS is divided into $L = 5$ sub-blocks. The number of elements in each sub-block is $Z = [25, 35, 45, 55, 65]$. In addition, we consider two more cases by varying the size of the sub-blocks of each RIS. We consider $K = \{5, 10\}$, with $L = \{4, 3, 5, 3, 4\}$ and $L = \{4, 3, 5, 3, 4, 2, 6, 5, 2, 3\}$ respectively. The number of elements in this case is also randomly chosen between 25 and 75. The horizon size, T, is between 10 000 and 500 000. Each result is averaged over 15 independent experiments over the selected horizon size. In each experiment, the positions of the transmitter, RIS, and receivers are selected randomly. Unless otherwise specified, the simulation parameters are set to the values mentioned in Table II. The equivalent noise power at the receiver is given as:$\sigma_n^2 = N_0 + 10\log(\text{BW}) + \text{Noise Figure[dBm]}$ where N_0 is the thermal noise power density.

The performance metrics used for comparison are regret (5), outage probability, ergodic capacity, and energy efficiency. The lower regret indicates a higher average SNR at the receivers, guaranteeing reliable communication. The outage probability is defined as the probability that the SNR of the system falls below a certain threshold. The achievable SNR can be affected by various factors, including RIS selection, channel fading, and receiver noise. Mathematically it can be expressed as

$$\text{P}_{\text{out}} = Pr(\text{SNR}_n^r < \text{SNR}_{\text{th}}) \tag{17}$$

TABLE II
Parameters

Parameters	Values
Location of transmitter	(0,0)
Location of receivers	Random
Location of RISs	Random
Transmit power [dBm]	[−40,40]
Amplitude reflection coef., A	1[47]
Number of RIS	5 and 20
Number of sub-blocks per RIS	5
Number of elements in sub-blocks	[25,35,45,55,65]
Threshold SNR (SNR_{th})	7dBm
Bandwidth	10 MHz [47]
Noise figure	10 [47]
Thermal noise power density,(N_0)	−174 [47]
Antenna gain[dB]	5 [47]
Carrier frequency [GHz]	3 [47]
Circuit dissipated power in RIS [mW]	7.8 [48]
Circuit dissipated power in TX and RX[mW]	10 [48]

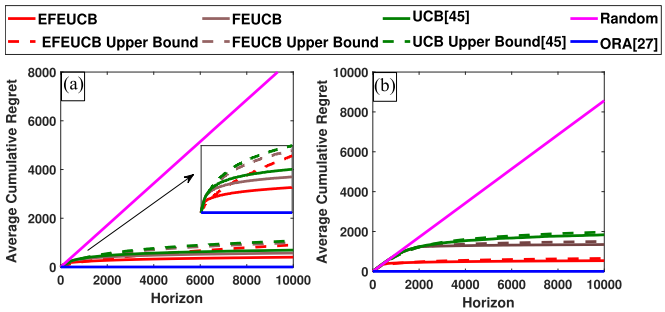


Fig. 2. Comparison of average cumulative regret for different algorithms for (a) $K = 5, L = 5$ and (b) $K = 20, L = 5$.

where SNR_{th}, is the minimum threshold SNR. The outage probability decreases with increased transmit power and the number of RIS elements.

Ergodic capacity refers to the maximum average rate at which the data can be transmitted reliably over a selected channel. Essentially, it represents the average capacity of a channel when both the sender and the recipient possess information on the channel state, and the channel varies randomly over time. Mathematically it can be expressed as

$$\text{EC} = \mathbb{E}[\log_2(1 + \text{SNR}_n^r)]. \tag{18}$$

The energy efficiency gauges the amount of data that can be sent for every unit of power used and can be expressed as

$$\eta = \frac{\text{BW} \times \text{Average Achievable Rate}}{\text{P}_C} \tag{19}$$

where the consumed power P_C is the sum of the circuit dissipated power at the transmitter, N receivers, and each element of the selected RIS block.

A. Regret Performance Analysis

We begin by comparing various algorithms for $K = 5$ (fewer number of RIS) and $K = 20$ (larger number of RIS), respectively, with $L = 5$ and $N = 4$. In Fig. 2(a) and (b), we compare the cumulative regret for the horizon size of 10 000. As expected, random RIS selection incurs high regret due to the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 09:44:39 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.

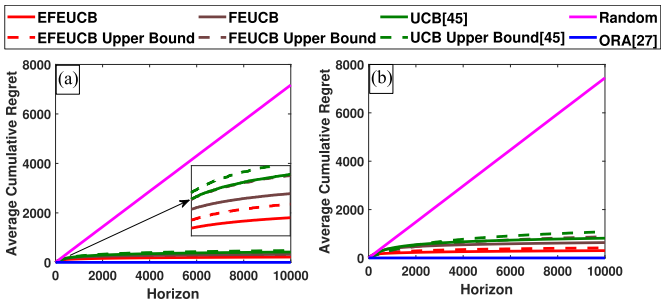


Fig. 3. Comparison of average cumulative regret for different algorithms for (a) $K = 5, L = [4, 3, 5, 3, 4]$ and (b) $K = 10, L = [4, 3, 5, 3, 4, 2, 6, 5, 2, 3]$.

frequent selection of the nonoptimal RIS blocks. The regret of the ORA scheme is zero as it has prior knowledge of the optimal RIS block. We have not included the ERA scheme for regret comparison since ERA selects all RIS blocks, while regret calculation requires selecting a single RIS block in each slot. Among learning-based UCB, FEUCB, and EFEUCB algorithms, the regret of the EFEUCB algorithm is lowest due to improved exploration, resulting in fewer selections of nonoptimal RIS blocks. It can be observed that the regret increases with the increase in K due to a large number of candidate RIS blocks, resulting in higher exploration time. However, the difference between the regret of the FEUCB/EFEUCB and UCB increases with the increase in K, which validates the superiority of the proposed focused exploration approach. The difference between FEUCB and EFEUCB increases as K increases, highlighting the impact of empirical thresholding-based enhancement in focused exploration. We have also compared the experimental regret and its desired upper bound, i.e., worst case performance. It is evident that the experimental results follow a similar trend as the theoretical results, and as expected, experimental results offer lower regret than the corresponding upper bound.

Next, we compare the regret performance for the case when all the RIS units are not identical in terms of element size and the number of sub-blocks. As shown in Fig. 3(a) and (b), the performance of the proposed algorithms is superior to the existing UCB algorithm, and EFEUCB offers improved performance as K increases.

B. Comparison of Achievable Rate, Outage Probability, and Ergodic Capacity

Next, we study the effect of the transmit power and total consumed power on the achievable rate in bits per second per Hertz (b/s/Hz). As shown in Figs. 4 and 5, different algorithms need different transmit power to achieve a given rate. For example, ERA [27] demands the lowest transmit power due to the use of all RIS, which in turn helps to improve the average SNR at the receiver. The random approach needs the highest transmit power due to the frequent selection of suboptimal RIS. All learning-based approaches need a similar transmit power as that of ORA [27], validating the successful learning and frequent selection of optimal RIS. EFEUCB offers better performance than FEUCB, which in turn offers better performance than UCB [44]. As K increases, the ERA performance improves due to increased SNR at the receiver. Furthermore, the difference between the performance of ORA and learning approaches is significant for large K, which is the penalty paid to learn optimal RIS block. As expected, focused exploration leads to a smaller penalty for EFEUCB and FEUCB than UCB.

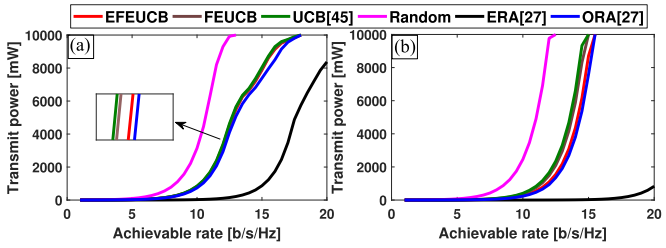


Fig. 4. Comparison of transmit power and achievable rate for different algorithms for (a) $K = 5, L = 5$, and (b) $K = 20, L = 5$.

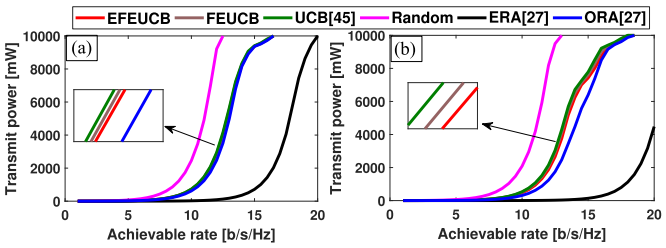


Fig. 5. Comparison of transmit power and achievable rate for different algorithms for (a) $K = 5, L = [4, 3, 5, 3, 4]$ and (b) $K = 10, L = [4, 3, 5, 3, 4, 2, 6, 5, 2, 3]$.

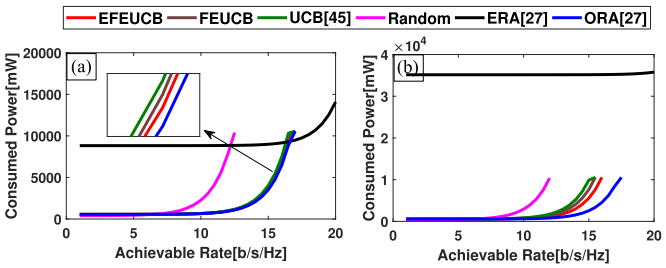


Fig. 6. Comparison of total consumed power and achievable rate for different algorithms for (a) $K = 5, L = 5$ and (b) $K = 20, L = 5$.

Next, we consider total consumed power P_C instead of transmit power in Figs. 6 and 7. As expected, ERA offers the poor performance, and the performance degrades substantially with the increase in the number of RIS. This confirms the practical challenges in the widely used ERA scheme, especially when K is large. Using the proposed learning approaches, consumed power can be reduced significantly by the appropriate selection of M.

Next, we compare the outage probability in Fig. 8(a) and (c) for different values of the transmit power. To avoid repetition of plots and discussion, the discussion is limited to $K = \{5, 20\}$ with $L = 5$. As expected, the outage probability decreases with the increase in the transmit power. The ERA scheme outperform others since it selects all RIS blocks, improving SNR at all

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 09:44:39 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.

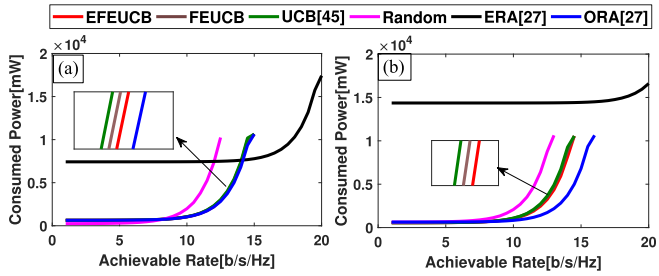


Fig. 7. Comparison of total consumed power and achievable rate for different algorithms for (a) $K = 5, L = [4, 3, 5, 3, 4]$ and (b) $K = 10, L = [4, 3, 5, 3, 4, 2, 6, 5, 2, 3]$.

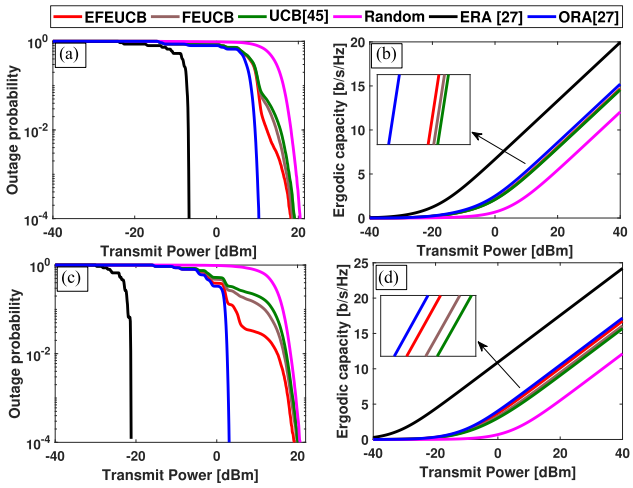


Fig. 8. Comparison of outage probability for (a) $K = 5$, and (c) $K = 20$, and ergodic capacity for (b) $K = 5$ and (d) $K = 20$ with $L = 5$ for different values of transmit power.

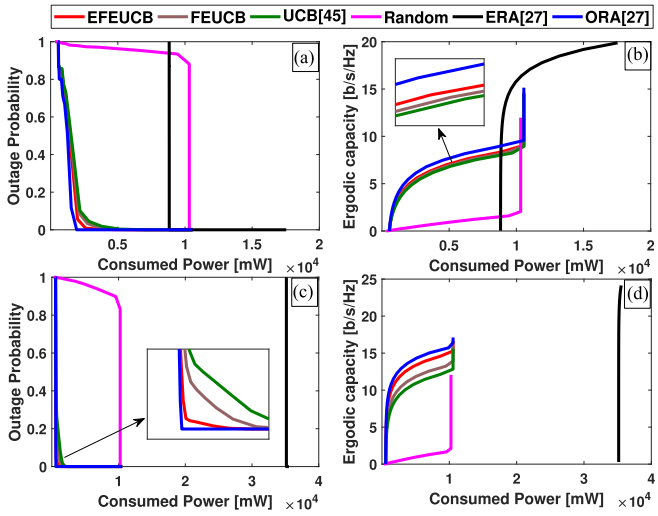


Fig. 9. Comparison of outage probability for (a) $K = 5$, and (c) $K = 20$, and ergodic capacity for (b) $K = 5$ and (d) $K = 20$ with $L = 5$ for different values of consumed power.

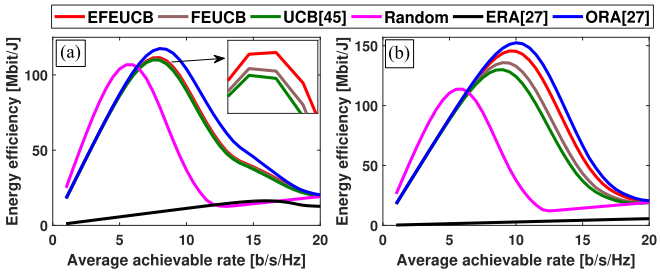


Fig. 10. Comparison of energy efficiency and average achievable rate for different algorithms for (a) $K = 5, L = 5$ and (b) $K = 20, L = 5$.

receivers. This is because the number of sub-blocks L in the ERA scheme is more as compared to the ORA scheme; hence, the total number of elements Z are more in the case of ERA scheme [27], because of which the outage probability of ERA decreases faster than that of ORA with a slight increase in transmit power. As expected, the outage probability of learning-based approaches is close to that of ORA since learning-based approaches eventually select the optimal RIS block after the initial exploration–exploitation tradeoff. The divergence at $\mathrm{P_{out}} = 10^{-2}$ is because of exploring other blocks while learning the optimal block. In Fig. 8(c), with a total of 20 RIS, ERA offers further improvement in performance as its outage probability starts decreasing at lower transmit power than in Fig. 8(a) with 5 RIS.

Similarly, in Fig. 8(b) and (d), we compare the ergodic capacity for different values of the transmit power. As expected, the ERA scheme offers the highest ergodic capacity due to the selection of all RIS elements, which allows the highest throughput for multiple users. The proposed EFEUCB offers a better outage probability among learning-based approaches. As K increases, exploration time increases; hence, the difference between the performance of ORA and learning-based approaches increases. This penalty is paid due to no prior knowledge of channel conditions and positions of RIS and receivers in the learning-based approach. Note that the difference between the performance of ERA and ORA increases with the increase in K, which validates the need for multi-RIS-aided wireless networks. In Fig. 9, we compare the outage probability and ergodic capacity for different values of consumed power. As expected, the performance of the ERA scheme is significantly poor due to the selection of all RIS, which results in high consumed power. Thus, the proposed approach addresses ERA's high consumed power drawback by learning and selecting the optimal RIS block instead of all RIS. It also outperforms UCB via focused exploration, especially when K is large.

C. Energy Efficiency Comparison

Next, we compare the energy efficiency of various algorithms for different achievable rates in Fig. 10(a) and (b). Here, we increase the transmit power to achieve the desired rate. As expected, the energy efficiency of the ERA scheme is abysmal due to the selection of all RIS elements. In contrast, the energy efficiency of ORA is highest due to the selection of optimal RIS, which results in lower consumed power. The energy efficiency of the learning-based scheme is significantly better than the ERA scheme and close to that of the ORA scheme. The difference between the energy efficiency of learning-based approaches and ORA is due to the higher transmit power needed

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 09:44:39 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.

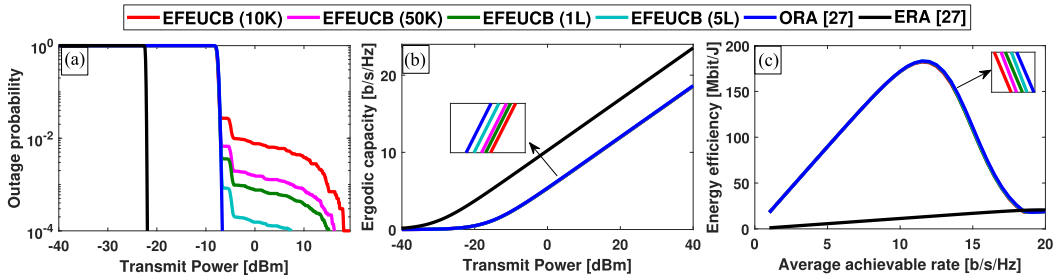


Fig. 11. Comparison of outage probability, ergodic capacity, and energy efficiency for different horizon time.

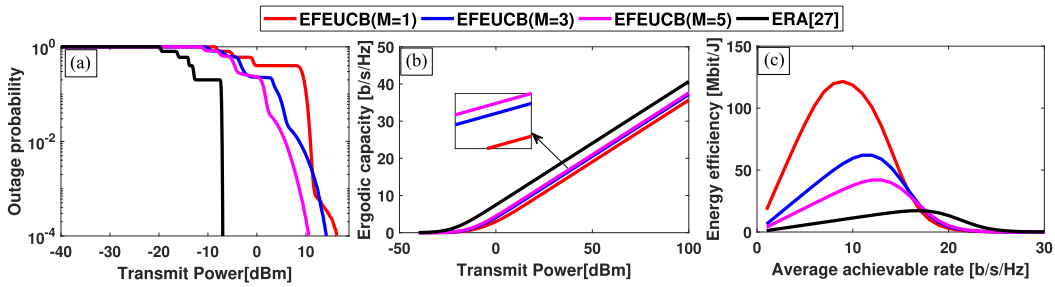


Fig. 12. Comparison of outage probability, ergodic capacity, and energy efficiency when different block sizes, M.

to meet the desired achievable rate whenever suboptimal RIS is selected due to the exploration–exploitation tradeoff. As the number of RIS increases, the energy efficiency of the proposed FEUCB and EFEUCB is better than that of UCB and random approaches, validating the superiority of the proposed focused exploration approach. The energy efficiency of the ORA and proposed learning-based algorithms decreases when the desired achievable rate goes beyond a certain value, as shown in Fig. 10. This is due to a significant increase in the desired transmit power, as shown in Fig. 4.

D. Effect of Horizon Size, T

Compared to ORA schemes, learning-based approaches do not have prior knowledge of the optimal RIS; hence, they need exploration time to learn and identify the optimal RIS block. In Fig. 11, we compare the effect of horizon size on the performance of learning algorithms. It can be observed that the difference between the learning and ORA schemes decreases with the increase in horizon time for a given K and L. This is because the penalty for selecting nonoptimal RIS during exploration time becomes insignificant as horizon time increases. This also validates the functional accuracy of the proposed learning-based approaches to identify the optimal RIS accurately and select it as many times as possible over the horizon.

E. Effect of Block Size, M

Each RIS block consists of M number of sub-blocks, and $M = 1$ corresponds to a single-RIS-aided system. We compare the effect of M on the performance of wireless systems in Fig. 12 for $M = \{1, 3, 5\}$. It can be observed that the outage probability and ergodic capacity of the EFEUCB improved with the increase in M and approaches that of ERA as M increases at the cost of poor energy efficiency. Thus, M should be carefully selected to achieve the desired tradeoff between performance and energy efficiency.

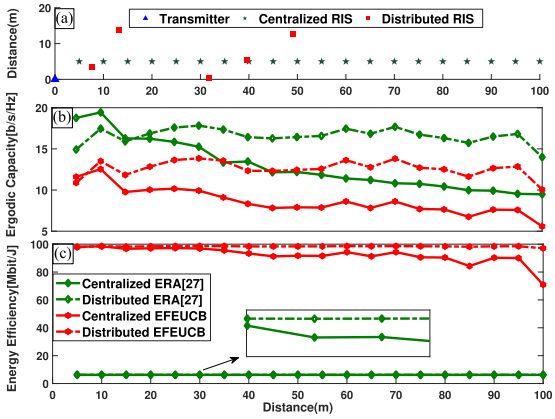


Fig. 13. Comparison between centralized and distributed RIS approaches: (a) Fixed locations of the five distributed RIS and different locations of centralized RIS. The locations of four receivers is selected randomly in each experiment. (b) Comparison of ergodic capacity for different position of centralized RIS. (b) Comparison of energy efficiency for different position of centralized RIS. The transmit power and achievable rate for energy efficiency is fixed at 20 dBm and 10 b/s/Hz, respectively.

F. Centralized Versus Distributed RIS Approach

Next, we compare the performance of the centralized and distributed RIS approaches in Fig. 13. In a centralized approach, we use a single RIS with the same number of elements as the total elements of K RIS in a distributed approach. We randomly select the position of the receivers in each experiment and average the results over multiple experiments. In Fig. 13,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 09:44:39 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.

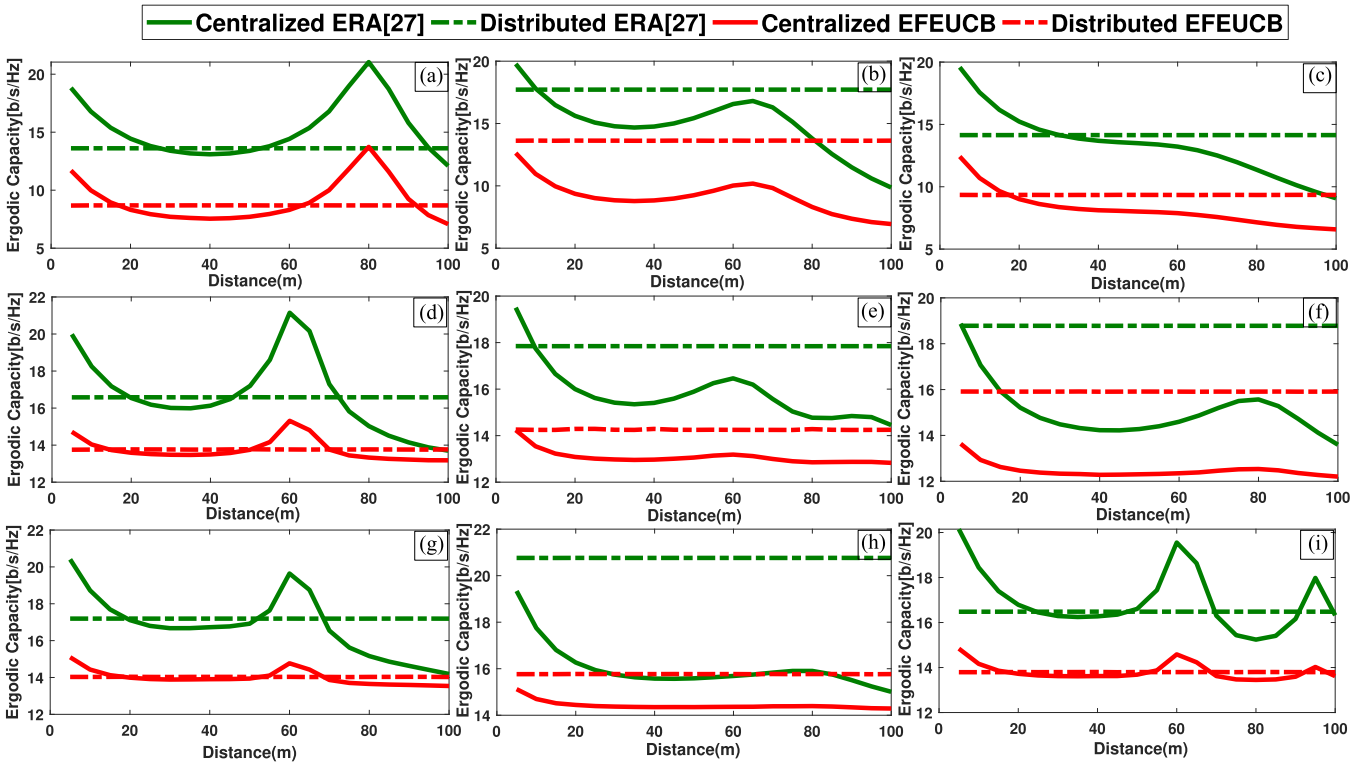


Fig. 14. Comparative results for ergodic capacity at transmit power 20 dBm, with different scenario. (a)–(c) Scenario 1 (d)–(f). Scenario 2. (g)–(i) Scenario 3.

we compare the ergodic capacity of the ERA and EFEUCB algorithms for different positions of the centralized RIS between transmitter and receiver. As expected, the distributed approach offers higher ergodic capacity than the centralized approach in both cases, thereby validating the need for a multi-RIS-aided distributed system. In Fig. 13, we also compare the energy efficiency of the centralized and distributed RIS approaches for ERA and EFEUCB algorithms. The distributed approach offers higher efficiency than the centralized one due to higher ergodic capacity.

Next, we make an in-depth comparison of the ergodic capacity of centralized and distributed approaches to analyze the effect of a number of receivers and their locations. In Table III, we consider three scenarios with 1, 2, and 4 receivers, respectively, along with the specific locations of the receivers in each case. The location of 5 distributed RIS is fixed in all scenarios. In Fig. 14(a)–(c), we compare the ergodic capacity for scenario 1 with a single receiver. The centralized RIS offers better performance than the distributed RIS in Fig. 14(a) as the receiver is near to the centralized RIS compared to any of the distributed RIS. On the other hand, the distributed approach significantly outperforms the centralized approach in Fig. 14(b) and (c) even though centralized RIS can be moved at any position between transmitter and receiver while the position of distributed RIS is fixed in advance. Similar observations can be made from various results presented in Fig. 14(d)–(i). These results validate the superiority of the distributed approach in Fig. 13, where receiver positions are selected randomly. Since the positions of receivers are not fixed in real networks, the distributed approach is more effective and practical.

TABLE III
PARAMETERS FOR EXPERIMENTS IN FIG. 14

Scenario	No. of RX	Figure	Location of Receiver(s)	Location of RIS
1	1	Fig.14 (a)	(80.35,0.65)	**Distributed:** RIS1 (16.95,3.58), RIS2 (94.12,24.38), RIS3 (81.53,21.45), RIS4 (57.10,10.55), RIS5 (63.74,12.33) **Centralized:** Moving along X axis, and Y coordinate is fixed at 5.
		Fig. 14 (b)	(66.71,16.95)	
		Fig. 14 (c)	(65.22,31.41)	
2	2	Fig. 14 (d)	(74.04,27.20), (61.99,1.41)	
		Fig. 14 (e)	(94.93,17.62), (62.25,16.73)	
		Fig. 14 (f)	(84.47,20.94), (79.96,18.58)	
3	4	Fig. 14 (g)	(45.15,20.15), (74.04,21.2), (62,1.5), (91.33,22.13)	
		Fig. 14 (h)	(84.47,20.94), (64.00,26.11), (79.96,18.58), (94.22,23.42)	
		Fig. 14 (i)	(62,1.5), (96,1.5), (86.96,20), (46.37,22.50)	

V. EXECUTION TIME COMPARISON ON EDGE PLATFORMS

In the MAB algorithm deployed in wireless networks, the execution time of the algorithm should be as short as possible. In time-slotted communication, a faster MAB algorithm leads to the early selection of the RIS block, allowing the transmitter more time to transmit the actual data and achieve high throughput. In Table IV and Fig. 15, we compare the execution time of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 09:44:39 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE IV
COMPARISON OF EXECUTION TIME IN MILLISECONDS OF THE VARIOUS ALGORITHMS ON EDGE PLATFORMS

S.No.	Sub Blocks	Algorithm	Cortex A9 (666 MHz)		Cortex A9 and NEON (666 MHz)		Cortex A9 (800 MHz)		Cortex A9 and NEON (800 MHz)		Cortex A53 (1.1 GHz)		Cortex A53 and NEON (1.1 GHz)	
			SPFL	DPFL	SPFL	DPFL	SPFL	DPFL	SPFL	DPFL	SPFL	DPFL	SPFL	DPFL
1	25	UCB[44]	12.73	12.95	9.34	9.21	12.42	12.61	9.11	8.98	5.86	5.91	4.17	4.43
2		FEUCB	8.49 (-33%)	8.60 (-33%)	5.84 (-37%)	5.59 (-39%)	8.29 (-33%)	8.39 (-33%)	5.70 (-37%)	5.45 (-39%)	3.89 (-33%)	3.92 (-33%)	2.69 (-37%)	2.81 (-37%)
3		EFEUCB	4.49 (-65%)	4.57 (-65%)	3.06 (-67%)	2.96 (-68%)	4.35 (-65%)	4.44 (-65%)	2.98 (-67%)	2.93 (-67%)	2.10 (-64%)	2.12 (-64%)	1.41 (-66%)	1.48 (-67%)
4	50	UCB[44]	25.22	25.76	18.56	18.40	24.59	25.14	18.12	17.96	11.64	11.73	8.30	8.81
5		FEUCB	13.55 (-46%)	13.78 (-46%)	9.64 (-48%)	9.37 (-49%)	13.23 (-46%)	13.45 (-46%)	9.40 (-48%)	9.14 (-49%)	6.32 (-46%)	6.37 (-46%)	4.39 (-47%)	4.60 (-48%)
6		EFEUCB	7.84 (-69%)	7.96 (-69%)	5.45 (-70%)	5.40 (-70%)	7.65 (-69%)	7.82 (-69%)	5.24 (-71%)	5.16 (-71%)	3.76 (-68%)	3.79 (-68%)	2.50 (-70%)	2.61 (-70%)
7	100	UCB[44]	49.43	51.25	36.96	36.60	48.24	50.02	36.08	35.72	23.11	23.28	16.48	17.49
8		FEUCB	17.05 (-66%)	17.52 (-66%)	12.46 (-66%)	12.32 (-66%)	16.64 (-66%)	17.10 (-66%)	12.16 (-66%)	12.02 (-66%)	8.08 (-65%)	8.14 (-65%)	5.68 (-65%)	5.98 (-66%)
9		EFEUCB	9.12 (-82%)	9.44 (-82%)	6.05 (-84%)	6.28 (-83%)	8.88 (-82%)	9.24 (-85%)	5.90 (-84%)	5.91 (-83%)	4.67 (-80%)	4.70 (-80%)	2.72 (-83%)	2.96 (-83%)

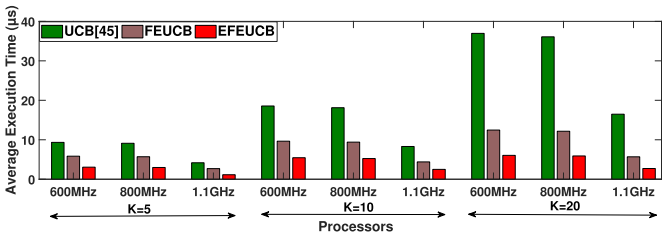


Fig. 15. Comparison of average execution time (μs) for different algorithms on different processors.

three MAB algorithms on the following three different types of processors used in edge platforms:

1) ARM Cortex A9 at 666 MHz;
2) ARM Cortex A9 at 800 MHz;
3) ARM Cortex A53 at 1.1 GHz.

We study the effect of SIMD NEON coprocessor and two different word lengths, single precision floating point (SPFL) and double precision floating point (DPFL), on the execution time. For $R = 25$ RIS sub-blocks, we can observe that FEUCB offers around 33%–39% lower execution time than UCB while EFEUCB offers around 64%–68% lower execution time than UCB. As the number of sub-blocks are increased from 25 to 50 and 100, further improvement in the performance of FEUCB and EFEUCB with respect to UCB is observed. The substantial degradation in the execution time of the UCB can be observed from Fig. 15. For 50 sub-blocks, FEUCB and EFEUCB offer around 46%–49% and 68%–71% lower execution time than UCB, respectively. This improvement is increased to 65%–66% and 80%–85% for 100 sub-blocks. The huge savings in execution time in proposed algorithms are due to the focused exploration approach, which reduces the number of computational arithmetic operations in each time slot. As expected, execution time with DPFL is slightly higher than SPFL, though we did not observe any degradation in performance with SPFL. Similarly, higher clock frequency and better processors lead to further reduced execution time, as shown in Table IV. The use of SIMD-based NEON coprocessor results in a significant reduction in execution time for all the algorithms. Further reduction in execution time can be achieved using multiple cores of these processors along with other coprocessors, such as graphic processing unit or field-programmable gate array.

VI. CONCLUSION

In this work, novel UCB-based distributed algorithms, i.e., FEUCB and EFEUCB, have been proposed to select optimal RIS sub-block. The theoretical regret analysis and the in-depth simulation results validate the effectiveness and superiority of our proposed algorithm over the existing state-of-the-art algorithms. The execution time comparison on the edge platforms demonstrates the compute and memory efficient architecture of the proposed algorithms. Future works include the hardware software codesign of the proposed algorithms and further extension in the quasi-stationary environment for vehicular applications. Furthermore, an important future direction is an efficient design of analog-front-end and antennas for directional communication over multiple RIS.

REFERENCES

[1] Q.-Y. Yu, H.-C. Lin, and H.-H. Chen, "Intelligent radio for next generation wireless communications: An overview," *IEEE Wireless Commun.*, vol. 26, no. 4, pp. 94–101, Aug. 2019.
[2] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Commun. Surv. Tut.*, vol. 15, no. 3, pp. 1136–1159, Third Quarter 2013.
[3] R. Liu, Q. Wu, M. Di Renzo, and Y. Yuan, "A path to smart radio environments: An industrial viewpoint on reconfigurable intelligent surfaces," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 202–208, Feb. 2022.
[4] H. Du, J. Zhang, J. Cheng, and B. Ai, "Millimeter wave communications with reconfigurable intelligent surfaces: Performance analysis and optimization," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2752–2768, Apr. 2021.
[5] H. Zhang, B. Di, L. Song, and Z. Han, "Reconfigurable intelligent surfaces assisted communications with limited phase shifts: How many phase shifts are enough?," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4498–4502, Apr. 2020.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 09:44:39 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.

[6] Z. Zhang et al., "Active RIS vs. passive RIS: Which will prevail in 6G?," *IEEE Trans. Commun.*, vol. 71, no. 3, pp. 1707–1725, Mar. 2023.

[7] E. Basar, "Reconfigurable intelligent surface-based index modulation: A new beyond MIMO paradigm for 6G," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 3187–3196, May 2020.

[8] S. Arzykulov, G. Nauryzbayev, A. Celik, and A. M. Eltawil, "RIS-assisted full-duplex relay systems," *IEEE Syst. J.*, vol. 16, no. 4, pp. 5729–5740, Dec. 2022.

[9] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 375–388, Jan. 2021.

[10] K. Keykhosravi, M. F. Keskin, S. Dwivedi, G. Seco-Granados, and H. Wymeersch, "Semi-passive 3D positioning of multiple RIS-enabled users," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 11073–11077, Oct. 2021.

[11] E. Shtaiwi, H. Zhang, S. Vishwanath, M. Youssef, A. Abdelhadi, and Z. Han, "Channel estimation approach for RIS assisted MIMO systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 452–465, Jun. 2021.

[12] H. Guo and V. K. N. Lau, "Uplink cascaded channel estimation for intelligent reflecting surface assisted multiuser MISO systems," *IEEE Trans. Signal Process.*, vol. 70, pp. 3964–3977, Jul. 2022.

[13] P. Zhang, J. Zhang, H. Xiao, H. Du, D. Niyato, and B. Ai, "RIS-aided 6G communication system with accurate traceable user mobility," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2718–2722, Feb. 2023.

[14] T. L. Nguyen, T. N. Do, G. Kaddoum, D. B. d. Costa, and Z. J. Haas, "Channel characterization for RIS-aided terahertz communications: A stochastic approach," *IEEE Wireless Commun. Lett.*, vol. 11, no. 9, pp. 1890–1894, Sep. 2022.

[15] T. Jiang, H. V. Cheng, and W. Yu, "Learning to reflect and to beamform for intelligent reflecting surface with implicit channel estimation," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 1931–1945, Jul. 2021.

[16] D. Shen and L. Dai, "Dimension reduced channel feedback for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7748–7760, Nov. 2021.

[17] X. Ma, Y. Fang, H. Zhang, S. Guo, and D. Yuan, "Cooperative beamforming design for multiple RIS-assisted communication systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10949–10963, Dec. 2022.

[18] K. Keykhosravi and H. Wymeersch, "Multi-RIS discrete-phase encoding for interpath-interference-free channel estimation," 2021, *arXiv:2106.07065*.

[19] Y. Zhao, W. Xu, X. You, N. Wang, and H. Sun, "Cooperative reflection and synchronization design for distributed multiple-RIS communications," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 980–994, Aug. 2022.

[20] M. He, W. Xu, H. Shen, G. Xie, C. Zhao, and M. Di Renzo, "Cooperative multi-RIS communications for wideband mmwave MISO-OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 10, no. 11, pp. 2360–2364, Nov. 2021.

[21] G. Zhou, C. Pan, H. Ren, K. Wang, and M. D. Renzo, "Fairness-oriented multiple RIS-aided mmwave transmission: Stochastic optimization methods," *IEEE Trans. Signal Process.*, vol. 70, pp. 1402–1417, 2022.

[22] K. Xu, S. Gong, M. Cui, G. Zhang, and S. Ma, "Statistically robust transceiver design for multi-RIS assisted multi-user MIMO systems," *IEEE Commun. Lett.*, vol. 26, no. 6, pp. 1428–1432, Jun. 2022.

[23] Y. Wang, W. Zhang, Y. Chen, C.-X. Wang, and J. Sun, "Novel multiple RIS-assisted communications for 6G networks," *IEEE Commun. Lett.*, vol. 26, no. 6, pp. 1413–1417, Jun. 2022.

[24] Z. Xie, W. Yi, X. Wu, Y. Liu, and A. Nallanathan, "Downlink multi-RIS aided transmission in backhaul limited networks," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1458–1462, Jul. 2022.

[25] Y. Huo, X. Dong, and N. Ferdinand, "Distributed reconfigurable intelligent surfaces for energy-efficient indoor terahertz wireless communications," *IEEE Internet Things J.*, vol. 10, no. 3, pp. 2728–2742, Feb. 2023.

[26] V. K. Chapala and S. M. Zafaruddin, "Multiple RIS-assisted mixed FSO-RF transmission over generalized fading channels," *IEEE Syst. J.*, vol. 17, no. 3, pp. 3515–3526, Sep. 2023.

[27] T. N. Do, G. Kaddoum, T. L. Nguyen, D. B. Da Costa, and Z. J. Haas, "Multi-RIS-aided wireless systems: Statistical characterization and performance analysis," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8641–8658, Dec. 2021.

[28] L. Yang, Y. Yang, D. B. d. Costa, and I. Trigui, "Outage probability and capacity scaling law of multiple RIS-aided networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 256–260, Feb. 2021.

[29] J. Tong, H. Zhang, L. Fu, A. Leshem, and Z. Han, "Two-stage resource allocation in reconfigurable intelligent surface assisted hybrid networks via multi-player bandits," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3526–3541, May 2022.

[30] E. M. Mohamed, S. Hashima, K. Hatano, and M. M. Fouda, "Cost-effective MAB approaches for reconfigurable intelligent surface aided millimeter wave relaying," *IEEE Access*, vol. 10, pp. 81642–81653, 2022.

[31] S. V. S. Santosh and S. J. Darak, "Multiarmed bandit algorithms on Zynq system-on-chip: Go frequentist or Bayesian?," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 2602–2615, Feb. 2024.

[32] M. K. Hanawal and S. J. Darak, "Multiplayer bandits: A trekking approach," *IEEE Trans. Autom. Control*, vol. 67, no. 5, pp. 2237–2252, May 2022.

[33] H. Tibrewal, S. Patchala, M. K. Hanawal, and S. J. Darak, "Distributed learning and optimal assignment in multiplayer heterogeneous networks," in *Proc. IEEE Conf. Comput. Commun.*, 2019, pp. 1693–1701.

[34] D. Ghosh, M. K. Hanawal, and N. Zlatanov, "Learning to optimize energy efficiency in energy harvesting wireless sensor networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1153–1157, Jun. 2021.

[35] Y. Zhang and R. W. Heath, "Multi-armed bandit for link configuration in millimeter-wave networks: An approach for solving sequential decision-making problems," *IEEE Veh. Technol. Mag.*, vol. 18, no. 2, pp. 39–49, Jun. 2023.

[36] Z. Kuai and S. Wang, "Thompson sampling-based antenna selection with partial CSI for TDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7533–7546, Dec. 2020.

[37] Y. Song, C. Liu, W. Zhang, Y. Liu, H. Zhou, and X. Shen, "Two stage beamforming in massive MIMO: A combinatorial multi-armed bandit based approach," *IEEE Trans. Veh. Technol.*, vol. 72, no. 5, pp. 6794–6799, May 2023.

[38] M.-J. Youssef, V. V. Veeravalli, J. Farah, C. A. Nour, and C. Douillard, "Resource allocation in NOMA-based self-organizing networks using stochastic multi-armed bandits," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 6003–6017, Sep. 2021.

[39] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, pp. 235–256, 2002.

[40] A. Garivier and O. Cappé, "The KL-UCB algorithm for bounded stochastic bandits and beyond," in *Proc. 24th Annu. Conf. Learn. Theory*, 2011, pp. 359–376.

[41] S. Agrawal and N. Goyal, "Further optimal regret bounds for thompson sampling," in *Proc. 16th Int. Conf. Artif. Intell. Statist.*, C. M. Carvalho and P. Ravikumar, Eds., 2013, pp. 99–107.

[42] E. M. Mohamed, S. Hashima, and K. Hatano, "Energy aware multiarmed bandit for millimeter wave-based UAV mounted RIS networks," *IEEE Wireless Commun. Lett.*, vol. 11, no. 6, pp. 1293–1297, Jun. 2022.

[43] A. M. Elbir, A. Papazafeiropoulos, P. Kourtessis, and S. Chatzinotas, "Deep channel learning for large intelligent surfaces aided mm-wave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1447–1451, Sep. 2020.

[44] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, pp. 235–256, 2002.

[45] S. Bubeck et al., "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Found. Trends Mach. Learn.*, vol. 5, no. 1, pp. 1–122, 2012.

[46] J. Kwon, V. Perchet, and C. Vernade, "Sparse stochastic bandits," in *Proc. Conf. Learn. Theory*, S. Kale and O. Shamir, Eds., 2017, pp. 1269–1270.

[47] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, Feb. 2020.

[48] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 09:44:39 UTC from IEEE Xplore. Restrictions apply.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/380402604

On-To-Methodology: Ontology Development Methodology

Article *in* SSRN Electronic Journal · June 2017

CITATIONS
0

READS
5

1 author:



Research Publication
University of the Cumberlands
795 PUBLICATIONS **106** CITATIONS

SEE PROFILE

All content following this page was uploaded by Research Publication on 08 May 2024.

The user has requested enhancement of the downloaded file.

On-To-Methodology: Ontology Development Methodology Magendra Singh

M.Tech Software Engineering
Department of Computer Engineering
Delhi Technological University, India

Dr. Daya Gupta
Professor & Head of Department
Department of Computer Engineering
Delhi Technological University, India

Abstract:

Ontologies are used to represent the domain knowledge. But the ontologies created by different creators are different and inconsistent. Huge difference exists between domains and currently there exists no single defined methodology for ontology development. In this work, we propose an ontology development methodology called *On-To-Methodology* that is comprehensive & generic. We elicit the domain knowledge in a structured manner and employ various documents to support the development process. We have also developed a tool that automates the development process.

Keywords: Ontology, Knowledge representation, Development methodology, Ontology design process, Ontology construction

1. Introduction

The American Heritage Dictionary defines "ontology" as "the branch of metaphysics that deals with the nature of being." Ontology, as a formal specification of a shared conceptualization [1], defines the thesaurus of the concept, relations between concepts and properties even relations between properties. Moreover, it describes axioms and individuals and relations between them, and provides sharing knowledge [2]. More specifically, it is an explicit specification of a conceptualization [3, 4].

Initial application of ontologies in computer science was that it acted as a means of providing the semantics for the semantic web i.e. web 2.0. By use of ontologies, the information retrieval could then be done based on the context (/meaning) rather than just simple string matching mechanisms.

The Ontologies created by different creators are different and inconsistent. There is a huge difference between different domains and currently there is no defined methodology for constructing Ontologies [5, 6, 7]. And inspite much being written in literature about Ontologies, the number and quality of actual, "non-toy" ontologies available on the Web today is remarkably low [7, 8].

Fernández et. al. proposed "Methontology" and defined the ontology development process and the ontology life cycle [9]. Farooq et. al. proposed "Design of Ontology in Semantic Web Engineering Process and illustrated their method with the help on a case study [10]. Gaoyun et. al. identified the reasons for the inconsistencies in ontology construction and put forward a method for Consistent Ocean Ontology Construction [7]. Zhang et. al. proposed a methodology namely MADRE and illustrated their

work by focusing on healthy housing [11]. Jia et. al. put forward "Automatic ontology construction approaches and its application on military intelligence" [12]. Chen et. al proposed Concept feature based ontology construction and maintenance which was backed up by a detailed framework [13]. A structured ontology construction by using data clustering and pattern tree mining was proposed by Yao-Tang et. al. [6].

All these proposals seem to suffer from following drawbacks [7]: (i) *Inconsistency of domain cognition*: this relates to the point that people have no consistent definition of the scope of the domain and about the contents of the domain. (ii) *Inconsistency of viewpoint*: ontologies of same scope and built by domain experts with the same background knowledge still have essential dissimilarities due to the differences in their viewpoint. (iii) *Ontology construction is affected by subjective factors*: Even in the same field with viewpoints, ontologies may be different from one another due to creators' interests, tastes, preferences, ideas, abilities etc. (iv) *Lexical inconsistency*: Conceptualization is independent from vocabulary, but ontology is language-related. Even if the language is unified, inconsistencies may still occur because the languages suffer from the problem of "polysemy" and "thesaurus".

Martin Hepp [8] classifies the ontology-related tasks into two main groups: building or contributing to the development of ontologies and committing to a particular ontology. Committing to a given ontology, explicitly or implicitly, means agreeing that it properly represents the domain's conceptual elements.

So to solve the aforesaid problems with ontology engineering, in this work, we propose a method called ***On-To-Methodology***, that is comprehensive, generic and makes it easy to commit and construct ontologies of various domains specifically transportation like bikes, cars, planes etc. We aim to produce documents similar to the documents in Software Engineering. Also, we provide a computer based tool which can automate the construction process. The proposed methodology will have following features:

- *Better domain cognition*

  As listed in above, there exists inconsistency in domain cognition. Our method handles this problem by adapting group oriented activities thus enabling experts and users to define the domain in better terms.

- *Unification of viewpoints*

  Inconsistency in viewpoints arises as different people have different perspectives and due to this contrary conclusions may be obtained. Our methodology resolves this problem by instrumenting the development process with evaluation activities after each phase of the development process. Also, the unification is achieved due to the adaptation of group oriented development activities.

- *Exhaustive documentation*

  On-To-Methodology provides an extensive and exhaustive documentation support thus making the development process more formal and helps in achieving traceability across phases. Documentation also simplifies the maintenance of the ontology by providing the backward traceability of various components of the ontology.

- *Extensive Verification & Validation*

  These two tasks of verification and validation are interleaved between various phases of the development process. This prevents introduction of bugs into the system right from the beginning of the development process and prevents faults till the maintenance phase.

We shall apply On-To-Methodology to domain of Automobile, specifically Bikes. The intended ontology of Bikes will be representation of bikes in the domain of ontology. The purpose of this ontology is to provide information on bikes based on the criteria specified by the users.

- The ontology would hold information to answer queries of customers based on single (/combination of) parameter(s) which are Make, Engine capacity, Power, Price, Fuel tank capacity, Mileage, Brake type, Weight, Wheel type, Ignition and Number of gears.
- Bike manufacturing organizations can use this ontology to identify the bike configurations that are suitable for a particular market and can also use it to analyze current sales and make future predictions. This will guide them to plan their production & inventory.
- This ontology can prove to be beneficial for bike retailers as they can use it to plan their inventory and analyze their sales.

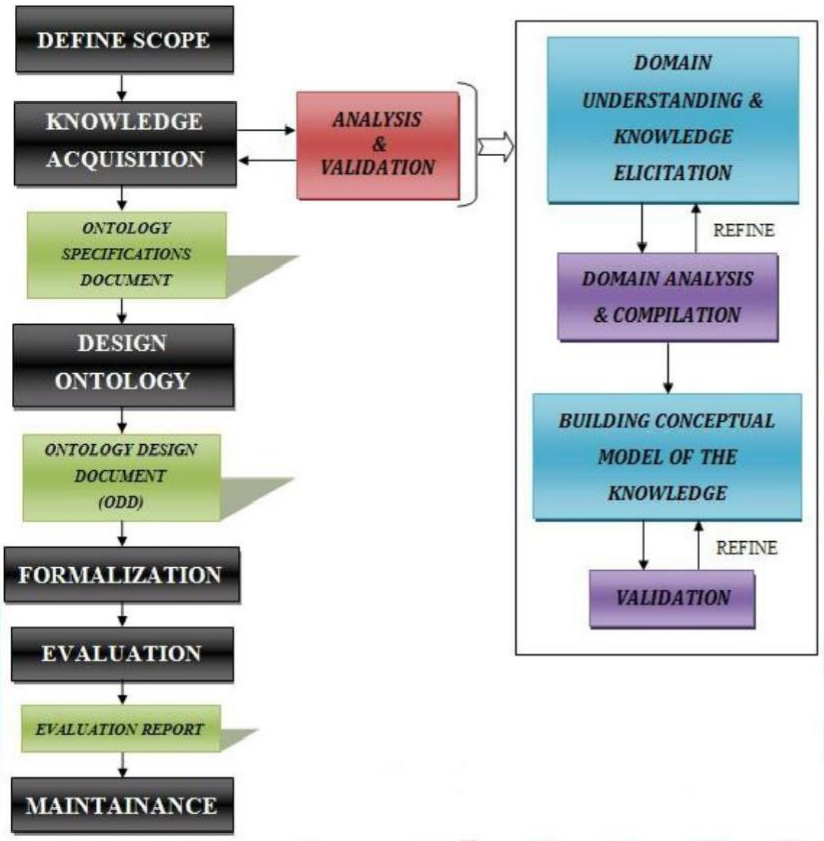
The advantage that this ontology would provide is its capability to answer the queries of the customers across a large information base of different bikes, based on multiple search criteria with complex inter-relations.

The rest of the paper is organized as follows. Section 2, proposes the methodology for ontology construction. Section 3, illustrates our methodology by constructing an Ontology of Bikes. In section 4, we evaluate the Ontology of Bikes for various competency questions and compare our approach with other existent methodologies. Section 6 concludes the thesis and proposes future work.

2. ON-TO-METHODOLOGY: Proposed methodology for ontology construction

This process is generic and extension of proposed methodology by Fernández et. al. [9] and Yao-Tang Yu et. al. [6]. The different phases in On-To-Methodology are described in Figure 1. These steps are generic and consist of four main activities namely Knowledge Acquisition, Design Ontology, Formalization and Evaluation. The output of Knowledge Acquisition is Ontology Specification document. This document is analogous to Software Requirements Specification (SRS) document IEEE standard 830. The output of Design Ontology is Ontology Design Document (ODD) and consists of logical design of the ontology being developed. The output of Evaluation is Evaluation Report. It consists of the evaluation of ontology for knowledge representation efficiency.

Fig. 1: On-To-Methodology: Ontology Development Methodology



2.1 Define Ontology Scope

The target of this phase is to establish the scope of the ontology in terms of purpose that the ontology would serve with associated advantages. It should provide an objective description of the intended users, various scenarios in which it would be used and the end users.

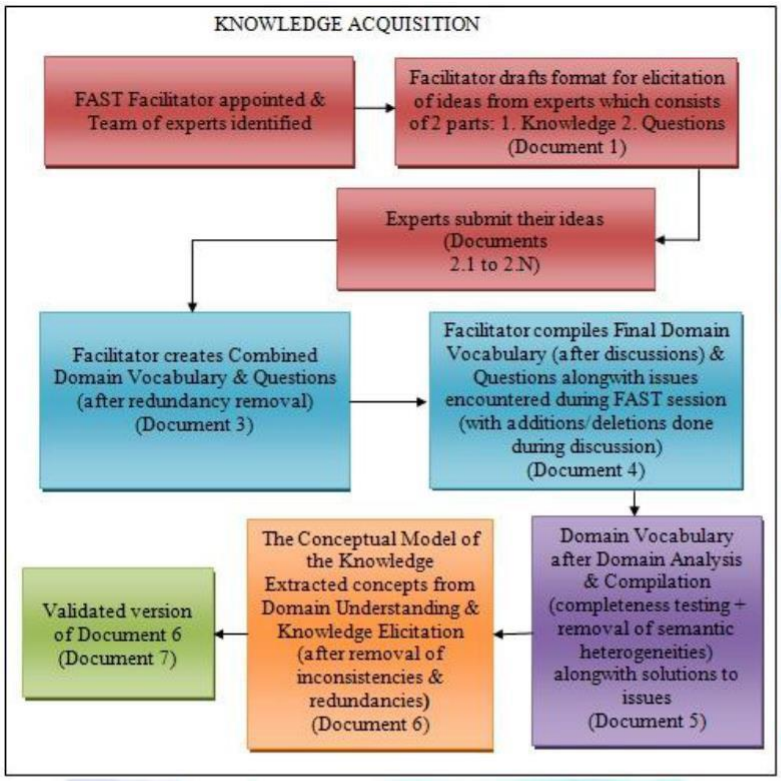
2.2 Knowledge Acquisition

This activity would require the identification of *knowledge source* from where the knowledge is to be acquired. It can include the following: Books, Experts, Figures, Tables, Files, Webpages, Other ontologies etc. The output of this phase is an *Ontology Specification Document*.

We propose the use of *FAST* (Facilitated Application Specification Technique) here. It is a team oriented Software Engineering approach for gathering of requirements. This approach encourages the creation of a joint team of domain experts and users who work together to understand the expectations and propose a set of requirements. We have adapted FAST to elicitate conceptual knowledge required to construct ontology in a formal way. We have designed various forms which are used to gather this knowledge. We identify the various stakeholders of the ontology and generate different forms to

acquire the required knowledge. Figure 2 illustrates the various forms that we have generated during the Knowledge Acquisition activity.

Fig. 2: Various forms generated during the Knowledge Acquisition



As presented in figure, in the first phase a facilitator & team of experts are identified. The facilitator controls the various activities of the FAST session and prepares an informal agenda to encourage the free flow of ideas. The facilitator drafts a form (Document 1) to elicit ideas from the experts which consists of two parts which are required to be filled by the experts. The two parts are Knowledge and Questions. This form is filled by the various experts and submitted back to the facilitator (Documents 2.1 to 2.N where N is the number of experts). The facilitator then compiles the ideas from these forms into a separate document (Document 3). The consolidated ideas are then discusses within the team and a final document i.e. document 4 is created which consists of the domain vocabulary and questions alongwith the issues encountered during the FAST session. The domain vocabulary listed in document 4 is analyzed for the given domain and thus document 5 is created after compilation. The concepts are then extracted from the Domain Understanding & Knowledge Elicitation as a result document 6 is obtained which is then validated and the final document consisting of the Conceptual model of the knowledge is obtained in the form of document 7.

We consider this activity consisting of four main tasks as described below. All these activities are carried out in a formal manner with the support of forms generated.

2.2.1 Domain Understanding & Knowledge Elicitation

In this phase, the basic elements of the ontology are identified and enumerated. The elements of ontology are Class, Relationships, Constraints, Forms, Instances, Constants, and Instance attributes. These are explained as follows [14]:

- *Class* or concepts for a domain such as location, city, travel, destination etc.
- *Relationships* are the properties between two classes such as 'isa'.
- *Constraints* are conditions that must be satisfied during the design.
- *Instance* is values for particular categories in ontology.

Here we don't categorize the domain keywords according to the above categories but the focus is on comprehensive list of constituents without worrying about redundancies or overlaps.

Competency questions are also elicitated in this phase. These questions are queries related to specific situations from real life problems that the ontology would help provide a solution. The competency questions are therefore used to evaluate the ontology thus developed.

2.2.2 Domain Analysis & Compilation

An expert for particular domain can easily examine the ontology in better way and changes suggested by expert will remove some of the inconsistencies at the earlier stage to make the design phase simpler. The expert review makes knowledge acquisition evolutionary [14].

So after drafting the basic model as described above as document 4, it is reviewed by the experts for following rules: [10]

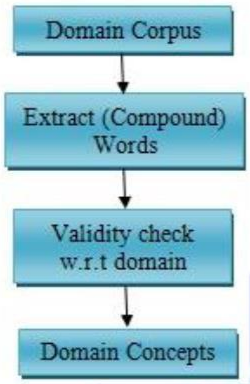
I. *Consistency rule*: various defects are analyzed for example usage of different names for a same concept.
II. C*ompleteness rule*: the defects such as omission of domain resources and the omission of relationship are diagnosed. The issues encountered during FAST sessions are also resolved here.

If any consistency or completeness check fails, it is refined by the expert and the output is called as document 5.

2.2.3 Building Conceptual Model of the Knowledge

A conceptual model is developed on the basis of the domain model which we call as the document 5. Here the knowledge is represented in form of trees and table of rules/ formulas. The inconsistencies and the redundancies are resolved. The domain vocabulary identified in the *Domain Understanding & Knowledge Elicitation* is classified as classes, sub-classes, properties, constraints and individuals. The mechanism as shown in figure 3 [15], can be employed.

Fig. 3: Framework for extracting the domain concepts



The definition of class hierarchy can be done in different ways [16] as following:

- Top-down development process begins with the definition of the concept of the field of the most common, followed by characterization of the concept.
- Bottom-up development process begins with the definition of the most specific classes, namely, class hierarchy the leaves, and then grouped into these classes more general concept.
- Combination of the two- defining more clearly the concept of the first, followed by their proper places generalization and characterization, it is the easier process.

The final result is document 6 which we call as the conceptual model of the ontology.

2.2.4 Validation

The conceptual model of the knowledge, developed in the last step is reviewed here. Expert review is again used here. An expert for particular domain will examine the ontology for the following rules and the changes suggested will be incorporated [10]:

I. *Correctness rule*: The use of incorrect relationship, aggregation & specialization of classes and cardinality of relationship are tested.
II. *Rule for Semantic heterogeneities*: Semantic heterogeneities are also determined and dissolved in this step.
III. *Constraints*: Constraints on domain and range values of each object property and datatype property are also verified in the testing activity.

The result of this activity is document 7 which is the validated version of document 6.

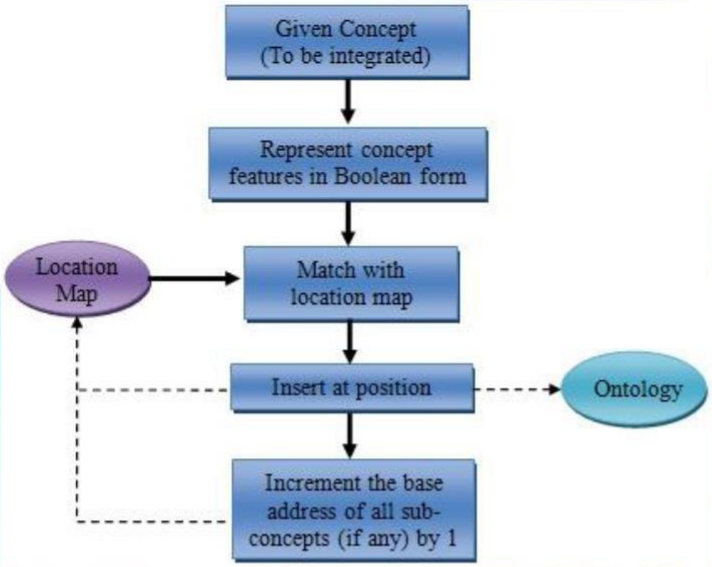
2.3 Design Ontology

This phase consists of two main tasks. One is the design where with the help of the Ontology Specification Document and the Conceptual model of the knowledge, the concepts are mapped to location in location map. In this phase the concepts in the conceptual model are structured so that knowledge can be accessed. Second task is to produce a design document called Ontology Design Document (ODD) which can be helpful in evolution of the ontology.

2.3.1 Design

This is the core task of building ontology. We have adapted an algorithm for this and apply the technique [13] as shown in figure 4 below:

Fig. 4: Addition of new concept



We have automated the above task by implementing the following algorithm:

i. Given an Ontology structure with a concept *Thing* which is super concept of all concepts.
ii. A concept with from Conceptual model of the knowledge is selected. Let it be X.
iii. All features for that concept are identified and labeled with a feature number as follows:

{X} → f^+ <feature no. from 1 to n>

iv. The concept is then represented in the Boolean form as follows: C (X): { $f^+ 1 \bullet f^+ 2 \bullet \dots \bullet f^+ n$ }

v. The concept represented in Boolean form is compared to other concepts in the location map one by one. Say the concept currently chosen from the Ontology structure be Y with address in location map as (a, b). Following cases can happen:

a. If {Y}: { $f^+ 1 \bullet f^+ 2 \bullet \dots \bullet f^+ (n-k)$ } where k<n, and we have
{X}: { $f^+ 1 \bullet f^+ 2 \bullet \dots \bullet f^+ (n-k) \bullet \dots \bullet f^+ n$ }

In this case there is match between X and Y for $f^+ 1$ to $f^+ (n-k)$. This means that X is a child of Y with $f^+ (n-k+1)$ to $f^+ n$ as additional features. Thus X is inserted in the Location map with address (a+1, xx).

b. If {Y}: { $f^+ 1 \bullet f^+ 2 \bullet \dots \bullet f^+ m$ }, and we have
{X}: { $f^+ 1 \bullet f^+ 2 \bullet \dots \bullet f^+ n$ }

If there exists a feature in Y that is not present in X, then X is not a sub-concept of Y thus it is either a brother concept of this concept or a sub-concept of some other concept.

vi. Steps ii to v are repeated for all concepts in the conceptual model and the end result is the complete ontology structure.

The following are the snapshots of our tool, fig 5(a) and fig 5(b):

Fig. 5(a): On-To-Methodology tool

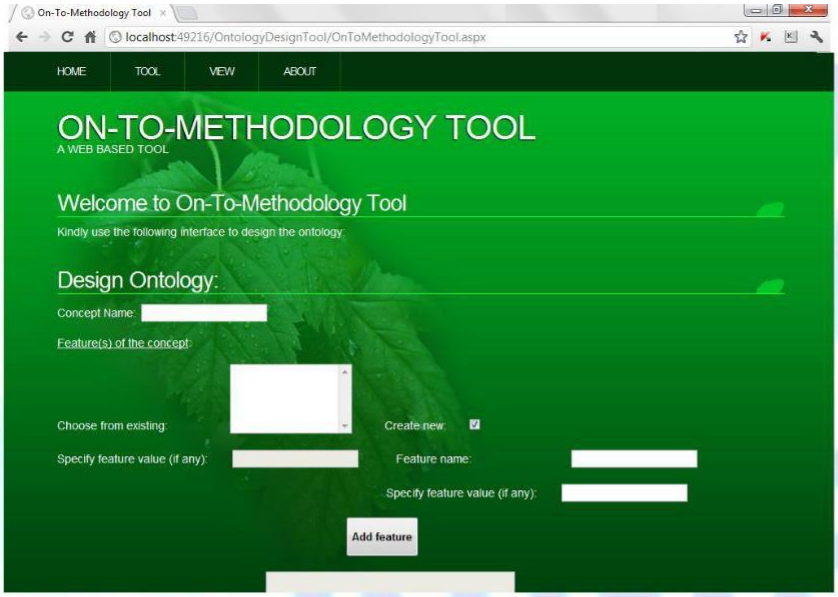
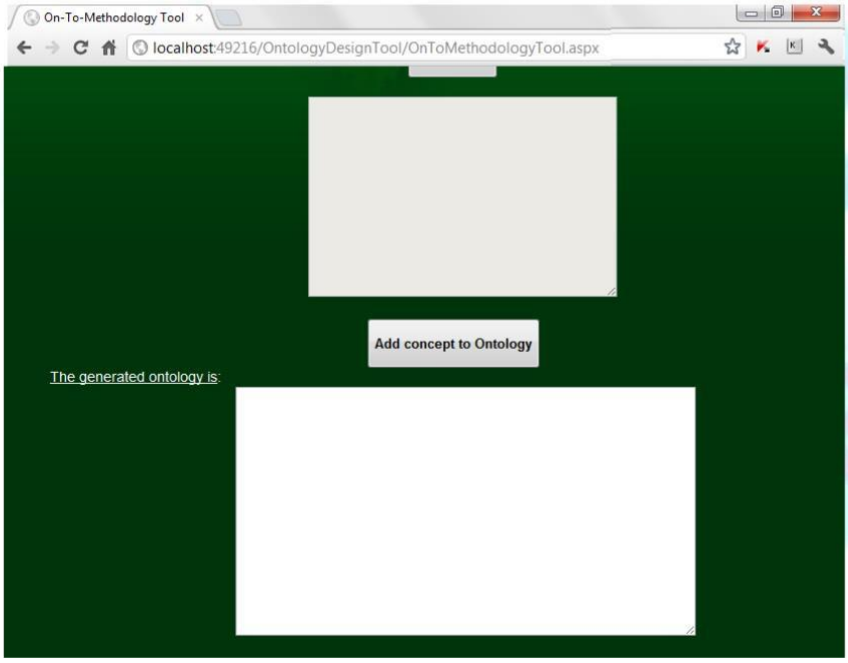


Fig. 5(b): On-To-Methodology tool



2.3.2 Ontology Design Document

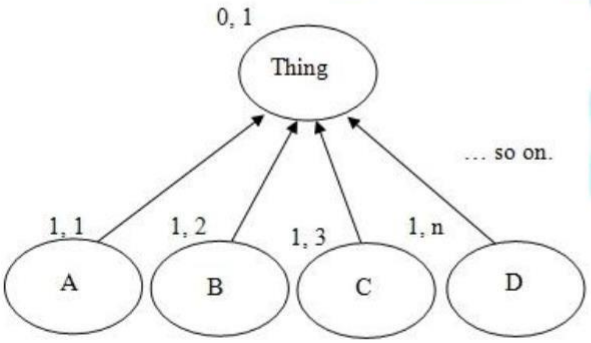
The output of this phase is the Ontology Design Document (ODD). This document is a formal record of the ontology structure and consists of the Location map alongwith the graphical representation of the ontology structure. The format for ODD is as follows [13]:

ONTOLOGY DESIGN DOCUMENT

The location map is as below:

CONCEPT ADDRESS	CONCEPT FEATURE
Thing (0, 1)	X
A (a1, a2)	f^+ 1 • f^+ 2• …
B (b1, b2)	f^+ x • f^+ y • …
…	…

The graphical representation of ontology structure is as follows:



2.4 Formalization

The main aim behind ontology development is to make the information to be understood by the machines. Thus, the ontology is finally formalized using the selected ontology languages and tools.

Semantic Web is not a single technology and it comprises of a number of components including ontology languages (RDF/RDFS, OWL etc.), editing tools (Protégé, Ontolingua etc.) and standards (WSMO, OWL-S etc.). As quoted in [14], different ontology languages provide different facilities. The most recent development in standard ontology languages is OWL from the W3C. OWL [17] makes it possible to describe concepts but it also provides new facilities. It has a richer set of operators - e.g. and, or and negation. It is based on a different logical model which makes it possible for concepts to be defined as well as described. Complex concepts can therefore be built up in definitions out of simpler

concepts. OWL ontologies may be categorized into three sub languages: OWL-Lite, OWL-DL and OWL-Full. Islam et. al. [18] have compared several editing tools as shown in Table 1 below.

Table 1: A comparison of ontology editing tools

Tools	Free	Open Source	Dot Net and Web Based	Import Languages	Export Languages	Inferencing Reasoning Tools
Protégé	√	√	×	RDF, OWL	RDF, OWL, FLogic, CLIPS	√
OntoEdit (Free)	√	×	×	DAML + OIL, RDFS	DAML + OIL, RDFS	×
Differential Ontology Editor (DOE)	√	×	×	RDFS, OWL	DAML+OIL, OIL, RDFS, OWL	×
IsaViz	√	√	×	RDF/XML, N-Triples	RDF/XML, N-Triples	√
Ontolingua	√	×	×	DAML+OIL, KIF,CLIPS	DAML+OIL, KIF, CLIPS	√
Altova Semantic Works™	×	×	×	N-triples, OWL, RDF and RDFS	N-triples, OWL, RDF and RDFS	×
WebODE	√	×	×	RDF(S), DAML+OIL, UML, OWL	CLIPS, DAML+OIL, UML, OWL	√
TopBraid Composer	×	×	×	RDBMS, OWL, RDF	OWL,RDF, XML	√
Morla	√	√	×	RDF, OWL	RDF,OWL	×
Hozo	√	√	×	RDF, OWL (subset)	OWL, RDF	×
TODE	√	×	√	RDF, OWL-Lite, N-3, RDBMS, N-Triple,	RDF, OWL-Lite, N-Triple, N-3, RDBMS	×

2.5 Evaluation

After the ontology has been drafted, it should be checked for quality and knowledge representation efficiency. If any constraints are violated then they are corrected. The ontology engineers also verify the developed ontology against the requirement document. The term 'Evaluation' encompasses both-verification & validation. Based on [11], we identify following two ways of evaluation:

- *Relation Evaluation*: Relation evaluation infers the new individual based on constructed rule set, and judges whether the relation is coincident with the professionals' knowledge. This is done using the *Reasoner* that comes inbuilt in the tool, like Fact++ in Protégé.
- *Evaluation of hierarchy*: Evaluation of hierarchy develops limits definition of created class and individual properties, makes use of ontology tool (like OntoGraf, OWLViz) to infer and create Ontology Graph automatically, and judge whether Subordination between the class and individual in the graph are coincident.

Also, in above steps, the answers to competency questions are verified by the experts. The output of this phase is the *Evaluation Report*. It consists of following:

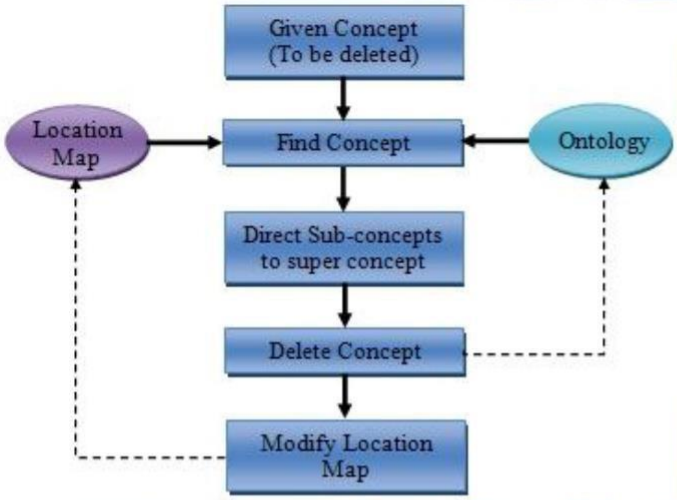
I. Evaluation of competency questions by the experts to check if the results are as per the expectation.
II. Evaluation of hierarchy using OntoGraf & OWLViz.

2.6 Maintenance

Ontologies need to adapt to the changing specifications to become more mature. This can be accomplished by adding, updating and removing concepts/ parts of the ontology. For this, following techniques can be employed:

- Addition of new concept: Technique as used in ontology design can be used.
- Deletion of a concept: Technique as shown in figure 6 can be used [13].

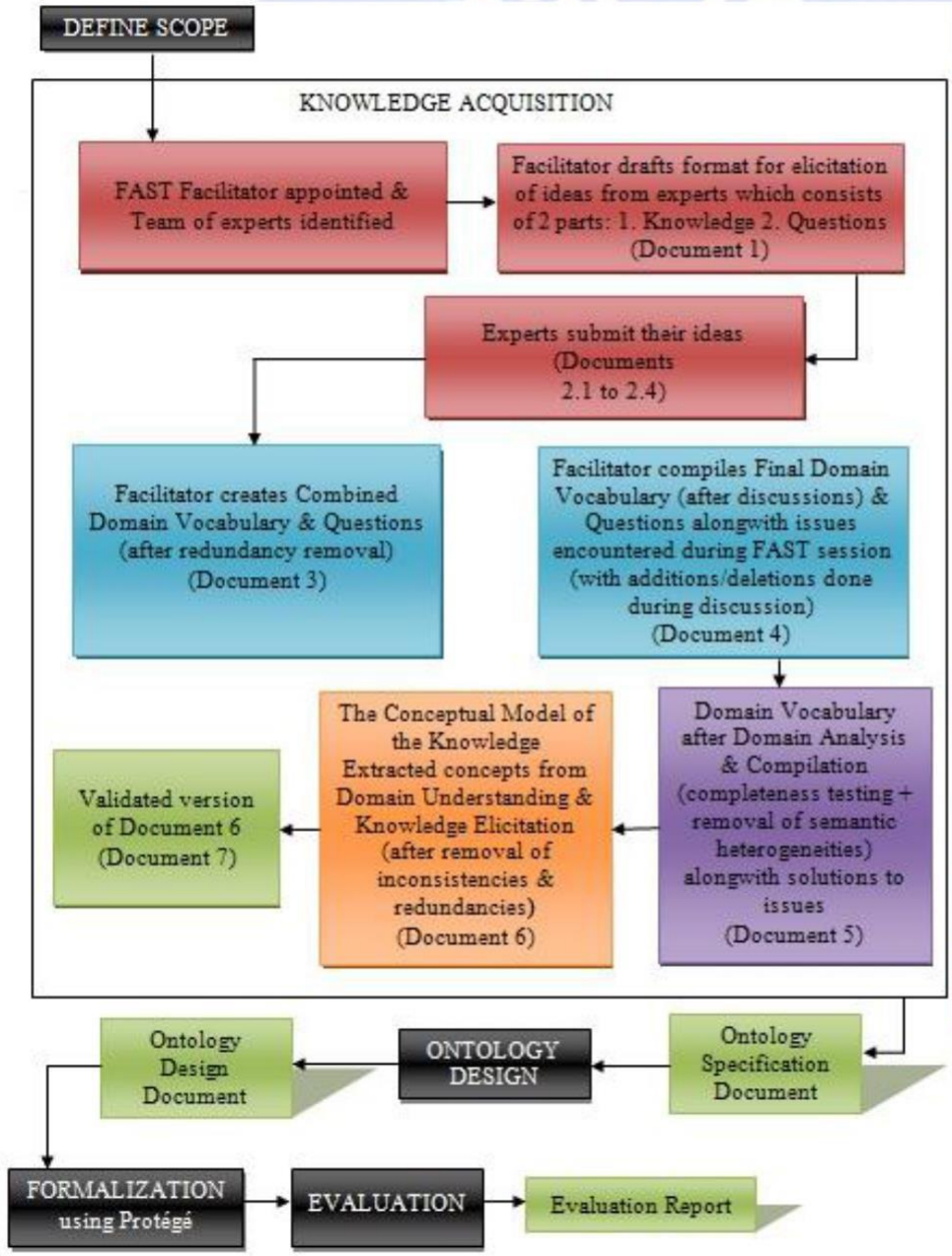
Fig. 6: Deletion of a concept



3. Case Study

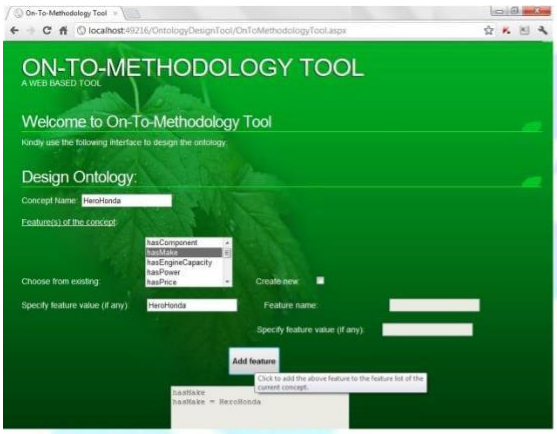
Here we illustrate our methodology by working out an Ontology of Bikes. The purpose of this ontology is to represent information related to bikes for parameters like Make, Engine capacity, Power, Price, Fuel Tank capacity, Mileage etc. Following is the workflow of the Bikes Ontology development process (figure 7):

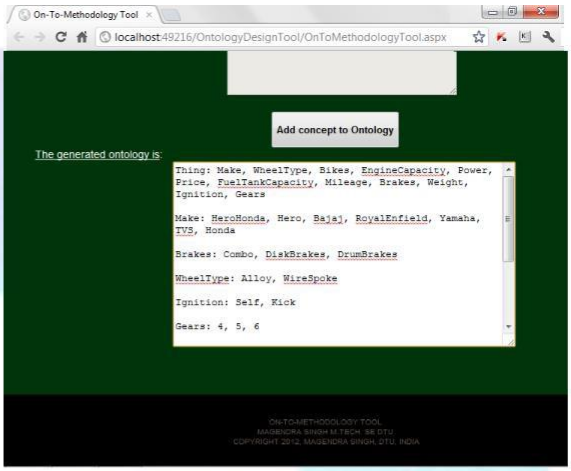
Fig. 7: Workflow of development of Ontology of Bikes



We employed our tool in the Ontology Design activity to design the ontology automatically. Following snapshots show the working of our tool in the designing process:

Fig. 8: Employment of our tool for design of Ontology of Bikes





The following is Ontology Design Document for the Ontology of Bikes.

Table 2: Final Location map

CONCEPT ADDRESS	CONCEPT FEATURE
Thing (0, 1)	X
Bikes (1, 1)	$f^+1 \bullet f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12$
Make (1, 2)	f^+2
EngineCapacity (1, 3)	f^+3
Power (1, 4)	f^+4
Price (1, 5)	f^+5
FuelTankCapacity (1, 6)	f^+6
Mileage (1, 7)	f^+7
Brakes (1, 8)	f^+8
Weight (1, 9)	f^+9
WheelType (1, 10)	f^+10
Ignition (1,11)	f^+11
Gears (1 ,12)	f^+12
NamedBikes (2, 1)	$f^+1 \bullet f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12$
HeroHonda (2, 2)	$f^+2 \bullet f^+13$
Hero (2, 3)	$f^+2 \bullet f^+14$
Bajaj (2, 4)	$f^+2 \bullet f^+15$
RoyalEnfield (2, 5)	$f^+2 \bullet f^+16$

Yamaha (2, 6)	$f^+2 \bullet f^+17$
TVS (2, 7)	$f^+2 \bullet f^+18$
Honda (2, 8)	$f^+2 \bullet f^+19$
HeroHondaBikes (3, 1)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+13$
HeroBikes (3, 2)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+14$
BajajBikes (3, 3)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+15$
RoyalEnfieldBikes (3, 4)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+16$
YamahaBikes (3, 5)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+17$
TVSBikes (3, 6)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+18$
HondaBikes (3, 7)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+19$
Combo (2, 9)	$f^+8 \bullet f^+20$
DiskBrakes (2, 10)	$f^+8 \bullet f^+21$
DrumBrakes (2, 11)	$f^+8 \bullet f^+22$
Alloy (2, 12)	$f^+10 \bullet f^+23$
WireSpoke (2, 13)	$f^+10 \bullet f^+24$
Self (2, 14)	$f^+11 \bullet f^+25$
Kick (2, 15)	$f^+11 \bullet f^+26$
6 (2, 16)	$f^+12 \bullet f^+27$
5 (2, 17)	$f^+12 \bullet f^+28$

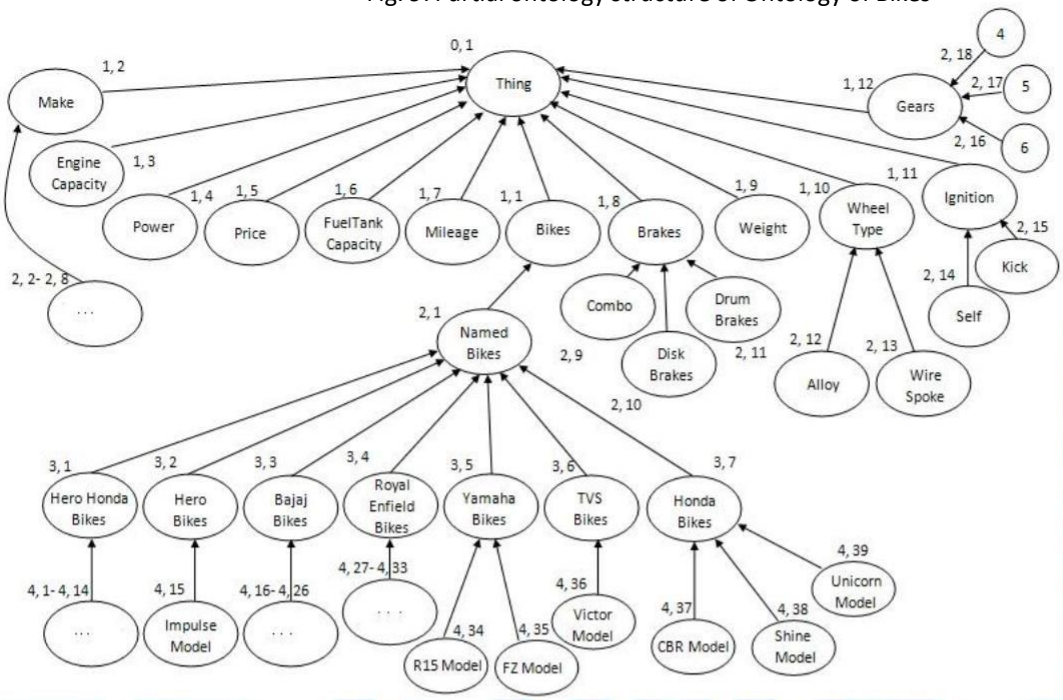
4 (2, 18)	$f^{+}12 \bullet f^{+}29$
Karizma(Normal)Model (4, 1)	$f^{+}2 \bullet f^{+}3 \bullet f^{+}4 \bullet f^{+}5 \bullet f^{+}6 \bullet f^{+}7 \bullet f^{+}8 \bullet f^{+}9 \bullet f^{+}10 \bullet f^{+}11 \bullet f^{+}12 \bullet f^{+}13 \bullet f^{+}20 \bullet f^{+}23 \bullet f^{+}25 \bullet f^{+}28$
Karizma(ZMR)Model (4, 2)	$f^{+}2 \bullet f^{+}3 \bullet f^{+}4 \bullet f^{+}5 \bullet f^{+}6 \bullet f^{+}7 \bullet f^{+}8 \bullet f^{+}9 \bullet f^{+}10 \bullet f^{+}11 \bullet f^{+}12 \bullet f^{+}13 \bullet f^{+}21 \bullet f^{+}23 \bullet f^{+}25 \bullet f^{+}28$
Splendor(Plus)Model (4, 3)	$f^{+}2 \bullet f^{+}3 \bullet f^{+}4 \bullet f^{+}5 \bullet f^{+}6 \bullet f^{+}7 \bullet f^{+}8 \bullet f^{+}9 \bullet f^{+}10 \bullet f^{+}11 \bullet f^{+}12 \bullet f^{+}13 \bullet f^{+}22 \bullet f^{+}24 \bullet f^{+}26 \bullet f^{+}29$
Splendor(NXG)Model (4, 4)	$f^{+}2 \bullet f^{+}3 \bullet f^{+}4 \bullet f^{+}5 \bullet f^{+}6 \bullet f^{+}7 \bullet f^{+}8 \bullet f^{+}9 \bullet f^{+}10 \bullet f^{+}11 \bullet f^{+}12 \bullet f^{+}13 \bullet f^{+}23 \bullet f^{+}23 \bullet f^{+}26 \bullet f^{+}29$
Splendor(Super)Model (4, 5)	$f^{+}2 \bullet f^{+}3 \bullet f^{+}4 \bullet f^{+}5 \bullet f^{+}6 \bullet f^{+}7 \bullet f^{+}8 \bullet f^{+}9 \bullet f^{+}10 \bullet f^{+}11 \bullet f^{+}12 \bullet f^{+}13 \bullet f^{+}22 \bullet f^{+}23 \bullet f^{+}25 \bullet f^{+}29$
Splendor(Pro)Model (4, 6)	$f^{+}2 \bullet f^{+}3 \bullet f^{+}4 \bullet f^{+}5 \bullet f^{+}6 \bullet f^{+}7 \bullet f^{+}8 \bullet f^{+}9 \bullet f^{+}10 \bullet f^{+}11 \bullet f^{+}12 \bullet f^{+}13 \bullet f^{+}22 \bullet f^{+}23 \bullet f^{+}25 \bullet f^{+}29$
PassionProModel (4, 7)	$f^{+}2 \bullet f^{+}3 \bullet f^{+}4 \bullet f^{+}5 \bullet f^{+}6 \bullet f^{+}7 \bullet f^{+}8 \bullet f^{+}9 \bullet f^{+}10 \bullet f^{+}11 \bullet f^{+}12 \bullet f^{+}13 \bullet f^{+}20 \bullet f^{+}23 \bullet f^{+}25 \bullet f^{+}29$
CD-DawnModel (4, 8)	$f^{+}2 \bullet f^{+}3 \bullet f^{+}4 \bullet f^{+}5 \bullet f^{+}6 \bullet f^{+}7 \bullet f^{+}8 \bullet f^{+}9 \bullet f^{+}10 \bullet f^{+}11 \bullet f^{+}12 \bullet f^{+}13 \bullet f^{+}22 \bullet f^{+}24 \bullet f^{+}26 \bullet f^{+}29$
CD-DeluxeModel (4, 9)	$f^{+}2 \bullet f^{+}3 \bullet f^{+}4 \bullet f^{+}5 \bullet f^{+}6 \bullet f^{+}7 \bullet f^{+}8 \bullet f^{+}9 \bullet f^{+}10 \bullet f^{+}11 \bullet f^{+}12 \bullet f^{+}13 \bullet f^{+}22 \bullet f^{+}23 \bullet f^{+}26 \bullet f^{+}29$

Glamour(Normal)Model *(4, 10)*	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+13 \bullet f^+20 \bullet f^+23 \bullet f^+25 \bullet f^+29$
Glamour(PGMFi)Model *(4, 11)*	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+13 \bullet f^+20 \bullet f^+23 \bullet f^+25 \bullet f^+29$
AchieverModel *(4, 12)*	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+13 \bullet f^+20 \bullet f^+23 \bullet f^+25 \bullet f^+28$
CBZXtremeModel *(4, 13)*	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+13 \bullet f^+20 \bullet f^+23 \bullet f^+25 \bullet f^+28$
HunkModel *(4, 14)*	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+13 \bullet f^+21 \bullet f^+23 \bullet f^+25 \bullet f^+28$
ImpulseModel *(4, 15)*	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+14 \bullet f^+20 \bullet f^+23 \bullet f^+25 \bullet f^+28$
CT100Model *(4, 16)*	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+15 \bullet f^+22 \bullet f^+24 \bullet f^+26 \bullet f^+29$
Pulsar135LSModel *(4, 17)*	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+15 \bullet f^+20 \bullet f^+23 \bullet f^+25 \bullet f^+28$
Pulsar150DTS-iModel *(4, 18)*	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+15 \bullet f^+20 \bullet f^+23 \bullet f^+25 \bullet f^+28$
Pulsar180DTS-iModel	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+15 \bullet f^+20 \bullet f^+23 \bullet f^+25 \bullet f^+28$

(4, 19)	
Pulsar220DTS-iModel (4, 20)	f+ 2 • f+ 3 • f+ 4 • f+ 5 • f+ 6 • f+ 7 • f+ 8 • f+ 9 • f+ 10 • f+ 11 • f+ 12 • f+ 15 • f+ 21 • f+ 23 • f+ 25 • f+ 28
Avenger220DTS-iModel (4, 21)	f+ 2 • f+ 3 • f+ 4 • f+ 5 • f+ 6 • f+ 7 • f+ 8 • f+ 9 • f+ 10 • f+ 11 • f+ 12 • f+ 15 • f+ 20 • f+ 23 • f+ 25 • f+ 28
Discover135Model (4, 22)	f+ 2 • f+ 3 • f+ 4 • f+ 5 • f+ 6 • f+ 7 • f+ 8 • f+ 9 • f+ 10 • f+ 11 • f+ 12 • f+ 15 • f+ 20 • f+ 23 • f+ 25 • f+ 29
Discover125Model (4, 23)	f+ 2 • f+ 3 • f+ 4 • f+ 5 • f+ 6 • f+ 7 • f+ 8 • f+ 9 • f+ 10 • f+ 11 • f+ 12 • f+ 15 • f+ 22 • f+ 23 • f+ 25 • f+ 28
Discover100Model (4, 24)	f+ 2 • f+ 3 • f+ 4 • f+ 5 • f+ 6 • f+ 7 • f+ 8 • f+ 9 • f+ 10 • f+ 11 • f+ 12 • f+ 15 • f+ 22 • f+ 23 • f+ 25 • f+ 29
Platina100Model (4, 25)	f+ 2 • f+ 3 • f+ 4 • f+ 5 • f+ 6 • f+ 7 • f+ 8 • f+ 9 • f+ 10 • f+ 11 • f+ 12 • f+ 15 • f+ 22 • f+ 23 • f+ 26 • f+ 29
Duke200Model (4, 26)	f+ 2 • f+ 3 • f+ 4 • f+ 5 • f+ 6 • f+ 7 • f+ 8 • f+ 9 • f+ 10 • f+ 11 • f+ 12 • f+ 15 • f+ 21 • f+ 23 • f+ 25 • f+ 27
BulletElectraTwinspark Model (4, 27)	f+ 2 • f+ 3 • f+ 4 • f+ 5 • f+ 6 • f+ 7 • f+ 8 • f+ 9 • f+ 10 • f+ 11 • f+ 12 • f+ 16 • f+ 20 • f+ 24 • f+ 25 • f+ 28
Bullet 350 Twinspark Model	f+ 2 • f+ 3 • f+ 4 • f+ 5 • f+ 6 • f+ 7 • f+ 8 • f+ 9 • f+ 10 • f+ 11 • f+ 12 • f+ 16 • f+ 20 • f+ 24 • f+ 25 • f+ 28

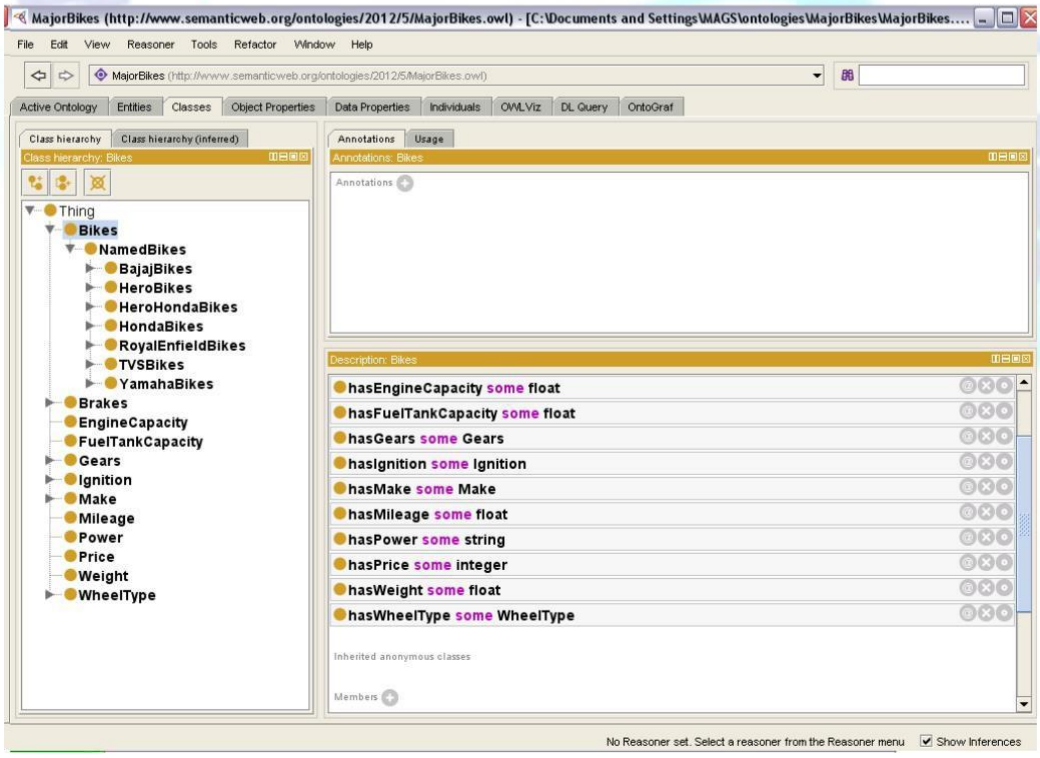
(4, 28)	
Bullet Electra EFI Model (4, 29)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+16 \bullet f^+20 \bullet f^+24 \bullet f^+25 \bullet f^+28$
Bullet Electra Deluxe Model (4, 30)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+16 \bullet f^+20 \bullet f^+24 \bullet f^+25 \bullet f^+28$
Royal Enfield Classic 500 Model (4, 31)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+16 \bullet f^+20 \bullet f^+24 \bullet f^+25 \bullet f^+28$
Royal Enfield Classic 350 Model (4, 32)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+16 \bullet f^+20 \bullet f^+24 \bullet f^+25 \bullet f^+28$
Thunderbird Twinspark Model (4, 33)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+16 \bullet f^+20 \bullet f^+24 \bullet f^+25 \bullet f^+28$
R15Model (4, 34)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+17 \bullet f^+21 \bullet f^+23 \bullet f^+25 \bullet f^+27$
FZModel (4, 35)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+17 \bullet f^+20 \bullet f^+23 \bullet f^+25 \bullet f^+28$
VictorModel (4, 36)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+18 \bullet f^+22 \bullet f^+24 \bullet f^+26 \bullet f^+29$
CBRModel (4, 37)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+19 \bullet f^+21 \bullet f^+23 \bullet f^+25 \bullet f^+27$
ShineModel (4, 38)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+19 \bullet f^+22 \bullet f^+23 \bullet f^+25 \bullet f^+29$
UnicornModel (4, 39)	$f^+2 \bullet f^+3 \bullet f^+4 \bullet f^+5 \bullet f^+6 \bullet f^+7 \bullet f^+8 \bullet f^+9 \bullet f^+10 \bullet f^+11 \bullet f^+12 \bullet f^+19 \bullet f^+20 \bullet f^+23 \bullet f^+25 \bullet f^+28$

Fig. 9: Partial ontology structure of Ontology of Bikes



The ontology was formalized using the Protégé tool from Stanford University [19], see figure 10.

Fig. 10: Class hierarchy for Ontology of Bikes



4. Evaluation

The developed ontology was evaluated by experts by querying it with competency questions and verifying the results. Figure 11(a) shows the following query:

" Average >=40 & price should be between Rs. 50000 & Rs. 60000, and engine capacity should be >=125"

Figure 11(b), shows the results after execution of the *Reasoner,* the results were verified by the experts and were found to be consistent.

Fig 11(a): Evaluation of competency question

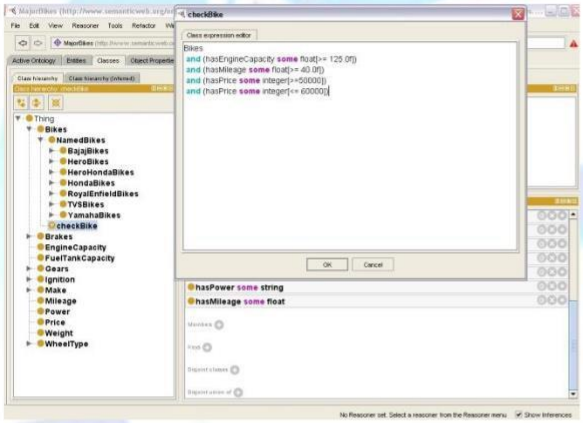
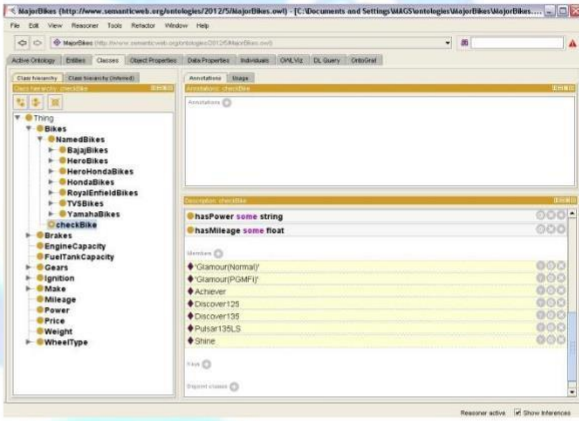


Fig 11(b): Evaluation of competency question



The hierarchy was also evaluated, figure 12 shows the output of OntoGraph and figure 13 shows the output of OWLViz.

Fig. 12: The output of OntoGraf

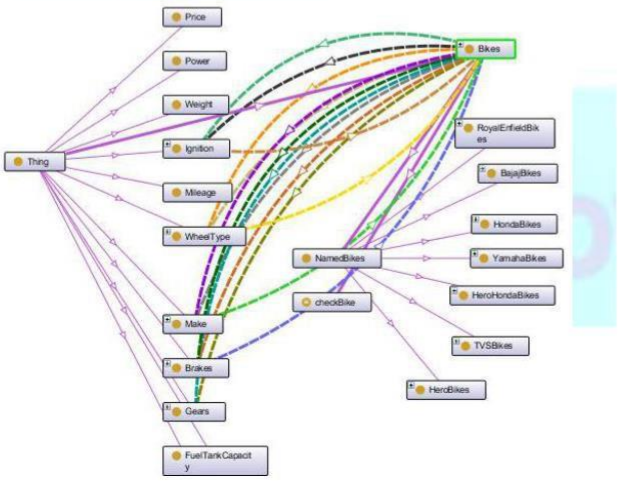


Fig. 13: The output of OWLViz

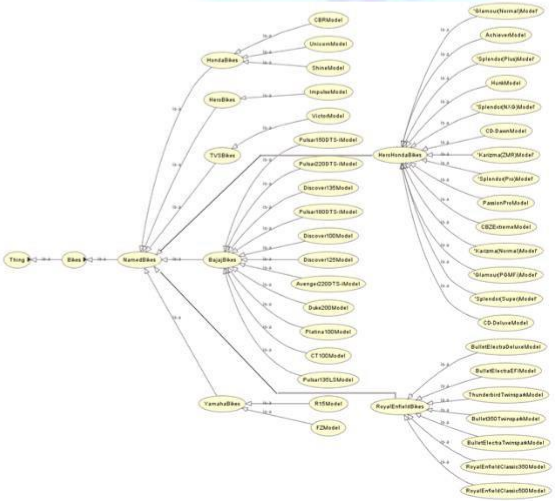


Table- 3 presents the mapping of main concepts of our On-To-Methodology to notions used by previously presented Ontology engineering methods.

Table 3: Mapping of concepts between On-To-Methodology & other methodologies

Variable Name	On-To-Methodology	METHONTOLOGY	Methodology by Farooq et. al.	Methodology by Gaoyun et. al.	MADRE	Automatic ontology construction approach by Jia et. al.	Concept feature based Ontology Construction and Maintenance	Structured Ontology Construction by using data clustering & Maintenance
TYPE	Semi Automatic	Manual	Manual	Manual	Manual	Automatic	Automatic	Semi Automatic
Documentation	✔	✓	X	X	X	X	X	✓
Validation/ Verification	✔/✔	✓/✓	✓/✓	X/✓	X/✓	✓/X	✓/X	–
	Define Scope	✓ =Specification	✓ =Adaptation at Specification level	✓ = Clarifying domain	✓ = Domain Analysis	--	–	–
	Knowledge Acquisition: Domain Understanding & Knowledge Elicitation	✓ =Knowledge Acquisition	✓ =Domain vocabulary declaration	✓	✓ =Ontology Knowledge Acquisition	✓ =Extraction from Free Text + Dictionary	✓	✓ =Relational analysis + Clustering (Ontology construction)
	Knowledge Acquisition: Building Conceptual Model of the Knowledge	✓ Conceptualize	✓ = Identifying and grouping resources + Identifying axioms, relationships, data-characteristics and naming them	✓	✓ =Phase of Ontology analysis & design (Construct Ontology Model)	✓ =Extraction from Knowledge Base	✓	✓ =Concept analysis (Ontology Construction)
	Analysis & Validation	X	✓ =Verification	–	X	✓ =Extract Relational patterns	X	–
	Design Ontology	✓ =Integration	✓ =Adaptation at design level	✓ =Identifying attributes	✓ =Phase of Ontology analysis & design		–	✓ = System architecture)
	Formalization	✓ = Implementation		✓	✓ =Phase of Ontology representation	✓ =Sub-ontology Extraction	–	–
	Evaluation	✓	X	X	✓	X	✓	–
	Maintenance	X	X	X	X	X	✓	X

5. CONCLUSION & FUTURE WORK

The work carried out in this report provides with strong base in Ontology Engineering. Here we have presented the theories and principles for construction of ontology according to various currently existent methodologies. We also noted the various issues related to the field of ontology development.

Since there exists no one methodology that can be applied in all scenarios of ontology development, our main emphasis here is on building a methodology that can be used as a standard method for ontology construction under various domains. For this we have proposed a method namely *On-To-Methodology*. Then we illustrated our methodology by working out an Ontology of bikes.

The future work entails development of a tool that will automatically generate the Conceptual model of the knowledge. Another future direction is to develop software that would

work on the ontology of Bikes that we created and has simpler graphical user interface so that it could be operated by non-technical users.

6. REFERENCES

[1] Guarino, N., Formal ontology in information systems. 1998, IOS Pr.

[2] Ishaq Azhar Mohammed. (2014). Economics of Identity and Access Management: Providing decision support for investments. International Journal of Managment, IT and Engineering (IJMIE), 4(2), 540–549. Retrieved from https://www.ijmra.us/project doc/2014/IJMIE_FEBRUARY2014/IJMIEFeb14-Ishaq.pdf

[3] Borst, P., H. Akkermans,and J. Top, "Engineering ontologies", International Journal of Human Computer Studies, 1997. 46: p. 365-406.

[4] Dhaya Sindhu Battina, "APPLICATION RESEARCH OF ARTIFICIAL INTELLIGENCE IN ELECTRICAL AUTOMATION CONTROL", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.3, Issue 4, pp.784-787, December 2015, Available at :http://www.ijcrt.org/papers/IJCRT1134266.pdf

[5] Gruber T R., "A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition", 1993,5(2): 199-220

[6] Gruber T R, 1993, "Towards principles for the design of ontologiesused for knowledge sharing", International Workshop on Formal Ontology in Conceptual analysis and Knowledge Representation, eds. G N & P R, Kluver Academic Publishers, Deventer, The Netherlands, Padova, Italy

[7] Zhang Qing,Tian Yuexin, "Requirements capturement and analysis",In China Science and Technology Information,Feb,2009:1-2.

[8] Ishaq Azhar Mohammed, "THE INTERACTION BETWEEN ARTIFICIAL INTELLIGENCE AND IDENTITY AND ACCESS MANAGEMENT: AN EMPIRICAL STUDY", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.3, Issue 1, pp.668-671, March 2015, Available at :http://www.ijcrt.org/papers/IJCRT1134113.pdf

[9] Ashok Kumar Reddy Nadikattu, "APPLICATION OF DATA SCIENCE IN HEALTHCARE AND THE IMPLICATIONS THAT ARISE", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.4, Issue 1, pp.64-69, January-2016, Available at :http://www.ijcrt.org/papers/IJCRT1134033.pdf

[10] Yao-Tang Yu, Chien-Chang Hsu, "A structured ontology construction by using data clustering and pattern tree mining", Proceedings of the 2011 International Conference on Machine Learning and Cybernetics, Guilin, 10-13 July, 2011, 45-50

[11] Gaoyun J, Jianliang W, Shaohua Y, "A Method for Consistent Ocean Ontology Construction", IEEE 201O 2nd International Conference on Industrial and Information Systems, 37-42

[12] Hepp M, "Possible Ontologies- How Reality Constrains the Development of Relevant Ontologies", 2007

[13] Fernández M, Gómez-Pérez A, Juristo N, "METHONTOLOGY: From Ontological Art Towards Ontological Engineering", AAAI Technical Report SS-97-06, available at *http://www.aaai.org/Papers/Symposia/Spring/1997/SS-97-06/SS97-06-005.pdf*

[14] Ishaq Azhar Mohammed, "THE IMPACT OF AI ON IDENTITY AND ACCESS MANAGEMENT: AN EMPIRICAL ANALYSIS", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.3, Issue 3, pp.672-675, September 2015, Available at :http://www.ijcrt.org/papers/IJCRT1134114.pdf

[15] Farooq A, Shah A, Asif K H, "Design of Ontology in Semantic Web Engineering Process" Zhang Y, tan L, Liu J, Yu CC, "A Domain Ontology Construction Method Towards Healthy Housing", 2010 IEEE

[16] Jia M, Zheng D, Liu L, Yang B, Sun W, Yang J, "Automatic Ontology Construction Approaches and its Application on Military Intelligence", 2009 Asia-Pacific Conference on Information Processing, 348-351

[17] Chen Y J, Chen Y M, Huang C Y, Chao C Y, "Concept Feature-based Ontology Construction and Maintenance", 2010 IEEE, 28-32

[18] Yeruva Ramana Reddy, "Modeling Load-Settlement Behavior Of Driven Piles Embedded In Cohesive Soils Using Artificial Neural Networks", IJRAR - International Journal of Research and Analytical Reviews (IJRAR), E-ISSN 2348-1269, P- ISSN 2349-5138, Volume.1, Issue 2, Page No pp.941-944, May 2014, Available at : http://www.ijrar.org/IJRAR19D3842.pdf

[19] Gupta D, Jat K, "Knowledge representation with Ontology and Relational database to RDF converter"

[20] Ishaq Azhar Mohammed, "HOW ARTIFICIAL INTELLIGENCE IS CHANGING CYBER SECURITY LANDSCAPE AND PREVENTING CYBER ATTACKS: A SYSTEMATIC REVIEW", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.4, Issue 2, pp.659-663, June 2016, Available at :http://www.ijcrt.org/papers/IJCRT1134111.pdf

[21] Dan Z, Li Z, Hao J, "Research on Semi-automatic Domain Ontology Construction", 2010 Seventh Web Information Systems and Applications Conference, 115-118

[22] Wei QUI, "Development and Application of Knowledge Engineering Based on Ontology", 2010 Third International Conference on Knowledge Discovery and Data Mining, 518-521

[23] N. F. Noy, and D. L. McGuinness, "Ontology Development 101: A guide to creating your first ontology", Technical Report SMI-2001-0880, Stanford Medical Informatics, 2001

[24] Dhaya Sindhu Battina, "AI-AUGMENTED AUTOMATION FOR DEVOPS, A MODEL-BASED FRAMEWORK FOR CONTINUOUS DEVELOPMENT IN CYBER-PHYSICAL SYSTEMS", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.4, Issue 3, pp.779-783, September 2016, Available at :http://www.ijcrt.org/papers/IJCRT1134265.pdf

[25] Islam N, Siddiqui MS, Shaikh ZA, "TODE : A Dot Net Based Tool for Ontology Development and Editing", 2010 2nd International Conference on Computer Engineering and Technology.

[26] http://protege.stanford.edu/ : Protégé tool from Stanford University

[27] Ibrahim Ali Mohammed, "An Exploratory Study on IoT-based Smart Technologies and Ongoing Research Trends", International Journal of Current Science (IJCSPUB), ISSN:2250-1770, Volume.2, Issue 4, pp.53-56, December 2012, Available at :http://www.ijcspub.org/papers/IJCSP12D1008.pdf

[28] VENKATESWARANAIDU KOLLURI, "AN IN-DEPTH EXPLORATION OF UNVEILING VULNERABILITIES: EXPLORING RISKS IN AI MODELS AND ALGORITHMS", IJRAR - International Journal of Research and Analytical Reviews (IJRAR), E-ISSN 2348-1269, P- ISSN 2349-5138, Volume.1, Issue 3, Page No pp.910-913, September 2014, Available at : http://www.ijrar.org/IJRAR19D5168.pdf

[29] Teja Reddy Gatla, "AN INNOVATIVE STUDY EXPLORING REVOLUTIONIZING HEALTHCARE WITH AI: PERSONALIZED MEDICINE: PREDICTIVE DIAGNOSTIC TECHNIQUES AND INDIVIDUALIZED TREATMENT", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.4, Issue 3, pp.585-589, August 2016, Available at :http://www.ijcrt.org/papers/IJCRT1135521.pdf

View publication stats

Check for updates

Received: 29 January 2024 | Revised: 14 April 2024 | Accepted: 19 April 2024

DOI: 10.1002/mma.10172

SPECIAL ISSUE PAPER

WILEY

Parametric Bernstein operators based on contagion distribution

Kanita | **Naokant Deo**

Department of Applied Mathematics, Delhi Technological University, Delhi, India

Correspondence
Naokant Deo and Kanita, Department of Applied Mathematics, Delhi Technological University, Delhi 110042, India.

Email: naokantdeo@dtu.ac.in and kanitatewari@gmail.com

Communicated by: D. Zeidan

Funding information
Delhi Technological University

This paper proposes the extended form of the α-Bernstein operators with the Pólya–Eggenberger distribution, also known as contagion distribution, using the integral operators for $m \geq 2$. We will refer to these sequence of operators as the "Parametric-Bernstein operators based on contagion distribution" or simply the modified form of α-Bernstein operators. These operators are defined with the help of two parameters, namely, α and γ. We observe the significance of both these parameters. Subsequently, we calculate their moments and on the basis of Korovkin's approximation theorem we assert that our defined operators converge for a real valued continuous function h. We give some basic properties, study the Voronovskaya type asymptotic result and discuss the A-statistical approximation of our operators. We also use the modulus of continuity to study the rate of convergence. Alongside our primary work, we have also defined the King-type modification that preserves the operators at e_0 and e_2. Further, using modulus of continuity, we compare the properties and behavior of both the operators.

KEYWORDS

Bernstein operators, contagion distribution, King-type modification, Pólya–Eggenberger distribution

MSC CLASSIFICATION

41A25, 41A36

1 | INTRODUCTION

The Bernstein operators hold significant importance in approximation theory, offering a powerful framework for constructing polynomial approximations used in computer-aided design and graphics through the creation of Bézier curves. The employment of Bernstein polynomials in spline construction serves as a linchpin connecting various disciplines. Additionally, their role in probability and statistics underscores their significance in Bayesian statistics. Over the past years, several studies have been made on these operators and their applications. In the year 1923, Pólya and Eggenberger formulated a discrete probability distribution that is associated with the Pólya–Eggenberger urn scheme. This distribution, commonly referred to as the contagion distribution or the Pólya distribution, was derived using an urn model [1]. The distribution represents a straightforward procedure: An urn holds U red balls and V blue balls, and one ball is selected randomly from it. The colour of the ball is recorded and is substituted along with with X identical, same colored balls. The aforementioned procedure is iterated m times. The occurrence of obtaining a red (or blue) ball in the jth draw is represented by the random variable Y_j, which takes the value 1 (or 0) to indicate the event. The probability of observing a total of k red drawings, denoted as $k = \sum Y_j$, can be determined by the following expression:

 © 2024 John Wiley & Sons Ltd.

$$\rho_{m,k} = {}^{m}C_k \frac{\prod_{i=0}^{k-1}(U+iX)\prod_{i=0}^{m-k-1}(V+iX)}{\prod_{i=0}^{m-1}(U+V+iX)}. \tag{1.1}$$

The year 1968 saw the introduction of linear positive operators $P_m^{\langle\gamma\rangle} : C[0,1] \to C[0,1]$, defined by Stancu [2, 3], based on the above defined contagion distribution (1.1) with density function as

$$P_m^{\langle\gamma\rangle}(f;x) = \sum_{k=0}^{m} {}^{m}C_k \frac{x^{[k,-\gamma]}(1-x)^{[m-k,-\gamma]}}{1^{[m,-\gamma]}} f\left(\frac{k}{m}\right), \tag{1.2}$$

where, γ is a non-negative parameter and ${}^{m}C_k$ is the binomial coefficient.

The factorial power is determined as $a^{[n,\nu]} = a(a-\nu)(a-2\nu)\ \ldots\ (a-(n-1)\nu)$ and $a^{[0,\nu]} = 1$. This represents a polynomial of degree n.

Later on, several mathematicians gave various modifications of Bernstein operators [4–9]. Apart from the modifications of Bernstein operators, a number of studies have also been made on their Durrmeyer variant which introduces a weight function into the Bernstein basis functions [10–14]. Year 2017 was a crucial year in the field of positive linear operators as Cheng et al. [9] defined the following operators:

For each $m \geq 2$ and fixed real α,

$$\mathrm{B}_{m,\alpha}(f;x) = \sum_{k=0}^{m} f_k \left[{}^{m-2}C_k\,(1-\alpha)\,x + {}^{m-2}C_{k-2}\,(1-\alpha)\,(1-x) + {}^{m}C_k\,\alpha x(1-x)\right] x^{k-1}(1-x)^{m-k-1}, \tag{1.3}$$

where $f_k = f\left(\frac{k}{m}\right)$ and f is defined on $[0,1]$. These operators, $\mathrm{B}_{m,\alpha}(f;x)$ are named as the α-Bernstein operators.

In recent studies, several mathematicians (see the literature [15–19]) gave various modifications of α-Bernstein operators (1.3). These modifications have been analyzed in terms of their auxiliary properties and used modulus of continuity to observe the rate of convergence, analyzed the direct local approximation theorem, function of bounded variation and Voronovskaya type asymptotic result.

Now we consider another modification of the α-Bernstein operators with the help of contagion distribution as density function, defined for $m \geq 2$ as

$$\begin{aligned} \mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x) = \sum_{k=0}^{m} \Bigg[& {}^{m-2}C_k\,(1-\alpha)\,\frac{x^{[k,-\gamma]}(1-x)^{[m-k-1,-\gamma]}}{1^{[m-1,-\gamma]}} \\ & - {}^{m-2}C_{k-2}\,(1-\alpha)\,\frac{x^{[k-1,-\gamma]}(1-x)^{[m-k,-\gamma]}}{1^{[m-1,-\gamma]}} \\ & + {}^{m}C_k\,\alpha\frac{x^{[k,-\gamma]}(1-x)^{[m-k,-\gamma]}}{1^{[m,-\gamma]}} \Bigg] h_k, \end{aligned} \tag{1.4}$$

where, $h_k = h\left(\frac{k}{m}\right)$.

Let us denote the polynomial inside summation as $p_{m,k}^{\langle\alpha,\gamma\rangle}(x)$, that is,

$$\begin{aligned} p_{m,k}^{\langle\alpha,\gamma\rangle}(x) = & {}^{m-2}C_k\,(1-\alpha)\,\frac{x^{[k,-\gamma]}(1-x)^{[m-k-1,-\gamma]}}{1^{[m-1,-\gamma]}} \\ & - {}^{m-2}C_{k-2}\,(1-\alpha)\,\frac{x^{[k-1,-\gamma]}(1-x)^{[m-k,-\gamma]}}{1^{[m-1,-\gamma]}} \\ & - {}^{m}C_k\,\alpha\frac{x^{[k,-\gamma]}(1-x)^{[m-k,-\gamma]}}{1^{[m,-\gamma]}}. \end{aligned}$$

10991476, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mma.10172 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

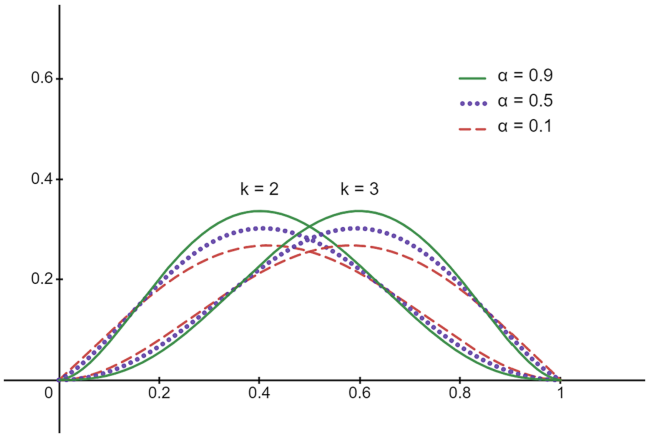


FIGURE 1 Graph of the α-Bernstein polynomial for $m = 5$, $k = 2$ and $k = 3$ and different values of α. [Colour figure can be viewed at wileyonlinelibrary.com]

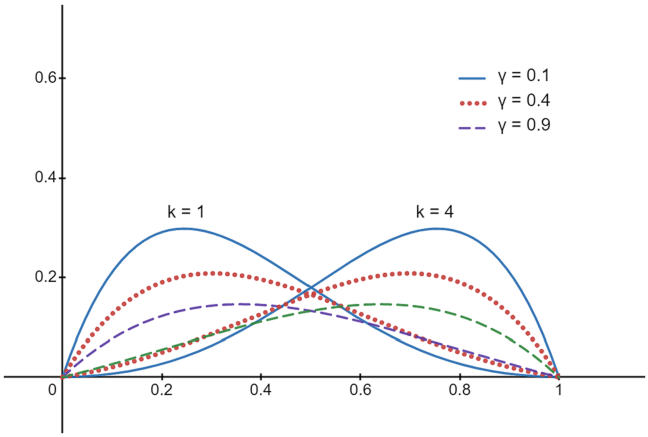


FIGURE 2 Graph of the contagion form of α-Bernstein polynomial for $m = 5$, $k = 1$ and $k = 4$, $\alpha = 0.5$ and different values of γ. [Colour figure can be viewed at wileyonlinelibrary.com]

Some basic calculations give us the alternate form of $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x)$:

$$\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x) = (1-\alpha)\sum_{k=0}^{m-1} {}^{m-1}C_k \frac{x^{[k,-\gamma]}(1-x)^{[m-k-1,-\gamma]}}{1^{[m-1,-\gamma]}} g_k - \alpha \sum_{k=0}^{m} {}^{m}C_k \frac{x^{[k,-\gamma]}(1-x)^{[m-k,-\gamma]}}{1^{[m,-\gamma]}} h_k, \tag{1.5}$$

where $h_k = h\left(\frac{k}{m}\right)$, $g_k = \left(1 - \frac{k}{m-1}\right) h_k + \frac{k}{m-1} h_{k+1}$
and ${}^{m}C_k$ is the binomial coefficient.

Special cases:

1. For $\gamma = 0$ and $\alpha = 1$, the proposed operators $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x)$ transform into the classical Bernstein operators.
2. For $\alpha = 1$, the proposed operators transform into the operators (1.2), proposed by Stancu.
3. For $\gamma = 0$, we get the operators (1.3) as defined by Cheng et al. for $m \geq 2$.

From Figures 1 and 2, it is clear that like in the classical Bernstein, α-Bernstein and the Stancu-Bernstein operators, the basis polynomial, $p_{m,k}^{\langle\alpha,\gamma\rangle}(x)$ as defined with the help of operators (1.4) are also symmetric in $[0, 1]$. Contagion distribution is particularly important and applicable in various real-world scenarios where the probability of an event depends on its past occurrences.

10991476, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mma.10172 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

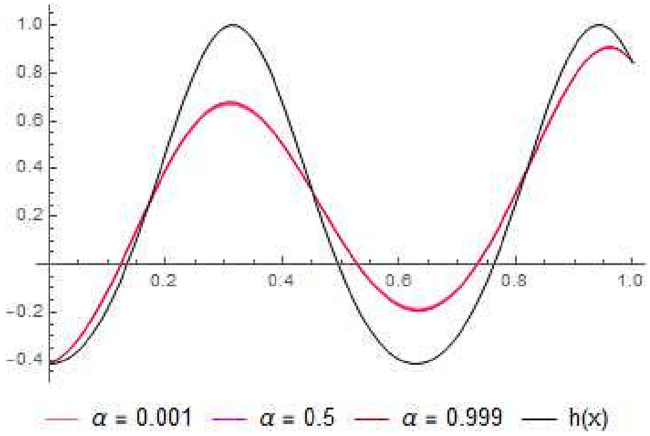


FIGURE 3 Approximation process of $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(\cos(2\cos 5t);x)$ for $m = 30$ when $\gamma = 0.01$ is fixed and α is varying in $[0,1]$. [Colour figure can be viewed at wileyonlinelibrary.com]

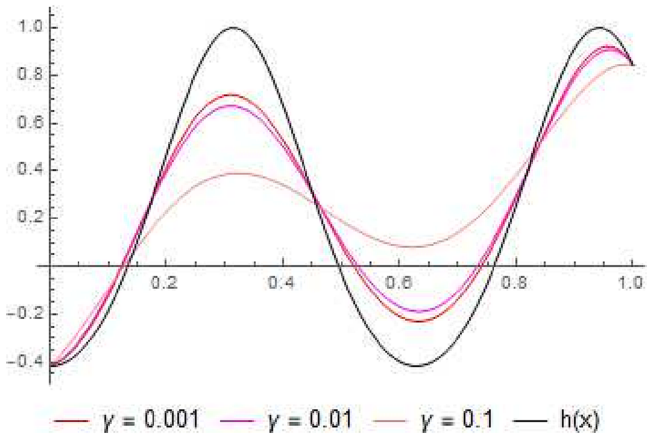


FIGURE 4 Approximation process of $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(\cos(2\cos 5t);x)$ for $m = 30$ when $\alpha = 0.5$ is fixed and γ is varying in $[0,1]$. [Colour figure can be viewed at wileyonlinelibrary.com]

1.1 | Comparison of the parameters α and γ

This paper aims to analyze the results of approximations for the sequence of operators (1.5), which have been defined by the utilization of two parameters, namely, α and γ. Both entities are restricted to exist within the closed interval $[0,1]$. In order to gain a better understanding of the impact of these parameters on the process of approximation, an analysis is performed using a specific example, with the function $h(x) = \cos(2\cos 5x)$.

Initially, we fix the value of γ and then observe how varying α affects the results (as shown in Figure 3). Remarkably, it is observed that despite substantial variations in α, its impact on the enhancement of the approximation process is not significant. In contrast, modifying the values of γ exerts a substantial impact on the precision of the approximations made by the operators. Notably, smaller values of γ results in more precise approximations, suggesting that as m approaches ∞, γ tends to converge toward 0 (Figure 4).

Thus, we can conclude a relationship between γ and m as $\gamma = O(1/m)$. This can also be studied in the upcoming section concerning the Korovkin theorem. Nevertheless, it is worth noting that there exists potential of further research by considering simultaneous operators for $\gamma = O(1/m^a)$, where $a > 0$, or any other function $\gamma(m) \to 0$, which opens door for analysis on the convergence properties in future. Our choice of $\gamma = O(1/m)$ reduces further complexities.

Lemma 1.1. *Let $\alpha \in [0,1]$ and $\gamma = O(1/m) \in [0,1]$. Then the moments of $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x)$ are as follows:*

(i) $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(1;x) = 1$

(ii) $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(t;x) = x$

(iii) $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(t^2;x) = \frac{1}{1+\gamma}\left[x^2 + \frac{m+2(1-\alpha)}{m^2}x(1-x)\right] + x\frac{\gamma}{1+\gamma}$

10991476, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mma.10172 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

(iv) $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}\left(t^3;x\right)=x\left[\frac{m+6(1-\alpha)}{m^3}\right]+x\left(\frac{x+\gamma}{1+\gamma}\right)\left[\frac{3m^2+3m(1-2\alpha)-18(1-\alpha)}{m^3}\right]+x\left(\frac{x+\gamma}{1+\gamma}\right)\left(\frac{x+2\gamma}{1+2\gamma}\right)\left[\frac{m^3-3m^2-2m(2-3\alpha)+12(1-\alpha)}{m^3}\right]$

(v) $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}\left(t^4;x\right)=x\left[\frac{m+14(1-\alpha)}{m^4}\right]+x\left(\frac{x+\gamma}{1+\gamma}\right)\left[\frac{7m^2+m(29-36\alpha)-86(1-\alpha)}{m^4}\right]+6x\left(\frac{x+\gamma}{1+\gamma}\right)\left(\frac{x+2\gamma}{1+2\gamma}\right)\left[\frac{m^3-m^2(1+2\alpha)-2m(7-8\alpha)+24(1-\alpha)}{m^4}\right]$
$+x\left(\frac{x+\gamma}{1+\gamma}\right)\left(\frac{x+2\gamma}{1+2\gamma}\right)\left(\frac{x+3\gamma}{1+3\gamma}\right)\left[\frac{m^4-6m^3-m^2(1-12\alpha)+6m(9-10\alpha)-72(1-\alpha)}{m^4}\right].$

Proof. To calculate the moments of the contagion form of α-Bernstein operators, we will use the moments of the contagion form of the classical Bernstein operators (1.2). For $h(x)=1$, we have $h_k=1=g_k$, and thus,

$$\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(1;x)=(1-\alpha)\sum_{k=0}^{m-1}{}^{m-1}C_k\frac{x^{[k,-\gamma]}(1-x)^{[m-k-1,-\gamma]}}{1^{[m-1,-\gamma]}}+\alpha\sum_{k=0}^{m}{}^{m}C_k\frac{x^{[k,-\gamma]}(1-x)^{[m-k,-\gamma]}}{1^{[m,-\gamma]}}=1.$$

Now for $h(x)=x$, $h_k=\frac{k}{m}$ and $g_k=\frac{k}{m-1}$, we have

$$\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(t;x)=(1-\alpha)\sum_{k=0}^{m-1}{}^{m-1}C_k\frac{x^{[k,-\gamma]}(1-x)^{[m-k-1,-\gamma]}}{1^{[m-1,-\gamma]}}\frac{k}{m-1}+\alpha\sum_{k=0}^{m}{}^{m}C_k\frac{x^{[k,-\gamma]}(1-x)^{[m-k,-\gamma]}}{1^{[m,-\gamma]}}\frac{k}{m}=x.$$

For $h(x)=x^2$, $h_k=\frac{k^2}{m^2}$ and $g_k=\frac{k^2}{(m-1)^2}\left[\frac{m^2-1}{m^2}\right]+\frac{k}{(m-1)m^2}$, we have

$$\begin{aligned}
\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}\left(t^2;x\right)&=(1-\alpha)\left[\sum_{k=0}^{m-1}{}^{m-1}C_k\frac{x^{[k,-\gamma]}(1-x)^{[m-k-1,-\gamma]}}{1^{[m-1,-\gamma]}}\frac{k^2}{(m-1)^2}\left(\frac{m^2-1}{m^2}\right)\right.\\
&\quad\left.+\sum_{k=0}^{m-1}{}^{m-1}C_k\frac{x^{[k,-\gamma]}(1-x)^{[m-k-1,-\gamma]}}{1^{[m-1,-\gamma]}}\frac{k}{(m-1)}\left(\frac{1}{m^2}\right)\right]\\
&\quad-\alpha\sum_{k=0}^{m}{}^{m}C_k\frac{x^{[k,-\gamma]}(1-x)^{[m-k,-\gamma]}}{1^{[m,-\gamma]}}\frac{k^2}{m^2}\\
&=(1-\alpha)\left\{\left(\frac{m^2-1}{m^2}\right)\frac{1}{1+\gamma}\left[\left(\frac{m-2}{m-1}\right)x^2+\left(\frac{(m-1)\gamma+1}{m-1}\right)x\right]+\frac{1}{m^2}x\right\}\\
&\quad-\frac{\alpha}{1+\gamma}\left[\left(\frac{m-1}{m}\right)x^2+\left(\frac{m\gamma+1}{m}\right)x\right]\\
&=\frac{x^2}{1+\gamma}\left[(1-\alpha)\left(1-\frac{1}{m}-\frac{2}{m^2}\right)+\alpha\left(1-\frac{1}{m}\right)\right]\\
&\quad-\frac{x}{1+\gamma}\left[(1-\alpha)\frac{(m+1)}{m^2}((m-1)\gamma+1)+\frac{\alpha}{m}(m\gamma+1)\right]+x\frac{(1-\alpha)}{m^2}\\
&=\frac{x^2}{1+\gamma}\left[1-\frac{1}{m}-2\frac{(1-\alpha)}{m^2}\right]+\frac{x}{1+\gamma}\left[(1-\alpha)\frac{(m+1)}{m^2}\right.\\
&\quad\left.+(1-\alpha)\gamma\frac{(m+1)(m-1)}{m^2}+\frac{\alpha}{m}+\alpha\gamma\right]+x\frac{(1-\alpha)}{m^2}\\
&=\frac{x^2}{1+\gamma}\left[1-\frac{1}{m}-2\frac{(1-\alpha)}{m^2}\right]+\frac{x}{1+\gamma}\left[\frac{1}{m}+(1-\alpha)\frac{1}{m^2}\right]\\
&\quad-x\frac{\gamma}{1+\gamma}\left[(1-\alpha)\left(1-\frac{1}{m^2}\right)+\alpha\right]+x\frac{(1-\alpha)}{m^2}\\
&=\frac{1}{1+\gamma}\left[x^2-\frac{x^2}{m}-(1-\alpha)\frac{x^2}{m^2}+\frac{x}{m}+(1-\alpha)\frac{x}{m^2}\right]-(1-\alpha)\frac{1}{1+\gamma}\frac{x^2}{m^2}+x\frac{\gamma}{1+\gamma}\\
&\quad-(1-\alpha)\frac{\gamma}{1+\gamma}\frac{x}{m^2}+x\frac{(1-\alpha)}{m^2}\\
&=\frac{1}{1+\gamma}\left[x^2+\frac{m+2(1-\alpha)}{m^2}x(1-x)\right]+x\frac{\gamma}{1+\gamma}.
\end{aligned}$$

Similarly, we can obtain the values for t^3 and t^4. □

10991476, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mma.10172 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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 result brings us to a conclusion that as $m \to \infty$, $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(e_s;x) \to e_s$, where $e_s = t^s$ for $s = 0,1,2$. From the Bohman–Korovkin theorem [20, 21], we can hence state that for a real valued continuous function h, the modified α-Bernstein operators $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x)$ converges to $h(x)$ uniformly. That is,

$$\lim_{m\to\infty} \mathcal{P}_{m,\alpha}^{\langle\gamma\rangle} h = h.$$

2 | APPROXIMATION RESULTS

Lemma 2.1. *Considering Lemma 1.1 and computing the central moments of (1.4), we obtain the following limits:*

(i) $\lim_{m\to\infty} m\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(\varphi_2(t);x) = \frac{2C}{1+\gamma}x(1-x)$

(ii) $\lim_{m\to\infty} m\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(\varphi_4(t);x) = 0,$

where $\varphi_m(t) = (t-x)^m$ *and* $C = \lim_{m\to\infty} \frac{1+m\gamma}{2}$.

Theorem 2.2. *Let* $0 \le \alpha \le 1$ *and* h *is a bounded function defined for all* $x \in [0,1]$. *Then for* $\gamma = O(1/m)$,

$$\lim_{m\to\infty} m\left[\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h(t);x) - h(x)\right] = \frac{C}{1+\gamma}x(1-x)\,h''(x),$$

where $C = \lim_{m\to\infty} \frac{1+m\gamma}{2}$. *The theorem holds true only for those* $x \in [0,1]$ *where* $h''(x)$ *exists.*

Proof. By means of Taylor's series expansion for $k \le m$, we have

$$h(t) = h(x) + (t-x)h'(x) + \frac{1}{2}(t-x)^2h''(x) + \text{ɀ}(t-x)(t-x)^2,$$

where $\lim_{t\to x} \text{ɀ}(t-x) = 0.$. Taking $t = k/m$, we get,

$$\begin{aligned}
&m\left[\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h(t);x) - h(x)\right] \\
&= m\sum_{k=0}^{m} p_{m,k}^{\langle\alpha,\gamma\rangle}(x)\left[\left(\frac{k}{m}-x\right)h'(x) + \frac{1}{2}\left(\frac{k}{m}-x\right)^2 h''(x) + \text{ɀ}\left(\frac{k}{m}-x\right)\left(\frac{k}{m}-x\right)^2\right] \\
&= \sum_{k=0}^{m}(k-mx)p_{m,k}^{\langle\alpha,\gamma\rangle}(x)h'(x) + \frac{1}{2m}\sum_{k=0}^{m}(k-mx)^2 p_{m,k}^{\langle\alpha,\gamma\rangle}(x)h''(x) + mR_m(x),
\end{aligned}$$

where, $R_m(x) = \sum_{k=0}^{m} \text{ɀ}\left(\frac{k}{m}-x\right)\left(\frac{k}{m}-x\right)^2 p_{m,k}^{\langle\alpha,\gamma\rangle}(x).$.

The first summation will be zero and for the second summation, we have

$$\begin{aligned}
\sum_{k=0}^{m}(k-mx)^2 p_{m,k}^{\langle\alpha,\gamma\rangle}(x) &= \sum_{k=0}^{m} k^2 p_{m,k}^{\langle\alpha,\gamma\rangle}(x) - 2mx\sum_{k=0}^{m} k p_{m,k}^{\langle\alpha,\gamma\rangle}(x) + m^2x^2\sum_{k=0}^{m} p_{m,k}^{\langle\alpha,\gamma\rangle}(x) \\
&= \frac{m^2}{1+\gamma}\left[x^2 + \frac{m+2(1-\alpha)}{m^2}x(1-x)\right] + m^2x\frac{\gamma}{1+\gamma} - 2m^2x^2 + m^2x^2 \\
&= \frac{m^2\gamma + m + 2(1-\alpha)}{1+\gamma}x(1-x).
\end{aligned}$$

10991476, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mma.10172 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

Hence,

$$m\left[\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h(t);x)-h(x)\right]=\frac{x(1-x)}{2m(1+\gamma)}\left\{m^2\gamma+m+2(1-\alpha)\right\}h''(x)+mR_m(x)$$
$$=\frac{1}{1+\gamma}\left(\frac{m\gamma}{2}+\frac{1}{2}+\frac{1-\alpha}{m}\right)x(1-x)\,h''(x)+mR_m(x).$$

Now, all that is left to prove is that, $mR_m(x)\to 0$ as $m\to\infty$.
We know, for every $\varepsilon>0$, $\exists\ \delta>0$: $\varepsilon>0,\ \exists\ \delta>0:\ |\not{e}(t-x)|<\varepsilon$.
Also, since $0\leqslant|\not{e}(t-x)|\leqslant m$, therefore $\exists\ M>0$ such that $|\not{e}(t-x)|\leqslant M$ whenever $|t-x|\geqslant\delta$. Hence, we get

$$|\not{e}(t-x)|\leq\varepsilon+M\frac{(t-x)^2}{\delta^2},\qquad\text{for every } t.$$

Thus,

$$m\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}\left(\not{e}(t-x)(t-x)^2;x\right)\leq m\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}\left(\left(\varepsilon+M\frac{(t-x)^2}{\delta^2}\right)(t-x)^2;x\right)$$
$$\leq m\varepsilon\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(\varphi_2;x)+m\frac{M}{\delta^2}P_{n,\alpha}^{\langle\gamma\rangle}(\varphi_4;x).$$

Thus, we can say $\lim_{m\to\infty} m\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}\left(\not{e}(t-x)(t-x)^2;x\right)=0$.
Hence, we get

$$\lim_{m\to\infty} m\left[\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h(t);x)-h(x)\right]=\frac{x(1-x)}{1+\gamma}\left(\lim_{m\to\infty}\frac{m\gamma}{2}+\frac{1}{2}+\frac{(1-\alpha)}{m}\right)h''(x)$$
$$=\frac{C}{1+\gamma}x(1-x)\,h''(x),$$

where $C=\lim_{m\to\infty}\frac{m\gamma+1}{2}$. □

Theorem 2.3. *Let the functions $g,h:(0,1)\to\mathbb{R}$. If $g,h\in C^2(0,1)$ then,*

$$\lim_{m\to\infty} m\left[\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(gh;x)-\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(g;x)\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x)\right]=\frac{2C}{1+\gamma}x(1-x)\,g'(x)h'(x),$$

where $C=\lim_{m\to\infty}\frac{m\gamma+1}{2}$.

Proof. We consider

$$\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(gh;x)-\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(g;x)\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x)$$
$$=\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(gh;x)-g(x)h(x)-\frac{1}{2}(gh)''\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(\varphi_2;x)$$
$$-h(x)\left[\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(g;x)-\frac{1}{2}g''(x)\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(\varphi_2;x)-g(x)\right]$$
$$-\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(g;x)\left[\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x)-\frac{1}{2}h''(x)\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(\varphi_2;x)-h(x)\right]$$
$$-\frac{1}{2}\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(\varphi_2;x)\left[2g'(x)h'(x)+g(x)h''(x)-h''(x)\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(g;x)\right].$$

Multiplying both sides by m, taking $m\to\infty$ and using the above proven identities, we get

$$\lim_{m\to\infty} m\left[\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(gh;x)-\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(g;x)\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x)\right]=\frac{2C}{1+\gamma}x(1-x)\,g'(x)h'(x),$$

where $C=\lim_{m\to\infty}\frac{m\gamma+1}{2}$. □

10991476, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mma.10172 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

We have seen that for any continuous real-valued function h, $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}h \to h$ uniformly on $[0,1]$, giving us an explicit approximation to h. Our next focus shifts on knowing how good this approximation will be. For that we calculate the error that would have occurred during the approximation process. One way to determine this is using the modulus of continuity.

Definition 2.4. The modulus of continuity, denoted by $\omega(h(\cdot);\delta)$, of a function h is defined as

$$\omega(h(t);\delta) = \sup_{\substack{|t_1-t_2|\le\delta \\ t_1,t_2\in[a,b]}} |h(t_1)-h(t_2)|,$$

where $t_1, t_2 \in [a,b]$ and $\delta > 0$.

For simplicity, we can denote $\omega(h(\cdot);\delta)$ with $\omega(\delta)$ as well.

Remark 2.5. For modulus of continuity, $\omega(\delta)$ of $h(t)$ on $[a,b]$, the following property holds: If $\upsilon > 0$, then $\omega(\upsilon\delta) \le (1+\upsilon)\,\omega(\delta)$.

Let us take into account the error bound

$$\left\|h - \mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}h\right\| = \max_{0\le x\le 1}\left|h(x) - \mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x)\right|,$$

where $\|.\|$ is the uniform norm over the interval $[0,1]$. Our next attempt is to attain an upper bound to estimate the approximation error committed during the process.

Theorem 2.6. *Let* $0 \le \alpha \le 1$ *and* $\gamma = O(1/m)$. *For a bounded function* h *on* $[0,1]$, *we deduce*

$$\left\|h - \mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}h\right\| \le \frac{3}{2}\omega\left(\frac{\sqrt{m^2\gamma + m + 2(1-\alpha)}}{m\sqrt{1+\gamma}}\right),$$

where $\omega(.)$ *denotes modulus of continuity.*

Proof. From operators (1.5), we get

$$\begin{aligned}\left|h(x) - \mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x)\right| &= \left|\sum_{k=0}^{m} p_{m,k}^{\langle\alpha,\gamma\rangle}(x)\left[h(x) - h\left(\frac{k}{m}\right)\right]\right| \\ &\le \sum_{k=0}^{m}\omega\left(\left|x-\frac{k}{m}\right|\right)p_{m,k}^{\langle\alpha,\gamma\rangle}(x).\end{aligned}$$

Using Schwarz's inequality, we can say

$$\begin{aligned}\sum_{k=0}^{m}\left|x-\frac{k}{m}\right|p_{m,k}^{\langle\alpha,\gamma\rangle}(x) &= \sum_{k=0}^{m}\left|x-\frac{k}{m}\right|\sqrt{p_{m,k}^{\langle\alpha,\gamma\rangle}(x)}.\sqrt{p_{m,k}^{\langle\alpha,\gamma\rangle}(x)} \\ &\le \left[\sum_{k=0}^{m}\left(x-\frac{k}{m}\right)^2 p_{m,k}^{\langle\alpha,\gamma\rangle}(x)\right]^{1/2}\left[\sum_{k=0}^{m}p_{m,k}^{\langle\alpha,\gamma\rangle}(x)\right]^{1/2} \\ &= \left[\frac{1}{1+\gamma}\left(\frac{m+2(1-\alpha)}{m^2}+\gamma\right)x(1-x)\right]^{1/2} \\ &\le \left[\frac{m^2\gamma + m + 2(1-\alpha)}{4m^2(1+\gamma)}\right]^{1/2}.\end{aligned}$$

10991476, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mma.10172 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

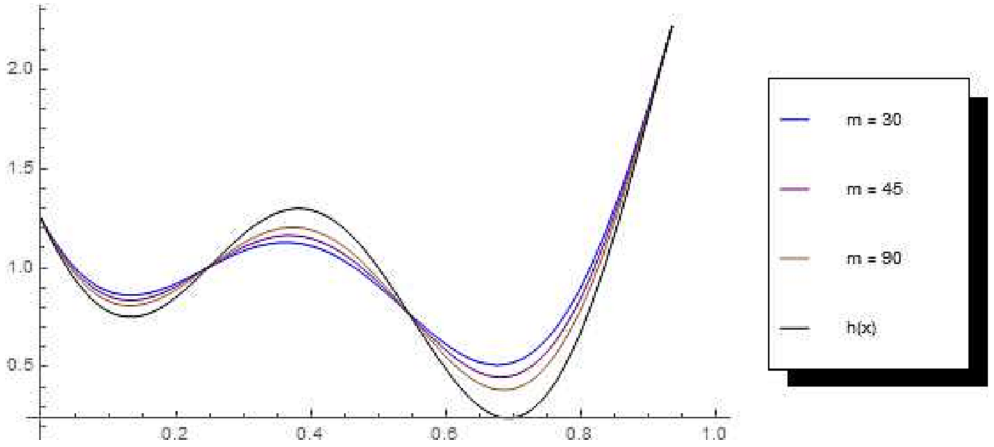


FIGURE 5 Convergence of $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x)$ for parameters $\alpha = 0.7$ and $\gamma = 0.01$, for different values of m. [Colour figure can be viewed at wileyonlinelibrary.com]

From Remark 2.5,

$$\omega\left(\left|x-\frac{k}{m}\right|\right) = \omega\left(\frac{m\sqrt{1+\gamma}}{\sqrt{m^2\gamma+m+2(1-\alpha)}}\left|x-\frac{k}{m}\right|.\frac{\sqrt{m^2\gamma+m+2(1-\alpha)}}{m\sqrt{1+\gamma}}\right)$$
$$\leq \left[1+\frac{m\sqrt{1+\gamma}}{\sqrt{m^2\gamma+m+2(1-\alpha)}}\left|x-\frac{k}{m}\right|\right]\omega\left(\frac{\sqrt{m^2\gamma+m+2(1-\alpha)}}{m\sqrt{1+\gamma}}\right).$$

Thus, we get

$$\begin{aligned}
&\left|h(x)-\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x)\right| \\
&\leq \sum_{k=0}^{m}\left[1+\frac{m\sqrt{1+\gamma}}{\sqrt{m^2\gamma+m+2(1-\alpha)}}\left|x-\frac{k}{m}\right|\right]\omega\left(\frac{\sqrt{m^2\gamma+m+2(1-\alpha)}}{m\sqrt{1+\gamma}}\right)p_{m,k}^{\langle\alpha,\gamma\rangle}(x) \\
&\leq \omega\left(\frac{\sqrt{m^2\gamma+m+2(1-\alpha)}}{m\sqrt{1+\gamma}}\right)\left[1+\frac{m\sqrt{1+\gamma}}{\sqrt{m^2\gamma+m+2(1-\alpha)}}\sum_{k=0}^{m}\left|x-\frac{k}{m}\right|p_{m,k}^{\langle\alpha,\gamma\rangle}(x)\right] \\
&\leq \frac{3}{2}\omega\left(\frac{\sqrt{m^2\gamma+m+2(1-\alpha)}}{m\sqrt{1+\gamma}}\right),
\end{aligned}$$

and thereafter, we arrive at our required result. □

Now that we have obtained an upper bound for the error function, we can also use the fact that $\lim_{\delta\to 0}\omega(\delta) = 0$ iff h is uniformly continuous on $[a,b]$ to see that our modified α-Bernstein operators, $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}h \to h$ uniformly on the interval $[0,1]$. The following graph (Figure 5) gives us a visual representation of the approximation of our operators $\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x)$ defined in (1.4) for $h(x) = (3x-1)\sin\frac{7\pi}{3}x + \frac{5}{4}$.

3 | A-STATISTICAL APPROXIMATION

In 1951, an alternative concept of convergence emerged, recognized as statistical convergence, introduced by Fast [22] and Steinhaus [23]. The term statistical convergence used henceforth is nothing but another name for "almost convergence," used by Antoni Zygmund in his book [24]. Unlike classical convergence, almost convergence accommodates divergent sequences, offering a framework to analyze and understand the behavior of sequences that do not converge in the traditional sense. One of the ways for summing sequences and series using an infinite matrix is the matrix summation method (see Braha et al. [25]). Taking an infinite matrix $A = (a_{mk})$ for $m,k \in \mathbb{N}$, a given sequence $x = \langle x_m\rangle$ can be transformed

10991476, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mma.10172 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

into the sequence $\langle Ax\rangle_m$.

$$\langle Ax\rangle_m = \sum_{k=1}^{\infty} a_{mk}x_k.$$

If $\lim_{m\to\infty}\langle Ax\rangle_m = L$ and the series on the RHS, $\sum_{k=1}^{\infty} a_{mk}x_k$, converges $\forall\ m$, then the sequence $\langle x_m\rangle$ is said to be summable by the matrix A and L will be called its limit. Further, if the limit is preserved then this method is referred to as regular matrix summability method and the infinite matrix A is called a regular matrix.

Definition 3.1. For every $\varepsilon > 0$, if $\lim_m \sum_{m:|x_m-L|\geq\varepsilon} a_{mk} = 0$, then the sequence, $x = \langle x_m\rangle$ is said to be A-statistically convergent to L and is denoted by $st_A - \lim x = L$.

Lemma 3.2 (Gadjiev and Orhan [26])**.** *Let $\{\mathcal{L}_m\}$ be a sequence of linear positive operators and $A = (a_{mk})$ be a regular positive summability matrix. Then,*

$$st_A - \lim_m \|\mathcal{L}_m h - h\| = 0; h \in C([a,b])$$

iff

$$st_A - \lim_m \|\mathcal{L}_m e_r - e_r\| = 0; \quad \textit{where}\ \ e_r = t^r \textit{ for } r = 0,1,2.$$

Theorem 3.3. *Let $A = (a_{mk})$ be a regular positive summability matrix. Then for each $h \in C([0,1])$, we get*

$$st_A - \lim_{m\to\infty} \left\|\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}h - h\right\| = 0.$$

Proof. From Lemma 3.2, it is enough to prove that

$$\left\|\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}e_r - x^r\right\| = 0;$$

where, $e_r = t^r$ for $r = 0,1,2$.

Using the moments of modified α-Bernstein operator (1.4), we have

$$st_A - \lim_{m\to\infty} \left\|\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}e_0 - 1\right\| = 0$$

and

$$st_A - \lim_{m\to\infty} \left\|\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}e_1 - x\right\| = 0;$$

also,

$$\begin{aligned}
&\left\|\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}e_2 - x^2\right\| \\
&\leq \sup\left\{\left|\frac{x^2}{1+x^2}\left[\frac{(m+1)(m-2)}{m^2}\frac{1-\alpha}{1+\gamma} + \frac{m-1}{m}\frac{\alpha}{1+\gamma} - 1\right]\right|\right. \\
&\quad \left. + \left|\frac{x}{1+x^2}\left[\frac{((m-1)\gamma+1)(m+1)}{m^2}\frac{1-\alpha}{1+\gamma} + \frac{m\gamma+1}{m}\frac{\alpha}{1+\gamma} + \frac{1-\gamma}{m^2}\right]\right|\right\} \\
&= \frac{(m+1)(m-2)}{m^2}\frac{1-\alpha}{1+\gamma} + \frac{m-1}{m}\frac{\alpha}{1+\gamma} - 1 \\
&\quad - \frac{((m-1)\gamma+1)(m+1)}{m^2}\frac{1-\alpha}{1+\gamma} + \frac{m\gamma+1}{m}\frac{\alpha}{1+\gamma} + \frac{1-\gamma}{m^2} \\
&= \frac{(1-\alpha)(-1-\gamma)+(1-\gamma)(1+\gamma)}{m^2(1+\gamma)} \\
&= \frac{\alpha-\gamma}{m^2}.
\end{aligned}$$

Taking the sets $M = \{k : \left\|\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}e_2 - x^2\right\| \geq \varepsilon\}$ and $M_1 = \left\{k : \frac{\alpha-\gamma}{m^2} \geq \varepsilon\right\}$.

10991476, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mma.10172 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

Since $M \subseteq M_1$, therefore

$$\sum_{m\in M} a_{mk} \leq \sum_{m\in M_1} a_{mk}.$$

Thus,

$$st_A - \lim_{m\to\infty} \left\| \mathcal{P}^{\langle\gamma\rangle}_{m,\alpha} e_2 - x^2 \right\| = 0 \ \ as \ \ k \to \infty.$$

□

4 | KING-TYPE MODIFICATION

Till date many operators have been defined in approximation theory. The majority of these operators preserve e_0 and e_1. That is, for a sequence of operators $\{\mathcal{L}_m\}$, we have

$\mathcal{L}_m(e_0;x) = e_0(x)$ and $\mathcal{L}_m(e_1;x) = e_1(x)$.

The same does not hold for e_2, that is, $\mathcal{L}_m(e_2;x) \neq e_2(x)$. Rather, $\mathcal{L}_m(e_2;x)$ tends to $e_2(x)$ as m tends to infinity. The modified α-Bernstein operators (1.4) defined in this paper is an example for the same.

In 2003, King [27] gave a method to modify any sequence of operators into such a form which preserves e_0 and e_2, and for e_1 we get an expression which tends to $e_1(x)$ as m tends to infinity. Recently, Finta [28] established the unique existence of functions r_n, as defined by King, on $[0,1]$ ensuring that the associated sequence of King operators provides an approximation to every continuous function and preserves the functions e_0 and e_t, where $t \in \{2,3,\dots\}$ is fixed.

On the same grounds, we define $\tilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}(h;x)$, an alteration of $\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}(h;x)$ such that $\tilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}(1;x) = 1$, $\tilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}\left(t^2;x\right) = x^2$ and $\tilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}(t;x) = r_m(x) \to x$ as $m \to \infty$. Then by Korovkin theorem, for all continuous functions h and $x \in [0,1]$, we have

$$\lim_{m\to\infty} \widetilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}(h;x) = h(x).$$

Now, from Lemma 1.1, we have

$$\begin{aligned}
&\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}(1;x) = 1, \ \mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}(t;x) = x, \ \text{ and} \\
&\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}\left(t^2;x\right) = \frac{1}{1+\gamma}\left[x^2 + \frac{m+2(1-\alpha)}{m^2}x(1-x)\right] + x\frac{\gamma}{1+\gamma} \\
&\qquad = x^2\frac{1}{1+\gamma}\left[1 - \frac{m+2(1-\alpha)}{m^2}\right] + x\frac{1}{1+\gamma}\left[\gamma + \frac{m+2(1-\alpha)}{m^2}\right] \\
&\qquad = a_m x^2 + b_m x
\end{aligned}$$

where

$$a_m = \frac{1}{1+\gamma}\left[1 - \frac{m+2(1-\alpha)}{m^2}\right] \ \text{ and } \ b_m = \frac{1}{1+\gamma}\left[\gamma + \frac{m+2(1-\alpha)}{m^2}\right].$$

We consider the modified operators as

$$\begin{aligned}
\tilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}(h;x) = (1-\alpha)\sum_{k=0}^{m-1} {}^{m-1}C_k \frac{(r_m(x))^{[k,-\gamma]}(1-r_m(x))^{[m-k-1,-\gamma]}}{1^{[m-1,-\gamma]}} g_k \\
- \alpha \sum_{k=0}^{m} {}^{m}C_k \frac{(r_m(x))^{[k,-\gamma]}(1-r_m(x))^{[m-k,-\gamma]}}{1^{[m,-\gamma]}} h_k,
\end{aligned} \tag{4.1}$$

where $h_k = h\left(\frac{k}{m}\right)$, $g_k = \left(1 - \frac{k}{m-1}\right)h_k + \frac{k}{m-1}h_{k+1}$, and

$$r_m(x) = \frac{-b_m + \sqrt{{b_m}^2 + 4a_m x^2}}{2a_m},$$

where a_m and b_m are as defined above.

10991476, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mma.10172 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

Lemma 4.1. *For the operators (4.1), we can prove the following identities:*

(i) $\tilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}(e_0;x) = 1$,
(ii) $\tilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}(e_1;x) = r_m(x)$,
(iii) $\tilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}(e_2;x) = x^2$,

where $e_r = t^r$ *for* $r = 0, 1, 2$.

Lemma 4.2. *From the above Lemma 4.1, we can also state*

- $\tilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}(\varphi_1(t);x) = r_m(x) - x$,
- $\tilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}(\varphi_2(t);x) = -2x\tilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}(\varphi_1;x)$,

where, $\varphi_m(t) = (t-x)^m$.

Lemma 4.3 (DeVore [29])**.** *For any sequence of linear positive operators* $\{\mathcal{L}_m\}$ *and* $h \in C[a,b]$, *we have the following relations:*

$$|\mathcal{L}_m(h;x) - h(x)| \le |h(x)|\,|1 - \mathcal{L}_m(1;x)| - \omega(h,\delta)\left[\mathcal{L}_m(1;x) + \frac{1}{\delta}(\mathcal{L}_m(1;x))^{1/2}\alpha_m(x)\right], \tag{4.2}$$

where $\omega(h,\delta)$ *is the modulus of continuity of* h *and* ${\alpha_m}^2(x) = \mathcal{L}_m\left((t-x)^2;x\right)$. *Further, if* h' *is continuous, then*

$$|\mathcal{L}_m(h;x) - h(x)| \le |h(x)|\,|1 - \mathcal{L}_m(1;x)| + \left|h'(x)\right|\,|\mathcal{L}_m((t-x);x)| - \omega\left(h',\alpha_m(x)\right)\left[1 + (\mathcal{L}_m(1;x))^{1/2}\right]\alpha_m(x). \tag{4.3}$$

Let us compare the contagion form of α-Bernstein operators (1.4) and its King-type modification (4.1), as defined above, on the basis of these relations.

Theorem 4.4. *Let* $\left\{\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}\right\}$ *and* $\left\{\tilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}\right\}$ *be the sequences of linear positive operators as defined in the above equations and* $h \in C([a,b])$. *Then for* $0 \le \alpha \le 1$ *and* $\gamma = O(1/m)$, *we have the following:*

$$\left|\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}(h;x) - h(x)\right| \le \omega(h,\delta)\left[1 + \frac{1}{\delta}\sqrt{b_m x(1-x)}\right]$$
$$\left|\tilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}(h;x) - h(x)\right| \le \omega(h,\delta)\left[1 + \frac{1}{\delta}\sqrt{2x(x - r_m(x))}\right].$$

Further, if $h \in C^2([a,b])$, *then*

$$\left|\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}(h;x) - h(x)\right| \le 2\omega\left(h',\alpha_m(x)\right)\sqrt{b_m x(1-x)}$$
$$\left|\tilde{\mathcal{P}}^{\langle\gamma\rangle}_{m,\alpha}(h;x) - h(x)\right| \le 2\sqrt{2}\omega\left(h',\alpha_m(x)\right)\sqrt{x(x - r_m(x))},$$

where $\omega(h,\delta)$ *is the modulus of continuity of* h, ${\alpha_m}^2(x)$ *is the respective second central moment of that operators and* $r_m(x)$ *and* $b_m(x)$ *are as defined above.*

Proof. Using Equation (4.2), Lemma 4.1 and Lemma 4.3, with $\varphi_m(t) = (t-x)^m$, we have

$$\begin{aligned}\left|\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}(h;x) - h(x)\right| &\le |h(x)|\left|\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}(1;x) - 1\right| \\ &\quad - \omega(h,\delta)\left[\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}(1;x) + \frac{1}{\delta}\sqrt{\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}(1;x)\,\mathcal{P}^{\langle\gamma\rangle}_{m,\alpha}(\varphi_2(t);x)}\right] \\ &= \omega(h,\delta)\left[1 + \frac{1}{\delta}\sqrt{\frac{1}{1+\gamma}\left(\frac{m+2(1-\alpha)}{m^2} + \gamma\right)x(1-x)}\right] \\ &= \omega(h,\delta)\left[1 + \frac{1}{\delta}\sqrt{b_m x(1-x)}\right],\end{aligned}$$

10991476, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mma.10172 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

and

$$\begin{aligned}\left|\tilde{\mathcal{P}}_{m,\alpha}^{\langle\gamma\rangle}(h;x)-h(x)\right| &\le |h(x)|\left|\tilde{\mathcal{P}}_{m,\alpha}^{\langle\gamma\rangle}(1;x)-1\right| \\ &\quad -\omega(h,\delta)\left[\tilde{\mathcal{P}}_{m,\alpha}^{\langle\gamma\rangle}(1;x)+\frac{1}{\delta}\sqrt{\tilde{\mathcal{P}}_{m,\alpha}^{\langle\gamma\rangle}(1;x)\,\tilde{\mathcal{P}}_{m,\alpha}^{\langle\gamma\rangle}(\varphi_2(t);x)}\right] \\ &= \omega(h,\delta)\left[1+\frac{1}{\delta}\sqrt{2x(x-r_m(x))}\right].\end{aligned}$$

Now, if h' is also continuous, then we have

$$\begin{aligned}\left|\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(h;x)-h(x)\right| &\le |h(x)|\left|1-\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(1;x)\right|+\left|h'(x)\right|\left|\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(\varphi_1(t);x)\right| \\ &\quad -\omega\left(h',\alpha_m(x)\right)\left[1+\left(\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(1;x)\right)^{1/2}\right]\left(\mathcal{P}_{m,\alpha}^{\langle\gamma\rangle}(\varphi_2(t);x)\right)^{1/2} \\ &= 2\omega\left(h',\alpha_m(x)\right)\sqrt{\frac{1}{1+\gamma}\left(\frac{m+2(1-\alpha)}{m^2}+\gamma\right)x(1-x)} \\ &= 2\omega\left(h',\alpha_m(x)\right)\sqrt{b_m x(1-x)},\end{aligned}$$

and

$$\begin{aligned}\left|\tilde{\mathcal{P}}_{m,\alpha}^{\langle\gamma\rangle}(h;x)-h(x)\right| &\le |h(x)|\left|1-\tilde{\mathcal{P}}_{m,\alpha}^{\langle\gamma\rangle}(1;x)\right|+\left|h'(x)\right|\left|\tilde{\mathcal{P}}_{m,\alpha}^{\langle\gamma\rangle}(\varphi_1(t);x)\right| \\ &\quad -\omega\left(h',\alpha_m(x)\right)\left[1+\left(\tilde{\mathcal{P}}_{m,\alpha}^{\langle\gamma\rangle}(1;x)\right)^{1/2}\right]\left(\tilde{\mathcal{P}}_{m,\alpha}^{\langle\gamma\rangle}(\varphi_2(t);x)\right)^{1/2} \\ &= \left|h'(x)\right|\left|r_m(x)-x\right|+2\sqrt{2}\omega\left(h',\alpha_m(x)\right)\sqrt{x(x-r_m(x))} \\ &\le 2\sqrt{2}\omega\left(h',\alpha_m(x)\right)\sqrt{x(x-r_m(x))}.\end{aligned}$$

□

AUTHOR CONTRIBUTIONS

Kanita was involved in conceptualization, methodology and in the original draft preparation. Naokant Deo contributed to review editing, investigation and supervision. All authors read and agreed to the final manuscript.

ACKNOWLEDGEMENTS

This paper's first author would like to express gratitude to Delhi Technological University for their support. We are also indebted to the editor and the anonymous reviewers for their constructive comments and suggestions, which led to an improvement in the original manuscript.

CONFLICT OF INTEREST STATEMENT

There are no competing interests. The authors have no relevant competing interests to disclose, either financial or otherwise.

CONSENT FOR PUBLICATION

Both the authors have given their consent to the final version that has been submitted.

ORCID

Kanita https://orcid.org/0009-0002-2837-9959
Naokant Deo https://orcid.org/0000-0001-7079-4211

10991476, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mma.10172 by Delhi Technological University, Wiley Online Library on [21/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

REFERENCES

1. F. Eggenberger and G. Pólya, *Uber die Statiatik Verketteter Vorgange*, Zeitschrift für Angewandte Mathematik und Mechanik **1** (1923), 279–289.
2. D. D. Stancu, *Approximation of functions by a new class of linear polynomial operators*, Rev. Roum. Math. Pures Appl. **13** (1968), 1173–1194.
3. D. D. Stancu, *Two classes of positive linear operators*, Anal. Univ. Timiş,oara (Ser. Ştin. Mat.) **8** (1970), 213–220.
4. A. Kajla and T. Acar, *Modified α-Bernstein operators with better approximation properties*, Ann. Funct. Anal. **10** (2019), no. 4, 570–582.
5. E. W. Cheney and A. Sharma, *On a generalization of Bernstein polynomials*, Riv. Mat. Univ. Parma (1964), 77–84.
6. H. S. Jung, N. Deo, and M. Dhamija, *Pointwise approximation by Bernstein type operators in mobile interval*, Appl. Math. Comput. **214** (2014), no. 1, 683–694.
7. K. Lipi and N. Deo, *lambda-Bernstein operators based on Pólya distribution*, Numer. Funct. Ana. Optim. **44** (2023), no. 6, 529–544.
8. N. Deo, M. A. Noor, and M. A. Siddiqui, *On approximation by a class of new Bernstein type operators*, Appl. Math. Comput. **201** (2008), 604–612.
9. X. Chen, J. Tan, Z. Liu, and J. Xie, *Approximation of functions by a new family of generalized Bernstein operators*, J. Math. Anal. Appl. **450** (2017), 244–261.
10. A. Kajla and T. Acar, *Blending type approximation by generalized Bernstein-Durrmeyer type operators*, Miskolc. Math. Notes. **19** (2018), no. 1, 319–336.
11. A. Kajla and T. Acar, *Bézier-Bernstein-Durrmeyer type operators, Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A*, Matemá,ticas **114** (2020), 1–11.
12. A. Kajla, M. Mursaleen, and T. Acar, *Durrmeyer-type generalization of parametric Bernstein operators*, Symmetry **12** (2020), no. 7, 1141.
13. R. Paltanea, *Durrmeyer type operators on a simplex*, Constr. Math. Anal. **4** (2021), no. 2, 215–228.
14. T. Acar, P. N. Agrawal, and T. Neer, *Bezier variant of the Bernstein-Durrmeyer type operators*, Results Math. **72** (2017), no. 3, 1341–1358.
15. J. Bustamante, *Directs estimates a Voronovskaja-type formula for Mihesan operators*, Constr. Math Anal. **5** (2022), no. 4, 202–213.
16. N. Deo and R. Pratap, *α-Bernstein-Kantotovich operators*, Afr. Mat. **31** (2020), 609–618.
17. Q. B. Cai and X. W. Xu, *Shape-preserving properties of a new family of generalized Bernstein operators*, J. Inequal. Appl. (2018), 241.
18. R. Pratap and N. Deo, *q-Analogue of generalized Bernstein-Kantorovich operators*, Math. Anal.-I Approx. Theory ICRAPAM - **2018** (2020), 67–75.
19. S. A. Mohiuddine, T. Acar, and A. Alotaibi, *Construction of a new family of Bernstein-Kantorovich operators*, Math. Methods Appl. Sci. **40** (2017), 7749–7759.
20. H. Bohman, *On approximation of continuous and of analytic functions*, Arkiv fö,r Matematik **2** (1952), no. 1, 43–56.
21. P. P. Korovkin, *Linear Operators and Approximation Theory*, 1960. translated from the Russian edition (1959), Russian Monographs and Texts on Advances Mathematics and Physics, Gordon and Breach Publishers, Inc. New York, Hindustan Publ. Corp. (India), Delhi, 3.
22. H. Fast and Sur la convergence statistique, Colloq. Math. **2** (1951), 241–244.
23. H. Steinhauss and Sur la convergence ordinaire et la asymptotique, Colloq. Math. **2** (1951), 73–74.
24. A. Zygmund, *Trigonometric Series*, The University Press, Cambridge, London, 1959.
25. N. L. Braha, T. Mansour, M. Mursaleen, and T. Acar, *Convergence of λ-Bernstein operators via power series summability method*, J. Appl. Math. Comput. **65** (2021), 125–146.
26. A. D. Gadjiev and C. Orhan, *Some approximation theorems via statistical convergence*, Rocky Mountain J. Math. **32** (2002), no. 1, 129–138.
27. J. P. King, *Positive linear operators which preserve* x^2, Acta Math. Hungarica **99** (2003), 203–208.
28. Z. Finta, *King operators which preserve* x^j, Constr. Math. Anal. **6** (2023), no. 2, 90–101.
29. R. A. DeVore, Fundamental approximation properties of positive linear operators, *The approximation of continuous functions by positive linear operators*, vol. **293**, Springer, 1972, pp. 22–47.

How to cite this article: K. and N. Deo, *Parametric Bernstein operators based on contagion distribution*, Math. Meth. Appl. Sci. (2024), 1–14, DOI 10.1002/mma.10172.

10991476, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mma.10172 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

Journal of the Brazilian Society of Mechanical Sciences and Engineering (2024) 46:354
https://doi.org/10.1007/s40430-024-04943-6

TECHNICAL PAPER

Performance evaluation of absorption cooling system for air conditioning-based novel trigeneration system using solar energy

Shobhit Mishra[1] · **Raj Kumar Singh**[1]

Received: 12 December 2023 / Accepted: 22 April 2024
© The Author(s), under exclusive licence to The Brazilian Society of Mechanical Sciences and Engineering 2024

Abstract

This work aims to exploit the waste heat from a basic helium Brayton cycle (HBC) driven by a solar power tower (SPT). The two subsystem heat recovery steam generator (HRSG) and vapour absorption cooling system (VACS) have been implemented to recover the waste heat, respectively, for heating to generate steam and cooling for air conditioning. A comprehensive exergy and energy analysis of this proposed trigeneration system was carried out with parametric analysis. For analysis purpose Engineering Equation Solver software has been used. It was concluded that energy and exergy efficiency and net work output of the proposed trigeneration system were observed as 44.96%, 34.15%, 14,562 kW, respectively. The heating production through the HRSG was obtained 8510 kW while the cooling production by VACS as 115.10 kW. Moreover, coefficient of performance (COP) of the VACS subsystem was observed as 0.8015. Energy and exergy efficiency of the SPT operated basic HBC using the subsystems was improved by 58.98% and 12.92%, respectively. Parametric analysis revealed that the helium turbine inlet temperature and solar heliostat field efficiency much affected the trigeneration system performance.

Keywords Performance assessment · Solar power tower · Trigeneration system · Vapour absorption cooling system · HRSG · Air conditioning

Abbreviations

CR	Central receiver
HBC	Helium Brayton cycle
SPT	Solar power tower
SHE	Solution heat exchanger
SCO_2	Supercritical CO_2
VACS	Vapour absorption cooling system
ORC	Organic Rankine cycle
IHE	Intermediate heat exchanger
HC	Helium compressor
HT	Helium turbine
COP	Coefficient of performance
CPR	Compressor pressure ratio
CIT	Compressor inlet temperature
HTIT	Helium turbine inlet temperature
DNI	Direct normal irradiation (W/m^2)
HRSG	Heat recovery steam generator

List of symbols

$\dot{EX}$	Rate of exergy (kW)
h	Specific enthalpy (kJ/kg)
$\dot{W}$	Power (kW)
N_{hel}	Number of heliostats
$\dot{Q}$	Rate of heat interaction (kW)
T	Temperature (K)
$\dot{ED}$	Exergy destruction rate (kW)
A	Area (m^2)
s	Specific entropy (kJ/kg K)
$\dot{m}$	Mass flow rate (kg/s)

Subscripts

c	Condenser
e	Exit/evaporator
g	Generator
j	Particular state
rec	Receiver
ref	Reference/reflectivity
hel	Heliostat
i	Inlet
0	Dead condition
tri	Trigeneration
co	Cogeneration

Technical Editor: Guilherme Ribeiro.

✉ Shobhit Mishra
shobhit721mishra@gmail.com

[1] Department of Mechanical Engineering, Delhi Technological University, Bawana Road, Delhi 110042, India

Published online: 12 May 2024

ex Exergy
en Energy

Greek letters

η Efficiency
ε Effectiveness

1 Introduction

Demand of energy in the world is increasing day by day. It seems to increase by 50% in near future i.e. by 2040. The generation of energy devoid of carbon dioxide is essential for sustainability and future security [1, 2]. Therefore, solar energy is the amazing source of energy that can provide carbon free and secure energy in future. The power production industry is currently showing a great deal of interest in the CSP (concentrated solar power) technology due to its high-temperature production that able to meet multi-generation purpose [3]. The supercritical carbon dioxide (sCO_2) Brayton cycle is more suitable to harvest the energy at that much high temperature due to its pressure quality. It works much higher pressure that of the steam Rankine cycle (SRC) or air Brayton cycle [4]. Also it is known that if pressure at the inlet of the compressor increased that increased the net work output necessitating a low compression pressure ratio [5]. Also sCO_2 is the affordable, environmentally safe, non-toxic, non-flammable, non-corrosive and also it is chemically stable. Due to these factors sCO_2 replaces the air and other gages in Brayton cycle. Due to the high density of sCO_2, extremely compact turbo-machinery designs are possible, which is particularly appealing for applications that recover waste heat inside of automobiles. Transcritical and supercritical CO_2 cycles have drawn more attention because of their enhanced functionality for high-temperatures waste heat recovery [6]. Currently, various power cycles use the CO_2 as the working fluid because it not only increases the efficiency of the cycle but also increases efficiencies of heat exchangers and turbines [7].

In order to convert heat into electricity, SPT facilities may use conventional Rankine cycle and closed Brayton cycles depending on the temperature of the heat source, because of sCO_2 Brayton cycle's advantageous characteristics such as its compact size and effective operation research on the subject has recently grown in popularity [8–10]. Gkountas et al. [11] discovered that by using Al_2O_3 nanofluid to enhance heat transfer, the heat exchanger's length decreased by 0.9%, which in turn decreased the pressure drop in the sCO_2 system by 14%. Chai and Tassuo [12] suggests that nickel-based alloys are used at higher temperatures, while stainless steel is considered as the best for heat exchangers material up to an operating temperature of 650 °C while analysing the helium and sCO_2 as the working fluid. After modelling the SPT system and comparing five different sCO_2 cycle designs, Al-Sulaiman and Atif [13] observed that the recompression configuration of sCO_2 Brayton was observed as the best performing cycle having highest efficiency of 52%. Additionally they observed that this advance technology is very suitable to harvest the solar heat from the CSP technology like SPT. Osorio [14] investigated various sCO_2 setups for SPT plants. They found that 21% of peak efficiency was reached at the variable seasonal situations. Using a standard sCO_2 cycle by itself was less efficient for waste recovery than combining it with an ORC or another bottoming cycle. When heated by solar energy, these coupled cycles improved in terms of sustainability, security, and environmental sensitivity.

Apart from sCO_2 fluid some researcher used the helium as the working fluid in the Brayton cycle such as Zare and Hasanzadeh [15] proposed a study in which HBC harvested the solar energy from the SPT system. They used two ORCs for utilizing the waste heat from the HBC system to enhance the thermal performance of the system. Finally they observed that the system performance increased to 30%. The working fluid for the recompression cycle in Miao et al.'s [16] optimisation study was a combination of nitrous oxide and helium, rather than sCO_2. In comparison with N_2O and CO_2, they found that the suggested method raised thermal efficiency by 5.1% and 6.5%, respectively. The optimal thermal efficiency under perfect circumstances was discovered to be 42.67%. Recently, Khan et al. [17] proposed a pre-compression configuration of the HBC for power production operated by the SPT and also transcritical cycle incorporated for waste heat recovery. They observed that HBC combined cycle produced thermal and exergy efficiency of 31.59% and 33.12%, respectively. Malik et al. [18] did the thermodynamic assessment of the Brayton cycle using the helium and mixture of helium–xenon compared the performance. They observed that compressor performance increased due to employment mixture instead of alone helium by reducing the compressor stages from 16 to 3. As a result, power plant that is working at the high-temperature xenon–helium mixture is more suitable working fluid in the turbo-machinery than that of the pure helium.

The VACS has been studied as the bottoming cycle to improve the overall performance of the plant, in addition to the combined power cycles. Recently, Zhou et al. [19] designed an SPT operated combined cycle in which the VACS is used as the bottoming cycle for the HBC's intake cooling and the HBC is used as the topping cycle. Their results show that the energy efficiency of the optimized combined cycle is 14.5% higher than that of the traditional cycle. Also, Zoghi [20] proposed a hybrid system of combined cycle driven of solar and natural gas in which they used gas turbine cycle and air Brayton cycle and ORC as the power generating cycle and to utilize the waste heat they also used HRSG and LiBr–H_2O single-effect VACS system to generate

the process heat and cooling, respectively. They found that using the HRSG and VACS, exergy and energy efficiency of the basic system enhanced by 4.43% and 13.39%, respectively. Mohtaram et al. [21] did exergy and energy analysis of the hybrid refrigerating system using the ejector transcritical CO_2 system and single-effect LiBr–H_2O VACS. They found that combined system enhances COP and exergy efficiency by 17.28% and 4.8%, respectively. Mishra and Singh [22] did a comprehensive parametric evaluation of the solar-operated combined system. The alone sCO_2 cycle and the VACS system are examples of waste heat recovery systems used for cooling applications like air conditioning. They came to the conclusion that the system's energy and exergy efficiencies were 41.89% and 72.20%, respectively. Furthermore, it was discovered that the COP of VACS for cooling and heating was approximately 0.46 and 1.43, respectively.

From the above-mentioned literature it can be concluded that it is being necessary to enhance the performance of the existing energy cycles; also very limited studies that based on helium as working fluid in the Brayton cycle for solar applications were performed. As a result, in order to close this research gap, the HRSG and VACS subsystems were added to the basic HBC system in order to maximize waste heat and improve system performance. The present study aims to achieve the following objectives:

- To develop a novel SPT-driven combined cycle implementing the HRSG and VACS subsystems
- To report the effects of the adding subsystems on total performance the basic HBC system.
- To perform the comprehensive exergy and energy analysis of the proposed novel combined system.
- To conduct a thorough parametric study of the currently under investigation combined cycle.

2 System description

Figure 1 depicts the proposed solar-operated trigeneration system. The SPT system contained the heliostats and volumetric air receiver as main components. The input and geometric parameters values of the SPT system are listed in Table 1. The volumetric air receiver is taken in this SPT plant where air is considered as the heat transfer fluid (HTF). Air is abundantly available and also obtained high temperature (about 1000 °C) [15]. Also air has no temperature stability limit as in molten salt. This proposed trigeneration system is designed and analysed at the peak load conditions that is why thermal energy storage system is not part of the present study. The topping cycle HBC is started from the state 1 where helium is compressed in the helium compressor (HC), where the temperature of compressed stream is raised. The heated stream passed through the recuperator where it takes heat from the hot flue gases which lead to rise in temperature (state 3). Then, heated helium stream passes through the intermediate heat exchanger (IHE), where it takes the heat from the SPT system by HTF. It gets the maximum temperature of 850 °C at the inlet of helium turbine (HT) (state 4). This hot stream expanded in turbine to generate the power. Expanded stream has enough pressure and temperature at state 5; then, it passes through the recuperater where its heat is utilized to heat the low-temperature stream of helium of state 2. After the recuperator it still has the sufficient amount of temperature at state 6. Therefore, to utilize this heat a subsystem HRSG is implemented to generate the medium-temperature (approximate 200 °C) and high-pressure (2000 kPa) steam for industrial processing. After the HRSG (state 7) low-temperature waste heat is available. To utilize this heat a subsystem VACS implemented to produce the cooling effect for room air conditioning. At the state 8 it goes to the precooler and then again goes to the HC.

In the VACS subsystem water is separated from the LiBr–H_2O in the generator which produces the cooling effect in evaporator passing through expansion valve (EXV) and the condenser. Generated vapour in the evaporator from the state 9 and LiBr–H_2O mixed at state 10 in the absorber and again pumped back to the generator through the solution pump. By using heat from the stream leaving the generator (state 13) to warm the mixture in state 11, the solution heat exchanger (SHE) helps the system work well.

3 Modelling of proposed system

A number of assumptions were taken into account when examining the current system: (1) every component ought to function in a stable condition. (2) It is assumed that the components' pressure loss is as indicated in Table 1. (3) There was a disregard for kinetic and potential energy. (4) Assumed and reported in Table 1 are the compressor, pump, and turbine isentropic efficiencies. The trigeneration system (HBC-HRSG-VACS) and the solar sub-system comprise the full system for thermodynamic analysis. For every component in the system, a thermodynamic model is supplied. The energy-to-exergy balance for a control volume in a steady-state process is expressed using the following equations:

$$\dot{Q}_{CV} - \dot{W}_{CV} + \sum \left(\dot{m}_i h_i\right) - \sum \left(\dot{m}_e h_e\right) = 0 \tag{1}$$

$$\dot{E}D = \dot{E}X_{in} - \dot{E}X_{out} \tag{2}$$

where $\dot{E}X_{in}$ and $\dot{E}X_{out}$ denote the rate at which exergy enters and leaves the thermodynamic system, respectively, and $\dot{E}D$ denotes the destruction in exergy within the system. The heat and work interactions from the control volume are denoted

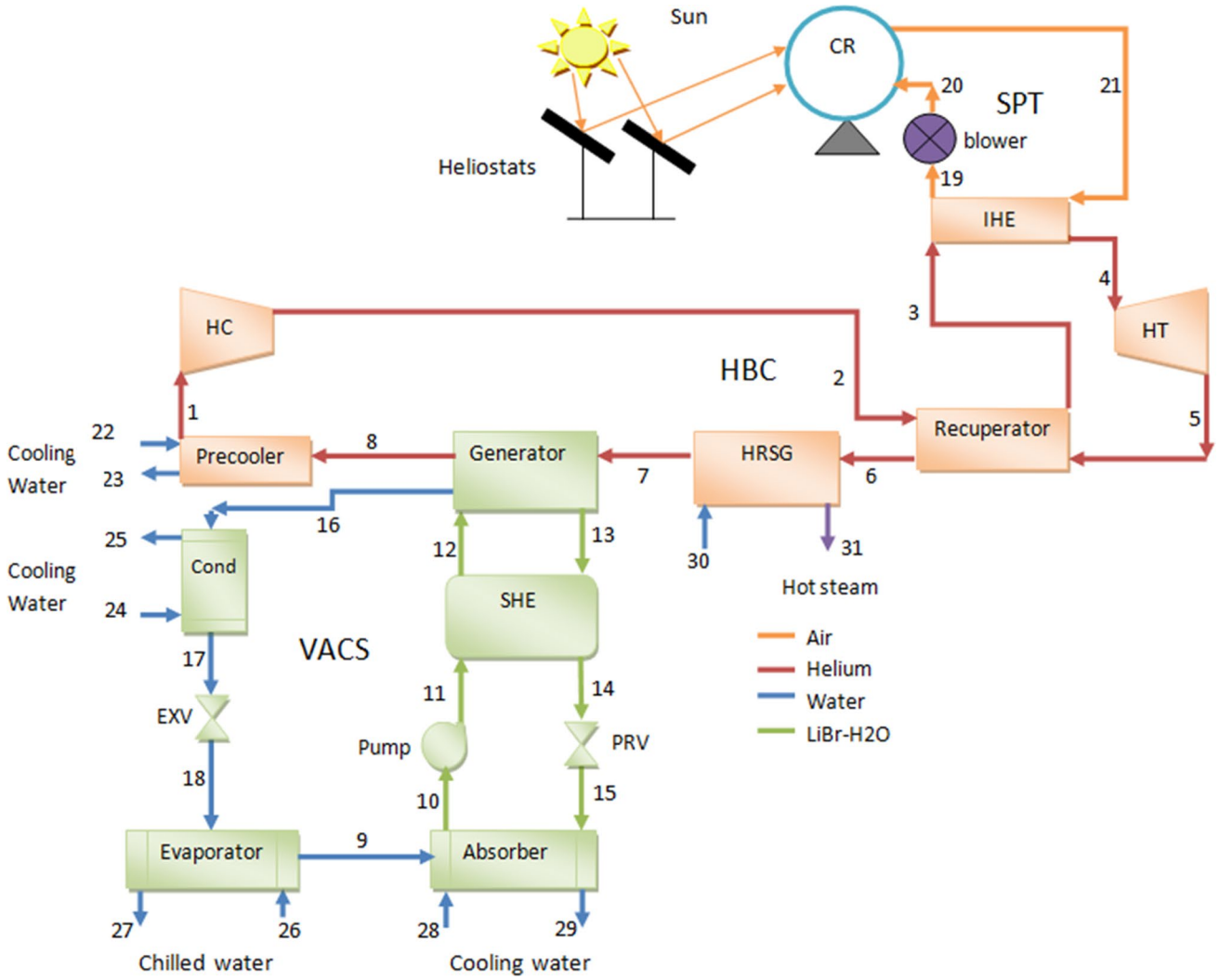


Fig. 1 Proposed SPT-based trigeneration system

by $\dot{Q}_{CV}$ and $\dot{W}_{CV}$, respectively. Without accounting for the energy associated with height, velocity, and chemical exergy (because there is no chemical concentration in the proposed system), the rate of exergy flow of a fluid stream as the physical exergy can be described as;

$$\dot{\underset{j}{\mathrm{EX}}} = \dot{m} \cdot \left[\left(h_j - h_0 \right) - T_0 \left(s_j - s_0 \right) \right] \tag{3}$$

where 'j' stands for a specific state and $\dot{\underset{j}{\mathrm{EX}}}$ represents the physical exergy at state j. The sun radiation reflects on to the thousands of heliostats through the DNI. The amount of solar radiation that actually receives by the receiver from the heliostats depends on the heliostats efficiency. Only amount of the solar energy received by the receiver, remaining amount was lost to the environment. Also the amount of the solar received depends on the geographical location i.e. amount of average DNI received. Therefore, mathematically the amount of solar thermal heat received by the receiver can be expressed as [15];

$$\dot{Q}_{\text{rec, in}} = \eta_{\text{field}} \cdot \dot{Q}_{\text{Sun}} = \eta_{\text{field}} \cdot \text{DNI} \cdot A_{\text{hel}} \cdot N_{\text{hel}} \tag{4}$$

where η_{field} is the efficiency of heliostats and expressed as [15];

$$\eta_{\text{field}} = \eta_{\cos} \cdot \eta_{s\&b} \cdot \eta_{\text{int}} \cdot \eta_{\text{att}} \cdot \eta_{\text{ref}} \tag{5}$$

where $\eta_{\text{ref}}, \eta_{\text{att}}, \eta_{\text{int}}, \eta_{\text{s\&b}}$ and$\eta_{\cos}$ denotes the efficiencies of heliostats reflectivity, atmospheric attenuation, interception efficiency, shading and blocking and cosine effect, respectively. These values have been taken directly from the operational SPT plant [15].

The receiver absorbs $\dot{Q}_{\text{rec,in}}$ amount of heat, while $\dot{Q}_{\text{rec,loss}}$ amount of heat is lost to the atmosphere due to the various modes of heat transfer. The remaining heat is transported

Table 1 Given data for base conditions

Parameters	Value
SPT system	
Heliostat field efficiency (η_{field})	0.75 [23]
Number of heliostat (N_{hel})	500 [24]
Direct normal irradiation (DNI)	850 W/m^2 [16, 24]
Heliostat reflective area (A_{hel})	9.45 × 12.84 m^2 [24]
Aperture area of receiver (A_{rec})	68.1 m^2 [24]
Efficiency of receiver (η_{rec})	0.9 [23]
HBC	
HT inlet temperature (T_4)	850 °C [16, 24]
HC isentropic efficiency (η_{HC})	0.88 [20]
Compressor pressure ratio (CPR)	2.5[25]
HT isentropic efficiency (η_{HT})	0.9[20]
HC inlet pressure (P_1)	2500 kPa [16]
Recuperator effectiveness (ε)	0.9 [16, 20]
HRSG	
Pinch point temperature difference	30 °C [21]
Steam exit pressure (P_{31})	2000 kPa [21]
VACS	
Generator temperature (T_{13})	80 °C [21]
Condenser temperature (T_{17})	35 °C [21]
Absorber temperature (T_{10})	35 °C [21]
Evaporator temperature (T_9)	6 °C [21]
SHE effectiveness $(\varepsilon_{\text{SHE}})(T_{24})(T_{25})$	0.7 [21]
Pump isentropic efficiency (η_{Pump})	0.9 [21]
Water temperature at condenser inlet	$T_{17} - 8$[21]
Water temperature at condenser outlet	$T_{17} - 3$ [21]
Chilled water temperature at evaporator outlet (T_{27})	$T_9 + 3$[21]
Water temperature at absorber inlet (T_{28})	$T_{10} - 8$[21]
Chilled water temperature at evaporator inlet (T_{26})	$T_9 + 8$[21]
Water temperature at absorber outlet (T_{29})	$T_{10} - 3$[21]
Atmospheric temperature (T_0)	25 °C [16, 25]
Atmospheric pressure (P_0)	101.3 kPa [16, 25]

to the HTF, which is air, in an amount of $\dot{Q}_{\text{rec,net}}$. Heat transport and receiver efficiency are stated as [15];

$$\eta_{\text{rec}} = \frac{\dot{Q}_{\text{rec,net}}}{\dot{Q}_{\text{rec,in}}} \tag{6}$$

$$\dot{Q}_{\text{rec,in}} = \dot{Q}_{\text{rec,net}} + \dot{Q}_{\text{rec,loss}} = \dot{m}_{\text{air}}\left(h_{30} - h_{31}\right) + \dot{Q}_{\text{rec,loss}} \tag{7}$$

when modelling the VACS subsystem, the following assumptions are taken into consideration. The system's components are considered during steady-state operations. The heat exchangers and pipes' pressure and heat loss were disregarded. LiBr–H_2O is regarded as being in equilibrium at the specific pressure and temperature in generator and absorber. Water is used to cool both the absorber and condenser. In the current analysis, saturated conditions are considered at state points 9 and 17. Table 1 lists and assumes the cooling water's pressure and temperature. Mathematical equations used for the exergy energy analysis for the each component are given in Table 2. These mathematical equations were programmed in the EES using together with the base data given in Table 1. Using the inbuilt property function in the EES the thermodynamic properties at each state were determined. Using these properties values, work and heat interaction and exergy at the each state were calculated.

3.1 The performance parameters

Net power of the proposed system $(\dot{W}_{\text{net}})$ neglecting the pump work is evaluated as:

$$\dot{W}_{\text{net}} = \dot{W}_{\text{HT}} - \dot{W}_{\text{HC}} \tag{8}$$

The COP of the VACS is defined as the ratio of cooling effect $(\dot{Q}_{\text{e}})$ at the evaporator to the heat interaction in VACS generator $(\dot{Q}_{\text{g}})$. After neglecting the pump work it can be expressed as [20];

$$\text{COP}_{\text{VACS}} = \frac{\dot{Q}_{\text{e}}}{\dot{Q}_{\text{g}}} \tag{9}$$

The energy proposed systems can be expressed as [20];

$$\eta_{\text{en, HBC}} = \frac{\dot{W}_{\text{net}}}{\dot{Q}_{\text{Sun}}} \tag{10}$$

$$\eta_{\text{en, co,h}} = \frac{\dot{W}_{\text{net}} + \dot{Q}_{\text{HRSG}}}{\dot{Q}_{\text{Sun}}} \tag{11}$$

$$\eta_{\text{en, co,c}} = \frac{\dot{W}_{\text{net}} + \dot{Q}_{\text{e}}}{\dot{Q}_{\text{Sun}}} \tag{12}$$

$$\eta_{\text{en, tri}} = \frac{\dot{W}_{\text{net}} + \dot{Q}_{\text{HRSG}} + \dot{Q}_{\text{e}}}{\dot{Q}_{\text{Sun}}} \tag{13}$$

The exergy proposed systems can be expressed as [20];

$$\eta_{\text{ex, HBC}} = \frac{\dot{W}_{\text{net}}}{\dot{Q}_{\text{Sun}} \cdot \left(1 - \frac{T_0}{T_{\text{ref,Sun}}}\right)} \tag{14}$$

Table 2 Exergy and energy balance equations for each component

Component	Energy balance equation	Exergy balance equation
Heliostat field	$\dot{Q}_{\text{rec, in}} = \eta_{\text{field}} \cdot \text{DNI} \cdot A_{\text{hel}} \cdot N_{\text{hel}}$	$\dot{Q}_{\text{Sun}} \cdot \left(1 - \frac{T_0}{T_{\text{ref,Sun}}}\right) = \dot{Q}_{\text{rec, in}} \cdot \left(1 - \frac{T_0}{T_{\text{ref,hel}}}\right) + \dot{\text{ED}}_{\text{hel}}$
Receiver	$\dot{Q}_{\text{rec, in}} = \dot{m}_{\text{air}}\left(h_{21} - h_{20}\right) + \dot{Q}_{\text{rec, loss}}$	$\dot{EX}_{20} + \dot{Q}_{\text{rec, in}} \cdot \left(1 - \frac{T_0}{T_{\text{ref,hel}}}\right) = \dot{EX}_{21} + \dot{Q}_{\text{rec, loss}} \cdot \left(1 - \frac{T_0}{T_{\text{rec}}}\right) + \dot{ED}_{\text{rec}}$
Intermediate heat exchanger	$\dot{Q}_{\text{IHE}} = \dot{m}_{\text{air}} \cdot \left(h_{21} - h_{19}\right) + \dot{m}_{\text{He}} \cdot \left(h_4 - h_3\right)$	$\dot{\text{EX}}_{21} - \dot{\text{EX}}_{19} = \dot{\text{EX}}_{4} - \dot{\text{EX}}_{3} + \dot{\text{ED}}_{\text{IHE}}$
Helium turbine	$\dot{W}_{\text{HT}} = \eta_{\text{HT}} \cdot \dot{m}_{\text{He}} \cdot \left(h_4 - h_{5\text{s}}\right)$	$\dot{\text{EX}}_{4} = \dot{\text{EX}}_{5} + \dot{W}_{\text{HT}} + \dot{\text{ED}}_{\text{HT}}$
Helium compressor	$\dot{W}_{\text{HC}} = \eta_{\text{HC}} \cdot \dot{m}_{\text{He}} \cdot \left(h_{2\text{s}} - h_1\right)$	$\dot{\text{EX}}_{1} = \dot{\text{EX}}_{2} - \dot{W}_{\text{HC}} + \dot{\text{ED}}_{\text{HC}}$
HRSG	$\dot{Q}_{\text{HRSG}} = \dot{m}_6\left(h_6 - h_7\right) = \dot{m}_{30} \cdot \left(h_{31} - h_{30}\right)$ $\Delta T_{\text{min}} = T_6 - T_{31}$	$\dot{\text{EX}}_{6} - \dot{\text{EX}}_{7} = \dot{\text{EX}}_{31} - \dot{\text{EX}}_{30} + \dot{\text{ED}}_{\text{HRSG}}$
Precooler	$\dot{m}_{\text{He}} \cdot \left(h_8 - h_1\right) = \dot{m}_{\text{water}} \cdot \left(h_{23} - h_{22}\right)$	$\dot{\text{EX}}_{8} - \dot{\text{EX}}_{1} = \dot{\text{EX}}_{22} - \dot{\text{EX}}_{23} + \dot{\text{ED}}_{\text{Preooler}}$
Recuperator	$\left(h_3 - h_2\right) = \left(h_5 - h_6\right)$ $\varepsilon_{\text{Recuperator}} = \frac{(T_3 - T_2)}{(T_5 - T_2)}$	$\dot{\text{EX}}_{5} - \dot{\text{EX}}_{6} = \dot{\text{EX}}_{3} - \dot{\text{EX}}_{2} + \dot{\text{ED}}_{\text{Recuperator}}$
Generator	$\dot{Q}_{\text{g}} = \dot{m}_{16} \cdot h_{16} + \dot{m}_{13} \cdot h_{13} - \dot{m}_{12} \cdot h_{12} \dot{m}_{12} = \dot{m}_{13} + \dot{m}_{16}$	$\dot{\text{EX}}_{13} + \dot{Q}_{\text{g}} \cdot \left(1 - \frac{T_0}{T_g}\right) = \dot{\text{EX}}_{16} + \dot{\text{EX}}_{13} + \dot{\text{ED}}_{g}$
Condenser	$\dot{Q}_c = \dot{m}_{16}\left(h_{16} - h_{17}\right) = \dot{m}_{water} \cdot \left(h_{25} - h_{24}\right)$	$\dot{\text{EX}}_{16} - \dot{\text{EX}}_{17} = \dot{Q}_{\text{c}} \cdot \left(1 - \frac{T_0}{T_c}\right) + \dot{\text{ED}}_{\text{c}}$
Expansion valve	$h_{17} = h_{18}$	$\dot{\text{EX}}_{17} - \dot{\text{EX}}_{18} = \dot{\text{ED}}_{\text{EXV}}$
Absorber	$\dot{Q}_a = \dot{m}_9 \cdot h_9 + \dot{m}_{15} \cdot h_{15} - \dot{m}_{10} \cdot h_{10}$ $\dot{m}_{10} = \dot{m}_9 + \dot{m}_{15}$ $x_{13} \cdot \dot{m}_{15} = x_{10} \cdot \dot{m}_{10}$	$\dot{\text{EX}}_{10} + \dot{Q}_a \cdot \left(1 - \frac{T_0}{T_a}\right) + \dot{\text{ED}}_{g} = \dot{\text{EX}}_{9} + \dot{\text{EX}}_{15}$
Pump	$\dot{W}_P = \dot{m}_{10}\left(h_{11} - h_{10}\right)$ $\eta_P = \frac{(h_{11\text{s}} - h_{10})}{(h_{11} - h_{10})}$	$\dot{\text{EX}}_{10} = \dot{\text{EX}}_{11} - \dot{W}_{\text{Pump}} + \dot{\text{ED}}_{\text{Pump}}$
Solution heat exchanger	$\dot{m}_{13} \cdot h_{13} - \dot{m}_{14} \cdot h_{14} = \dot{m}_{12} \cdot h_{12} - \dot{m}_{11} \cdot h_{11}$ $\varepsilon_{\text{SHE}} = \frac{(T_{13} - T_{14})}{(T_{13} - T_{11})}$	$\dot{\text{EX}}_{13} + \dot{\text{EX}}_{11} = \dot{\text{EX}}_{12} + \dot{\text{EX}}_{14} + \dot{\text{ED}}_{\text{SHE}}$
Pressure reducing valve	$h_{14} = h_{15}$	$\dot{\text{EX}}_{14} - \dot{\text{EX}}_{15} = \dot{\text{ED}}_{\text{PRV}}$
Evaporator	$\dot{Q}_{\text{e}} = \dot{m}_9\left(h_9 - h_{18}\right) = \dot{m}_{\text{chilled water}} \cdot \left(h_{26} - h_{27}\right).$	$\dot{\text{EX}}_{9} - \dot{\text{EX}}_{18} = \dot{Q}_{\text{e}} \cdot \left(1 - \frac{T_0}{T_{\text{e}}}\right) + \dot{\text{ED}}_{\text{e}}$

$$\eta_{\text{ex, co,h}} = \frac{\dot{W}_{\text{net}} + \dot{EX}_{\text{HRSG}}}{\dot{Q}_{\text{Sun}} \cdot \left(1 - \frac{T_0}{T_{\text{ref,Sun}}}\right)} \tag{15}$$

$$\eta_{\text{ex, co,c}} = \frac{\dot{W}_{\text{net}} + \dot{EX}_{\text{e}}}{\dot{Q}_{\text{Sun}} \cdot \left(1 - \frac{T_0}{T_{\text{ref,Sun}}}\right)} \tag{16}$$

$$\eta_{\text{ex, tri}} = \frac{\dot{W}_{\text{net}} + \dot{EX}_{\text{HRSG}} + \dot{EX}_{\text{e}}}{\dot{Q}_{\text{Sun}} \cdot \left(1 - \frac{T_0}{T_{\text{Sun}}}\right)} \tag{17}$$

where T_{Sun} denotes the apparent Sun's temperature and taken as 4500 K [15, 26].

4 System validation

Results of the developed models such as HBC and the VACS have been validated with the existing literature to check the correctness of the mathematical modelling procedure. The results of the alone HBC system have been validated with the existing literature Zare et al. [25]. It can be seen from Fig. 2 that very less difference in the data has been reported i.e. 0.25%. Therefore, it is acceptable. The correctness of the subsystem VACS is evaluated under the identical input conditions using research findings from the Ref. [27]. To

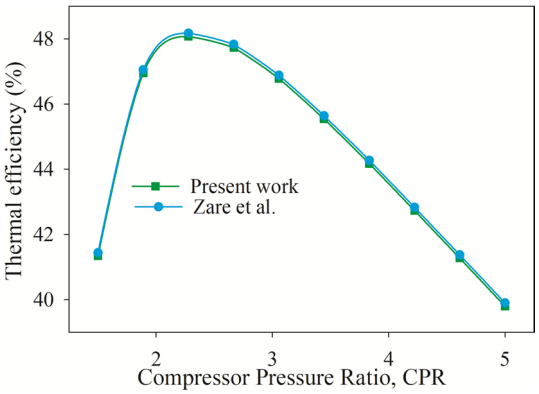


Fig. 2 Helium Brayton cycle validation

Table 3 Validation of VACS model

Results T (°C)	Present study	Ref. [27]	Deviation (%)
T_9	5.69	5.67	0.35
T_{10}	33.12	33.14	−0.06
T_{11}	33.21	33.15	0.18
T_{12}	59.25	59.25	0
T_{13}	80.87	81.39	−0.63
T_{14}	50.51	50.51	0
T_{15}	44.12	43.99	0.29
T_{16}	80.87	81.39	−0.63
T_{17}	38.42	38.4	0.05
T_{18}	5.69	5.67	0.35

conduct an adequate comparison, the produced model is appropriately modified. On temperature at all state points, the results of Ref. [27] and the results of the current model are compared. The largest difference was discovered to be − 0.63%, a negligible amount that exhibits high accuracy, according to Table 3's comparison data. The model that was developed as a result has been verified and may be trusted.

5 Results and discussion

This paper presents the novel trigeneration system using new subsystem HRSG and VACS to produce the extra process heat and cooling for the air conditioning, respectively. The results were calculated using the EES software.

5.1 Results at base conditions

First, as shown in Table 1, the trigeneration system's results were computed under the base conditions. Exergy and energy analysis parameters at the base conditions of the proposed trigeneration are indicated in Table 4. At the inlet of the HT the helium temperature is 850 °C. The power output from the trigeneration system was obtained as 14,562 kW. It was worth mentioning that exergy and energy efficiency of the trigeneration system was observed as 34.15% and 44.96%, respectively. The heating production through the HRSG was obtained 8510 kW while the cooling production by VACS as 115.10 kW. The exergy and energy efficiency of the basic HBC system was obtained as 30.24% and 28.28%, respectively, which can be improved to 34.15% and 44.96% and, respectively, by using the subsystems accordingly as can be observed in Table 4. As the exergy and energy production in the HRSG system is more than the VACS cooling system, therefore the energy and exergy efficiency of the cogeneration for heating is more than that of cooling. The energy and exergy of cogeneration for heating were observed as 44.74% and 34.14%, respectively, and corresponding values for the cogeneration for cooling were 30.26% and 28.46%, respectively. Therefore, it was observed that the improvement in the exergy and energy efficiency of the basic HBC using the subsystems is 12.92% and 58.98%, respectively. Apart from the exergy and energy efficiency of the proposed system, the COP of the VACS subsystem is found as 0.8015.

Table 4 Output results of the SPT-based trigeneration system

Parameters	Values
$\eta_{\text{en, HBC}}$	28.24%
$\eta_{\text{ex, HBC}}$	30.24%
$\eta_{\text{en, tri}}$	44.96%
$\eta_{\text{ex, tri}}$	34.15%
$\eta_{\text{en, co,h}}$	44.74%
$\eta_{\text{ex, co,h}}$	34.14%
$\eta_{\text{en, co,c}}$	28.46%
$\eta_{\text{ex, co,c}}$	30.26%
$\dot{m}_{\text{Air}}$	54.34 kg/s
$\dot{m}_{\text{Helium}}$	20.49 kg/s
$\dot{m}_{\text{Steam}}$	12.84 kg/s
$\dot{Q}_{\text{HRSG}}$	8510 kW
$\dot{Q}_{\text{e}}$	115.1 kW
$\dot{W}_{\text{net}}$	14,562 kW

The effects on the exergy destruction of the combined cycle of the subsystem were also investigated. The basic HBC, cogeneration for heating, cogeneration for cooling, and trigeneration systems are responsible for the exergy destruction values 28,911 kW, 30,064 kW, 29,089.96 kW, and 30,242.96 kW, respectively. Therefore, it can be said that implementing the both subsystems in the basic SPT operated HBC system, exergy destruction is improved by 4.6% only. Exergy destruction has additionally been determined and is presented in Table 5 for each part of the proposed trigeneration system. It was seen from the exergy analysis that

Table 5 Componentwise exergy destruction and exergy efficiency

Component	Exergy efficiency (%)	Exergy destruction (kW)
Heliostats	64.27	12,038
Solar receiver	58.28	10,713
Intermediate heat exchanger	93.01	1774
Helium turbine	96.1	1258
Recuperator	93.35	1308
HRSG	.92	1153
Precooler	27.83	499.3
Helium compressor	91.98	1320
Generator	84.25	7.027
Condenser	42.81	2.416
Expansion valve	3.071	0.3068
Evaporator	68.76	15.67
Absorber	35.98	49.69
Pump	5.23	0.11
Solution heat exchanger	60.5	51.18
Pressure reducing valve	56.87	49.68

heliostats field was responsible for highest exergy destruction. It is due to the solar irradiation that was received from the high-temperature source i.e. 4500 K. This energy has a significant irreversibility because it is received by the receiver at a temperature of about 1125 °C [15]. The energy efficiency of the heliostat was determined to be 64.27%. The IHE's helium and HTF temperatures change very little since there is not any combustion, which is the primary factor in conventional power systems' irreversibility. Consequently, if the SPT is ignored, the combined cycle has a high energy efficiency of almost 66.31%. However, the exergy efficiency of the power plant is quite low, at 34.15%, due to the significant exergy destructions in the receiver and heliostat field. The componentwise exergy destruction rate and the exergy efficiency were analysed, and results are shown in Table 5. Highest and lowest exergy destruction component of the proposed plant were heliostats and pump values 12,038 kW (39.80% of total exergy destruction) and 0.11 kW, respectively. Second highest exergy destruction component is receiver in which exergy destruction occurred as 10,713 kW contributing 35.42% of the total exergy destruction. The helium turbine has the best exergy efficiency (96.1%) of all the system's components when combined.

5.2 Parametric analysis

This section presents the results of a parametric study of the proposed system to examine the impact of the primary parameter on the output. As indicated in Table 1, other parameters were held constant while examining the impact of one variable.

5.2.1 Effect of the helium turbine inlet temperature

The exergy, energy efficiency, heating load, and net work output of the trigeneration system were continuously increased in order to improve the helium turbine inlet temperature (HTIT). The enthalpy differences within the turbine would increase as the HTIT grew, as a result net work output increases. Therefore, energy efficiency and exergy efficiency were also increased. Also the as HTIT increased, the helium gas temperature at the inlet of the HRSG was increased consequently heating production increased as can be observed in Fig. 3. Energy and exergy efficiency was increased from 42.35 to 45.67% and 26.25 to 36.29%, respectively, as HTIT increased from 600 to 950 °C. However, net work output and heating production were improved by 52.31% and 47.63%, respectively, as HTIT increased from 600 to 950 °C. On the hand, the cooling production or cooling effect did not affected by the HTIT, while COP of VACS decreased with the HTIT. Increased HTIT improved the generator

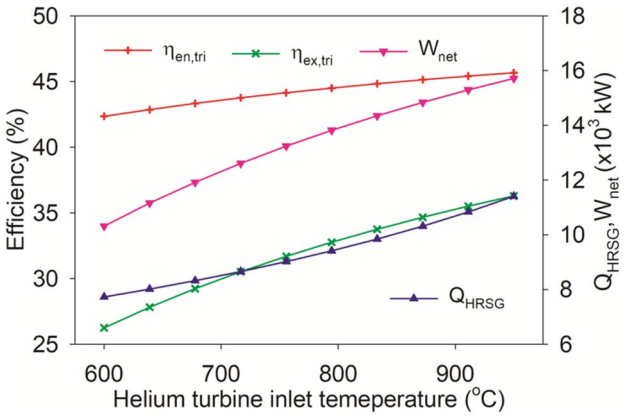


Fig. 3 Trigeneration system performance variations with HTIT

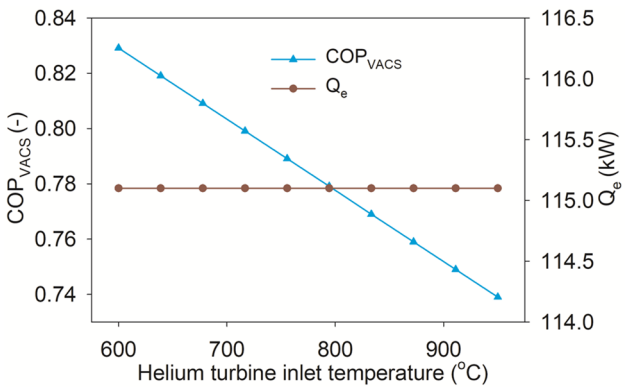


Fig. 4 COP and cooling production variation with HTIT

temperature that leads to the improvement in the generator load however no effect on the cooling production that leads to the decrement in the COP of the VACS. COP decreased by 10.86% when HTIT increased from 600 to 950 °C as revealed in Fig. 4. Therefore it can be said that the HTIT is the main critical parameters that much affected the trigeneration system performance.

5.2.2 Effect of the compressor pressure ratio

Figure 5 displays that the net power output and exergy efficiency of the trigeneration system the first increased and then decreased. They have inflation point at which they attained theirs maximum values. Therefore the optimal point is 2.278 of CPR. This pattern can be discussed by remembering the optimum CPR values: when the CPR climbed, the expansion and compression work both increased. However, in this area, the expansion work improvement rate predominated over the compression work improvement rate. Consequently, network output will increase. As a result, the combined cycle energy efficiency increases. After the CPR results' optimum values were displayed, the reverse was done. The exergy efficiency and the net work output were attained its values 34.89% and 14,610 kW, respectively, at the optimal values of CPR. However, the energy efficiency and the heating production were decreased with the CPR. Energy efficiency and the heating production decreased by 35.33% and the 59.42%, respectively, as the CPR increased from 1.5 to 5. It can be seen the improvement in the CPR values highly affected the energy efficiency and the net work output. Figure 6 displays that cooling production and COP were slightly increased with CPR. As CPR increased the output temperature at the inlet of generator decreased, it means that COP is decreased. The COP and cooling production slightly decreased by 1.05% and 8.37%, respectively.

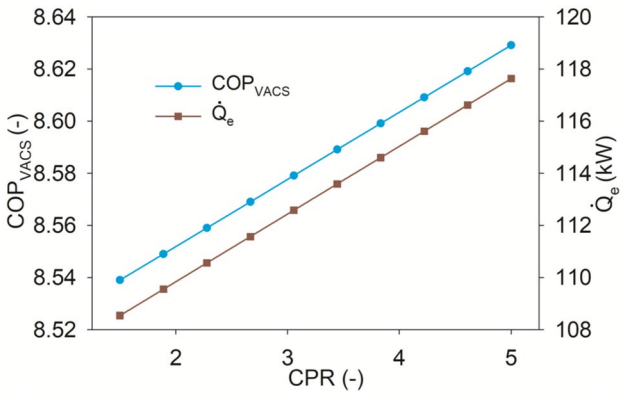


Fig. 6 COP and cooling production variation with CPR

5.2.3 Effect of the HC inlet temperature

Figure 7 shows that as the compressor inlet temperature (CIT) increased, the energy, exergy efficiency, and net work production somewhat dropped. It can be accounted for by noting that when CIT grew, across the compressor the enthalpy difference also increased, improving compressor performance. Net work output dropped as a result. Consequently, efficiencies of the trigeneration dropped with CIT. The exergy efficiency, net work output, and energy efficiency were decreased by 2.14%, 3.16%, and 0.18%, and, respectively, as CIT increased from 25 to 35 °C. However, the heating production slightly increased with the CIT. Heating production increased by 0.45% only as CIT increased from 25 to 35 °C. Therefore, heating production and energy efficiency were not much affected by CIT. On other hand the cooling production and COP were increased by the CIT. Cooling production ($\dot{Q}_e$) increased from 110.539 to 119.6291 kW as CIT increased from 25 to 35 °C as shown in Fig. 8 which is the result of the compressor removing a lot of heat from the generator side

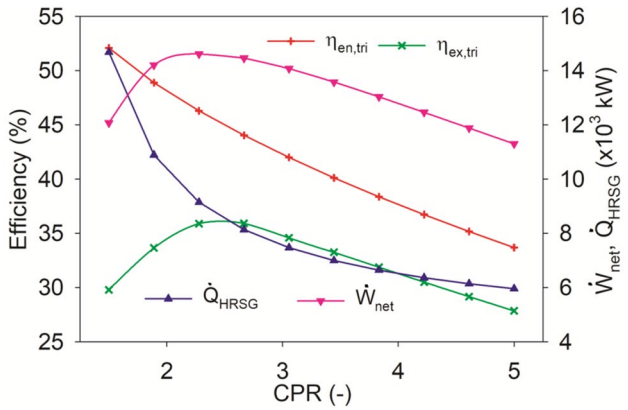


Fig. 5 Trigeneration performances variation with CPR

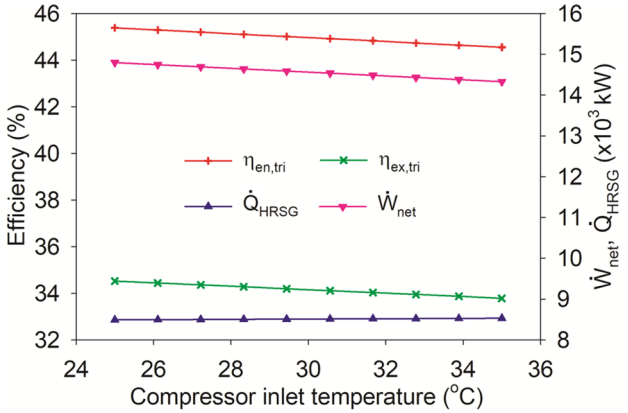


Fig. 7 Trigeneration performance variations with the CIT

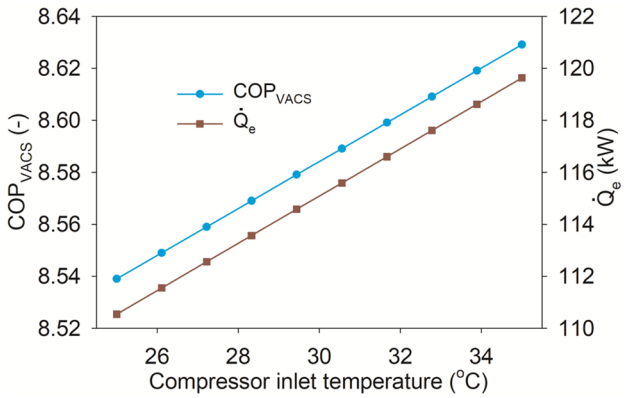


Fig. 8 COP and cooling production variation with CIT

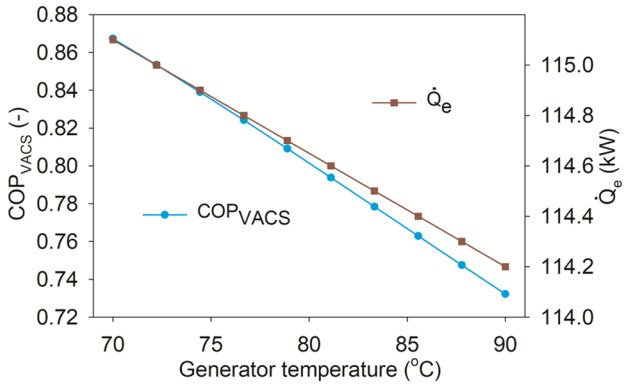


Fig. 10 COP and cooling production variation with generator temperature

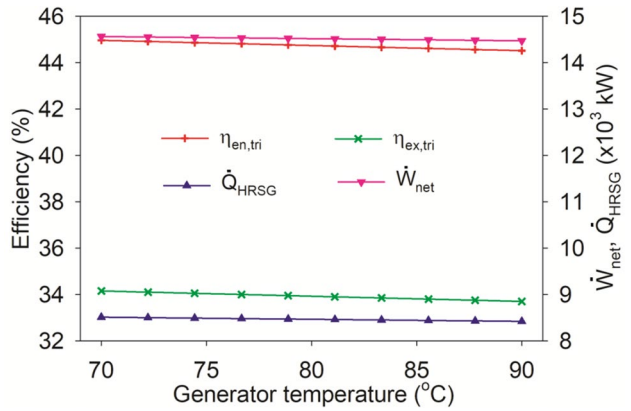


Fig. 9 Trigeneration system performance variations with generator temperature

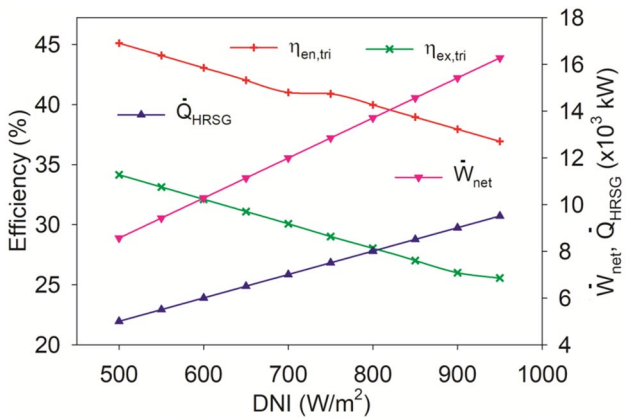


Fig. 11 Trigeneration system performance variations with the DNI

by slowing down the heat transfer from the heating coil to the generator as CIT tends to rise, which causes generator temperature to fall concurrently. Also COP rises from 8.539 to 8.6291. As the COP strongly depends on $\dot{Q}_e$, therefore reason for increasing COP is also same as discussed above for improvement in $\dot{Q}_e$.

5.2.4 Effect of the generator temperature

In this part, the effect of bottoming cycle parameters, such as VACS, on the functionality of the trigeneration system cycle, as opposed to topping cycle parameters, has also been covered. Figure 9 demonstrates that the exergy, energy efficiency of the trigeneration system, and heating output all slightly decrease from 34.15 to 33.7%, 44.96 to 44.51%, and 8510 to 8410 kW, respectively, as the generator temperature climbs from 70 to 90 °C. But as shown in Fig. 10, the COP_{VACS} falls from 0.8673 to 0.7322. The reason for the declining efficiency and COP is the rising temperature of the generator's flowing refrigerant, which can increase the average temperature of the condenser and absorber unit in the VACS and cause more heat transfer losses or energy destruction as the generator's temperature rises. As a result, cooling production drops from 115.1 to 114.2 kW and net output work slightly declines from 14,562 to 14,470 kW.

5.2.5 Effect of the solar irradiation

Figure 11 shows how changes in DNI affect the output parameters of the system. According to Eqs. (13) and (17), as DNI increases from 500 to 950 W/m^2, the receiver's outlet air temperature rises, increasing the total solar energy input. This causes a decrease in energy and exergy efficiency of 18.12% and 25.17%, respectively. However the inlet condition of the turbine improves and the net output work increases. Additionally, heating productions experience improvement from 5006 to 9511 kW,

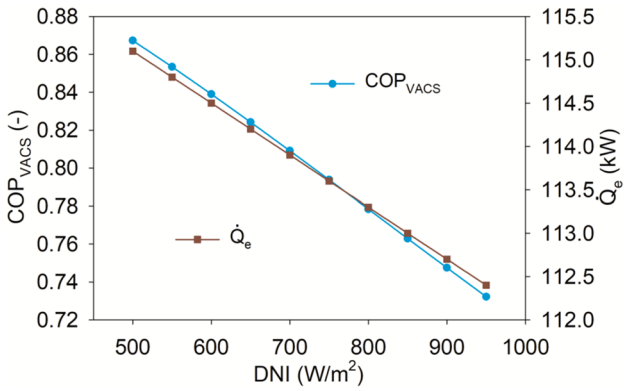


Fig. 12 COP and cooling production variation with DNI

respectively. Apart from this the COP_{VACS} and cooling production decreased with the DNI. The temperature at the generator increased that lowered the flow rate of water through evaporator as refrigerant as the DNI increased. Therefore the cooling production decreases consequently which leads to the COP_{VACS}. Cooling production and the COP_{VACS} decreased from 0.8673 to 0.7322 and 115.1 to 112.4 kW, respectively, as DNI improves from 500 to 950W/m^2, respectively, as can be observed in Fig. 12.

5.2.6 Effect of the efficiency of heliostat field

Another important factor influencing the overall cycle performance is the heliostat field efficiency. Exergy efficiency, energy efficiency, net output work, and the heating production increased with the heliostat efficiency. The performance of the under consideration trigeneration system is shown in relation to the effects of heliostat field efficiency in Fig. 13. It is evident that rising has a positive impact on system performance. For this reason, when field efficiency rises, energy

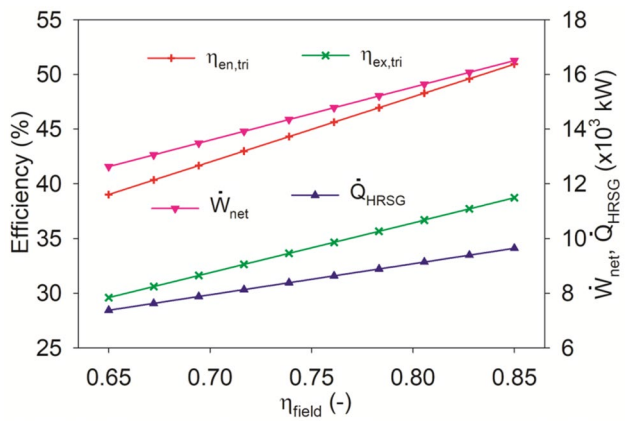


Fig. 13 Trigeneration system performance variations with heliostat field efficiency

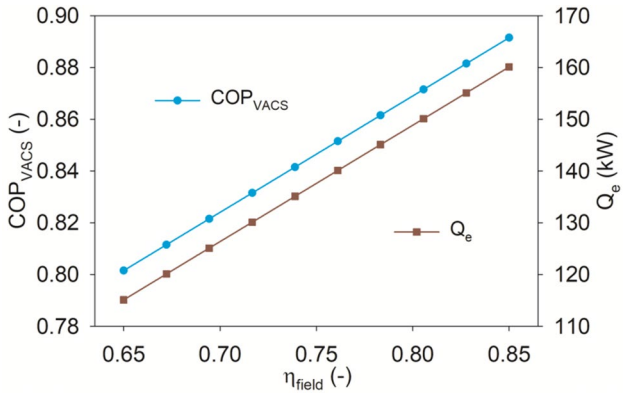


Fig. 14 COP and cooling production variation with heliostat field efficiency

loss decreases and energy transmission to the receiver improves. The much energy the receiver can absorb, the faster the flow rate of HTF and helium, which boosts power production as depicted in Fig. 13. Given the direct correlation between net output work and exergy efficiency, exergy efficiency improves along with net output work. As field efficiency improved from 0.65 to 0.85, energy efficiency, heating production, net output work, and energy efficiency grew from 29.6 to 38.71%, 39 to 50.93%, 7376 to 9645 kW, and 12,621 to 16,504 kW, respectively. Figure 14 shows that when field efficiency climbed from 0.65 to 0.85, cooling production and COP also increased, from 115.1 to 160.2 kW and 0.8015 to 0.8915, respectively.

6 Conclusions

A novel solar-operated trigeneration system has been proposed in this study where the solar energy from the SPT system is harvested using HBC in order to produce power. Additionally, to recover the waste heat from HBC the HRSG and VACS two subsystems were used to generate the heating and cooling production, respectively. Further, a comprehensive energy and exergy analysis has been conducted of the proposed trigeneration system. The following conclusions were made from this study:

- The net work output, energy, and exergy efficiency the proposed trigeneration system were observed as 14,562 kW, 34.15%, and 44.96%, respectively. The heating production through the HRSG was obtained 8510 kW while the cooling production by VACS as 115.10 kW. Additionally, COP of the VACS subsystem is found as 0.8015.

- Energy and exergy efficiency of the SPT-operated basic HBC using the subsystems is improved by 58.98% and 12.92%, respectively.
- The exergy and energy efficiency, net work output, and heating production of the proposed trigeneration system improved with the HTIT, while COP of VACS decreased and cooling production did not affected much.
- CPR and CIT much affected the proposed system's energy and exergy efficiency, net work output, and heating production however the COP and cooling production affected marginally.
- While generator temperature reduced the COP and cooling production greatly on other hand, it is marginally affected the exergy and energy efficiency, net work output, and heating production of the proposed trigeneration system.
- Solar irradiation decreased the exergy and energy efficiency and improved the net work output and heating production, while COP and cooling production marginally decreased with DNI.
- However, heliostat field efficiency has positive impact on the performance of the proposed trigeneration system.
- The cost effective analysis of the proposed novel trigeneration system is scope of the future study.

Declarations

Conflict of interest The authors confirm that they do not now have any recognized economic or interpersonal disputes that could have affected the findings of this study.

References

1. Devarajan MM, Kumaraguruparan G (2024) Thermal analysis of solar panel with phase change material: experimental and numerical study. J Braz Soc Mech Sci Eng 46(4):222. https://doi.org/10.1007/s40430-024-04792-3
2. Khan Y, Raman R, Rashidi MM et al (2023) Thermodynamic analysis and experimental investigation of the water spray cooling of photovoltaic solar panels. J Therm Anal Calorim 148(12):5591–5602. https://doi.org/10.1007/s10973-023-12119-y
3. Khan Y, Mishra RS (2021) Thermo-economic analysis of the combined solar-based pre-compression supercritical CO_2 cycle and organic Rankine cycle using ultra-low GWP fluids. Therm Sci Eng Progress 23:100925
4. Geneiting T, Lerch S, Schulz B (2023) Probabilistic solar forecasting: benchmarks, post-processing, verification. Sol Energy 252:72–80. https://doi.org/10.1016/j.solener.2022.12.054
5. Khan Y, Mishra RS, Singh AP (2024) Performance comparison of organic Rankine cycles integrated with solar based combined cycle: a thermodynamic and exergoenvironmental analysis. Proc Inst Mech Eng Part C: J Mech Eng Sci 238(1):233–248. https://doi.org/10.1177/09544062231167069
6. Al-Sulaiman FA (2014) Exergy analysis of parabolic trough solar collectors integrated with combined steam and organic Rankine cycles. Energy Convers Manage 77:441–449. https://doi.org/10.1016/j.enconman.2013.10.013
7. Khan Y, Caliskan H, Hong H (2023) A comparative study of combined cycles for concentrated solar power for efficient power generation using low global warming potential (GWP) fluids to reduce environmental effects. IET Renew Power Gener 17(16):3741–3754. https://doi.org/10.1049/rpg2.12866
8. Zhang Y, Shi X, Zhang H et al (2022) Review on deep learning applications in frequency analysis and control of modern power system. Int J Electr Power Energy Syst 136:107744. https://doi.org/10.1016/j.ijepes.2021.107744
9. Li H, Hou K, Xu X et al (2022) Probabilistic energy flow calculation for regional integrated energy system considering cross-system failures. Appl Energy 308:118326
10. Du Q, Wang Y, Zhang D (2021) Numerical investigation on the performance of impeller back seal configuration in a supercritical CO2 radial-inflow turbine. Proc Inst Mech Engin Part C: J Mech Eng Sci 235(22):6152–6169. https://doi.org/10.1177/09544062211009564
11. Gkountas AA, Benos LT, Nikas KS et al (2020) Heat transfer improvement by an Al2O3-water nanofluid coolant in printed-circuit heat exchangers of supercritical CO_2 Brayton cycle. Therm Sci Eng Progr 20:100694
12. Chai L, Tassou SA (2020) A review of printed circuit heat exchangers for helium and supercritical CO_2 Brayton cycles. Therm Sci Eng Prog 18:100543
13. Al-Sulaiman F, Atif M (2015) Performance comparison of different supercritical carbon dioxide Brayton cycles integrated with a solar power tower. Energy 82:61–71. https://doi.org/10.1016/j.energy.2014.12.070
14. Osorio JD, Hovsapian R, Ordonez JC (2016) Dynamic analysis of concentrated solar supercritical CO2-based power generation closed-loop cycle. Appl Therm Eng 93:920–934. https://doi.org/10.1016/j.applthermaleng.2015.10.039
15. Zare V, Hasanzadeh M (2016) Energy and exergy analysis of a closed brayton cycle-based combined cycle for solar power tower plants. Energy Convers Manage 128:227–237. https://doi.org/10.1016/j.enconman.2016.09.080
16. Miao X, Zhang H, Sun W et al (2022) Optimization of a recompression supercritical nitrous oxide and helium Brayton cycle for space nuclear system. Energy 242:123023
17. Khan Y, Raman R, Said Z et al (2024) Sustainable power generation through solar-driven integration of brayton and transcritical co_2 cycles: a comprehensive 3e (energy, exergy, and exergoenvironmental) evaluation. Global Chall 8(2):2300223. https://doi.org/10.1002/gch2.202300223
18. Malik A, Zheng Q, Lin A (2019) The design and performance analysis of highly loaded compressor of closed brayton cycle HTGR power plant with helium xenon gas mixture as working fluid. Prog Nucl Energy 117:103084
19. Zhou J, Ali MA, Zeki FM et al (2023) Thermoeconomic investigation and multi-objective optimization of a novel efficient solar tower power plant based on supercritical Brayton cycle with inlet cooling. Therm Sci Eng Prog 39:101679
20. Zoghi M, Habibi H, Chitsaz A et al (2021) Exergoeconomic and environmental analyses of a novel trigeneration system based on combined gas turbine-air bottoming cycle with hybridization of solar power tower and natural gas combustion. Appl Therm Eng 188:116610
21. Mohtaram S, Wu W, Castellanos HG et al (2023) Enhancing energy efficiency and sustainability in ejector expansion transcritical CO_2 and lithium bromide water vapour absorption refrigeration systems. Therm Sci and Eng Prog 43:101983

22. Mishra RS, Singh H (2018) Detailed parametric analysis of solar driven supercritical CO_2 based combined cycle for power generation, cooling and heating effect by vapor absorption refrigeration as a bottoming cycle. Therm Sci Eng Prog 8:397–410. https://doi.org/10.1016/j.tsep.2018.09.013
23. Farges O, B´ezian JJ, Hafi El M (2018) Global optimization of solar power tower systems using a monte carlo algorithm: application to a redesign of the PS10 solar thermal power plant. Renew Energy 119:345–353. https://doi.org/10.1016/j.renene.2017.12.028
24. Sun L, Wang D, Xie Y (2021) Energy, exergy and exergoeconomic analysis of two supercritical CO_2 cycles for waste heat recovery of gas turbine. Appl Therm Eng 196:117337
25. Zare V, Mahmoudi SMS, Yari M (2015) On the exergoeconomic assessment of employing Kalina cycle for GT-MHR waste heat utilization. Energy Convers Manag 90:364–374. https://doi.org/10.1016/j.enconman.2014.11.039
26. Xu C, Wang Z, Li X et al (2011) Energy and exergy analysis of solar power tower plants. Appl Therm Eng 31(17–18):3904–3913. https://doi.org/10.1016/j.applthermaleng.2011.07.038
27. Kerme ED et al (2017) Energetic and exergetic analysis of solar-powered lithium bromide-water absorption cooling system. J Cleaner Prod 151:60–73. https://doi.org/10.1016/j.jclepro.2017.03.060

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.

Positional Analysis and Comparison of Different Charge Generation Layer based Blue OLEDs

Sugandha Yadav
Electronics and Communication Engineering
Delhi Technological University
Delhi, India
sugandhayadav555@gmail.com

Poornima Mittal
Electronics and Communication Engineering
Delhi Technological University
Delhi, India
poornimamittal@dtu.ac.in

Shubham Negi
Electronics and Communication Engineering
Chandigarh University
Mohali, India
shubhamnegi0192@gmail.com

2024 2nd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT) | 979-8-3503-7284-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/DICCT61038.2024.10532975

***Abstract*—This paper includes three blue OLED devices which are D_1, D_2 and D_3 where D_1 is a non-CGL (charge generation layer) device, and D_2 and D_3 are CGL based devices. In all the devices, the CGL contains two materials 1,4,5,8,9,11-hexaazatriphenylene-hexacarbonitrile (HAT-CN) and 1,1-bis[(di-4-tolyamino)phenyl)]cyclohexane (TAPC) which act as electron and hole generators, respectively. As CGL provides additional charge carriers to D_2 and D_3 devices, these devices exhibit noteworthy improvement on the basis of luminescence and current. Among all the three devices, D_3 shows the best performance. The luminescence of D_3 is 2112.2 cd/m^2 which is 10 times and 1.3 times enhanced in comparison with D_1 and D_2, correspondingly. Additionally, the current of D_3 is 0.256 A which is 10 times and 1.4 times higher on comparing with D_1 and D_2, respectively. Moreover, the internal analysis of the devices is also performed using cutline and structure-based analysis to understand the in-depth performance improvements of D_1, D_2 and D_3. The reported CGL based devices are useful for bio-medical and sensor applications.**

Keywords—Luminescence, Organic LED current, charge generation layer (CGL), cutline analysis

I. INTRODUCTION

Over the past few years, full color displays with organic materials are gaining lot of attention because of their salient features like easy fabrication, light weight, wide viewing angle, better lifetime and most important flexibility. Despite having these attractive features, the OLEDs (organic light emitting diode) have made their strong presence in the display industry. On the other hands, there is need of refinement in these displays in terms of luminesces, current and lifetime. Previously, to improve these parameters various layers like EIL/HIL (electron/hole injection layer), ETL/HTL (electron/hole transport), HBL (hole blocking layer) etc. were being used widely [1-4]. These layers helped a lot for enhancing device performance to very good extent. Nevertheless, in recent times, some novel have been introduced to further improve the performance of OLED devices. These novel layers are mixed interlayers [5], spacer layers [6], and charge generation layers (CGL) [7]. To make mixed interlayers, different materials are mixed in a ratio and a single layer formed which can be used between the layers of OLED. In case of the CGL, one or two layers which are used to generate additional free electrons/holes. The CGL layers generally work as hole generator/electron generator [8-10].

In this paper, the three non-CGL and CGL based devices (D_1, D_2 & D_3) are analyzed and compared together in terms of sevral performance parameters such as luminescence and current. Herein, CGL is the mixture of two materials, one is TAPC and other is HAT-CN. TAPC plays as hole generator and HAT-CN works as electron generator. Moreover, a cutline analysis is performed here to study in-depth performance variation of all the devices. Additionally, thickness optimization of the best device is also included.

II. PARAMETERS EXTRACTION AND CHARACTERISTIC ANALYSIS OF BLUE OLED

First of all, a reference device D_1 is validated to set few boundary conditions of the modified devices D_2 and D_3 in terms of materials and thickness. The reference device represents a multi-layered blue OLED device which does not contain any charge generation layer. There are basically seven layers of cathode, ETL, EML, HTL and anode. The device contains the total length and width as 400 and 1000 μm, correspondingly. The materials and thickness of the layers are highlighted in Table I. Fig. 1 is depicting the basic structure of D_1.

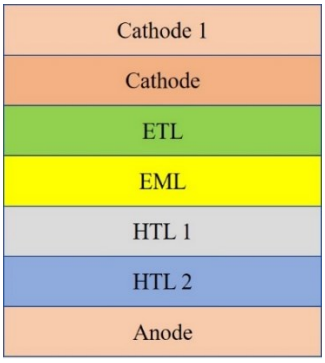


Fig. 1. Reference Blue OLED Device D_1

TABLE I. MATERIALS AND THEIR THICKNESS FOR D_1 DEVICE

Layer	Material	Thickness (nm)
Cathode1	Al	150
Cathode	LiF	1
ETL	BPhen	30
EML	DPVBi & BCzVBi	20
HTL2	NPB	60
HTL1	PEDOT:PSS	90
Anode	ITO	20

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:41:47 UTC from IEEE Xplore. Restrictions apply.

To validate the reference device, Langevin's recombination rate model is utilized in Silvaco Atlas 2-D device simulator. The equation for the model is given below:

$$R_L(n, p) = r_l(x, y, t)(np - n_i^2) \quad (1)$$

Here, r_l, p, n and n_i are for the recombination rate coefficient, hole concentration, electron concentration and intrinsic concentration, correspondingly. Additionally, Poole- Frenkel mobility model is opted to investigate the carrier's transportation. The expression for this model is as follow:

$$\mu(E) = \mu_0 e^{\left[-\frac{\Delta}{kT}+\left(\frac{\beta}{kT}-\gamma\right)\sqrt{|E|}\right.} \quad (2)$$

where E is electric field. Moreover, μ_0 and μ are zero field mobility and field dependent mobility, consequently. Other symbols which are k, Δ, β, T and γ depict Boltzmann constant, activation energy at zero electric field, , hole Pool–Frenkel factor, effective temperature for charge carriers, and fitting parameter, correspondingly. Fig. 2 is showing the simulated structure of the reference device.

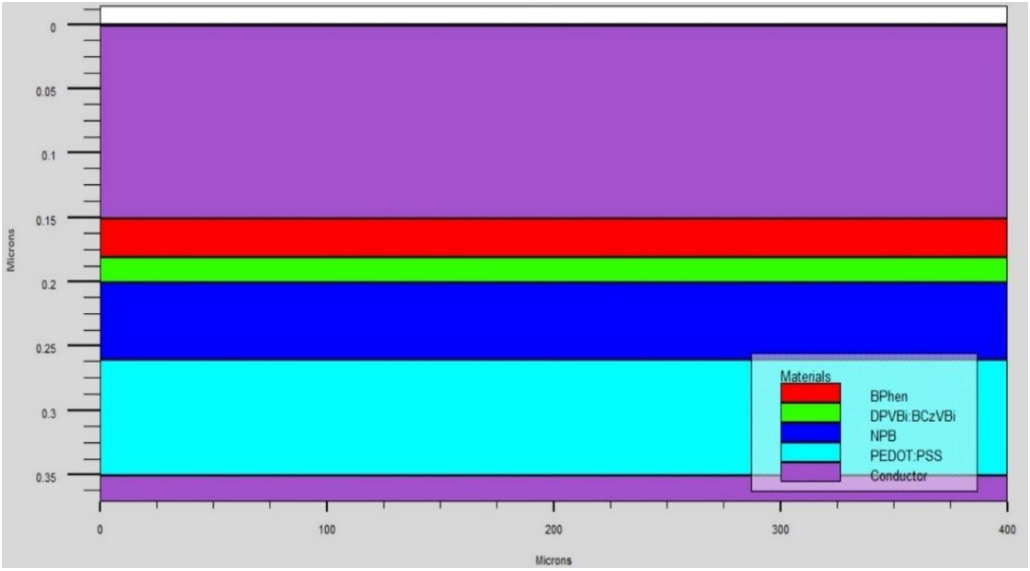


Fig. 2. Simulated Device structure of D_1

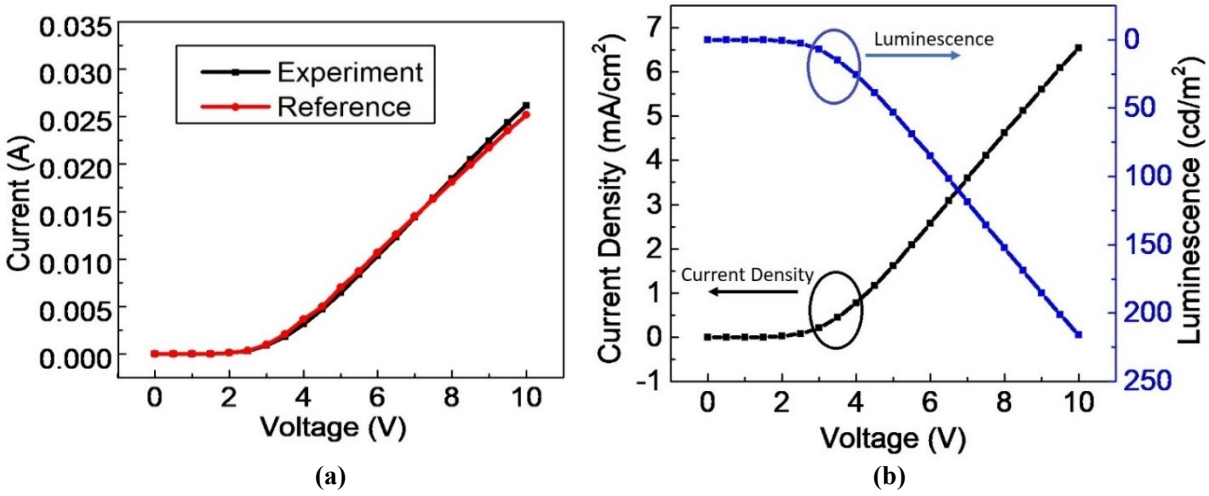


Fig. 3. (a) Experimental and Reference results of current & (b) Curves for luminescence and current density of device D_1

Fig. 3 (a) is outlining the comparison of reference and experiment data of D_1. In figure, it can be seen that experiment results are showing only less than 5% deviation in comparison with reference device as current values of the reference and experement devices are 0.025 and 0.026 A, correspondingly. Additonally, Fig. 3 (b) is illustrating the combined curve for current density and luminsecence with maximum values of 6.5 mA/cm^2 and 215 cd/m^2 and, respectively.

III. Comparison of Three Non-CGL and CGL Based Devices (D_1, D_2 & D_3)

In this section, three OLED devices D_1, D_2, and D_3 are analyzed and compared. Among these devices D_1 is the reference device which is a non-CGL device. Rest two devices D_2 and D_3 are CGL based devices which contain CGL at different positions. Herein, D_2 contains CGL just below the cathode layer and D_3 contain CGL just above the anode layer as mentioned in Fig. 4. All the devices are simulated using

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:41:47 UTC from IEEE Xplore. Restrictions apply.

Silvaco Atlas tool. The performance of D_2 is better than D_1 as it contains CGL which is able to produce a greater number of free charge carriers. These charge carriers eventually contribute in the recombination inside the emissive layer.

Among all these devices, D_3 is providing the best results. Let us understand the reason behind this. In case of D_3, the CGL is situated bottom side near anode. Here, free electrons inside HAT-CN get attracted from anode side and on the other hand, free holes in TAPC get pulled away from anode which will travel towards EML. Also, in organic electronics hole's mobility is better as compared to mobility of electrons. Hence, hole concentration will be increased in this case more significantly in comparison with D_2 where CGL is placed near cathode side. And so, recombination rate in D_3 will be higher than D_2 for the given reasons.

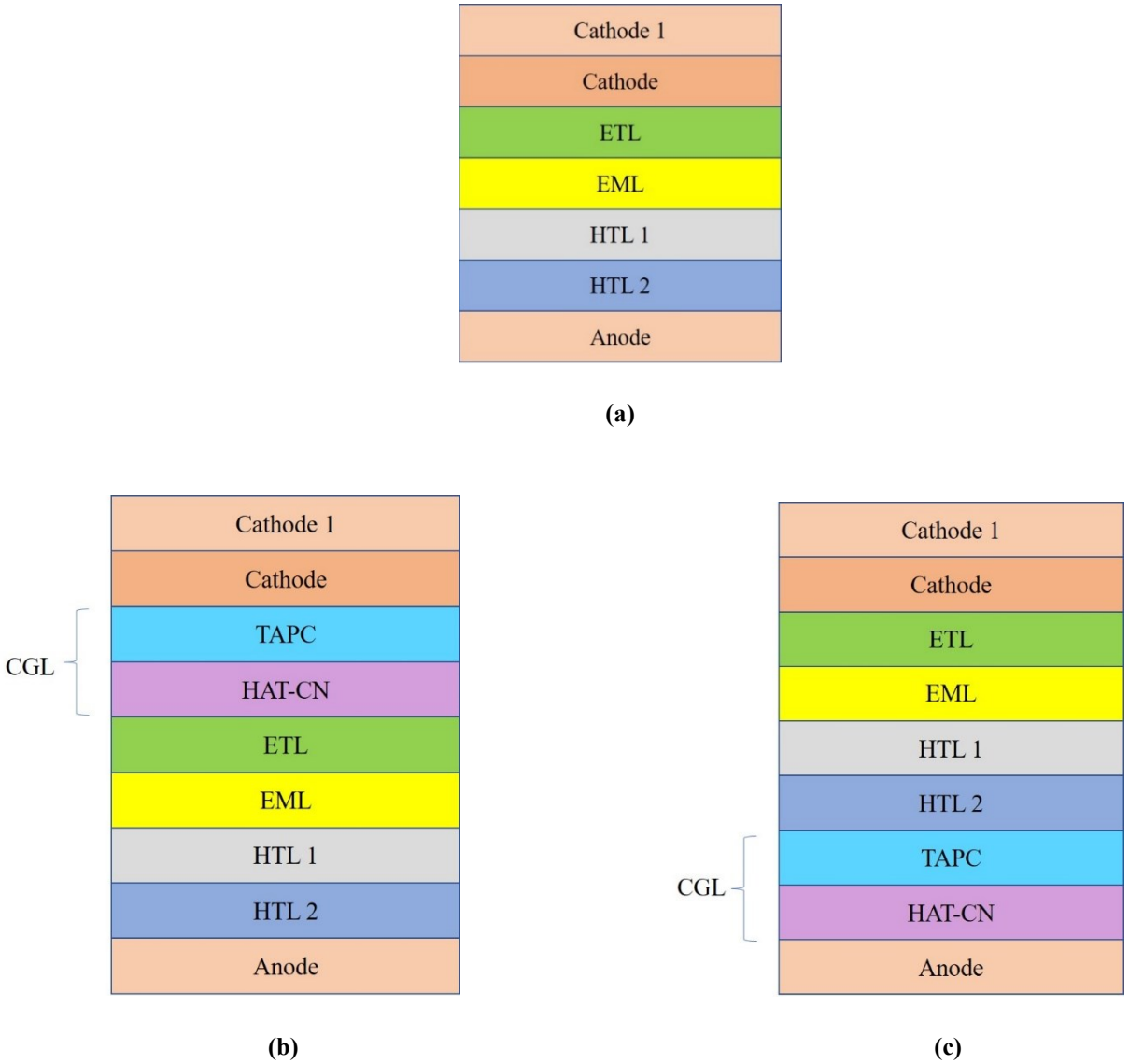


Fig. 4. (a) Reference device D_1 with no CGL, (b) Device D_2 with CGL below cathode and (c) Device D_3 with CGL above anode

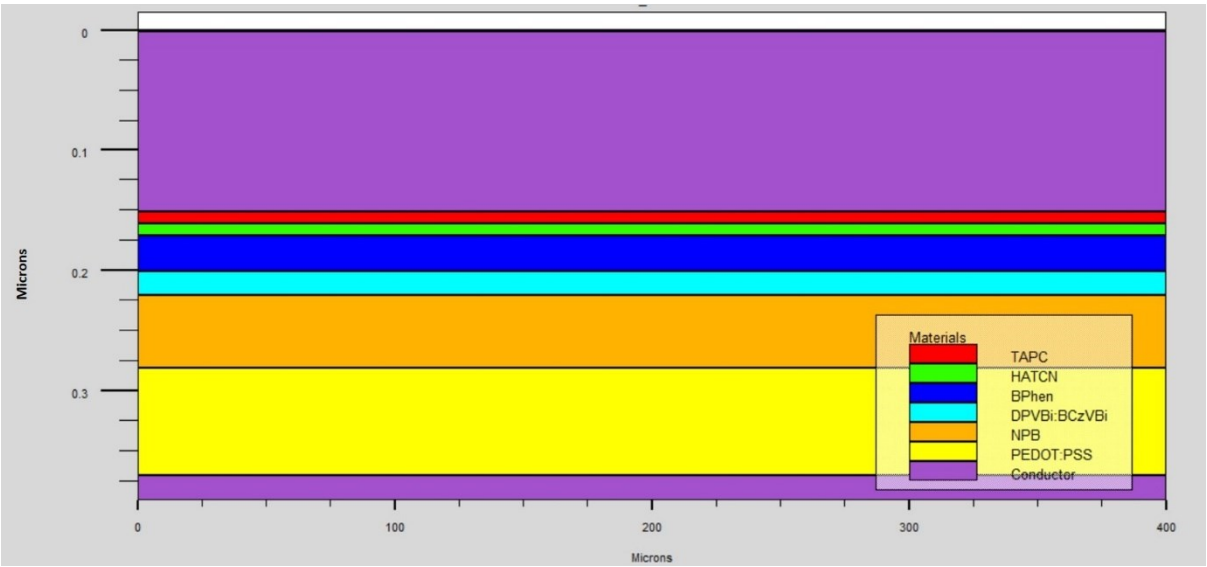
TABLE II. PARAMETRIC COMPARISON OF LUMINESCENCE AND CURRENT OF VARIOUS BLUE OLEDS

Device	Current (A)	Luminescence (cd/m^2)
D_1 (Reference)	0.026	215.9
D_2	0.193	1592.3
D_3	0.256	2112.2

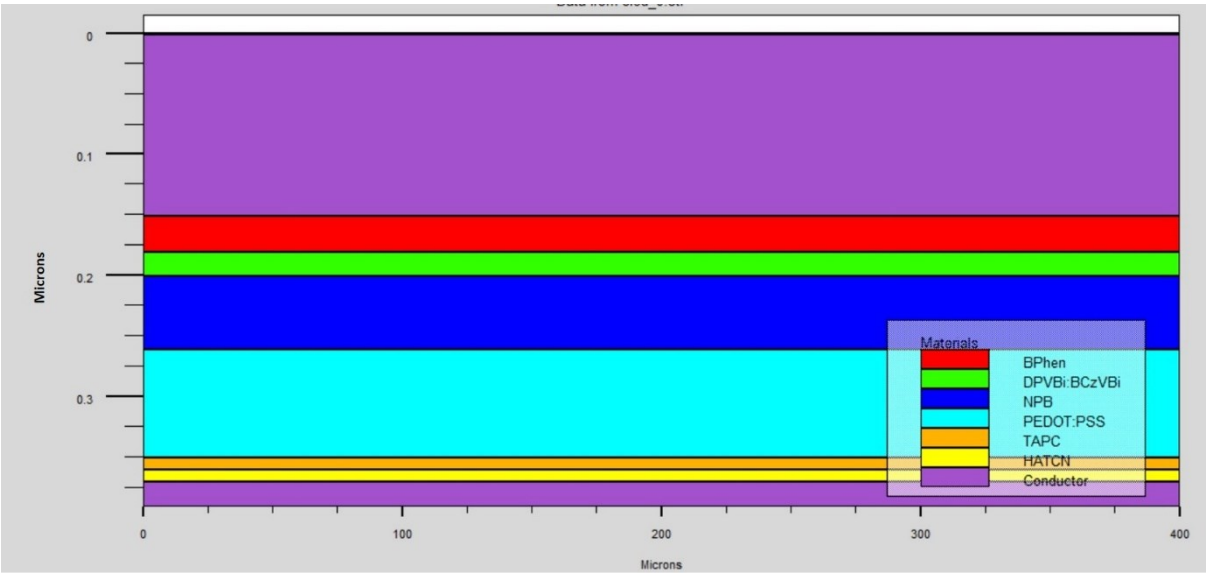
Table 2 is illustrating the parametric comparison on the basis of luminescence and current of all three devices. Moreover, Fig 5 (a) and (b) are representing the simulated structures of D_2 and D_3, correspondingly.

Fig. 6 is demontrating the comparison of all the three devices in terms of current and luminesence. Device D_3 is almost 10 times and 1.3 times improved in terms of current on comaring with D_1 and D_2, correspondingly. Moreover, D_3 is almost 10 and 1.4 times better in comarison with D_1 and D_2, respectively, in terms of luminescence.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:41:47 UTC from IEEE Xplore. Restrictions apply.

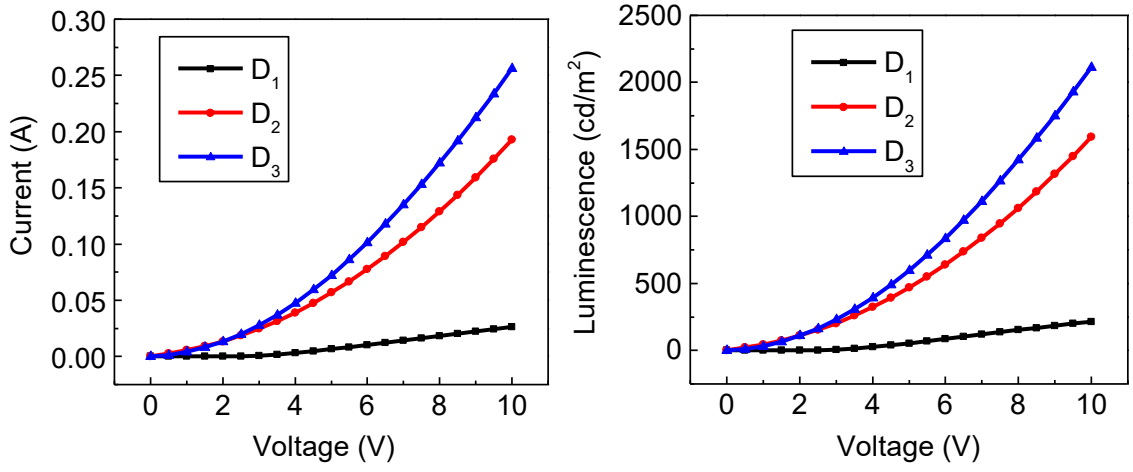


(a)



(b)

Fig. 5. Simulated structures of devices (a) D_2 and (b) D_3



(a) **(b)**

Fig. 6. Comparison of all three devices D_1, D_2 and D_3 on the basis of (a) Current and (b) Luminesence

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:41:47 UTC from IEEE Xplore. Restrictions apply.

IV. Internal Analysis of the Blue OLEDs Devices

In this section, the internal analysis is performed of all the mentioned device D_1, D_2 and D_3. This analysis is incorporated using two ways: 1) using cutine analysis and 2) using device structures. Both menthods are discussed in details in the below sub-sections.

A. Internal analysis using vertical cutline

Herein, a vertical cutline analysis is incorporated on the simulated structures of D_1, D_2 and D_3 by drawing a vertical cutline at 200 µm. Three parameters which are electron concentration, hole concentration and recombination rate are investigated to understand the in-depth performance of all the devices.

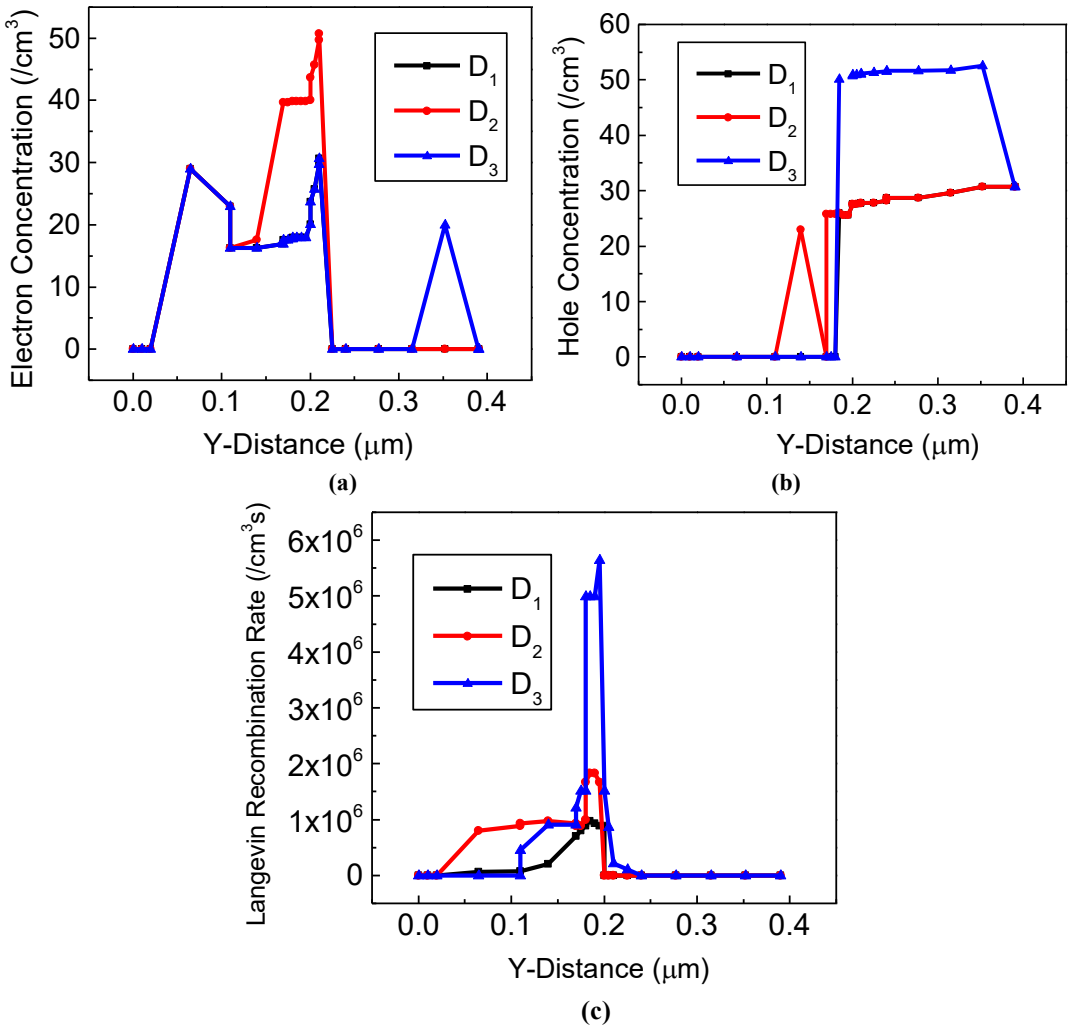


Fig. 7. Parametric comparison of D_1, D_2 and D_3 in terms of (a) Electron concentration, (b) Hole concentration and (c) Langevin recombination rate

It can be cleary observed from the Fig 7 (a) and (b) that D_3 device is outlining the best results for hole and electron concentration, respectively. This is only because of the position of CGL. In Device D_1 there is no CGL and in D_2 the CGL is placed just below the cathode layer. As discussed previously, the CGL close to anode is more benificial which is cleary visible in above figures. This increment in electrons and hole concentration directly enhances the recombination rate of device D_3 as illustrated in Fig. 7 (c). In this figure, D_3 is highlighting the maximum recombination rate among all the devices.

B. Internal Analysis using device structure

In this sub-section, the internal analysis is done using complete device structure of D_3 as shown in Fig 8. Herein, Fig. 8 (a) is demonstrating the potential distribution of device D_3. It ranges from 0 to 10 V which is in increasing fashion from cathode to anode. Moreover, Fig 8 (b) is representing the Langevin recombination rate of D_3. It can be noticed here that maximum recombination is near the emissive layer where electrons and holes recombine together and produce light.

V. Conclusion

In this paper, three devices D_1, D_2 and D_3 are analyzed and compared. Herein, D_1 is a non-CGL based device whereas, D_2 and D_3 are CGL based devices. Among all these devices, D_3 exhibits best performance. It shows the current 10 times and 1.3 times higher as compared with D_1 and D_2, correspondingly. Moreover, the luminescence of D_3 is about 1.4 and 10 times increased on comparing with D_2 and D_3, respectively. All the devices are analyzed using Silvaco Atlas tool. Furthermore, the internal analysis of the all the devices is also performed using two methods. One is cutline analysis where three parameters which are hole concentration, electron concentration and Langevin recombination rate are

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:41:47 UTC from IEEE Xplore. Restrictions apply.

compared together. For all the parameters, D_3 exhibits the best performance because of the given reasons in the previous sections. The other method is structure based internal analysis. In this method, the potential distribution and Langevin recombination rate of the D_3 is highlighted.

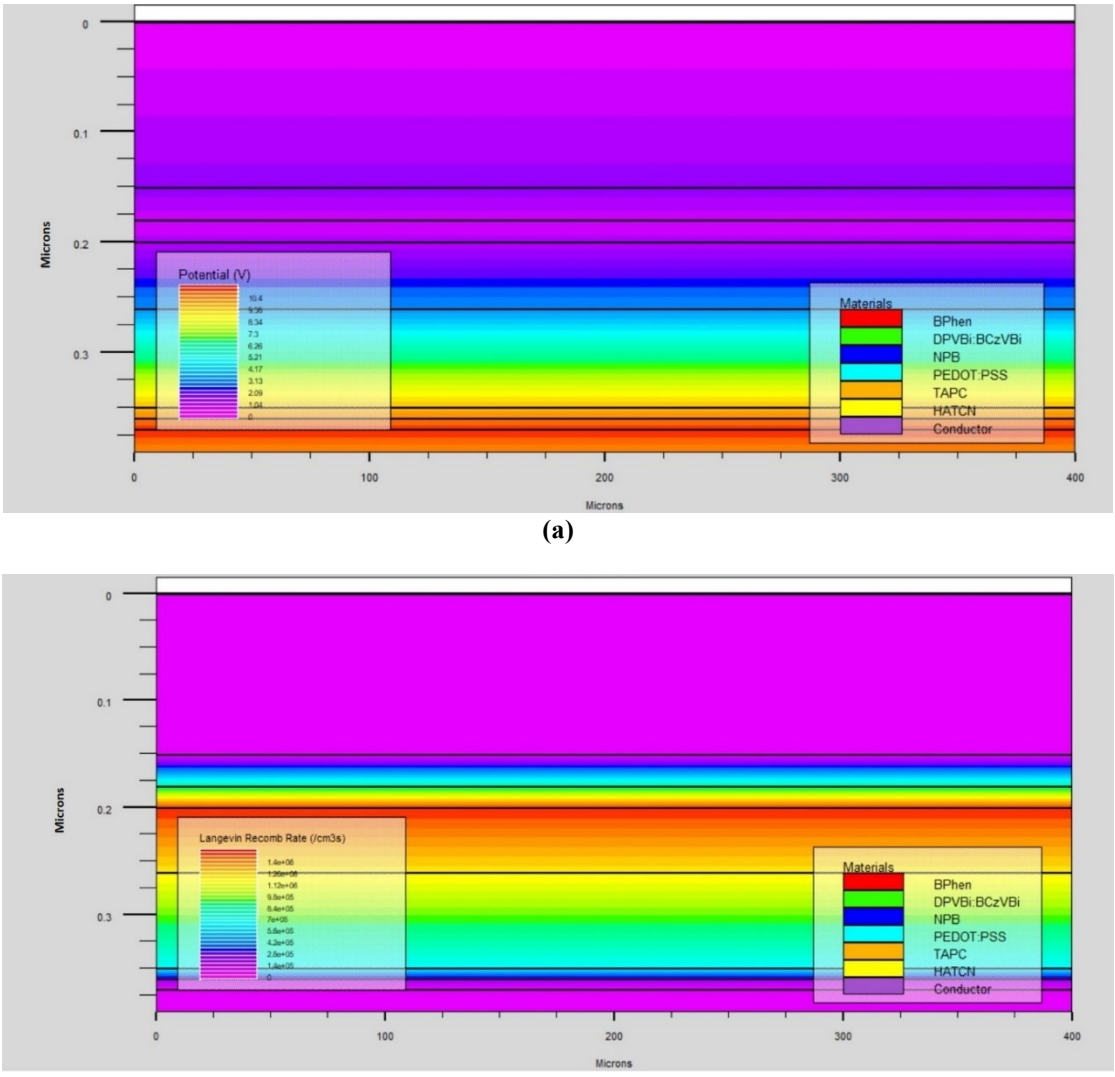


Fig. 8. Structural internal analysis of D_3 in terms of (a) Potential and (b) Recombination Rate

REFERENCES

[1] P. Chamola and P. Mittal, "Flexible organic solar cell to power modern cardiac pacemakers: Versatile for all age groups, skin types and genders", Physica Scripta, vol 98, pp. 035018, 2023.

[2] S. Negi, P. Mittal and B. Kumar, "Characteristic performance of OLED based on hole injection, transport and blocking layers", Recent Patents on Engineering. Vol 14, pp. 373, 2020.

[3] S. Negi, P. Mittal and B. Kumar, "Impact of different layers on performance of OLED", Microsystem Technologies, 24, 498, 2018.

[4] S. Yadav, P. Mittal and S. Negi, "Recent advancements over a decade for organic light-emitting diodes: from structural diversity, role of layers, colour emission, material classification, performance improvement, fabrication to applications", Bulletin of Materials Science, vol 45, pp. 109, 2022.

[5] C. H. Gao, X. B. Shi, D. Y. Zhou, L. Zhang, Z. K. Wang and L. S. Liao, "Highly efficient white organic light-emitting diodes with controllable excitons behavior by a mixed interlayer between fluorescence blue and phosphorescence yellow-emitting layers", International Journal of Photoenergy, 2013.

[6] Y. Chen, J. Chen, D. Ma, D. Yan, L. Wang and F. Zhu, "High power efficiency tandem organic light-emitting diodes based on bulk heterojunction organic bipolar charge generation layer", Applied Physics Letters, vol 98, pp. 114, 2011.

[7] Y. M. Kim, J. W. Lee, J. H. Jung, K. K. Paek, M. Y. Sung, J. K. Kim and B. K. Ju, "Enhanced brightness and efficiency of organic light-emitting diodes with an LiF in the Alq/sub*",* IEEE Electron Device Letters*, vol* 27, pp 558, 2006.

[8] H. Sun, Q. Guo, D. Yang, Y. Chen, J. Chen and D. Ma, "High efficiency tandem organic light emitting diode using an organic heterojunction as the charge generation layer: an investigation into the charge generation model and device performance", Acs Photonics, vol 2, pp. 271, 2015.

[9] S. Yadav, P. Mittal and S. Negi, "Recent advancements over a decade for organic light-emitting diodes: from structural diversity, role of layers, colour emission, material classification, performance improvement, fabrication to applications", Bulletin of Material Science, vol 45, pp. 109, 2022.

[10] Y. Liu, X. Wu, Z. Xiao, J. Gao, J. Zhang, H. Rui, X. Lin, N. Zhang, Y. Hua and S. Yin, "Highly efficient tandem OLED based on C60/rubrene: MoO3 as charge generation layer and LiF/Al as electron injection layer", Applied Surface Science, vol 413, pp. 302, 2017.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:41:47 UTC from IEEE Xplore. Restrictions apply.

J. Chem. Sci. (2024) 136:39
https://doi.org/10.1007/s12039-024-02282-6
© Indian Academy of Sciences

REGULAR ARTICLE

Potassium *tert*-butoxide promoted a direct one-pot synthesis of nitriles from aldehydes at room temperature

SAURAV KUMAR[a], RITIKA KUBBA[a], NITYANANDA AGASTI[b], ANITHA SELVARAJ[c] and ANIL KUMAR[a,*]

[a]Department of Applied Chemistry, Delhi Technological University, Delhi 110042, India
[b]Department of Chemistry, Deen Dayal Upadhyaya College, University of Delhi, Delhi 110078, India
[c]Department of Chemistry, Sri Sai Ram Engineering College, West Tambaram, Chennai 600044, India
*Corresponding author. E-mail: anil_kumar@dce.ac.in

MS received 4 March 2024; revised 8 April 2024; accepted 9 April 2024

Abstract. Potassium *tert*-butoxide mediated a direct one-pot synthesis of diversely substituted nitriles from aldehydes *via* the sequential addition of hydroxylamine and benzoyl chloride is reported. A wide range of aromatic, heteroaromatic, and vinyl aldehydes were converted into corresponding nitriles in good to excellent yields. Room temperature reactions, transition metal-free conditions, wide functional group tolerance, simple operation, and short reaction time are the salient features of the developed methodology.

Keywords. nitrile; benzaldehyde; hydroxyl amine; benzoyl chloride; potassium *tert*-butoxide.

1. Introduction

Nitrile (cyano) is one of the key functional groups in organic synthesis.[1] Nitrile groups are found in many natural products, drugs, and bioactive molecules (Figure 1).[2] For instance, anastrozole is used against breast cancer, while verapamil and tofacitinib are used for treating high blood pressure and arthritis, respectively. On the other hand, functional group transformation of nitrile with different reagents allows easy access to primary amines, amides, carboxylic acid, ketones, etc.[3] Hence, compounds with nitrile functional group have been used as a precursor for preparing various bioactive molecules, natural products, pharmaceuticals, agrochemicals, polymers, dyes, etc.[1–4]

Considering the importance of nitrile groups in organic synthesis, numerous methods have been developed for their preparation.[5] The conventional approaches for synthesizing nitriles include Kolbe's nitrile synthesis, Sandmeyer reaction, Rosenmund-von Braun reaction, dehydration of amides, cyanide-halide exchange reactions, etc.[6] However, this approach requires complex operations, toxic metal and metal-free reagents, harsh reaction conditions, etc. In this context, the synthesis of nitriles from aldehydes received considerable attention owing to their commercial availability and inexpensiveness.[7] A diverse range of nitrogen sources has been employed in transforming aldehydes into nitriles.[5,7] However, some of the developed methods under this approach suffer due to limited substrate scope, use of strong oxidants and metal catalysts, high reaction temperature, tedious operation procedure, low yields, and poor functional group tolerances. Therefore, developing a new method for directly synthesizing nitriles from aldehydes under mild reaction conditions is still desirable.

In this context, we disclosed ^{t}BuOK-mediated one-pot synthesis of nitriles from aldehydes via the sequential addition of hydroxylamine and benzoyl chloride at room temperature. This method relies on the generation of aldoxime *in situ* from aldehyde and the sequential transformation of aldoxime into the nitrile group using benzoyl chloride and potassium *tert*-butoxide at room temperature (Scheme 1).

2. Results and discussion

In the beginning, optimization of the reaction conditions was investigated using benzaldehyde as the model substrate. Initially, benzaldehyde **1a** was

Supplementary Information: The online version contains supplementary material available at https://doi.org/10.1007/s12039-024-02282-6 and https://www.ias.ac.in/chemsci.

Published online: 23 May 2024

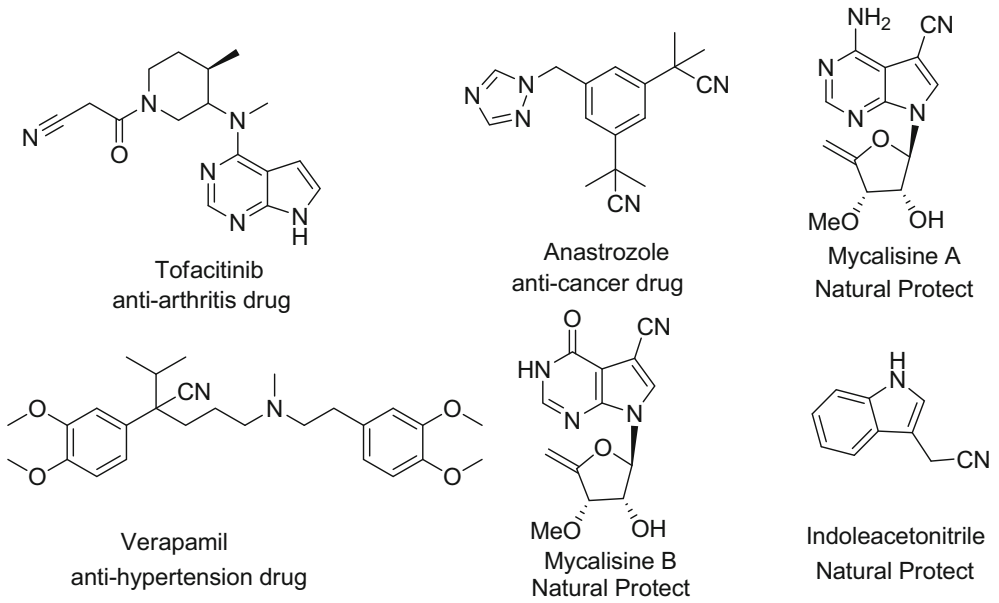
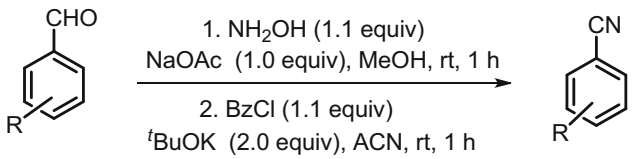


Figure 1. Nitrile groups on natural products and bioactive molecules.



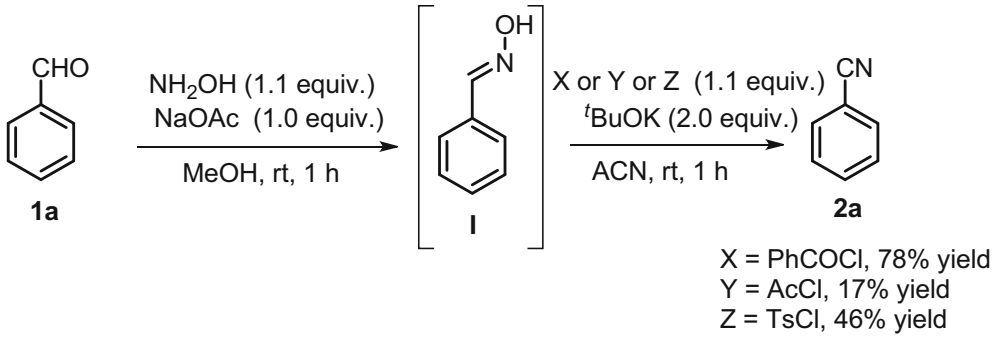
Scheme 1. Synthesis of nitrile to aldehyde.

converted into benzaldoxime **I** in the presence of hydroxylamine hydrochloride and sodium acetate in methanol. The reaction was carried out at room temperature for 1 h to obtain quantitative yield aldoxime **I**. After that, the methanol was evaporated to dryness, and the crude mixture was dissolved in acetonitrile. The resulting mixture was treated with benzoyl chloride (1.1 equiv.) followed by base (1.0 equiv.). Various organic and inorganic bases, including Et_3N, DBU, DABCO, KOH, NaOH, K_2CO_3, Cs_2-CO_3, K_3PO_4, K_2HPO_4, ^{t}BuONa, and ^{t}BuOK were investigated. The desired product **2a** was not obtained with Et_3N, DABCO, KOH, NaOH, K_2CO_3, and Cs_2CO_3. However, DBU, K_3PO_4, and K_2HPO_4 gave product **2a** in 16–24% yields. On the other hand, sodium *tert*-butoxide gave the desired nitrile **2a** in 36% yield in 3 h, while 46% was obtained with potassium *tert*-butoxide in 1 h. Further, we investigated the reaction with additional equiv. of potassium *tert*-butoxide. To our delight, the desired product **2a** was obtained with a 78% yield with two equiv. of potassium *tert*-butoxide at room temperature. However, a further increase in the equiv. of potassium *tert*-butoxide did not improve the yield of the product.

The optimization study was further continued by screening acyl chlorides in the presence of potassium *tert*-butoxide (Scheme 2). The reaction provided **2a** only in 17% yield when acetyl chloride was used instead of benzoyl chloride. On the other hand, the reaction with sulfonyl chloride gave the product a 46% yield, indicating the superiority of benzoyl chloride over other acyl chlorides. The reason sulfonyl chloride gave a low yield might be due to the potassium ion coordinate with the nitrogen atom of imine and the oxygen atom of the carbonyl group, which accelerates the reaction.

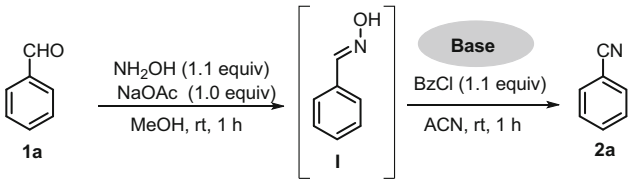
After identifying optimized reaction conditions (Table 1, entry 12), the scope of the methodology was studied using different aromatic, heteroaromatic, and vinyl aldehydes 1a-v (Scheme 3). Initially, aromatic aldehydes bearing functional groups at the *para*-position were investigated to understand the electronic effects of substituents on the reaction progress.

Benzaldehydes bearing electron-donating groups such as OMe and SCH_3 (**1b** and **1c**) gave the desired products **2b** and **2c** with 77% and 80% yields, respectively. Notably, a heterocyclic compound such as morpholine functionalized benzaldehyde **1d** gave the desired product **2d** in 74% yield. On the other hand, halogens such as Br and CF_3 functionalized benzaldehydes **1e** and **1f** were successfully converted to corresponding nitriles **2e** and **2f** in 70% and 72% yields, respectively. Moreover, carboxylic esters and carboxylic acid functionalized benzaldehydes **1g** and **1h** delivered the desired products **2g** and **2h** in 66% and 64% yields, respectively. Interestingly, the strong



Scheme 2. Screening of acyl chlorides.

Table 1. Optimization of reaction conditions.[a]



Entry	Base (equiv.)	Time (h)	Yield (%)[b] **2a**
1	Et_3N	24	NR
2	DBU	24	16
3	DABCO	24	NR
4	KOH (1.0)	24	NR
5	NaOH (1.0)	24	NR
6	K_2CO_3 (1.0)	24	NR
7	Cs_2CO_3 (1.0)	24	NR
8	K_3PO_4 (1.0)	24	24
9	K_2HPO_4 (1.0)	24	20
10	^{t}BuONa (1.0)	3	36
11	^{t}BuOK (1.0)	1	46
12	**^{t}BuOK (2.0)**	**1**	**78**
13	^{t}BuOK (3.0)	1	76

[a]Reactions were performed using 1.0 mmol of 1a, 1.1 equiv. of NH_2OH and 1.0 equiv. of sodium acetate in 2.0 mL of methanol at RT. Then, methanol was removed under reduced pressure. For the second step, reactions were carried out using 1.1 equiv. of BzCl and 2.0 equiv. of the base in 2.0 mL of ACN at RT. [b]Isolated yield.

electron deficient group, such as NO_2-functionalized aldehyde **1i**, gave the corresponding nitrile compound **2i** in 57% yield. On the other hand, the substrate bearing functional groups at the *meta*-position of benzaldehyde **1j** proceeded well under optimized reaction conditions to afford **2j** in 65% yields. Further, we investigated various aldehydes bearing Cl, Br, OH, and Me substituents at the *ortho*-position (**1k-r**). These substrates successfully converted into corresponding nitriles **2k-r** with 62–82% yields. Notably, 1-naphthylbenzaldehyde **1s**, cinnamaldehyde **1t**, and dimethylamino cinnamaldehyde **1u** gave desired products **2s-u** in 65–82% yields under the optimized conditions. More interestingly, the inorganic compound ferrocene carboxaldehyde **1v** was successfully converted into **2v** with a 72% yield.

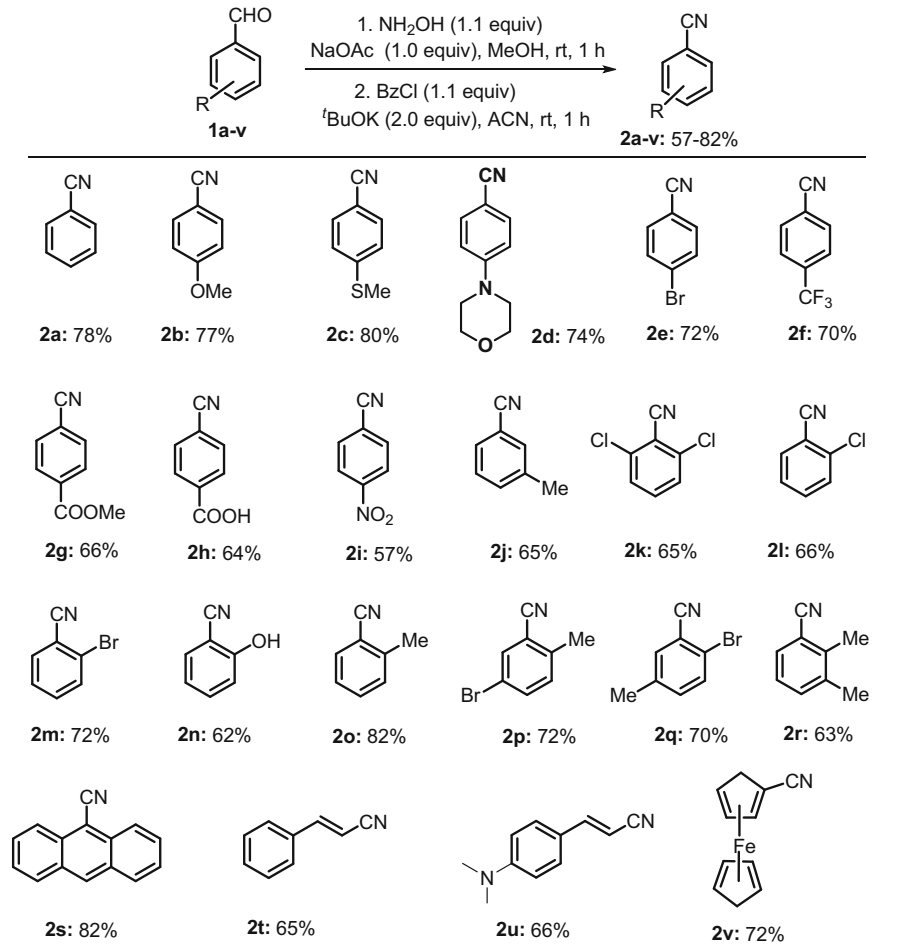
Considering the important role of heterocyclic scaffolds in medicinal and pharmaceutical chemistry, we further examined the transformation of heteroaromatic aldehydes into corresponding nitriles (Scheme 4). A range of nitrogen, sulfur, and oxygen-containing heteroaromatic aldehydes **3a-e** were smoothly converted into the corresponding nitriles **4a-e** in 75–84% yields.

To demonstrate the practicality of the protocol, a gram-scale reaction was performed using model substrate **1a** under the optimized reaction conditions (Scheme 5). The reaction proceeded very well to give the product **2a** in 76% yield.

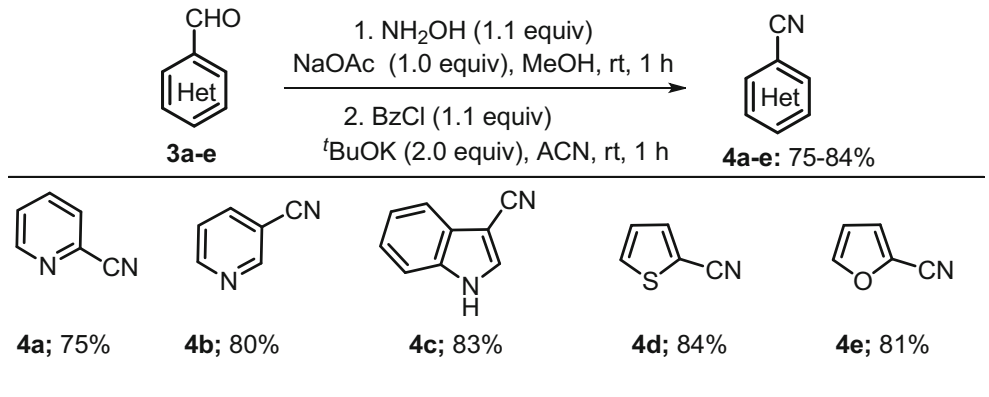
Based on the observed results and the available literature,[5–8] a plausible mechanism for the reaction has been proposed as shown in Scheme 6. Initially, aldehyde **1** reacts with hydroxylamine to afford the aldoxime **A**. *O*-Acylation of aldoxime **A** in the presence of benzoyl chloride and ^{t}BuOK leads to the formation of compound **B.** The intermediate **B** is converted into desired compound **2** in the presence of ^{t}BuOK via an elimination reaction.

3. Conclusion

An efficient base-mediated direct synthesis of aromatic and heteroaromatic nitriles from corresponding aldehydes was reported as the starting ingredients for the

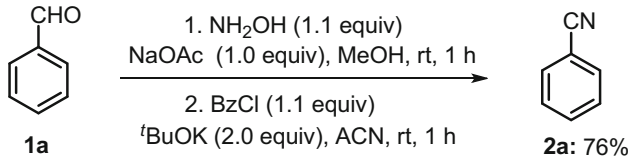


Scheme 3. Scope for substituted aromatic and vinyl aldehydes.

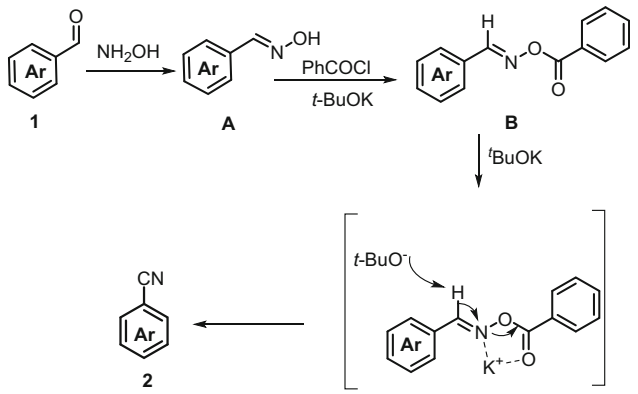


Scheme 4. Scope for substituted heteroaromatic aldehydes.

synthesis; aldehyde and hydroxylamine were combined with benzoyl chloride in a series of reactions. Notably, this process may be carried out at room temperature without needing transition metal catalysts, making it a practical and effective process. Various aromatic and heteroaromatic aldehydes gave the products good to excellent yields. The important features of the developed methodology include broad substrate scope, easy operation, quick reactions, tolerance of different functional groups, and room temperature reactions.



Scheme 5. Gram-scale synthesis.



Scheme 6. Proposed reaction mechanism.

Acknowledgements

AK and SK gratefully acknowledge Delhi Technological University for providing instrumental facilities. SK thanks CSIR for the SRF fellowship (file no. 08/133/(0032)/2019-EMR-I).

References

1. (a) Rappoport Z 1970 The Chemistry of the Cyano Group (London: Interscience Publishers); (b) Fatiadi A J, in: Patai S, Rappoport Z (eds) 1983 Preparation and Synthetic Applications of Cyano Compounds (New York: Wiley) p. 1057.
2. (a) Fleming F F 1999 Nitrile-containing natural products *Nat. Prod. Rep.* **16** 597; (b) Fleming F F, Yao L, Ravikumar P C, Funk L and Shook B C 2010 Nitrile-Containing Pharmaceuticals: Efficacious Roles of the Nitrile Pharmacophore *J. Med. Chem.* **53** 7902; (c) Frizler M, Lohr F, Furtmann N, Kläs J and Gütschow M 2011 Structural Optimization of Azadipeptide Nitriles Strongly Increases Association Rates and Allows the Development of Selective Cathepsin Inhibitors *J. Med. Chem.* **54** 396
3. (a) Larock R C 1989 Comprehensive Organic Transformations: A Guide to Functional Group Preparations (New York: VCH); (b) Smith M B and March J 2007 March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure, 6th edn (Hoboken, NJ: Wiley)
4. (a) Sruthi P R and Anas S J 2020 An overview of synthetic modification of nitrile group in polymers and applications *Polym. Sci*. **58** 1039; (b) Sathyendran S and Senadi G C 2023 Progression on Oxidative Decyanation Approaches from Secondary Nitriles to Ketones and α-Aminonitriles to Amides. *Asian. J. Org. Chem*. **12** e202300433; (c) Cohen M A, Sawden J and Turner N J 1990 Selective hydrolysis of nitriles under mild conditions by an enzyme *Tetrahedron Lett.* **31** 7223; (d) Kumar S, Dixit S K and Awasthi S K 2014 An efficient one pot method for synthesis of carboxylic acids from nitriles using recyclable ionic liquid [bmim]HSO_4 *Tetrahedron Lett.* **55** 3802
5. (a)Yan G, Zhang V and Wang J 2017 Recent Advances in the Synthesis of Aryl Nitrile Compound *Adv. Synth. Catal.* **359** 4068; (b) Patel R I, Sharma S and Sharma A 2021 Cyanation: a photochemical approach and applications in organic synthesis *Org. Chem. Front.* **8** 3166; (c) Mowry D T 1948 The Preparation of Nitriles *Chem. Rev*. **42** 189; (d) Anbarasan P, Schareina T and Beller M 2011 Recent developments and perspectives in palladium-catalyzed cyanation of aryl halides: synthesis of benzonitriles *Chem. Soc. Rev*. **40** 5049
6. (a) Wang Z 2010 Kolbe Nitrile Synthesis. In Comprehensive Organic Name Reactions and Reagents (Germany: Wiley) p. 1661; (b) Sandmeyer T 1884 Ueber die Ersetzung der Amid-gruppe durch Chlor, Brom und Cyan in den aromatischen Substanzen *Ber. Dtsch. Chem. Ges.* **17** 2650; (c) Nielsen M A, Nielsen M K and Pittelkow A 2004 Scale-Up and Safety Evaluation of a Sandmeyer Reaction *Org. Process Res. Dev*. **8** 1059; (d) Wang D P, Kuang L P, Li Z W and Ding K L 2008 Proline-Promoted Rosenmund-von Braun Reaction *Synlett.* **2008** 69; (e) Pradal A and Evano G 2014 A vinylic Rosenmund–von Braun reaction: practical synthesis of acrylonitriles *Chem. Commun.* **50** 11907; (f) Enthaler S 2011 Straightforward Uranium-Catalyzed Dehydration of Primary Amides to Nitriles *Chem.– Eur. J.* **17** 9316; (g) Sueoka S, Mitsudome T, Mizugaki T, Jitsukawa K and Kaneda K 2009 Supported monomeric vanadium catalyst for dehydration of amides to form nitriles *Chem. Commun*. **46** 8243; (h) Zhang X, Xia A, Chen H and Liu Y 2017 General and Mild Nickel-Catalyzed Cyanation of Aryl/Heteroaryl Chlorides with $Zn(CN)_2$: Key Roles of DMAP *Org. Lett.* **19** 2118
7. (a) Wu Q, Luo Y, Lei A and You J 2016 Aerobic Copper-Promoted Radical-Type Cleavage of Coordinated Cyanide Anion: Nitrogen Transfer to Aldehydes to form Nitrile**s** *J. Am. Chem. Soc*. **138** 2885; (b) Hyodo K, Togashi K, Oishi N, Hasegawa G and Uchida K 2017 Brønsted Acid Catalyzed Nitrile Synthesis from Aldehydes Using Oximes via Transoximation at Ambient Temperature *Org. Lett*. **19** 3005; (c) An X-D and Yu S 2015 Direct Synthesis of Nitriles from Aldehydes Using an *O*-Benzoyl Hydroxylamine (BHA) as the Nitrogen Source *Org. Lett*. **17** 5064; (d) Laulhe S, Gori S S and Nantz M H 2012 A Chemoselective, One-Pot Transformation of Aldehydes to Nitrile *J. Org. Chem*. **77** 9334; (e) Kelly C B, Lambert K M, Mercadante M A, Ovian J M, Bailey W F and Leadbeater N E 2015 Access to Nitriles from Aldehydes Mediated by an Oxoammonium Salt *Angew. Chem., Int. Ed*. **54** 4241; (f) Dornan L M, Cao Q, Flanagan J C A, Crawford J J, Cook M J and Muldoon M 2013 Copper/TEMPO catalysed synthesis of nitriles from aldehydes or alcohols using aqueous ammonia and with air as the oxidant *J. Chem. Commun*.

49 6030; (g) Noh J-H and Kim J 2015 Aerobic Oxidative Conversion of Aromatic Aldehydes to Nitriles Using a Nitroxyl/No_xCatalyst System *J. Org. Chem*. **80** 11624; (h) Rokade B V and Prabhu K R 2012 Chemoselective Schmidt Reaction Mediated by Triflic Acid: Selective Synthesis of Nitriles from Aldehydes *J. Org. Chem*. **77** 5364

8. (a) Rapeyko A, Climent M J, Corma A, Concepcion P and Iborra S 2016 Nanocrystalline Ceo_2 as a Highly Active and Selective Catalyst for the Dehydration of Aldoximes to Nitriles and One-Pot Synthesis of Amides and Esters *ACS Catal*. **6** 4564; (b) Hyodo K, Kitagawa S, Yamazaki M and Uchida K 2016 Brønsted acid catalyzed transoximation reaction: synthesis of aldoximes and ketoximes without the use of hydroxylamine salts *Green. Chem.* **11** 5788; (c) Yamaguchi K, Fujiwara H, Ogasawara Y, Kotani M and Mizuno N A 2007 Tungsten–Tin Mixed Hydroxide as an Efficient Heterogeneous Catalyst for Dehydration of Aldoximes to Nitriles *Angew. Chem., Int. Ed*. **46** 3996; (d) Ishihara K, Furuya Y and Yamamoto H 2002 Rhenium(VII) Oxo Complexes as Extremely Active Catalysts in the Dehydration of Primary Amides and Aldoximes to Nitriles *Angew. Chem., Int. Ed.* **41** 3109; (e) Yang S H and Chang S 2001 Highly Efficient and Catalytic Conversion of Aldoximes to Nitriles *Org. Lett*. **3** 4209

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/s00202-024-02432-6

ORIGINAL PAPER

Power quality enhancement in single phase two level/five level converters using adaptive-RBFNN algorithm

Sudhanshu Mittal[1] · **Alka Singh**[1] · **Prakash Chittora**[1]

Received: 4 August 2023 / Accepted: 22 April 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
This paper presents Adaptive Radial Basis Function Neural Network (A-RBFNN) algorithm for the control of two level and five level converters. This paper investigates the performance of five level cascaded H-bridge (CHB) multilevel converter as compared to two level converters in power distribution system to improve power quality of connected system. The CHB multilevel inverter is a popular configuration because of its wide application for high power medium voltage requirement. The power system including two level (2L) and five level (5L) converter is modeled in MATLAB environment. Experimental setup is also developed in the laboratory for hardware verification purpose and both the 2L and 5L CHB converters are integrated to single phase system as compensators. The A-RBFNN algorithm is able to regulate the DC link voltages to its rated value for two H-bridges in 5L configuration. The designed algorithm is able to compensate for PQ issues of the connected system. It is also tested under steady state and dynamic conditions. Further, A-RBFNN is used to extract individual harmonics of any desired order. Finally, comparison is presented based on the performance results of the proposed algorithm with 2L as well as 5L converters.

Keywords Two level/five level converter · Cascaded H-Bridge (CHB) · Multilevel converter (MLC) · Total harmonic distortion (THD) · Shunt active power filter (SAPF)

1 Introduction

Research in the modern power system is concentrated mainly around the negative effects of power quality events (PQEs), for providing clean and sustainable power to industry sectors and residential complexes [1]. Significant increase in non-linear loads, such as power sources for electronics devices, adjustable speed drives, uninterruptible power supplies and rectifiers increase the burden of power quality issues [2]. Reduction in harmonic content is an important prerequisite of good power quality. Harmonics are classified into two types: voltage harmonics and current harmonics [3]. Current harmonics are typically generated due to the harmonics in the voltage supply and also depend on the type of load, such as resistive, capacitive or inductive. Various topologies starting from two level inverter to multilevel inverters have been studied for improvement of power quality [4].

There is an increasing demand for highly efficient, power-dense DC–AC converters to reduce harmonics in distribution systems. The conventional two-level converter suffers from poor harmonic distortion, higher Direct Current (DC) link voltage requirement, larger dv/dt and rotor heating etc. So, due to the above mentioned issues 2L converter is not used for medium/high voltage and high power applications. In multilevel inverters, the switching stress and harmonic distortions are less in comparison to the conventional two-level converter [5]. Moreover, the voltage stress on the devices is less. Multilevel converters are being realised for high power applications as well as renewable energy integration. They are capable of providing high voltage while reducing harmonic distortions due to their unique circuit topology. Multilevel converters can generate a staircase output waveform that appears to be closer to sinusoidal and produces very low harmonics as compared to a conventional two level converter output voltage. Thus, multilevel converters have

✉ Alka Singh
alkasingh.eed@gmail.com

Sudhanshu Mittal
sudhanshu15mittal@gmail.com

Prakash Chittora
prakashchittora@gmail.com

[1] Electrical Engineering, Delhi Technological University, Delhi, India

Published online: 21 May 2024

shown to be an efficient and practical solution for reducing switching stress in high power, high voltage applications. There are mainly three types of multilevel inverters viz. Diode clamped, Flying capacitor and Cascaded H- bridge multi-level converter. Diode clamped multilevel inverter was firstly introduced in 1981 by Nabae et al. [6]. The additional clamping diodes are used in the diode clamped multilevel converter to produce AC voltage waveforms with multiple levels. Another modification in mid-1990s was developed in the form of the flying capacitor based converter as presented by Lavieville et al. along with Maynard et al. [7]. The output of this converter is half of the DC voltage input which is a disadvantage of this configuration. The third configuration which is cascaded H-bridge multilevel converter is a much simpler multilevel converter topology. The advantage of this type of multilevel converter is that it requires fewer components than diode clamped or flying capacitor converters and the modules can be combined easily with the possibility of renewable energy integration also [8].

A Shunt Active Power Filter (SAPF) is a power electronic device that is used to mitigate power quality issues caused by harmonic distortions, reactive power burden and unbalances in electrical systems. Different control algorithms can be used to address PQ issues like reactive power, harmonic unbalance, power fluctuations, etc. Among the various control algorithms, reported in literature Savitzky Golay Filter (SGF) based Phase Locked Loop controller [9] is a classical technique. New algorithms such as genetic algorithm optimized fuzzy logic [10], ANOVA kernel Kalman filter [11], adaptive normalized Huber control algorithm [12] have been recently proposed. A Phase-Locked Loop (PLL) is also required in SAPF to perform synchronization and control functions. Some PLL based algorithms are Second Order Generalized Integrator (SOGI) [13], PLL-less SRF [14], Enhanced Phase Locked Loop (EPLL) [15], Leaky Least Mean Fourth (LLMF) algorithm [16], NARLMMN algorithm [17] etc.

Neural network (NN) is a new and powerful tool for solving optimization problems. The use of neural networks has grown significantly over the past two decades due to improved computer models and development in sensor technology [18]. NN help in system identification and development of valid models which are required for system analysis. Various NN techniques are reported in literature for shunt compensation viz Hermite function-based ANN [19], Bernstein Functional Blending Neural Network [20], functional link artificial neural network controller [21], Immune RBF neural network [22], Neural Network (NN) based Improved Linear Sinusoidal Tracer (ILST) [23] etc.

An RBF neural network is an artificially generated neural network that uses the radial basis functions as activation function. RBF Neural Networks are generally applied to approximate functions, classify data, and recognize patterns. The RBF architecture has three layers with input layer, hidden layer and output layer.

In this paper, both two level (2L) and five level (5L) converter configurations are being implemented as SAPF in single phase system. A new control technique viz Adaptive Radial Basis Function Neural Network (A-RBFNN) is used in this paper. The highlights of this work are:

1. Development of an Adaptive A-RBFNN algorithm with detailed mathematical analysis.
2. The application of the developed algorithm and comparison is tested on both 2L and 5L converter topologies.
3. An application of A-RBFNN for segregation of different harmonic orders is also added.
4. Extensive steady state and dynamic results are presented for a fair comparison of two converter topologies.

2 System configuration of proposed system

Figure 1 shows the circuit diagram of the proposed system with five level cascaded H-bridge configuration. A single-phase AC source supplies power to a rectifier load having R-L branch connected at DC side. Source resistance and source inductance parameters are denoted as R_S and L_S. Two converter configurations (2L/5L CHB) are connected one by one in shunt configuration to compensate the PQ problems. The control of the proposed system requires input variables such as source voltage v_s, source current i_s, load current i_L, total dc link voltage ($V_{dc1} + V_{dc2}$). These are sensed with LEM based current and voltage sensors. These sensed voltage and current signals are fed through ADC channel of DSP (dSPACE 1104). The DSP generates the required gating pulses for SAPF. An interfacing inductor (L_C) is connected with converter to suppress the AC output ripples.

3 Description of 2-level/5-level CHB converter

The two level and five level converters being employed for harmonic mitigation and load compensation are shown in Fig. 2a, b. At the nominal frequency and voltage, the DC to AC conversion produces a two level output voltage in conventional 2L converter. However, this converter has some demerits like higher switching stresses, higher distortion and reduced efficiency. 5L-CHB converter generates sinusoidal AC with reduced harmonics as compared to conventional 2L converter [24].

A single H-bridge converter with four switches and one DC link voltage generates a three-level output voltage i.e. + V_{dc}, − V_{dc}, and 0 V [25]. This multilevel converter (MLC)

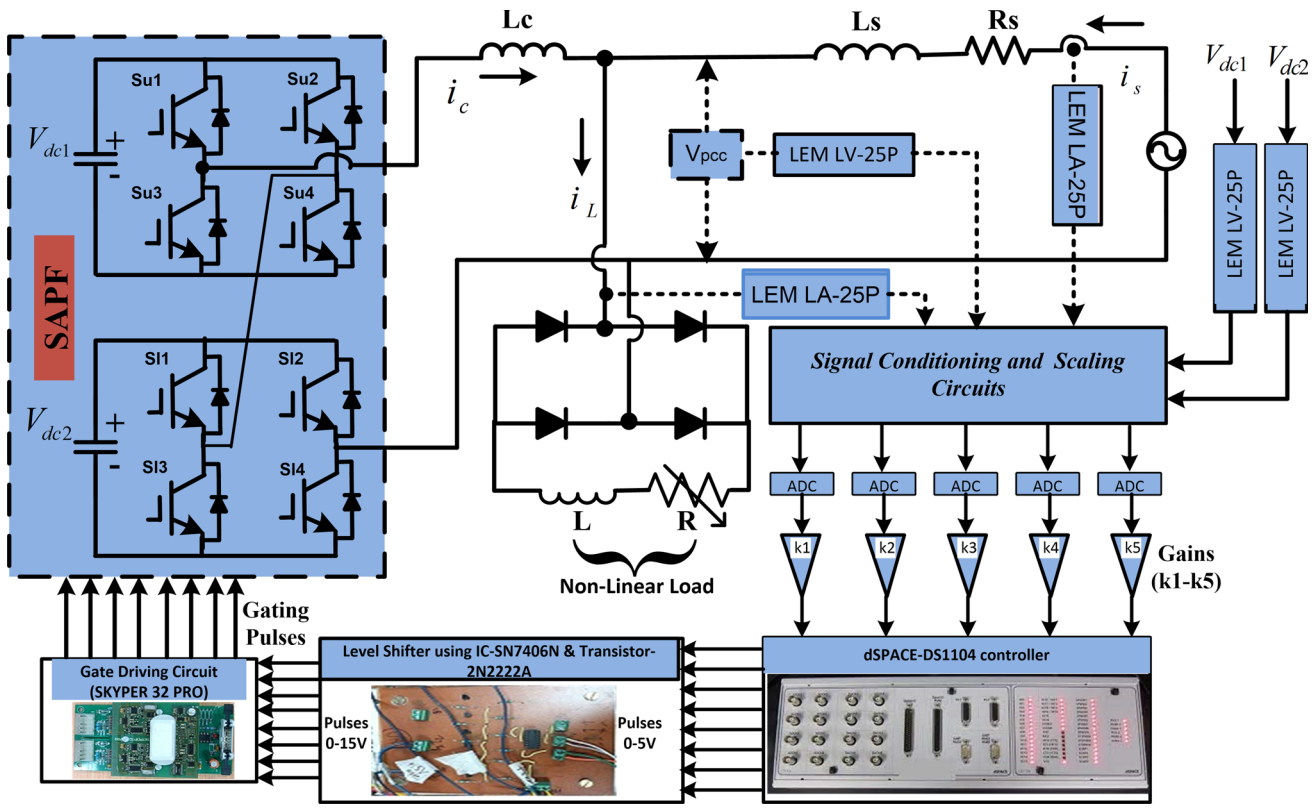


Fig. 1 System configuration of proposed system

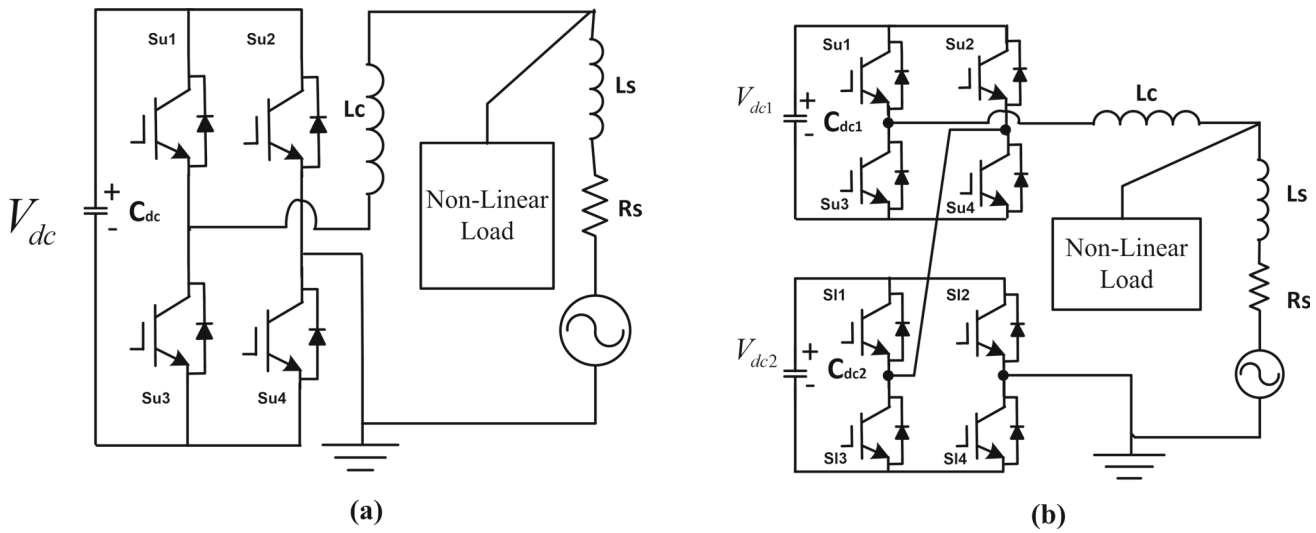


Fig. 2 **a** Two level SAPF based systems **b** Five level SAPF based systems

configuration in contrast has higher number of levels which results in stepped output voltage waveform having lesser harmonics. A full-bridge converter is also known as H-bridge cell as shown in Fig. 2a. A H-bridge circuit comprises four Insulated Gate Bipolar Transistor (IGBT) switches equipped with four freewheeling diodes.

CHB converters have been widely used in many applications such as grid interconnection due to its modular structure, high quality output waveforms and high power application in industries etc. Since the technology is advantageous to increase the converter capability as well as to improve the output voltage quality. Various phase-shifting techniques are used to reduce harmonic distortions in the output voltage waveform and bring it close to sinusoidal waveform [26].

The 5-level CHB converter consists of double H-bridge configuration each having its own DC link and IGBT switches. The number of H-bridges can be increased to increase the number of levels. However, this leads to increased power switch requirement. The design of MLC becomes complex for higher voltage steps and it is more costly and bulky also. So, in this paper, our scope is limited to 5L CHB configuration. In Fig. 2b, two H-bridges with separate DC link voltage are connected in series to generate higher voltage levels and reduced total harmonic distortion (THD) [27] in the output voltage. The 5L CHB converter is switched on so as to obtain possible voltage levels + $2V_{dc}$, + V_{dc}, 0, − V_{dc}, − $2V_{dc}$.

4 Mathematical modeling and formulation of control algorithm

In this section, Conventional Radial Basis Function Neural Network, Adaptive Radial Basis Function Neural Network and Harmonic extraction are discussed. The mathematical modeling and derivations are presented.

4.1 Conventional radial basis function neural network (C-RBFNN)

The Conventional Radial Basis Function (CRBF) Neural Network has the capability to approximate any continuous function with arbitrary accuracy if there are enough hidden neurons. The CRBF Neural Network also has the benefit of being a universal approximator, which means it can approximate any function regardless of its complexity or dimensionality. The CRBFNN structure is shown in Fig. 3a. The structure contains three layers and it is a feed-forward network. The first layer is the input layer that receives the input data, which is typically a vector of values that represent the characteristics of the data under analysis and the input signals (x) travel from this layer to the second layer. The second layer is a hidden layer made up of RBF neural units. A Gaussian function's centre is represented by each neuron in the hidden layer. Each neuron's level of activation indicates how near the input vector is to the centre point that neuron represents. The third layer is known as the output layer and it is a linear combination of activation function from the hidden layer. The weights are updated in the output layer with LMS control technique.

Neural networks are proven to have good convergence, simple structure and faster learning capability [28–30]. In Fig. 3b, the circuit diagram of CRBFNN controller is shown. The load current (i_L) is passed through activation function (f_k) to estimate the actual output (L_{OUT}) which is compared with desired output d(n) to estimate error e(n). The updated weight is calculated by LMS control technique.

Figure 3a shows, the input layer of the p_{th} node X_p = [x_1, x_2,….x_n], the centre vector of the k_{th} node g_k = [g_1, g_2,…..g_m], Euclidean distance between the input layer and RBF layer is ||dist||, bias width of kth node is b_k.

The activation function of kth node in the RBF layer is

$$f_k = \exp(-S_k^2/2b_k^2) \tag{1}$$

where,

$$S_k = \|X_p - g_k\|$$
$$f_k = \exp(-||X_p - g_k||^2/2b_k^2) \quad k = 1, 2 \ldots\ldots m,\ p = 1, 2 \ldots\ldots n \tag{2}$$

Now, L_{OUT} is the actual output of CRBFNN which is generated after the hidden layer and weight updation,

$$L_{out} = w_{p1} f_1 + w_{p2} f_2 + \ldots\ldots + w_{pm} f_m \tag{3}$$

$$L_{out} = \sum_{r=1}^{m} w_{pr} f_r \tag{4}$$

where, w_{p1}, w_{p2},……., w_{pm} are the updated weights.

In CRBFNN, the weight is updated with the help of Least Mean Square (LMS) controller and the error as shown in Eqs. (5–6).

$$w(t+1) = w(t) - \gamma \frac{\partial C(t)}{\partial w(t)} \tag{5}$$

$$e(n) = [d(n) - L_{OUT}] \tag{6}$$

where C(t) = [1/2 e^2(t)] denotes the minimum cost function and γ denotes the convergence factor having value less than 1.

4.2 Adaptive radial basis function neural network (A-RBFNN)

This neural network has been realized due to its fast self-learning ability, good generalization ability and the property of universal approximation which has been used for system control and development of nonlinear system model. Figure 3c shows the circuit diagram of A-RBFNN and it is used to extract fundamental weight of load current.

The proposed A-RBFNN technique, the dynamics of the weight is control with the help of Adaptive varying gain G as shown in Eq. (7).

$$G(n+1) = G(n) - \gamma \frac{\partial\, w_{pr}}{\partial G} \tag{7}$$

In circuit diagram of A-RBFNN as shown in Fig. 3c in which only one neuron (n = 1) is considered, the output

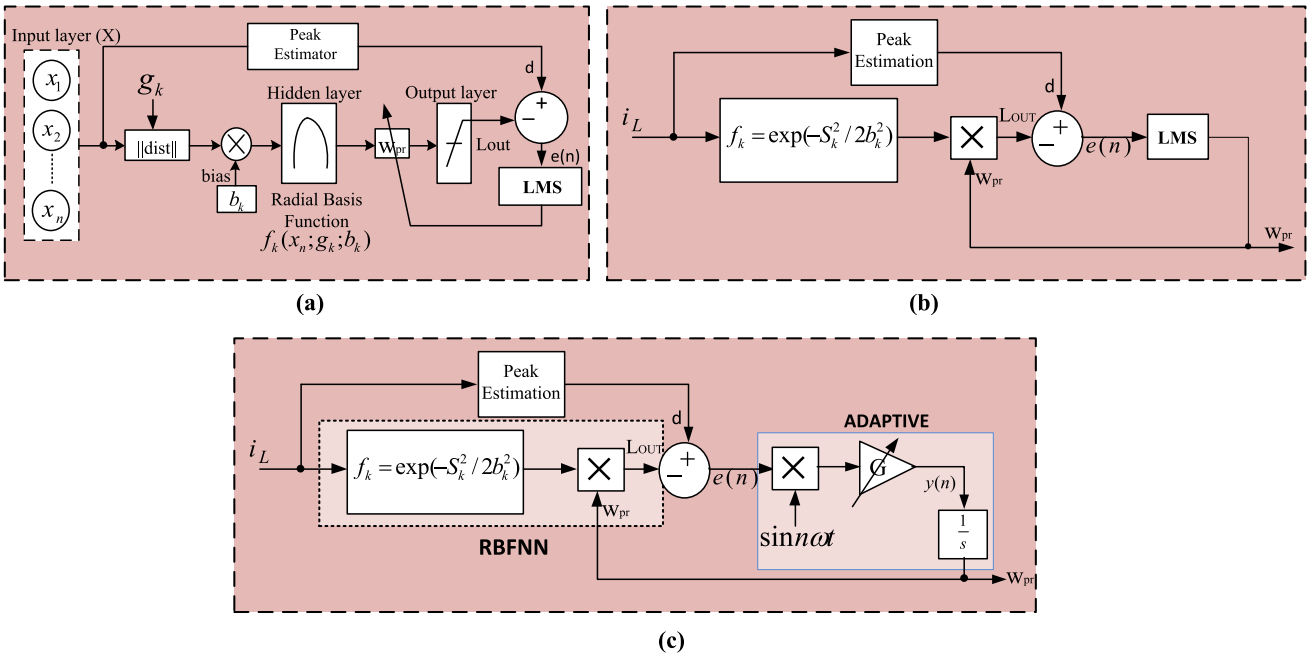


Fig. 3 **a** Structure of CRBFNN [31] **b** Circuit diagram of CRBFNN **c** Circuit diagram of A-RBFNN

'w_{pr}' is calculated as

$$w_{pr} = \int G(n)\, e\, \sin\omega t\; d\omega t \tag{8}$$

$$\therefore w_{pr} = -G(n)\, e\, \cos\omega t \tag{9}$$

The partial derivative of the cost function with respect to the weight vector is

$$\frac{\partial w_{pr}}{\partial G(n)} = -e\, \cos\omega t \tag{10}$$

Using Eqs. (7), (9) and (10)

$$G(n+1) = G(n) + \gamma\, e\, \cos\omega t \tag{11}$$

Thus, the gain varies adaptively in A-RBFNN technique until convergence is obtained.

4.3 Harmonic extraction

The distortion in current waveform is due to the harmonics injected from the nonlinear load. The circuit diagram of A-RBFNN for harmonic extraction is as shown in Fig. 3c. The load current (i_L) contains the sum of both fundamental and harmonic components, which can be represented as

$$i_L(t) = i_{Lh1} + i_{Lh2} + i_{Lh3} + i_{Lh4} + \ldots\ldots\ldots + i_{Lhn} \tag{12}$$

where i_{Lh1}, i_{Lh2}, i_{Lh3}..... i_{Lhn} denote the fundamental, 2nd, 3rd….nth order harmonic components of the load current respectively.

$$\begin{aligned} i_L(t) = {} & i_{L1}\sin(\omega t+\theta) + i_{L2}\sin(2\omega t+\theta) \\ & + i_{L3}\sin(3\omega t+\theta) + \cdots\cdots\cdots \\ & + i_{Ln}\sin(n\omega t+\theta) \end{aligned} \tag{13}$$

where ω is the angular frequency of the fundamental component; θ is the phase angle; i_{L1}, i_{L2}, i_{L3} are the peak values of the fundamental, 2nd, 3rd current components respectively. Now, the Eq. (13) can be expressed as

$$\begin{aligned} i_L(t) = {} & i_{L1}[\sin\omega t\cos\theta + \sin\theta\cos\omega t] \\ & + i_{L2}[\sin 2\omega t\cos\theta_2 + \cos 2\omega t\sin\theta_2] \\ & + \cdots\cdots + i_{Ln}[\sin n\omega t\cos\theta_n + \cos n\omega t\sin\theta n] \end{aligned} \tag{14}$$

$$\begin{aligned} i_L(t) = {} & [(i_{L1}\cos\theta)\sin\omega t + (i_{L1}\sin\theta)\cos\omega t] \\ & + [(i_{L2}\cos\theta_2)\sin 2\omega t + (i_{L2}\sin\theta_2)\cos 2\omega t] \\ & + \cdots\cdots + [(i_{Ln}\cos\theta_n)\sin n\omega t + (i_{Ln}\sin\theta_n)\cos n\omega t] \end{aligned} \tag{15}$$

The load current contains both the active and reactive components. The fundamental active load current component is $i_{Lp1} = i_{L1}\cos\theta$ and fundamental reactive load current component is $i_{Lq1} = i_{L1}\sin\theta$ and similar equations can be written for second, third and other higher order harmonics components.

Replacing $i_{L1}\cos\theta$ by w_{p1}, $i_{L1}\sin\theta$ by w_{q1}; $i_{L2}\cos\theta_2$ by w_{p2}, $i_{L2}\sin\theta_2$ by w_{q2}; and $i_{L3}\cos\theta_3$ by w_{p3}, $i_{L3}\sin\theta_3$ by w_{q3} respectively.

$$i_L(t) = [w_{p1}\sin\omega t + w_{q1}\cos\omega t] + [w_{p2}\sin 2\omega t$$

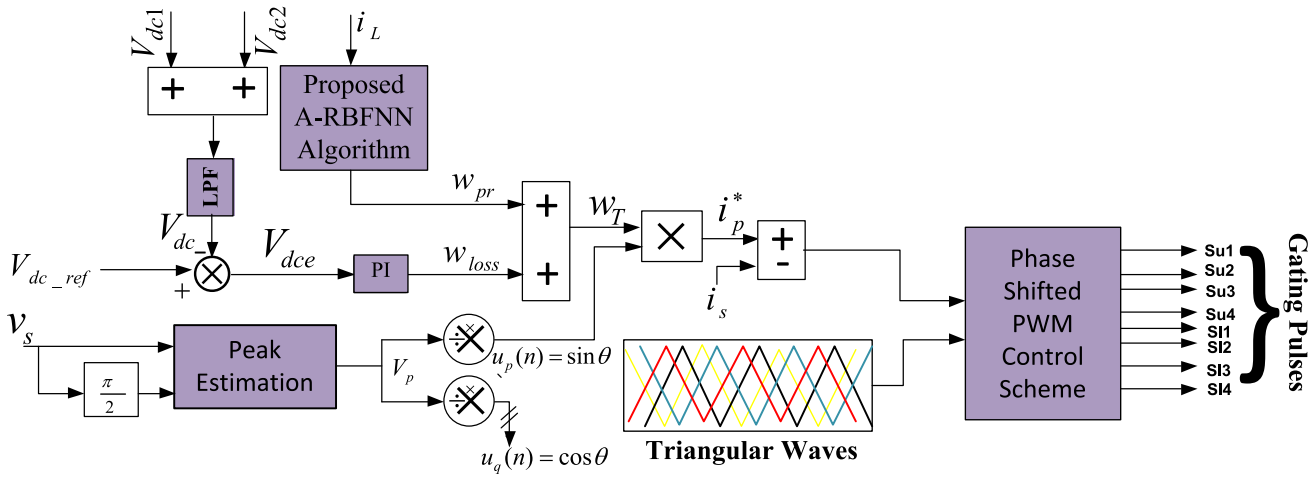


Fig. 4 Control diagram of proposed Adaptive-RBFNN based algorithm

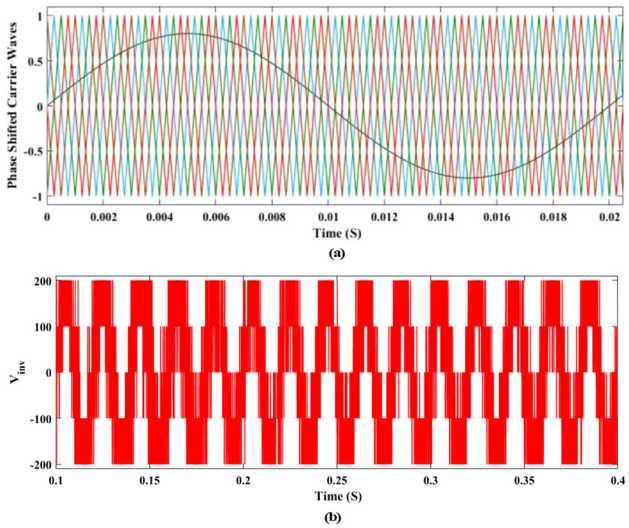


Fig. 5 Waveform showing **a** Phase Shifted-PWM techniques of 5L-CHB converter **b** five level converter voltage output of CHB

$$+ w_{q2}\cos 2\omega t] + [w_{p3}\sin 3\omega t + w_{q3}\cos 3\omega t] + \cdots\cdots\cdots + n\,terms \quad (16)$$

The reactive current components are w_{q1}, w_{q2}, w_{q3} and even harmonic component are neglected and active current component is considered to calculate reference current.

Where $w(h) = [w_{p1}, w_{p3}, w_{p5}, \ldots\ldots, w_{pn}]$ are the estimated active harmonic components of load current.

$$i_L(t) = w_{p1}\sin\omega t + w_{p3}\sin 3\omega t + w_{p5}\cos 5\omega t + \cdots\cdots\cdots + w_{pn}\cos n\omega t \quad (17)$$

To estimate w(h) the value of 'n' is updated in Fig. 3c as per harmonic order of the load current. Further; Eq. (18) can be used to calculate estimated load current upto 7th harmonics (n = 7) shown as,

Let, i_L be the actual nonlinear load current and i_{est} be an estimated current to evaluate the error signal e(h).

$$i_{est}(n) = [w_{p1}\sin\omega t + w_{p3}\sin 3\omega t + w_{p5}\sin 5\omega t + w_{p7}\sin 7\omega t] \quad (18)$$

$$e(h) = i_L - i_{est} \quad (19)$$

5 Control algorithm of the proposed system

The control diagram of proposed algorithm is shown in Fig. 4 for 5L converter topology. In this section, unit template, active weight loss component and the reference grid current calculation are discussed. The proposed algorithm is used to update the estimated weight signals output which is the fundamental component of the load current. The weight loss component is calculated by processing the DC-link voltage through the PI controller. The unit template is multiplied with the sum of the fundamental weights and the PI controller's output to produce the reference current, i_p^*.

5.1 Estimation of unit templates

In order to determine unit templates, the grid voltage (v_s) and its quadrature phase shifted component (v_j) as shown in Fig. 4, are passed through the peak estimation block. In order to calculate peak voltage (V_p) Eqs. (20–21) are used.

$$v_s = V_p\sin\omega t \quad (20)$$

$$v_j = V_p(\sin\omega t + \frac{\pi}{2}) \quad (21)$$

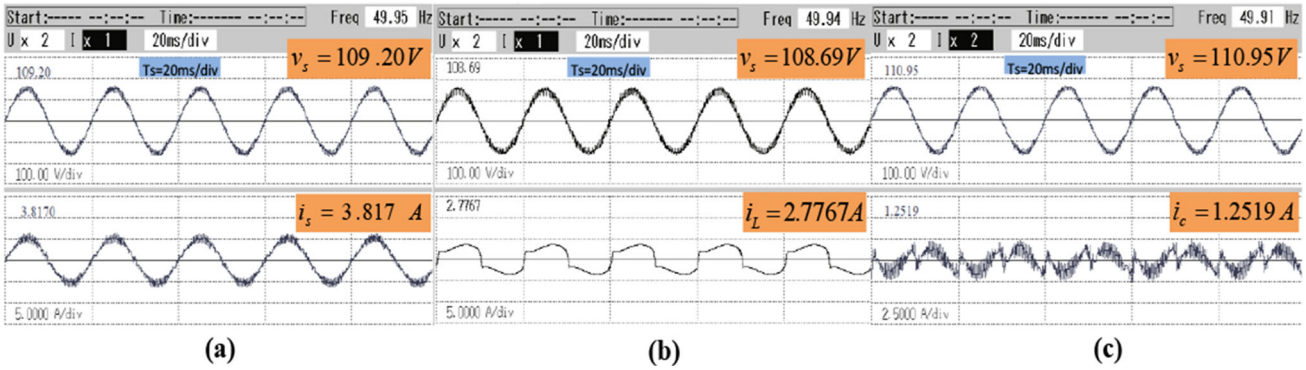


Fig. 6 Steady State experimental results for 2L converter configuration **a** Source current i_s **b** Load Current i_L **c** Compensating current i_c

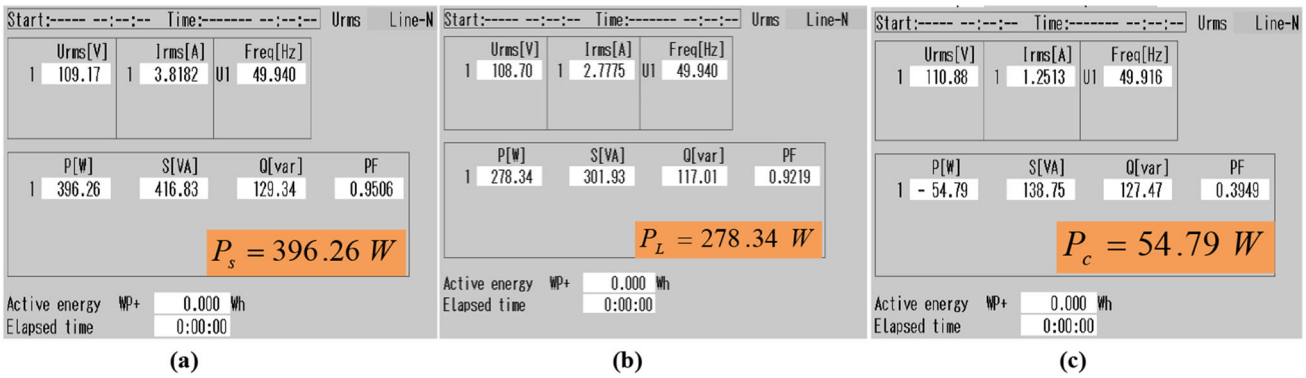


Fig. 7 **a–c** Steady State experimental results of active and reactive power for 2L converter configuration **a** source side (Ps) **b** load side (P_L) **c** compensating side (Pc)

Now, the peak voltage is computed as,

$$V_p = \sqrt{v_s^2 + v_j^2} \tag{22}$$

The unit template is computed as,

$$u(n) = \frac{v_s}{V_p} = \sin \omega t \tag{23}$$

5.2 Estimation of the active weight loss component

The DC link voltage needs to be maintained to a minimum reference value for active shunt filtering. The DC link reference voltage V_{dc_ref} and the total DC link voltage V_{dc} is compared to generate error V_{dce} as shown by Eq. (24). The Proportional Integral (PI) controller is now fed with the error V_{dce} in order to estimate the active weight loss component w_{loss} as per Eq. (25).

$$V_{dce} = V_{dc_ref} - V_{dc} \tag{24}$$

$$w_{loss} = w_{pr} + [k_P\{V_{dce}(n+1) - V_{dce}(n)\} + k_i\{V_{dce}(n+1)\}] \tag{25}$$

where k_p denotes the proportional constant gain and k_i is the integral controller gain.

For the considered system, each DC link was designed for 100 V. Hence, total DC link voltage was selected as 200 V.

5.3 Generation of grid reference current and gating pulses

The total weight w_T as shown in Eq. (26) is computed by the addition of both fundamental weight w_{pr} and active weight loss component w_{loss}. The fundamental weight w_{pr} is calculated by the proposed A-RBFNN algorithm. Now, the product of the total weight w_T and unit template u(n) is used to generate reference current i_p^* as shown in Eq. (27).

$$w_T = w_{pr} + w_{loss} \tag{26}$$

$$i_p^* = w_T * u(n) \tag{27}$$

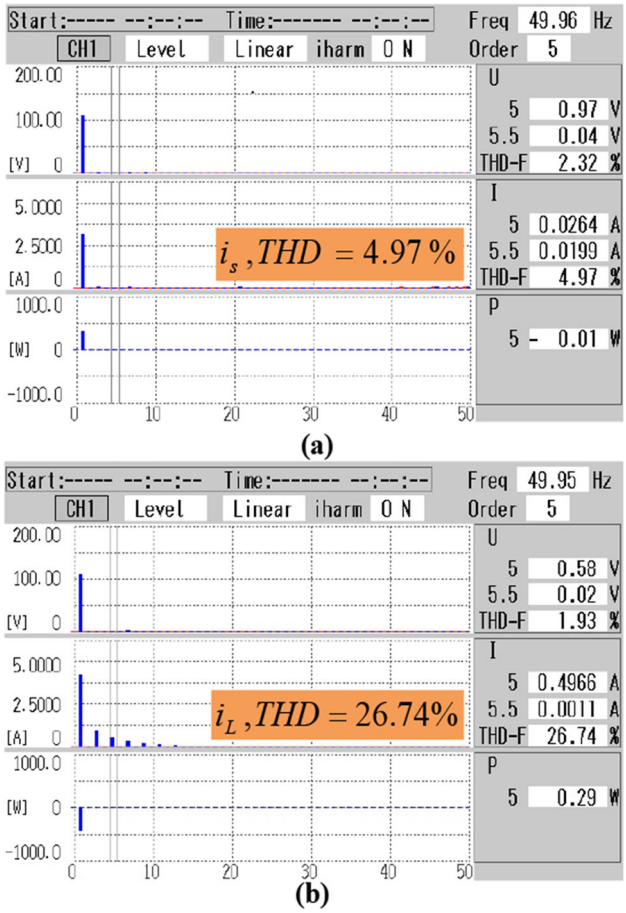


Fig. 8 Steady State experimental results for 2-level converter configuration **a** i_s, THD 4.97% **b** i_L, THD 26.74%

Eight switching pulses need to be generated so as to regulate the 5L-CHB converter. Hence, the reference current ip* is compared with triangular carrier wave of 5 kHz. Figure 5a shows the comparison of reference current signal and carrier waveform, using Phase Shifted Pulse Width Modulation (PS-PWM) technique to generate gating pulses. Here, four carrier waveforms are required in 5L CHB converter and these are compared with grid reference current (i_p*) so as to generate eight gating pulses. Figure 5b shows the obtained five level converter voltage output (V_{inv}) of CHB.

6 Experimental results

In the laboratory, single phase system incorporating 2L as well as 5L CHB converter is developed to verify the feasibility of proposed technique. The system rating is 110 V rms and 50 Hz frequency. The hardware setup for 2L converter uses two voltage sensors (LV-25P) for sensing (v_s, V_{dc}) and two current sensors (LA-25P) for sensing (i_s, i_L). The hardware setup for 5L converter features three voltage sensors (LV-25P) for sensing (v_s, V_{dc1}, V_{dc2}) and two current sensors (LA-25P) for sensing (i_s, i_L). These sensed voltage and current signals are processed by DSP (dSPACE 1104) to generate four gating pulses for 2L and eight gating pulses for 5L-CHB converter. The 2L and 5L CHB converters are controlled to inject compensating current at point of common coupling. Hardware results for steady state and dynamic states are captured using Digital Signal Oscilloscope (DSO) and Mixed Signal Oscilloscope (MSO).

6.1 Steady state results

In this section, steady state experimental results for single phase system incorporating 2-level and 5-level SAPF converters are discussed in detail.

6.1.1 Two level converters configuration

Figure 6a–c show the supply voltage (v_s) along with source current i_s, load current i_L and compensating current i_c. The load current is quite distorted while supply current is sinusoidal. The supply current is also observed to be in phase with supply voltage as shown in Fig. 6a whereas load current i_L is distorted as shown in Fig. 6b. The steady state active and reactive powers are shown in Fig. 7a–c. The active and reactive power demand of load is 278.34 W and 117.01 VAR. The grid supplies 396.26 W active and 129.34 VAR reactive power respectively. The compensator provides 54.79 W and 127.47 VAR active and reactive power respectively. The system active power demand is met by grid whereas most of the reactive power is provided by compensator. Thus, power factor at the supply end is improved from 0.92 to 0.95.

Figure 8a, b show the THD analysis of source current and load current. The THD of source current (i_s) is 4.97% in Fig. 8a and THD of i_L is 26.74% as shown in Fig. 8b. The load current is nonlinear in nature (rectifier with RL load) having high THD. The THD has been reduced by connecting SAPF at PCC and observed to be within the limits prescribed by IEEE standard.

6.1.2 Five level converter configuration

The results for the 5L converter configuration are now discussed. Figure 9a–c show the supply voltage (v_s) along with source current i_s, load current i_L and compensating current i_c. After compensation, the supply current (i_s) is observed to be in phase with supply voltage and sinusoidal as shown in Fig. 9a while the load current i_L is distorted as shown in Fig. 9b. The steady state active and reactive powers are shown in Fig. 10a–c. The active and reactive power demand of load is 287.16 W and 117.64 VAR. The grid supplies 307.79 W active and 68.72 VAR reactive power respectively. The compensator powers are 19.50 W and 120.48 VAR respectively.

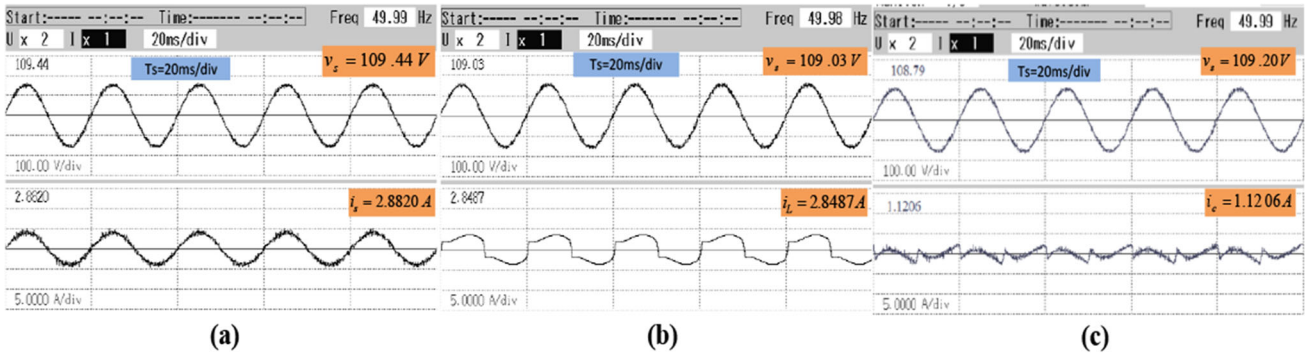


Fig. 9 Steady State experimental results for five level converter configuration **a** source current i_s **b** load Current i_L **c** compensating current i_c

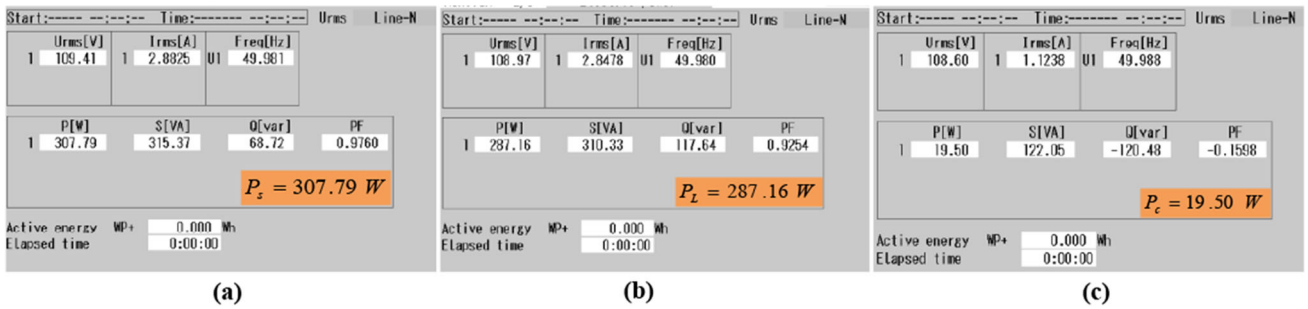


Fig. 10 **a–c** Steady State experimental results of active and reactive power for five level converter configurations **a** source side (Ps) **b** load side (P_L) **c** compensating side (Pc)

It is observed that the system active power demand is met by grid whereas most of the reactive power is provided by compensator. Thus, power factor at the supply end is improved from 0.92 to 0.97.

Figure 11a, b shows the THD analysis of source current and load current. The THD of i_s is 3.77% as shown in Fig. 11a and THD of i_L is 27.80% as shown in Fig. 11b. The load current is nonlinear (rectifier with RL load) having high THD. The source current THD has been reduced due to SAPF action and brought within prescribed IEEE standard.

6.2 Dynamic state results

In this section, the experimental results are discussed under dynamic load variations. The waveform of source voltage, source current, load current, compensating current, fundamental current, DC link voltage of first and second converter, total DC link voltage are observed under load variation for two level as well as five level converter configurations.

6.2.1 Two level converter configuration

Figure 12a, b show the waveforms of source voltage v_s, source current i_s, load current i_L, compensating current i_c and DC link voltage V_{dc}. Figure 12a shows when the load current is increased, the supply current also increases and remains sinusoidal in shape. The supply voltage and current are in same phase after compensation. In Fig. 12b, when there is a sudden increase in load, the DC link voltage dips but soon settles to its steady state value. In Fig. 12c, the waveform of v_s, i_L, w_{pr}, i_{ref} are shown. When there is an increase in load, the fundamental active weight (w) also increases and attains new steady state value quickly. Figure 12d shows waveform of i_L, w_{pr}, i_{ref}, i_s. The reference current i_{ref} also follows the change in load. Figure 12e shows the waveform of i_L, e, f_k, w_{pr}. It is observed that during, a sudden increase in load, the intermediate parameters (e, f_k, w_{pr}) also change and attain new steady state value.

6.2.2 Five level converter configurations

Figure 13a, b shows the waveforms of source voltage v_s, source current i_s, load current i_L, DC link voltage V_{dc} and

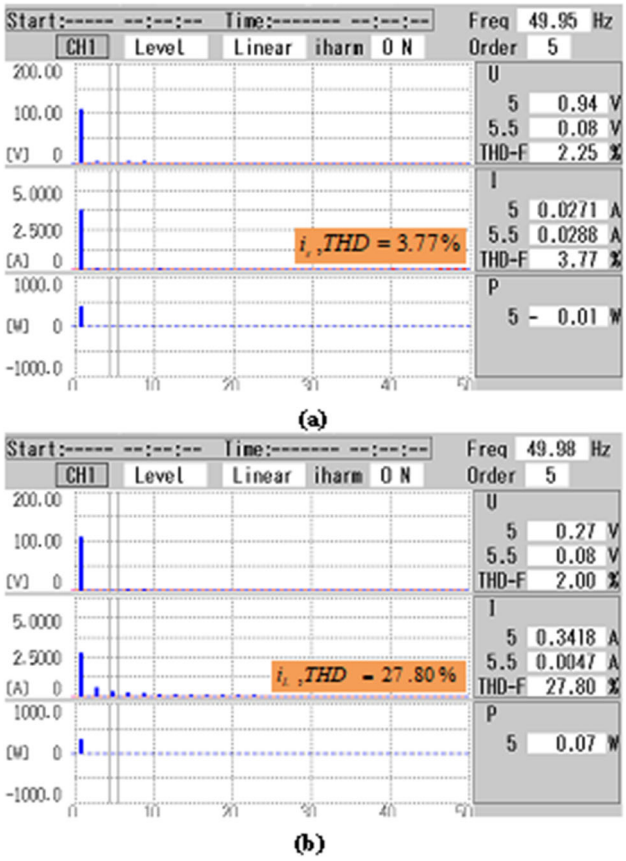


Fig. 11 Steady State experimental results for 5L converter configuration **a** i_s, THD 3.77% **b** i_L, THD 27.80%

v_s, i_s, i_L, compensating current i_c under load increase condition. The DC link voltage is well regulated to its reference value of 200 V under load dynamics as shown in Fig. 13a. Figure 13b shows, the supply voltage and current are in same phase before and after load increase. The waveform of v_s, i_L, w_{pr}, i_{ref} are shown in Fig. 13c. It is to be observed that, the load and source current both increase as expected. The reference current also changes with change in load current. In less than one cycle, the fundamental active weight also increases and reaches a new steady state value due to effective control action of A-RBFNN algorithm.

Figure 13d shows the waveforms of i_L, V_{dc1}, V_{dc2}, V_{dc}. It is observed that, there is sudden increase in load, the DC link voltage drops but soon settles to its steady state value. Figure 13e shows the intermediate signal waveforms of i_L, e, f_k, w_{pr}. It is seen that the intermediate parameters (e, f_k, w_{pr}) also change whenever there is a sudden increase in load.

Figure 14a shows the dynamic waveforms for i_L, $w_{p1}\sin\omega t$, $w_{p3}\sin 3\omega t$, $w_{p5}\sin 5\omega t$, $w_{p7}\sin 7\omega t$, i_{est}. At load increase, the fundamental, third, fifth and seventh harmonics also increases proportionally. The estimated load current i_{est} is obtained by the sum of fundamental to seventh harmonics.

Figure 14b depicts the waveform of i_L, w_{p1}, w_{p3}, w_{p5}, w_{p7}, error (e). As the load increases, the load current fundamental, third, fifth and seventh weight component also increases. The error e is calculated by the load current deducted to the sum of fundamental to seventh harmonics and observed that the error obtained is near to zero.

7 Comparison of two level and five level CHB converter

The proposed control algorithm was tested on two level and 5-level CHB converter connected as SAPF at PCC. The comparative analysis of 2L and 5L CHB converters are shown in Table 1. The topology and performance results for both 2-level and 5-level converters are thoroughly analyzed. It is concluded that 5L-CHB converter configuration shows better response in terms of THD of source current, lower switching losses and fast transient response. Although, number of IGBT switches increases from four (in 2L) to eight in 5L CHB configuration but this can be justified as the source current has much lower THD (3.77%) at similar value of load current than for 2L (4.97%). Moreover, the voltage stress on IGBT switches is less in 5L-CHB configuration. The output waveform of 5-L converter is more near too sinusoidal as compared to 2L.

8 Conclusion

In this paper, an Adaptive-RBFNN algorithm is applied to a single phase system for mitigation of power quality issues. The proposed control technique is able to estimate the magnitude of fundamental load current effectively. The control algorithm is tested with both two level and five level converter configurations along with a non-linear load. It is observed that, the performance of the hardware system with 5L converter configuration is better in terms of source current THD (3.77%), reduced voltage stress etc. The THD obtained using 2L is 4.97%. The 5L-CHB also shows faster dynamic response along with improved steady state response. A detailed comparison is also discussed in this paper. The proposed algorithm A-RBFNN is successfully implemented in this paper.

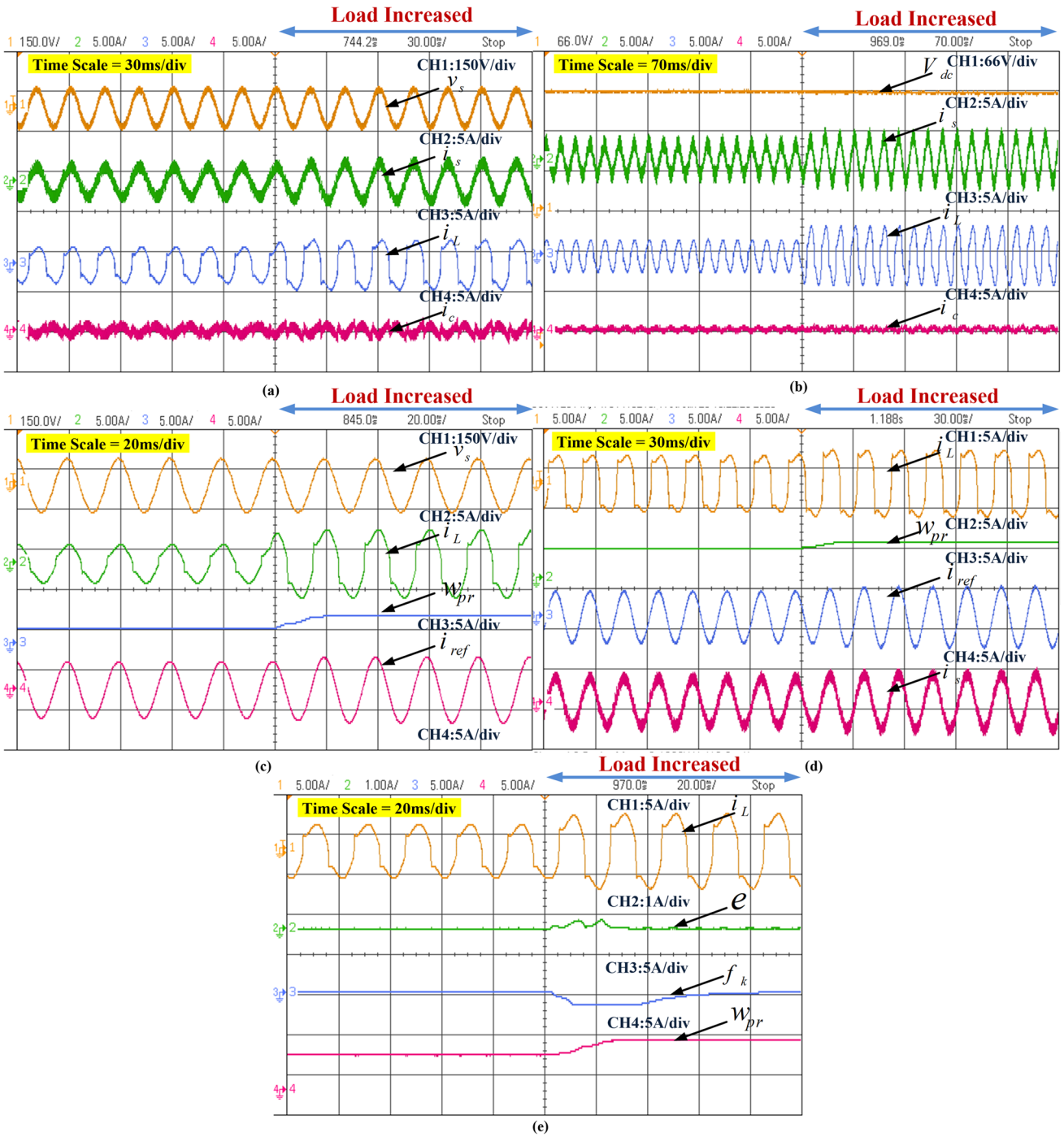


Fig. 12 Dynamic State experimental waveforms for two level SAPF based systems under load change **a** v_s, i_s, i_L, i_c **b** V_{dc}, i_s, i_L, i_c **c** v_s, i_L, w_{pr}, i_{ref} **d** i_L, w_{pr}, i_{ref}, i_s **e** i_L, e, f_k, w_{pr}

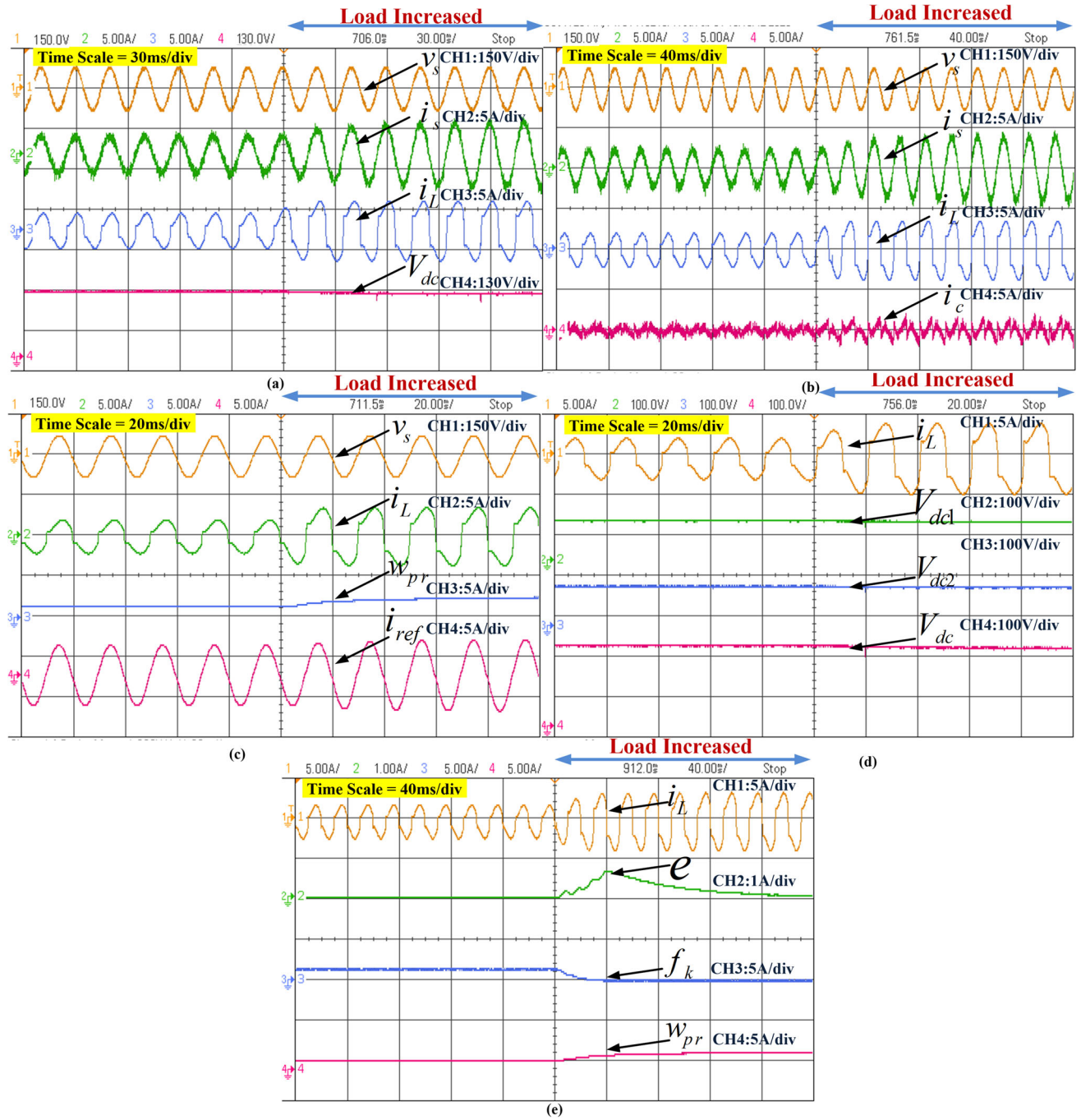


Fig. 13 Dynamic State experimental waveforms for five level SAPF based systems under load change **a** v_s, i_s, i_L, V_{dc} **b** v_s, i_s, i_L, i_c **c** v_s, i_L, w_{pr}, i_{ref} **d** i_L, V_{dc1}, V_{dc2}, V_{dc} **e** i_L, e, f_k, w_{pr}

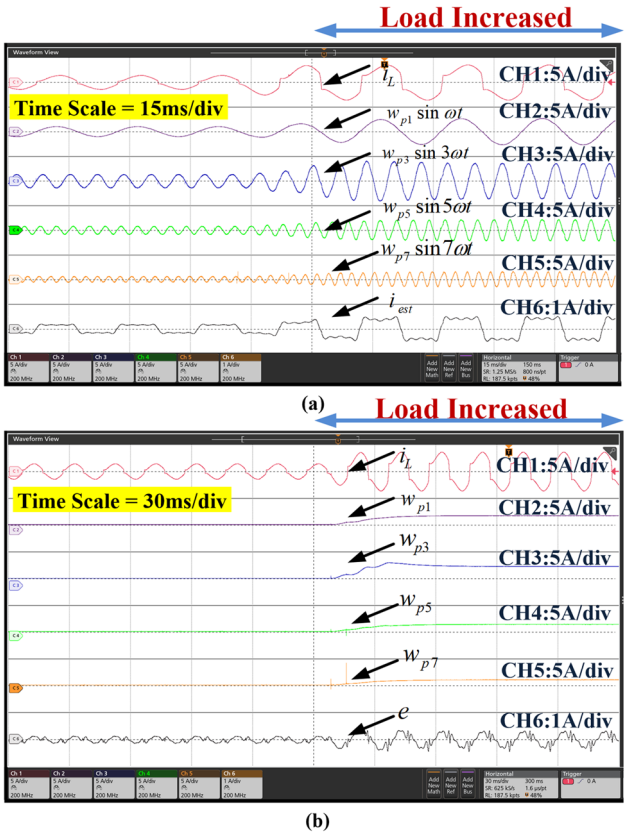


Fig. 14 Dynamic State experimental waveforms for five level inverter configuration under load change **a** i_L, $w_{p1}\sin\omega t$, $w_{p3}\sin 3\omega t$, $w_{p5}\sin 5\omega t$, $w_{p7}\sin 7\omega t$, i_{est} **b** i_L, w_{p1}, w_{p3}, w_{p5}, w_{p7}, e

Table 1 Comparison aspects of 2L and 5L converters

Parameters	2L Converter	5L CHB Converter
Number of IGBT switches	Four	Eight
Voltage stress	High (100 V)	Less (50 V)
THD, i_s	4.97%	3.77%
THD, i_L	26.74%	27.80%
Switching frequency	5 kHz	5 kHz
No. of DC link capacitor	One	Two
Power factor, i_s	0.9506	0.9760
Power factor, i_L Load side	0.9219	0.9254
Nature of output voltage waveform	Square	Stepped
Number of levels in output voltage	Two level	Five level

Acknowledgements The authors are thankful to the DST for the FIST project SR/FST/ET-I/2022/1026. Author declares that the manuscript is original, has not been published before and is not currently being considered for publication elsewhere. Author confirms that the manuscript has been read and approved for submission by all the named authors.

Author contributions The highlights of the researcher's contributions are: 1.Development of an Adaptive A-RBFNN algorithm with detailed mathematical analysis. 2.The application of the developed algorithm and comparison is tested on both 2L and 5L converter topologies. 3.An application of A-RBFNN for segregation of different harmonic orders is also added. 4.Extensive steady state and dynamic results are presented for a fair comparison of two converter topologies.

Declarations

Conflict of interest The authors declare no competing interests.

Appendix A

Experimental parameters

Grid supply voltage (v_s) = 110 V, nonlinear load resistance and inductance, R = 125 Ω and L = 80 mH, DC link voltage capacitance (C_{dc}) = 2400 μF, $V_{dc\text{-}ref}$ = 200 V, interfacing inductor (L_c) = 5 mH, switching frequency (f_{sw}) = 5 kHz, sampling time = 50 μs.

References

1. Montoya FG et al (2018) Power quality: scientific collaboration networks and research trends. Energies 11(8):2067
2. Michalec Ł et al (2021) Impact of harmonic currents of nonlinear loads on power quality of a low voltage network-review and case study. Energies 14(12):3665
3. Wu Z et al (2021) Transfer mechanism analysis of injected voltage harmonic and its effect on current harmonic regulation in FOC PMSM. IEEE Trans Power Electron 37(1):820–829
4. Hassan A et al (2020) A state of the art of the multilevel inverters with reduced count components. Electronics 9(11):1924
5. Wang Q et al (2020) Comprehensive comparison between two-level, three-Level, and hybrid three-level SiC inverter for high power high speed drive system. In: 2020 IEEE 9th international power electronics and motion control conference (IPEMC2020-ECCE Asia). IEEE
6. Abd Halim W et al (2016) Review of multilevel inverter topologies and its applications. J Telecommun Electron Comput Eng (JTEC) 8(7):51–56
7. Aishwarya V et al (2021) Review of reduced-switch multilevel inverters for electric vehicle applications. Int J Circuit Theory Appl 49(9):3053–3110
8. Mahato B et al (2018) Generalized symmetrical/asymmetrical single-phase MLI topology. In: 2018 International conference on recent trends in electrical, control and communication (RTECC). IEEE
9. Kundu S, Singh M, Giri AK et al (2024) Robust and fast control approach for islanded microgrid system and EV charging station applications. Electr Eng. https://doi.org/10.1007/s00202-024-02291-1

10. Shi JA et al (2014) Genetic algorithm optimized fuzzy control of DSTATCOM for improving power quality. In: Proc. Australasian Uni. Power Eng. Conf., Sept. 28–Oct. 1 2014, pp 1–6
11. Kumar N et al (2019) ANOVA kernel Kalman filter for multi-objective grid integrated solar photovoltaic-distribution static compensator. IEEE Trans Circuits Syst I Regul Pap 66(11):4256–4264
12. Bansal P et al (2022) Nonlinear adaptive normalized Huber control algorithm for 5-level distribution static compensator. Electr Eng 104(3):1635–1648
13. Mittal S, Singh A, Chittora P (2022) EV Control in G2V and V2G modes using SOGI controller. In: 2022 IEEE 3rd Global conference for advancement in technology (GCAT), Bangalore, India, pp 1–6.https://doi.org/10.1109/GCAT55367.2022.9972182
14. Hoon Y et al (2018) PLL-less SRF technique for operation of SAPF under unbalanced and distorted supply voltage, pp 20–28
15. Agrawal S et al (2018) Performance evaluation of 3-phase 4-wire SAPF based on synchronizing EPLL with fuzzy logic controller. In: 2018 2nd IEEE international conference on power electronics, intelligent control and energy systems (ICPEICES). IEEE
16. Mittal S, Singh A, Chittora P (2024) Design and development of leaky least mean fourth control algorithm for single-phase grid-connected multilevel inverter. Int J Circuit Theor Appl 52(1):328–345. https://doi.org/10.1002/cta.3746
17. Mittal S, Singh A, Chittora P (2023) Solar PV array based grid-connected bi-directional EV charger controlled using NARLMMN algorithm. In: 2023 9th IEEE India international conference on power electronics (IICPE), Sonipat, India, pp 1–6. https://doi.org/10.1109/IICPE60303.2023.10475089
18. Abdolrasol MGM et al (2021) Artificial neural networks based optimization techniques: a review. Electronics 10(21):2689
19. Rai KB et al (2022) Design and analysis of Hermite function-based artificial neural network controller for performance enhancement of photovoltaic-integrated grid system. Int J Circuit Theory Appl
20. Arora A et al (2022) Design and analysis of quadratic bernstein functional blending neural network for shunt compensation and phase locked loop. Electr Eng 104(5):3631–3647
21. Arora A et al (2019) Design and analysis of functional link artificial neural network controller for shunt compensation. IET Gener Transm Distrib 13(11):2280–2289
22. Arthy G et al (2016) Immune RBF neural network algorithm for DSTATCOM. In: 2016 International conference on computer communication and informatics (ICCCI). IEEE
23. Jayachandran J et al (2017) Neural network based ILST control strategy for a DSTATCOM with Solar photovoltaic system for power quality improvement. Int Rev Autom Control (IREACO) 10(5):399–409
24. Patel HK et al (2021) A new seven-level asymmetrical inverter for reduced component with lesser number of power electronics devices. In: Proceedings of integrated intelligence enable networks and computing: IIENC 2020. Springer Singapore
25. Castillo R et al (2018) Single-phase hybrid cascaded H-bridge and diode-clamped multilevel inverter with capacitor voltage balancing. IET Power Electron 11(4):700–707
26. Mahato B et al (2019) Multilevel inverter with optimal reduction of power semi-conductor switches. In: Renewable energy and its innovative technologies: proceedings of ICEMIT 2017, vol 1. Springer Singapore
27. Mahato B, Mittal S, Nayak PK (2018) Novel topology of multilevel inverter for higher voltage steps. In: 2018 International conference on recent trends in electrical, control and communication (RTECC). IEEE
28. Mirzaeinejad H (2018) Robust predictive control of wheel slip in antilock braking systems based on radial basis function neural network. Appl Soft Comput 70:318–329
29. Kamel AH et al (2021) RBFNN versus GRNN modeling approach for sub-surface evaporation rate prediction in arid region. Sustain Comput Inform Syst 30:100514
30. Jagtap AD et al (2020) Adaptive activation functions accelerate convergence in deep and physics-informed neural networks. J Comput Phys 404:109136
31. Konakoglu B (2021) Prediction of geodetic point velocity using MLPNN, GRNN, and RBFNN models: a comparative study. Acta Geod Geoph 56(2):271–291

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. Hydrol. Hydromech., 72, 2024, 2, 170–184
https://doi.org/10.2478/johh-2024-0004
© 2024 Vijay Kaushik et al., published by Sciendo. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Licence.

sciendo Journal of Hydrology and Hydromechanics Institute of Hydrology SAS

Prediction of shear stress distribution in compound channel with smooth converging floodplains

Vijay Kaushik[1*], Munendra Kumar[2]

[1] Department of Civil Engineering, Delhi Technological University, Delhi, 110042, India. E-mail: vijaykaushik_2k20phdce01@dtu.ac.in
[2] Department of Civil Engineering, Delhi Technological University, Delhi, 110042, India. E-mail: munendrakumar@dtu.ac.in
[*] Corresponding author.

Abstract: Climate change can have a profound impact on river flooding, leading to increased frequency and severity of floods. To mitigate these effects, it is crucial to focus on enhancing early warning systems and bolstering infrastructure resilience through improved forecasting. This proactive approach enables communities to better plan for and respond to flood events, thereby minimizing the adverse consequences of climate change on river floods. During river flooding, the channels often take on a compound nature, with varying geometries along the flow length. This complexity arises from construction and agricultural activities along the floodplains, resulting in converging, diverging, or skewed compound channels. Modelling the flow in these channels requires consideration of additional momentum transfer factors. In this study, machine learning techniques, including Gene Expression Programming (GEP), Artificial Neural Networks (ANN), and Support Vector Machines (SVM), were employed. The focus was on a compound channel with converging floodplains, predicting the shear force carried by the floodplains in terms of non-dimensional flow and hydraulic parameters. The findings indicate that the proposed ANN model outperformed GEP, SVM, and other established approaches in accurately predicting floodplain shear force. This research underscores the efficacy of utilizing machine learning techniques in the examination of river hydraulics.

Keywords: Compound channel; Converging floodplains; Shear stress distribution; Non-dimensional parameters; Machine learning approaches.

INTRODUCTION

The dynamics of flow within a river are influenced by turbulence and the presence of three-dimensional objects. Consequently, selecting an appropriate modelling technique becomes a challenge when aiming to estimate various flow characteristics, including discharge, velocity, and shear stress. Mean velocities in both the main channel and floodplain fluctuate due to variations in hydraulic conditions. The determination of velocity profiles and fluid fields is contingent upon the presence of boundary shear stress. Beyond the bankfull stage, the flow within the primary channel exerts a propulsive or accelerating force on the flow across floodplains, inherently inducing a resistive or decelerating impact on the flow within the primary channel. The pioneering work on boundary shear distributions in open channels was conducted by Leighly (1932). Numerous scholars have conducted investigations on the resolution of boundary shear distributions in streams, using either direct or indirect methodologies. The analysis of shear forces at the bed helps elucidate the mechanism of bedload transfer, while the examination of shear forces at the walls offers a comprehensive perspective on variations in the channel pattern. Several researchers have conducted studies on the flow, velocity, and shear stress distribution on compound sections (Bhattacharya, 1995; Kar, 1977; Khatua et al., 2011a, b; Knight and Demetriou, 1983; Mohanty, 2013; Myers, 1987; Patra, 1999; Patra and Kar, 2000; Patra et al., 2004; Yang and Lim, 2005). Traditional channel division techniques are used for the assessment of discharge capacity in compound channels. It is postulated that the compound section is delineated by hypothetical interface planes located at the confluence of the main channel and the floodplain. The absence of shear stress in these regions suggests a lack of momentum transmission. The momentum transfer phenomena were first examined by Sellin (1964) via laboratory research. Subsequently, several researchers discovered that the uneven distribution of boundary shear stress along the perimeter of straight compound channels could be attributed to momentum transfer. As a result, they developed an empirical equation for quantifying shear stress, as evidenced by the works of Knight and Demetriou (1983), Knight and Hamed (1984), Myers (1987), Stephenson and Kolovopoulos (1990), Patra et al. (2004), and Yang and Lim (2005). In their seminal work, Knight and Hamed (1984) proposed a theoretical framework to analyse the distribution of boundary shear stress in a compound channel with homogenous characteristics. Their model specifically focused on width ratio (α) values ranging up to 4. In their work, Khatua and Patra (2007) further investigated the subject matter by conducting more experimental observations. They subsequently formulated a model that encompasses channels with width ratios (α) ranging up to 5.25. In their study, Mohanty and Khatua (2014) proposed a novel model for a channel characterized by values of α ranging from 6.67 to 11.96. Extensive investigations were conducted in laboratory flumes to study both prismatic and meandering compound channel designs. Nevertheless, a comparison between the compound section data of prismatic compound channels and non-prismatic compound channels revealed significant discrepancies in the estimation of shear force carried by floodplains. These errors can be attributed to the omission of additional mass and momentum transfer, as elucidated by Bousmar and Zech (2002), Bousmar et al. (2004), Rezaei (2006), and Rezaei and Knight (2009). The additional momentum exchange associated with this phenomenon must be considered when doing flow modelling for non-prismatic compound channels. The study conducted by Proust et al. (2006) investigated the phenomenon of asymmetric geometry,

specifically focusing on its enhanced convergence rate. An augmented convergence angle of 22° leads to heightened levels of mass transfer and head loss. The flow behaviour of skewed, two-stage converging and diverging channels was investigated by Chlebek et al. (2010). The researchers noted an elevation in head losses resulting from the transfer of mass and momentum, as well as the homogeneity of velocity on shrinking floodplains. Additionally, they discovered an augmented velocity gradient on the expanding floodplains. The observed variations in flow distribution may be attributed to alterations in the flow force between the main channel and floodplains within different subsections. Rezaei and Knight (2011) conducted a study on compound channels with floodplains that are non-prismatic. The authors examined the depth-averaged velocity, local velocity distributions, and boundary shear stress distributions at different convergence angles. The analysis of depth-averaged velocities provides insights into the impact of contractions on velocity distributions. Notably, these contractions lead to an augmentation of velocity in close proximity to the walls of the main channel. This effect is particularly pronounced in the latter portion of the convergence reach. Additionally, for cases where the relative depth is high, the lateral flow entering the main channel also contributes to these velocity variations. The study conducted by Yonesi et al. (2013) examines the influence of roughness in floodplains on the occurrence of overbank flow in compound channels with non-prismatic floodplains. Raising the depth ratio or reducing the roughness ratio effectively reduces the velocity gradient between the main channel and the floodplain at the middle and end of the divergence stretch. The rate of change in velocity, as shown by the gradient, exhibited a rise in magnitude as the angle of divergence saw an increase. The magnitude of shear stress increases with the augmentation of surface roughness on the floodplain. The researchers Naik and Khatua (2016) proposed a novel equation that allows for the estimation of the distribution of boundary shear stress by using non-dimensional geometric and flow variables.

The use of soft computing approaches has seen a growing trend in the estimation of parameters inside compound channels, as shown by the studies conducted by Zahiri and Azamathulla (2014), Najafzadeh and Zahiri (2015), and Parsaie et al. (2015, 2017). GEP has a notable advantage over other soft computing approaches due to its ability to generate simpler equations without making assumptions about the pre-existing structure of the connection. However, the current body of literature on river engineering using the GEP technique is limited, with just a few studies available (Azamathulla et al., 2013; Guven and Gunal, 2008; Guven and Aytek, 2009; Pradhan and Khatua, 2017). Wang et al. (2010) examine three techniques for modelling and predicting water level and ice jam thickness. These techniques are based on data collected from field observations of river ice jams in the Quyu Reach of the Yellow River in China. To be more precise, we compare the simulation results produced by the conventional multi-variant regression approach with those obtained from the backpropagation neural network and the support vector machine approaches. Based on the findings from simulation and field observations, it has been determined that both the backpropagation neural network approach and the support vector machine method outperform the multi-variant regression method in terms of accuracy. In this study, Khuntia et al. (2018) used an ANN methodology to forecast the distribution of wall shear stress in straight compound channels. Mohanta et al. (2021) used several AI techniques, such as MARS, GMDH-NN, and GEP, to formulate model equations for compound channels characterized by meandering patterns and relative roughness. In comparison to the GEP and MARS models, the findings indicate that the proposed GMDH-NN model exhibited a high degree of accuracy in predicting the values. In their study, Mohseni and Naseri (2022) used ANN and SVM as computational techniques to predict the water surface profile in compound channels with vegetated floodplains. Based on the obtained findings, it can be concluded that the SVM method exhibited superior performance compared to both the ANN and regression models. In their recent study, Naik et al. (2022) introduced a unique equation that utilizes GEP to forecast the water surface profile in converging compound channels by including geometric factors. In their work, Yonesi et al. (2022) investigated the modelling of discharge in compound open channels with non-prismatic floodplains (CCNPF) using soft computation models, namely MARS and GMDH. The results were afterward compared with those produced from MLPNN. The statistical error indices demonstrate that the suggested models exhibit satisfactory accuracy in estimating the flow discharge in the CCNPF. Cheng et al. (2022) conducted measurements of water temperature and environmental variables at many locations in the primary channel of the middle part of China's South-to-North Water Transfer Project. The heat balance status of this water transfer route has been analysed using the provided temperature data. A novel method has been devised using support vector machine (SVM) theory to determine the amount of heat lost or gained by flowing water in a channel during winter. This innovative technique uses the Gaussian radial basis as its kernel function. By using supervised machine learning, the analysis of observed water temperature data reveals that the air-water temperature differential and thermal conditions are the primary variables influencing the heat loss or heat absorption of a water body. In their research, Bijanvand et al. (2023) employed soft computing methodologies, including the MLPNN, GMDH, NF-GMDH, and SVM, to forecast water surface elevation in compound channels featuring converging and diverging floodplains. The SVM model demonstrated exceptional performance throughout the testing phase, as substantiated by the statistical metrics. Kaushik and Kumar (2023a) used machine learning methodologies to predict the water surface profile of a compound channel with converging floodplains in their research. The predictions were developed by considering a mix of geometric and flow factors. The findings suggest that the ANN model, when compared to GEP, SVM, and existing methods, demonstrated a notable level of precision in predicting the water surface profile. In their study, Kaushik and Kumar (2023b) used GEP as a methodology to develop an innovative equation that describes compound channels with converging floodplains. The equation is formulated using non-dimensional geometric and flow variables and was used to predict the boundary shear force transmitted by floodplains. The findings indicate that the estimated fraction of shear force borne by floodplains determined by GEP has a good degree of agreement with the actual data.

Previous research has shown that there has been little focus on investigating the impact of flow characteristics on the distribution of shear stress in non-prismatic compound channels, as well as the development of novel methods for predicting shear stress. Consequently, further investigations are being conducted at elevated flow rates on compound channels with converging floodplains that do not have a uniform cross-sectional shape. In this investigation, three non-linear machine learning techniques, namely GEP, SVM, and ANN, are employed to forecast the shear force carried by the floodplains (S_{fp}). These predictions are based on both the geometric and hydraulic characteristics of the compound channel with converging floodplains. This research aims to identify a robust methodology, namely among the proposed techniques for accurately determining the proportion of shear force distribution on the floodplain (S_{fp}) in compound channels including convergent floodplains. The calculation of

the distribution of shear force on the floodplain, as determined by different researchers using various methodologies, is estimated. The efficacy of the generated models is evaluated by analysing their performance and errors.

MATERIALS AND METHODS

Data source

The present study incorporates the works of Rezaei (2006), Naik and Khatua (2016), as well as the experimental data collected by the author. Table 1 presents data about the experimental channel dimensions of compound channels with different geometric features.

Experimental setup

The tests were conducted in the Hydraulics laboratory, which is situated under the Department of Civil Engineering at Delhi Technological University. The trials were carried out inside a masonry flume with dimensions of 12 m in length, 1.0 m in width, and 0.8 m in depth. In this study, a compound cross section was created using brick masonry, with a main channel width of 0.5 m and a depth of 0.25 m (Figure 1). Figure 1 provides a comprehensive depiction of the geometric attributes associated with a two-stage channel. The construction of the converging portion of the channel included the use of brick masonry, which was employed to achieve a converging angle denoted as $\theta = 4°$. The compound channel consists of a prismatic segment of 6 m in length, followed by a non-prismatic section spanning 3.6 m, with the remaining portion being the downstream section. The experiment included conducting the flume at six distinct flow rates. Flow properties, including the stage-discharge relationship, water surface profile, velocity distribution, and shear stress distribution, were measured for each discharge in both the prismatic section (PS) and different non-prismatic portions (NPS), as seen in Figure 2. The parts denoted as NPS1 and NPS5 represent the first and final segments of the converging component, respectively. On the other hand, NPS3 corresponds to the central region of the converging section. The NPS2 serves as an intermediary segment situated between NPS1 and NPS3. NPS4 may be regarded as the intermediary segment situated between NPS3 and NPS5.

Table 1. Experimental channel dimensions (Kaushik and Kumar 2023a).

Verified test channel	Type of channel	Angle of convergent (θ)	Longitudinal slope (S)	Cross-sectional geometry	Total channel width (B) (m)	Main channel width (b) (m)	Main channel depth (h) (m)	Converging length (L) (m)	Aspect ratio (δ)
Channel 1 Rezaei (2006)	Converging	11.31°	0.002	Rectangular	1.2	0.398	0.05	2	7.96
Channel 2 Rezaei (2006)	Converging	3.81°	0.002	Rectangular	1.2	0.398	0.05	6	7.96
Channel 3 Rezaei (2006)	Converging	1.91°	0.002	Rectangular	1.2	0.398	0.05	6	7.96
Channel 1 Naik and Khatua (2016)	Converging	5°	0.0011	Rectangular	0.9	0.5	0.1	2.28	5
Channel 2 Naik and Khatua (2016)	Converging	9°	0.0011	Rectangular	0.9	0.5	0.1	1.26	5
Channel 3 Naik and Khatua (2016)	Converging	12.38°	0.0011	Rectangular	0.9	0.5	0.1	0.84	5
Present Channel	Converging	4°	0.001	Rectangular	1.0	0.5	0.25	3.60	2

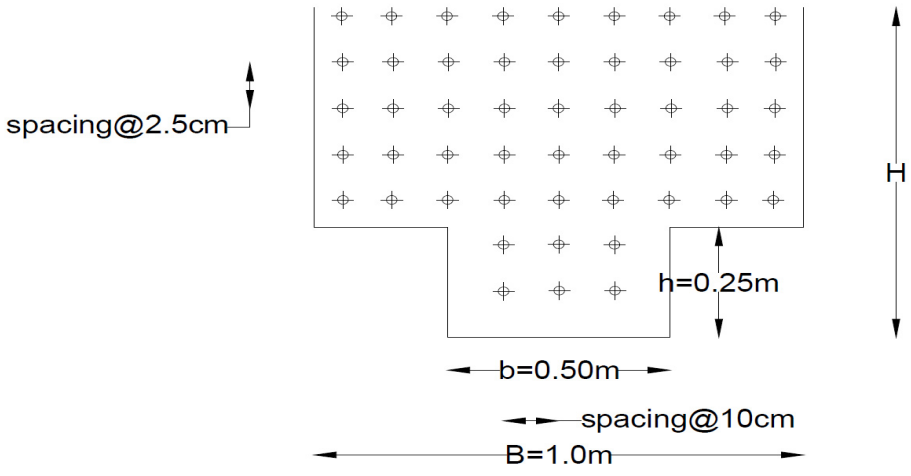


Fig. 1. Compound channel cross section.

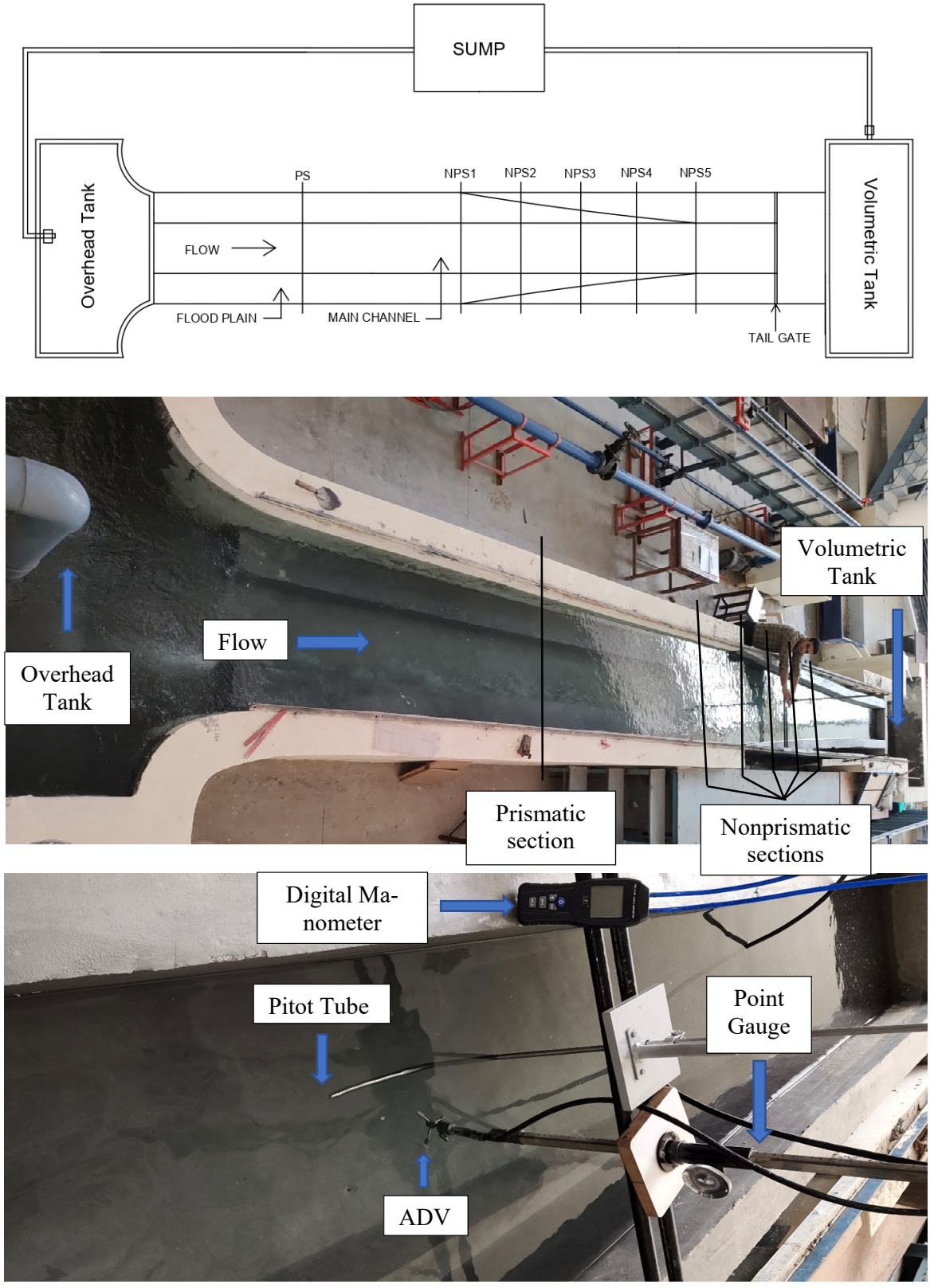


Fig. 2. Experimental setup.

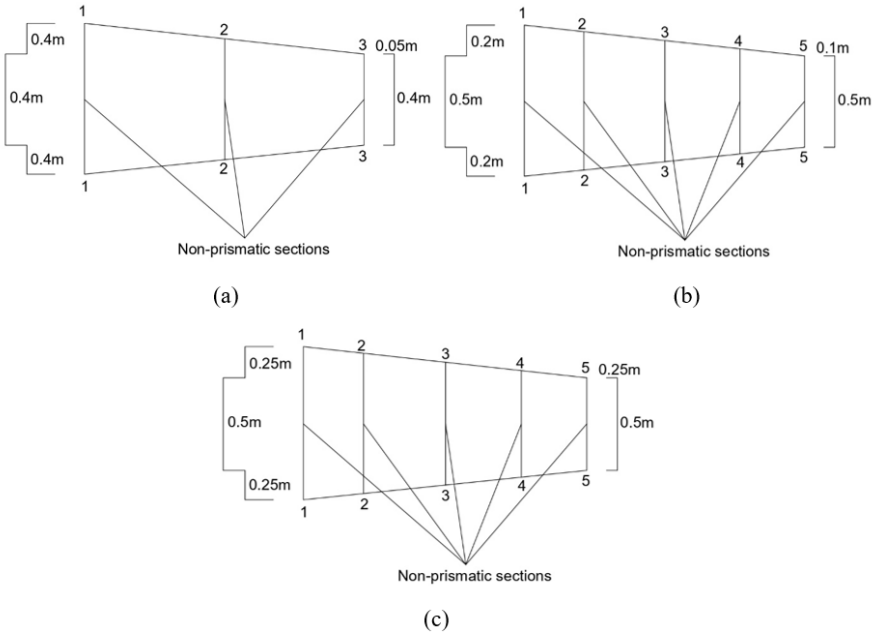


Fig. 3. Nonprismatic sections of (a) Rezaei (2006) (b) Naik and Khatua (2016) (c) Present channel.

The subcritical flow regime was seen under various circumstances in the two-stage channel, which had a longitudinal bed slope of 0.001. The estimation of Manning's n value was conducted based on data obtained from both in-bank and overbank flows in the floodplains and main channel. The water supply in the experimental channel is sourced from an underground sump and transferred to an above tank by the system. The water originating from the channel is collected inside a volumetric tank that is equipped with a v-notch. The v-notch was appropriately adjusted in order to accurately quantify the discharge emanating from the experimental channel. Subsequently, it facilitates the return of the fluid to the sump situated underneath. Figure 2 depicts the experimental configuration and the assortment of apparatus used in the investigation. Figure 3 illustrates the plan view of the non-prismatic cross-sections of Rezaei (2006), Naik and Khatua (2016), and the present channel. A tailgate was constructed at the downstream end of the flume to regulate the water surface profile and enforce a predetermined flow depth inside that section of the flume. A precision point gauge with a precision of 0.1 mm was used to measure the profile of the water surface at distances of 1.0 m and 0.3 m in the prismatic and non-prismatic sections, respectively. The researchers used an ADV to measure the mean velocity of the cross-section and the three-dimensional velocity distributions around the wetted perimeter. The measurements were taken at vertical intervals of 2.5 cm and horizontal intervals of 10 cm, as seen in the grid format presented in Figure 1. The data obtained via the use of the ADV were processed and refined using the Horizon ADV program. The measurement of lateral distributions of boundary shear stress was conducted using a Preston tube with an outer diameter of 5 mm at the identical portions where the velocity distributions were examined. The pressure differential was measured using a digital manometer. Subsequently, the calibration equations proposed by Patel (1965) were used to compute the shear stress values.

Theoretical background

Compound open channels have been the subject of investigation by several researchers, who have conducted simulations to analyse the boundary shear stress. In their seminal work, Knight and Demetriou (1983) introduced Equation (1) to describe compound channels, with values of α ranging up to 4. Knight and Hamed (1984) proposed Equation (2) as a means to include the presence of channels with non-uniform roughness in the analysis. Khatua and Patra (2007) introduced Equation (3) to represent the α within the range of 0 to 5.25. In their study, Khatua et al. (2012) derived Equation (4) to calculate the value of α, with a maximum limit of 6.67. In their study, Mohanty and Khatua (2014) introduced a new equation, denoted as Equation (5), which is applicable for values of α that are less than or equal to 12. In their study, Devi et al. (2016) developed a mathematical model, represented by Equation (6), to analyse a range of α between 3 and 12, relative flow depths between 0.10 and 0.50, and aspect ratios of the main channel between 3 and 10. Equation (7) was formulated by Naik and Khatua (2016) to describe the behaviour of converging compound channels with a width ratio of $\alpha = 1.8$ and an aspect ratio of $\delta = 5$. In their study, Kaushik and Kumar (2023b) proposed a new Equation (8) using the GEP approach to analyse compound channels with different floodplain angles.

$$S_{fp} = 48(\alpha - 0.8)^{0.289}(2\beta)^{m} \quad (1)$$

$$S_{fp} = 48(\alpha - 0.8)^{0.289}(2\beta)^{m}\{1 + 1.02\sqrt{\beta}log\gamma\} \quad (2)$$

where $m = \frac{1}{[0.75e0.38\alpha]}$

$$S_{fp} = 1.23(\beta)^{0.1833}(38ln\alpha + 3.6262)\{1 + 1.02\sqrt{\beta}log\gamma\} \quad (3)$$

$$S_{fp} = 4.1045\left(\frac{100\beta(\alpha-1)}{1+\beta(\alpha-1)}\right)^{0.6917} \quad (4)$$

$$S_{fp} = 3.3254\left(\frac{100\{\beta\alpha\delta-\beta(\delta+2s)\}}{\beta\alpha\delta+(1-\beta)(\delta+s)}\right)^{0.7467}\{1 + 1.02\sqrt{\beta}log\gamma\} \quad (5)$$

$$S_{fp} = 3.576\left\{\frac{100\beta\left(\alpha-1-\frac{2.5s}{\delta}+\frac{0.5s}{\delta^*}\right)}{\left(1+\frac{s}{\delta^*}\right)+\beta\left(\alpha-1-\frac{2s}{\delta}\right)}\right\} \quad (6)$$

$$S_{fp} = 18.505 + 62.140(\beta)^{0.631} - 24.42(X_r) + 1.38(\theta) \quad (7)$$

$$S_{fp}=\left[\left(\frac{(\alpha+\beta)\times(X_r+7.77)}{4}\right)-\left(\frac{\alpha-3.95}{2}+\frac{\alpha+0.196}{2}\right)\right]^2+\left[\frac{14.56X_r^2-38.14X_r}{\alpha^2}\right]+\tan\left[\mathrm{Max}\left((\beta+2.62),(\theta-7.23)\right)+1.50\right]^2+10.25 \quad (8)$$

However, while comparing the aforementioned equations with compound open channels that have narrowing parts, significant discrepancies are seen. Hence, the current work has examined the variation in the floodplain shear force (S_{fp}) with respect to various non-dimensional geometric and flow factors in a non-prismatic compound channel. Both the primary channel and the floodplain subsections of these compound channels have a consistent level of roughness. Manning's n values of 0.013 are assigned to all of the aforementioned smooth surfaces. In a complex channel characterized by limited floodplains, the distribution of boundary shear varies across different sections. The boundary shear force in a non-prismatic channel is believed to be influenced by various dimensionless factors, including the width ratio (α), relative depth ratio (β), aspect ratio (δ), converging angle (θ), relative distance (X_r), longitudinal slope (S_o), energy slope (S_e), discharge ratio (Q_r), Froude number (F_r), and Reynolds number (R_e). A diverse range of parameters is carefully chosen to ensure that the final model equations may be effectively applied across many channels. In order to streamline the equation for a straight compound channel, it is proposed to consider the converging angle (θ) as negligible and thereafter simplify the equation for the straight compound channel.

The equation that is necessary to be dimensionless may be expressed as:

$$S_{fp}=F(\alpha,\beta,\delta,\theta,X_r,S_o,S_e,Q_r,F_r,R_e) \quad (9)$$

SVM model

The SVR method is a frequently deployed machine learning approach utilized for regression analysis. The approach being discussed works on similar principles to SVMs used for classification but with the aim of predicting a continuous output variable based on input factors. The core principle that underlies SVR is to ascertain a function that can proficiently establish a mapping between input variables and output variables, with low error, while simultaneously adhering to a pre-established margin. The achievement of this aim necessitates the transformation of the input variables into a higher-dimensional space using a kernel function. Drucker et al. (1997) subsequently calculate the regression line that maximizes the margin while ensuring that the error remains within a preset tolerance threshold. SVR demonstrates a significant advantage in its ability to efficiently handle non-linear relationships between input and output variables. The achievement of this goal is helped by using kernel functions, such as polynomial, radial basis function, and sigmoid functions, among others. The aforementioned functions facilitate the representation of complex interdependencies among variables (Smola and Schölkopf, 2004). The use of SVR has shown to be beneficial in several fields, such as finance (Sharma et al., 2012), transportation (Yeh et al. 2010), and medical imaging (Borges et al., 2016). The careful selection and fine-tuning of the kernel function and its related parameters are crucial in achieving optimum performance. A commonly used technique for converting a linear classifier into a non-linear classifier is the application of a non-linear function to transfer the input space x onto a feature space F. A different methodology entails using a non-linear function to facilitate the process of mapping. The separating function in space F can be expressed as:

$$f(x)=w^T\varphi(x)+b \quad (10)$$

The statistical approach represented by the algebraic function f (x, w) may be described inside the subspace as follows:

$$w=\sum_{i=1}^{n}\alpha_i x_i \quad (11)$$

$$f(x,w)=\sum_{j=1}^{n}\alpha_i x_i\varphi_i(x)+b \quad (12)$$

$$f(x)=\sum_{i=1}^{n}\alpha_i x_i^T x+b \quad (13)$$

In the feature space, F, this expression takes the form:

$$f(x)=\sum_{i=1}^{n}\alpha_i\varphi(x_i)^T\varphi(x)+b \qquad 0\le\alpha_i\le C \quad (14)$$

$$k(x,x')=\varphi(x)^T\varphi(x') \quad (15)$$

$$f(x)=\sum_{i=1}^{n}\alpha_i k(x,x_i)+b \quad (16)$$

On the other hand, there is a wide array of additional kernel functions available that may be used in different contexts to improve overall performance.

Linear kernel: $k(x_i,x_j)=x_i^T x_j$

Polynomial kernel: $k(x_i,x_j)=(\gamma x_i^T x_j+r)^d,\ \gamma>0$

RBF kernel: $k(x_i,x_j)=exp(-\gamma\|x_i-x_j\|)^2,\ \gamma>0$

Sigmoid kernel: $k(x_i,x_j)=tanh(\gamma x_i^T x_j+r),\ \gamma>0$

The SVM model utilizes the parameters C, γ, r, and d. The effectiveness of the model relies on the quality of the chosen settings for its parameters, as well as the kernel parameters. The careful choice of values for the parameters C, γ, and r is of utmost importance in effectively controlling the complexity of the regression model. Nevertheless, the task of identifying the most suitable parameters is further hampered by the sophisticated complexity involved in their selection. Kernel functions are used to decrease the dimensionality of the input space and assist the process of categorization. The method of refining the SVM model involves exposing it to data sets, which is similar to the refinement process seen in other neural network models. The assessment of the proportion of shear stress borne by floodplains in a compound channel with converging floodplains was conducted using a dependable and high-caliber experimental dataset. Before constructing the SVM model, the dataset is divided into training and testing subsets using a 70:30 ratio in MATLAB R (2019). The model is then created using a cubic kernel function, which was determined to be the most optimal choice among the other kernel functions available.

ANN model

The ANN is a prominent computational approach that is now seeing significant breakthroughs in its progress. An artificial neural network is a computational model composed of linked units known as neurons or nodes, arranged in a hierarchical structure to facilitate manipulation and analysis. The ANN is structured according to the backpropagation neural network (BPNN) model and consists of three layers in a feedforward configuration. The three layers consist of a set of input neurons, denoted as I, a set of hidden neurons, denoted as m, and a set of output neurons, denoted as n. The input layer functions as the first stage for data to enter the neural network. The hidden layer, also known as the intermediate layer, receives input data from the previous layer, known as the input layer, which is largely responsible for processing the data. The second layer, referred to as the output layer, receives the processed data from the network.

The data obtained from the output is sent to external receptors. ANNs consist of interconnected layers that are sequentially linked to one another. The aforementioned relationships are often referred to as weights and weighted values, respectively. The augmentation of connectivity weights is a strategy for regulating the behaviour of specified cost functions inside an ANN (Kaushik and Kumar, 2023a).

The current study used an ANN methodology to predict the shear force borne by floodplains in a compound channel with converging floodplains. A feedforward neural network is implemented in MATLAB R (2019) using the backpropagation training algorithm with the objective of prediction. Before applying the network to any issue, the network must undergo a training phase. During this phase, the adjustment of weights and biases for each output neuron is performed in line with the training approach, with the objective of restricting the intended output of each output neuron. The training of ANNs encompasses three key components: inter-neuron weights, which dictate the significance of input sources; a sigmoid transfer function, which regulates the output level of neurons; and a collection of learning laws, which govern the adjustment of weights during the training process. During the training phase, it is common practice to use a non-linear function, often in the form of a sigmoid function (Kaushik and Kumar, 2023a):

$$F(a) = \frac{1}{1 + e^{-a}} \tag{17}$$

The result is sent to the succeeding layer of nodes for further processing. The BPNN techniques consist of four distinct steps. The aforementioned processes may be delineated as follows:

Sum the weighted input:

$$Nod_z = \sum_{i=1}^{n} (W_{xz} k_x) + \epsilon_z \tag{18}$$

Transform the weighted input:

$$Out_z = \frac{1}{1 + e^{-Nod_z}} \tag{19}$$

Sum the hidden node output:

$$Nod_y = \sum_{j=1}^{m} (W_{zy} Out_y) + \epsilon_y \tag{20}$$

Transform the weighted sum:

$$Out_y = \frac{1}{1 + e^{-Nod_y}} \tag{21}$$

The architecture of the neural network was designed with ten neurons in the input layer, ten neurons in the hidden layers, and one neuron in the output layer. Figure 4 depicts the fluctuation of neural network parameters, including the gradient, mean, and validation tests. During the training phase, the gradient and mean exhibit fluctuations at certain phases, ultimately decreasing, while the validation checks rise as the number of generations increases. This is an outcome of the optimization process, particularly gradient descent, which occurs during the training of ANNs. The repeated modifications to the weights, led by the gradient, strive to discover a collection of weights that minimizes the loss function and generates precise predictions on the training data. Figure 5 illustrates an error detected during the training, testing, and validation phases. The training error first had a rapid decline, then moved to a more consistent deceleration. After 30 generations, the network achieved convergence and demonstrated the lowest possible error value. The reduction in error seen during the training of an ANN is a direct result of the optimization process, which is led by the techniques of gradient descent and backpropagation. The parameters of the model are fine-tuned to discover the most effective setup that minimizes the error on the training data, with the ultimate objective of obtaining strong generalization performance.

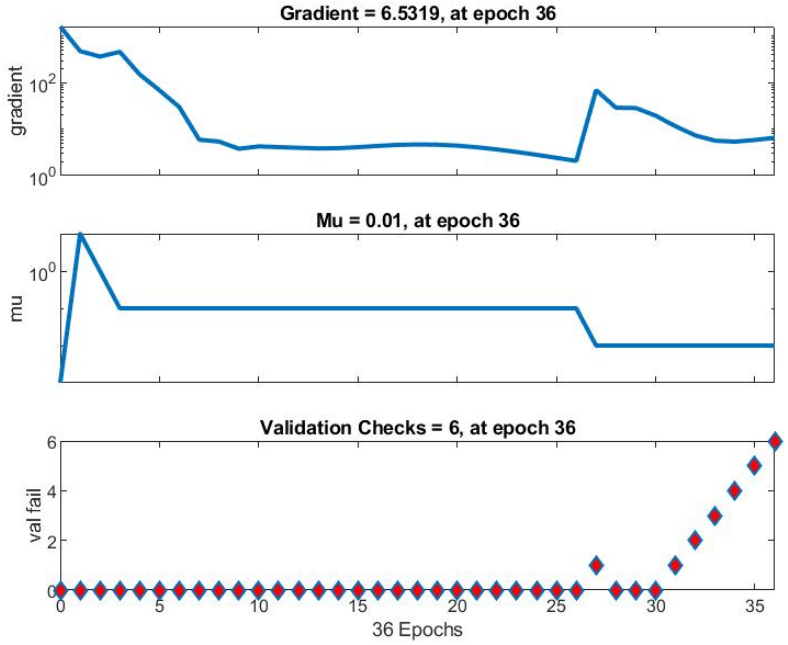


Fig. 4. ANN training parameters.

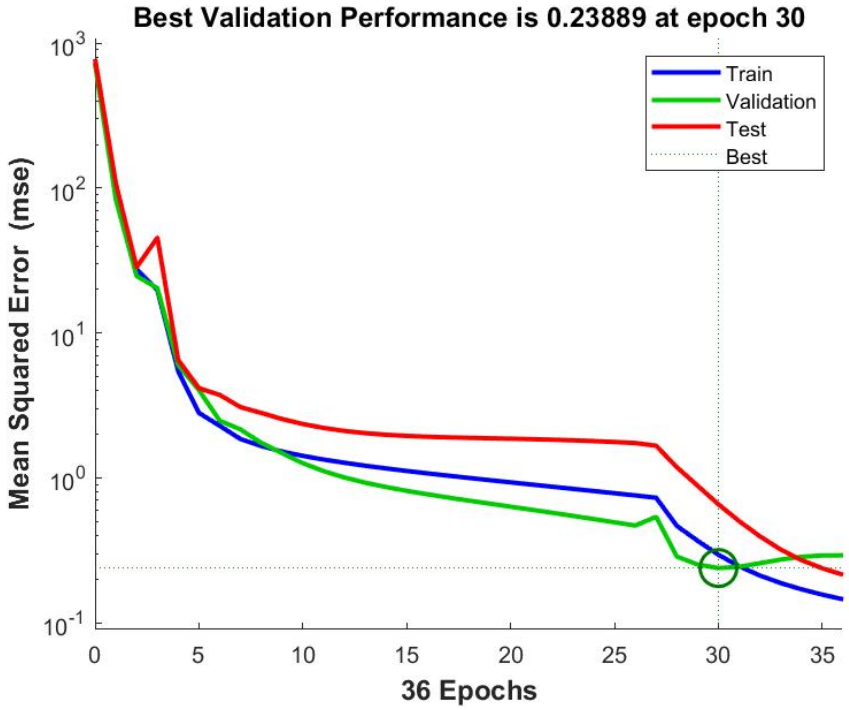


Fig. 5. BPNN convergence curve.

GEP model

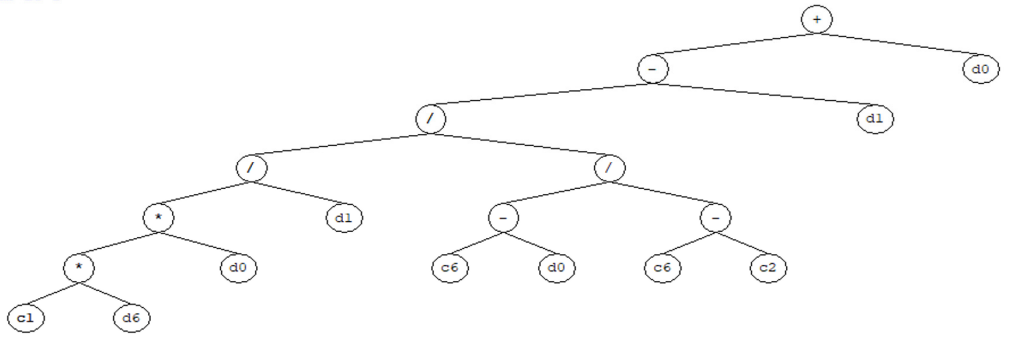
The use of GEP, a computational approach, enables the evolution of computer programs capable of solving complex problems. The first idea was introduced by Dr. Cândida Ferreira in 1992, as a component of her PhD dissertation at the University of Coimbra, situated in Portugal. The technique of the GEP is based on the basic principles of genetics and the process of natural selection. The operating mechanism of this system is the iterative development of a collection of computer programs through a sequence of mutation and selection procedures. The GEP technique encompasses the representation of individual programs as symbol sequences, which may be subjected to a range of genetic operations, such as combining and mutation. The selection of programs with higher performance on a specific task in each succeeding generation leads to an improvement in their overall fitness. The ability of GEP to effectively handle many goals simultaneously makes it an important tool in the field of optimization. The GEP has been widely used across several fields, including but not limited to data analysis, categorization, and modelling. According to Ferreira (2001), empirical data indicates that GEP has enhanced efficacy when compared to other evolutionary computation methods such as genetic programming and evolutionary strategies, particularly within certain problem areas. Furthermore, researchers have extended the GEP technique to support complex data structures such as trees and graphs. This has been achieved by the incorporation of modified genetic operators, as described by Otero et al. (2016).

The current study utilizes a modelling methodology in which the goal value is denoted as $\%S_{fp}$, while the ten independent components are regarded as input variables, as described in Equation (9). The construction of the model's structure included the use of the four basic arithmetic operators (+, −, ×, /) in combination with GeneXproTools 5.0 (2014). In general, the modelling procedure used high-calibre datasets that were randomly allocated during two discrete phases. Several models were created for the present investigation, using different ratios of training and testing datasets, program dimensions, and generations. The characteristics of the best model are shown in Table 2. The current investigation used the root-mean-squared error (RMSE) as the fitness function, which was designated as E_i. The fitness (f_i) was determined by using an equation developed from an expression tree, which took into consideration the cumulative total of mistakes with respect to the target value. The technique of addition was used as the method for joining the genetic components. Figure 6 illustrates the expression tree (ET) representing the percent shear force of the compound channel predicted from the GEP model. In the present representation, the input parameters are denoted as d0 to d9, whilst the constant values for gene one are assigned as G1c1, G1c2, and G1c6. The expression tree may be used to derive a mathematical formula that establishes a correlation between the input and output variables.

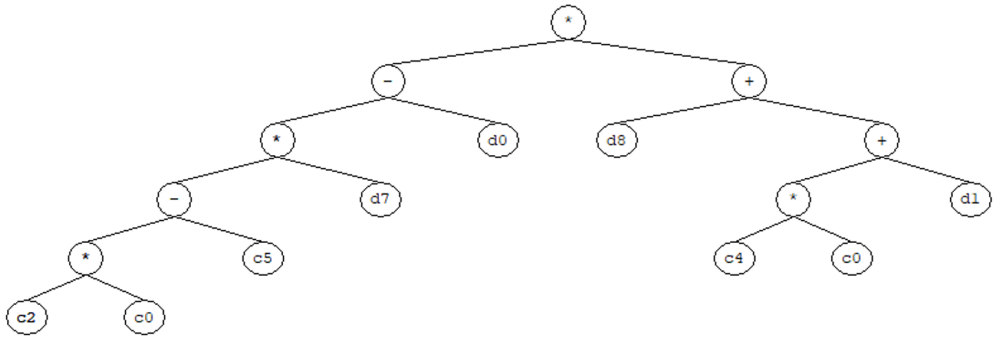
Table 2. Selection standards and attributes for the GEP model.

Description of parameter	Parameter setting
Function set	+, −, ×, /
Number of chromosomes	30
Head size	12
Number of genes	6
Gene size	38
Linking function	Addition
Fitness function	RMSE
Program size	114
Literals	36
Number of generations	476664
Constants per gene	10
Data type	Floating-point
Mutation	0.00138
Inversion	0.00546
Gene recombination rate	0.00277

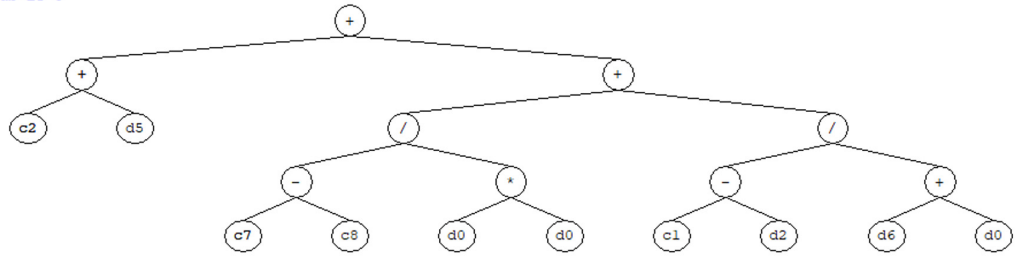
Sub-ET 1



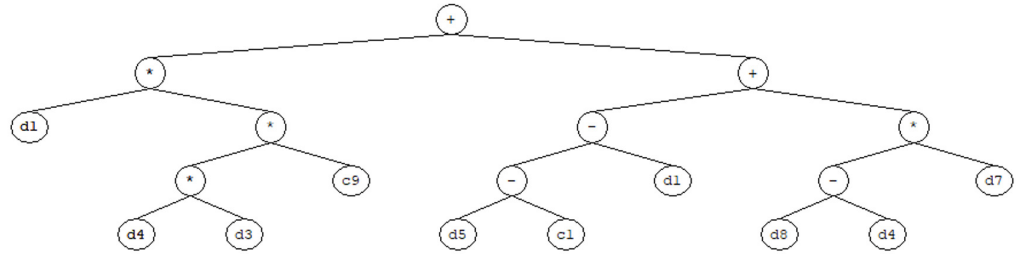
Sub-ET 2



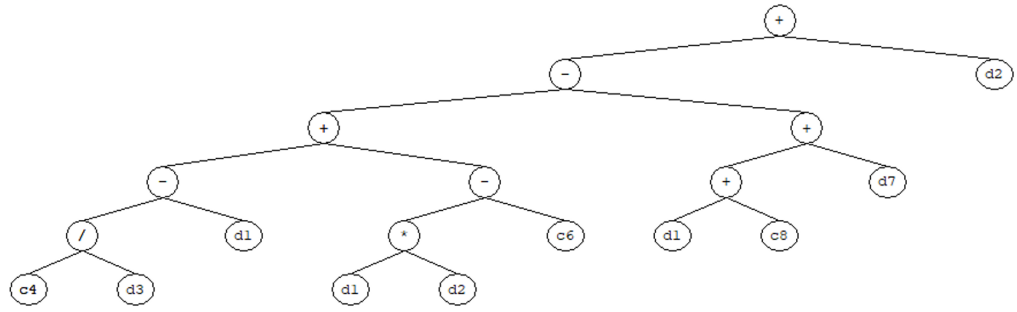
Sub-ET 3



Sub-ET 4



Sub-ET 5



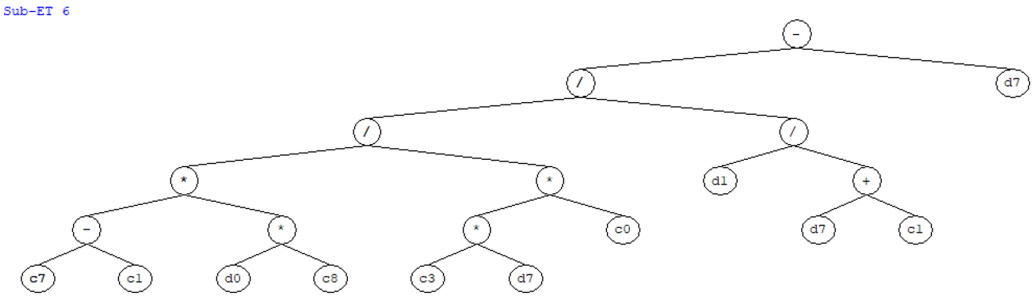


Fig. 6. GEP formulated expression tree.

Statistical measures

Various error analysis approaches were used to evaluate the accuracy of the machine learning models. The aforementioned metrics included the coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), scatter index (SI), coefficient of efficiency (E), and Akaike Information Criterion (AIC). The relevant equations were used to compute these measures (Kaushik and Kumar 2023a).

$$R^2 = \frac{\sum_{i=1}^{N}(a_i-\bar{a})^2\,(p_i-\bar{p})^2}{\sum_{i=1}^{N}(a_i-\bar{a})^2\ \sum_{i=1}^{N}(p_i-\bar{p})^2} \quad (23)$$

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(p_i - a_i)^2 \quad (24)$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - a_i)^2} \quad (25)$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|p_i - a_i| \quad (26)$$

$$\text{MAPE}(\%) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{|p_i-a_i|}{a_i} \times 100\right) \quad (27)$$

$$SI = \sqrt{\frac{\sum_{i=1}^{N}[(p_i-\bar{p})-(a_i-\bar{a})]^2}{\sum_{i=1}^{N} a_i^2}} \quad (28)$$

$$E = 1 - \frac{\sum_{i=1}^{N}(p_i-a_i)^2}{\sum_{i=1}^{N}(a_i-\bar{a})^2} \quad (29)$$

$$\text{AIC} = N\log\left(\frac{1}{N}\sum_{i=1}^{N}(p_i - a_i)^2\right) + 2k \quad (30)$$

RESULTS AND DISCUSSION

The connection between bed shear stress and longitudinal distance for various relative depths (range from 0.10 to 0.60) in a compound channel with converging floodplains (θ = 4°) is shown in Figure 7. The bed shear stress shows a linear augmentation in relation to the longitudinal distance till reaching the prismatic section. Nevertheless, it is observed that there is a sudden increase in shear stress along the longitudinal distance in the initial half of the channel's converging region. Additionally, in the latter half of this section, there is an abrupt rise in bed shear caused by the convergence of the channel's geometry, resulting in accelerated flow within the channel. The bed shear stress in both the main channel and floodplains exhibits an increase in magnitude as the longitudinal distance progresses. This increase is mostly attributed to the rise in relative depth, with greater relative depths seeing a more pronounced effect. The primary driver behind this phenomenon is the transfer of momentum from the main channel to the floodplains. The boundary shear stress distributions throughout the wetted perimeter of the compound channel are depicted in Figure 8. The figure includes both prismatic and non-prismatic cross-sections and the relative flow depth β is set at 0.60. The distribution of boundary shear stress exhibits symmetry and expands progressively from section 1 to section 5. The largest value of boundary shear stress is seen near the centre of the main channel, and this value rises as the flow transitions from a prismatic portion to a non-prismatic segment. This increase in boundary shear stress may be attributed to the boundary resistance resulting from the converging shape of the channel. The magnitude of the boundary shear stress diminishes as it approaches the floodplain interface across all sections. At lower relative depths, in close proximity to the interface between the main channel and the floodplain, the boundary shear stress experiences a rapid fall, followed by a gradual decrease until reaching a minimum at both ends of the floodplain. At elevated relative flow depth, in the vicinity of the interface between the main channel and the floodplain, there is a rapid decline in shear stress. This is followed by a gradual increase towards the centre of both floodplains and subsequently a decrease towards the boundaries of the channel. This pattern arises due to the transfer of momentum from the floodplains to the main channel in the compound channel configuration.

Figure 9 displays scatter plots that compare the anticipated and actual values of $\%S_{fp}$ for each of the several machine learning algorithms. The proximity of the values to the line representing excellent agreement serves as compelling evidence for the predictive accuracy of the developed GEP, ANN, and SVM models. When comparing the three models, it can be seen that the predictions made by the ANN exhibit a higher degree of proximity to the best-fitting line in comparison to the GEP and SVM models. The ANN model demonstrates a more concentrated distribution of values along the fitted line. Hence, it can be concluded that the ANN model exhibits the highest level of agreement with the experimental data, as shown by the significant R^2 values. Figure 10 presents a comparison of the created models used to forecast the proportion of shear force carried by floodplains, using different approaches. The built models have shown remarkable proximity to the best-fitted line in comparison to earlier methodologies, indicating a robust potential for generalization. There is an absence of observable indications of the overtraining phenomena in their behaviour. Several of the suggested methodologies exhibit a clear bias toward estimating shear stress in floodplains. These models incorporate all the relevant parameters, such as discharge ratio, energy slope, Froude number, and Reynolds number, which were previously overlooked in earlier approaches. The evaluation of the GEP, ANN, and SVM models' performance was conducted using various statistical metrics, including the R^2, MSE, RMSE, MAE, MAPE, SI, E, and AIC. These metrics are presented in Table 3. The table reveals that the ANN model demonstrates superior performance in comparison to other models constructed

using SVM, GEP, and previously developed methodologies. This is evident from the highest R^2 value of 0.998, lowest MSE value of 0.375, lowest RMSE value of 0.613, lowest MAE value of 0.338, and lowest MAPE value of 0.391 associated with the ANN model. The SI serves as a standardized measure for the quantification of inaccuracy. A lower SI value is indicative of a higher level of model performance. The ANN model has a significantly reduced SI value of 0.023 in both the training and testing stages, indicating a minimum level of error. The efficiency measure, represented as E, is assigned a value of 1 when there is a complete correspondence between the model and the observed data. The interval of E extends from negative infinity to 1. The ANN model has an E value of 0.998 for both the training and testing datasets. The AIC is a widely used criterion in statistical model selection. It is used to choose the best suitable model by evaluating the statistical likelihood function. The model that has the lowest AIC value is regarded as the most optimum model. The ANN model has the lowest AIC score of -148.470 when compared to the other models proposed. The statistical study of several indices indicates that ANN exhibits superior performance compared to other machine learning techniques and established methodologies in forecasting the proportion of shear force borne by floodplains in compound channels with narrowing floodplains.

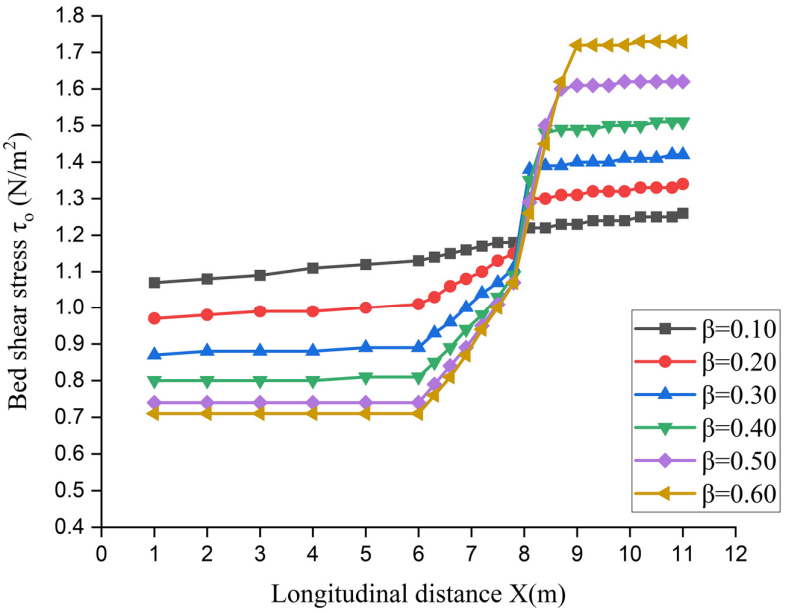


Fig. 7. Variation of bed shear stress along the longitudinal distance of the channel.

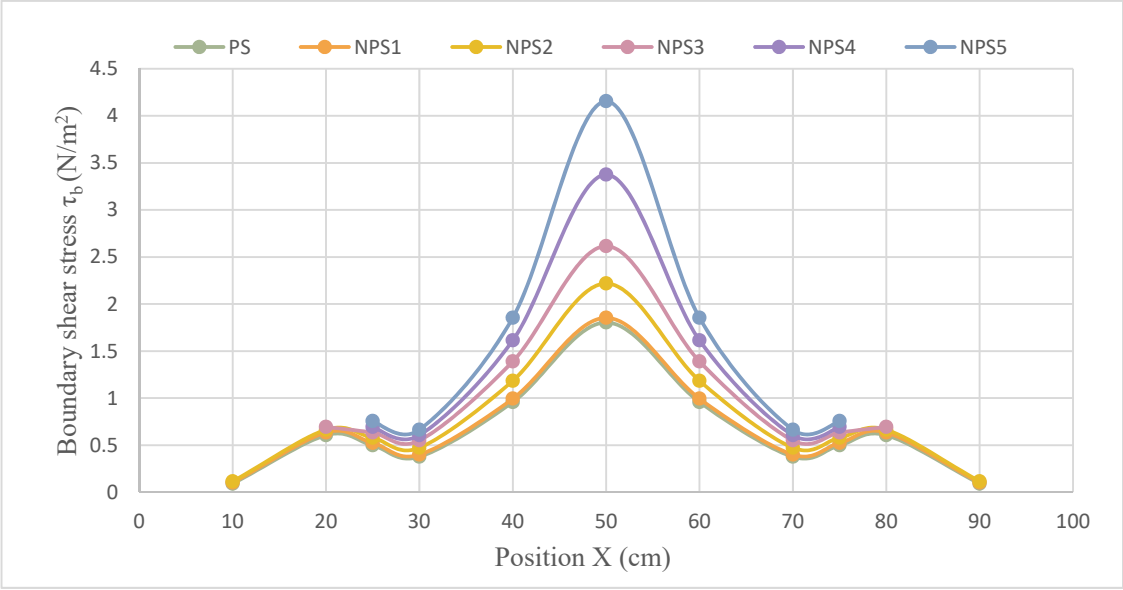


Fig. 8. Variation of boundary shear stress along the wetted perimeter.

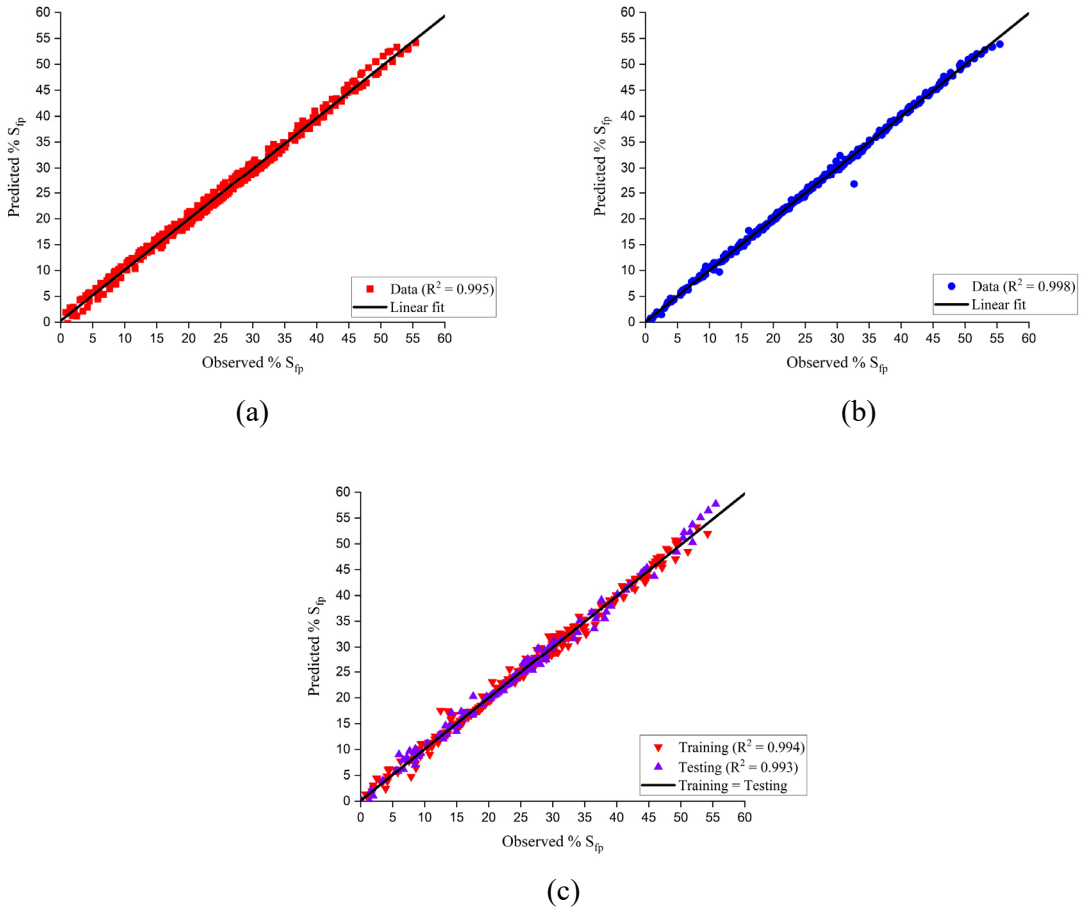


Fig. 9. Scatter plots of observed and predicted %S_{fp} for various ML models (a) SVM (b) ANN (c) GEP.

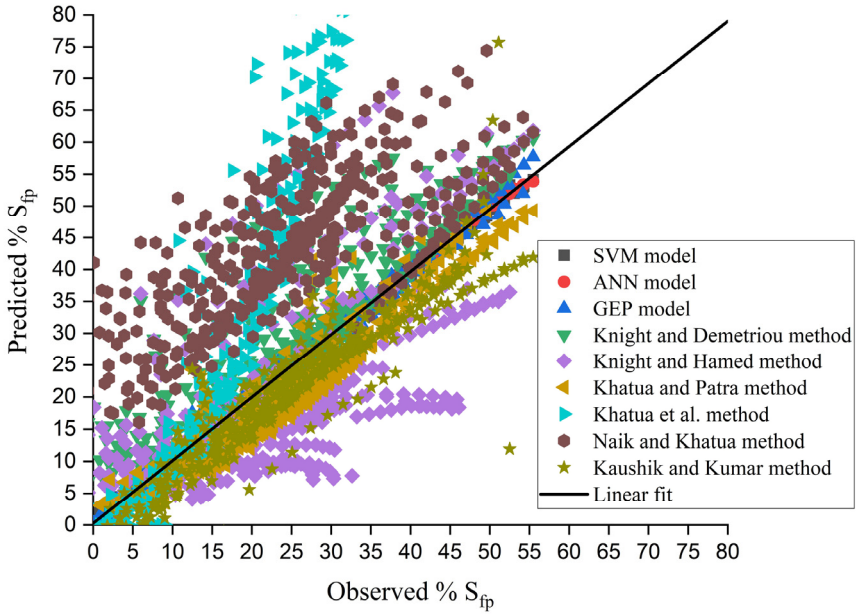


Fig. 10. Comparison of predicted value of %S_{fp} for different models.

Table 3. Statistical analysis of predicted $\%S_{fp}$ by different approaches.

Parameters	ANN model	GEP model	SVM model	Knight and Demetriou method	Knight and Hamed method	Khatua and Patra method	Khatua et al. method	Naik and Khatua method	Kaushik and Kumar method
R^2	0.998	0.993	0.995	0.767	0.416	0.895	0.847	0.568	0.961
MSE	0.375	1.234	0.869	71.704	144.986	29.911	4519.4	459.406	13.855
RMSE	0.613	1.111	0.932	8.468	12.041	5.469	67.226	21.434	3.722
MAE	0.338	0.807	0.827	5.884	9.705	4.631	45.777	19.416	2.990
MAPE	0.391	2.855	8.269	71.933	83.331	22.923	240.870	211.354	12.120
SI	0.023	0.041	0.034	0.313	0.442	0.199	2.472	0.790	0.285
E	0.998	0.993	0.995	0.618	0.228	0.841	–23.054	–1.445	0.961
AIC	–148.470	56.146	–4.158	754.796	875.886	604.429	1467.412	1074.230	105.920

CONCLUSIONS

This work presents an evaluation of the shear stress exerted on floodplains in a compound channel with converging floodplains. The assessment is conducted using machine learning techniques, including ANN, GEP, and SVM. The suggested models have been built using 396 meticulously gathered laboratory datasets from both current and previous work. These datasets pertain to non-prismatic compound channels characterized by varying converging angles and relative depths. The suggested model has apparent sensitivity to several factors, including width ratio, relative flow depth, aspect ratio, converging angle, relative distance, longitudinal bed slope, energy slope, discharge ratio, Froude number, and Reynolds number. The bed shear stress exhibits an increasing trend as one moves along the length of the channel and with respect to the relative flow depth. Nevertheless, a rapid increase in shear stress was seen in the converging portion, particularly in the latter half of the channel, due to the interaction and transfer of momentum between the main channel and floodplains. The largest wall shear stress was observed in the central region of the primary channel, with a rise in magnitude as the flow progresses downstream within the channel. At the junction between the main channel and the floodplain, the shear stress values experience a drop as a result of the momentum shift phenomenon. This decrease is seen to be at its lowest near the extremities of the channel for lower relative depths, while it rises towards the centre of the floodplain for greater relative depths. The connection between the percent shear force and the non-dimensional geometric and hydraulic variables of a converging compound channel is nonlinear for all relative flow depths. In contrast to past methodologies, the three proposed models provide superior outcomes in relation to statistical parameters. The results indicated that, in line with the established evaluation criteria, all three models ANN, GEP, and SVM exhibited a reasonable ability to estimate the shear force carried by floodplains of a compound channel with converging floodplains. The ANN model had superior performance in comparison to other models and earlier methodologies, as shown by its greatest R^2 value and lowest values for MSE, RMSE, MAE, MAPE, SI, and AIC.

In order to mitigate the negative impacts of flooding, it is crucial to have a comprehensive understanding of the flow dynamics in both prismatic and non-prismatic sections of rivers. This understanding should include the incorporation of extra momentum transfer in flow models. Such knowledge is essential for the effective construction of flood control and diversion structures. The models proposed in this paper possess practical implications for non-prismatic rivers, such as the River Main in Northern Ireland, the Brahmaputra River in India, and other analogous river systems. The models used to predict the shear stress have some limitations, including high computational expenses, lack of interpretability, susceptibility to overfitting, and their black box character. Additionally, these models have specific drawbacks such as demanding data requirements, high processing intensity, sensitivity to parameters, restricted scalability, and being primarily designed for binary classification tasks. The computational cost of machine learning models is a crucial factor that affects their practicality, effectiveness, and capacity to handle larger datasets. Various elements contribute to the computational expense of machine learning models, including the complexity of the model, the size of the dataset, the complexity of the algorithm, the use of hardware acceleration, the tuning of hyperparameters, the application of feature engineering, the implementation of regularisation techniques, the number of model training iterations, and the need for real-time inference. Performance parameters are essential measures used to assess the efficacy and proficiency of machine learning models. These parameters provide valuable information about the performance of a model on a specific job and help in the selection and refinement of models. Several essential performance metrics are often used to evaluate machine learning models, including accuracy, precision, sensitivity, specificity, mean squared error, R-squared, and the confusion matrix. The results of this research will be valuable for informing the design of such structures, leading to a reduction in both economic and human losses.

Acknowledgement. The authors express their gratitude for the help received from the Department of Civil Engineering at Delhi Technological University in Delhi, India.

Disclosure statement. The authors reported no potential conflicts of interest.

Funding. Not applicable.

Data Availability Statement. All relevant data are included in the manuscript or its Supplementary Information.

NOMENCLATURE

The following symbols are used in this paper:
B = total width of compound channel [m]
b = width of the main channel [m]
h = height of the main channel [m]
H = flow depth [m]

α = width ratio (B/b)
β = relative flow depth [(H – h)/H]
δ^* = flow aspect ratio (b/H)
δ = main channel aspect ratio (b/h)
θ = converging angle [degree]
X_r = relative distance (x/L)
L = converging length [m]
x = distance between two consecutive sections [m]
S_o = longitudinal bed slope [radian]
S_e = energy slope [radian]
Q_r = discharge ratio (Q/Q_b)
Q = discharge at any depth [m^3/sec]
Q_b = bankfull discharge [m^3/sec]
F_r = Froude number
R_e = Reynolds number
S_{fp} = shear force carried by floodplains [percent]
a = actual values
p = predicted values
$\bar{a}$ = mean of actual values
$\bar{p}$ = mean of predicted values
N = number of datasets
k = number of variables
a = sum of weighted input value plus bias
Nod_z = sum for the zth hidden node
n = total number of input nodes
W_{xz} = connection weight between the xth input and the zth hidden node
k_z = normalized input at the xth input node
$\in_z$ = bias value at the zth hidden node
Out_z = output from the zth hidden node
Nod_y = sum of yth output node
m = total number of hidden nodes
W_{zy} = connection weights between the zth hidden node and the yth output node
$\in_y$ = bias value at the yth output node
Out_y = output at the yth output node

ABBREVIATIONS

ANN = Artificial neural networks
GEP = Gene expression programming
SVM = Support vector machines
AI = Artificial intelligence
MARS = Multivariate adaptive regression splines
GMDH-NN = Group method of data handling neural network
CCNPF = Compound open channels with nonprismatic floodplains
MLPNN = Multilayer perceptron neural networks
NF-GMDH = Neuro-fuzzy group method of data handling
PS = Prismatic section
NPS = Nonprismatic sections
ADV = Acoustic Doppler velocimeter
SVR = Support vector regression
RBF = Radial basis function
BPNN = Back-propagation neural network
ET = Expression tree

REFERENCES

Azamathulla, H.M., Ahmad, Z., Ghani, A.A., 2013. An expert system for predicting Manning's roughness coefficient in open channels by using gene expression programming. Neural Comput. Appl., 23, 5, 1343–1349.

Bhattacharya, A.K., 1995. Mathematical model of flow in a meandering channel. IIT Kharagpur.

Bijanvand, S., Mohammadi, M., Parsaie, A., 2023. Estimation of water's surface elevation in compound channels with converging and diverging floodplains using soft computing techniques. Water Supply, 23, 4, 1684–1699. https://doi.org/10.2166/ws.2023.079

Borges, L.M., Mekitarian Filho, E.R., Paiva, A.C., 2016. Support vector regression applied to magnetic resonance imaging: An approach to predicting hepatic iron concentration. Journal of Digital Imaging, 29, 1, 70–77.

Bousmar, D., Zech, Y., 2002. Periodical turbulent structures in compound channels. In: Proc. River Flow International Conference on Fluvial Hydraulics, Louvain-la-Neuve, Belgium, pp. 177–185.

Bousmar, D., Wilkin, N., Jacquemart, J.H., Zech, Y., 2004. Overbank flow in symmetrically narrowing floodplains. J. Hydraul. Eng. ASCE, 130, 4, 305–312.

Cheng, T., Wang, J., Sui, J., Zhao, H., Hao, Z., Huang, M., Li, Z., 2022. A new approach for assessing heat balance state along a water transfer channel during winter periods. Water, 14, 20, 3269. https://doi.org/10.3390/w14203269

Chlebek, J., Bousmar, D., Knight, D.W., Sterling, M.A., 2010. Comparison of overbank flow conditions in skewed and converging/diverging channels. In: Proc. River Flows International Conference, pp. 503–511.

Devi, K., Khatua. K.K., Das, B.S., 2016. Apparent shear in an asymmetric compound channel. In: Proc. Int. Conf. on Fluvial Hydraulics (River Flow 2016). CRC Press, Boca Raton, pp. 48–56.

Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A.J., Vapnik, V., 1997. Support vector regression machines. Advances in Neural Information Processing Systems, 9, 155–161.

Ferreira, C., 2001. Gene expression programming: A new adaptive algorithm for solving problems. Complex Systems, 13, 2, 87–129.

Gepsoft, GeneXproTools 5.0, 2014. Data Modeling & Analysis Software. (n.d.). https://www.gepsoft.com/.

Guven, A., Aytek, A., 2009. New approach for stage-discharge relationship: gene-expression programming. J. Hydrol. Eng., 14, 8, 812–820.

Guven, A., Gunal, M., 2008. Genetic programming approach for prediction of local scour downstream of hydraulic structures. J. Irrig. Drain. Eng., 134, 2, 241–249.

Kar, S.K., 1977. A study of distribution of boundary shear in meander channel with and without floodplain and river floodplain interaction. PhD Thesis. Indian Institute of Technology Kharagpur, Kharagpur, India.

Kaushik, V., Kumar, M., 2023a. Assessment of water surface profile in nonprismatic compound channels using machine learning techniques. Water Supply, 23, 1, 356–378. https://doi.org/10.2166/ws.2022.430

Kaushik, V., Kumar, M., 2023b. Sustainable gene expression programming model for shear stress prediction in nonprismatic compound channels. Sustainable Energy Technologies and Assessments, 57, 103229. https://doi.org/10.1016/j.seta.2023.103229

Khatua, K.K., Patra, K.C., 2007. Boundary shear stress distribution in compound open channel flow. ISH J. Hydraul. Eng., 13, 3, 39–54.

Khatua, K.K., Patra, K.C., Mohanty, P.K., 2012. Stage-discharge prediction for straight and smooth compound channels with wide floodplains. J. Hydraul. Eng., 138, 1, 93–9. https://doi.org/10.1061/(asce)hy.1943-7900.0000491

Khatua, K.K., Patra, K.C., Mohanty, P.K., 2011a. Stage-discharge prediction for straight and smooth compound channels with wide floodplains. J. Hydraul. Eng., 138, 1, 93–99.

Khatua, K.K., Patra, K.C., Nayak, P., 2011b. Meandering effect

for evaluation of roughness coefficients in open channel flow. WIT Trans Ecol Environ, 146, 213–224.
Khuntia, J.R., Devi, K., Khatua, K.K., 2018. Boundary shear stress distribution in straight compound channel flow using artificial neural network. J. Hydrol. Eng., 23, 5, 04018014. https://doi.org/10.1061/(asce)he.1943-5584.0001651
Knight, D.W., Demetriou, J.D., 1983. Flood plain and main channel flow interaction. J. Hydraul. Eng., 109, 8, 1073–1092.
Knight, D.W., Hamed, M.E., 1984. Boundary shear in symmetrical compound channels. J. Hydraul. Eng., 110, 10, 1412–1430.
Leighly, J.B., 1932. Toward a theory of the morphologic significance of turbulence in the flow of water in streams. University of California Press.
MATLAB R, 2019. Computer Software. MathWorks, Natick, MA.
Mohanta, A., Pradhan, A., Mallick, M., Patra, K.C., 2021. Assessment of shear stress distribution in meandering compound channels with differential roughness through various artificial intelligence approach. Water Resour. Manage., 35, 13, 4535–4559. https://doi.org/10.1007/s11269-021-02966-5
Mohanty, P.K., 2013. Flow analysis of compound channels with wide flood plains prabir. PhD Thesis. National Institute of Technology Rourkela, Odisha, India.
Mohanty, P.K., Khatua, K.K., 2014. Estimation of discharge and its distribution in compound channels. J. Hydrodyn., 26, 1,144–54.
Mohseni, M., Naseri, A., 2022. Water surface profile prediction in compound channels with vegetated floodplains. Proceedings of the Institution of Civil Engineers - Water Management, 1–12. https://doi.org/10.1680/jwama.21.00005
Myers, W.R.C., 1987. Velocity and discharge in compound channels. J Hydraul Eng, 113, 6, 753–766.
Naik, B., Khatua, K.K., 2016. Boundary shear stress distribution for a converging compound channel. ISH. J Hydraul Eng, 22, 2, 212–9. https://doi.org/10.1080/09715010.2016.1165633
Naik, B., Kaushik, V., Kumar, M., 2022. Water surface profile in converging compound channel using gene expression programming. Water Supply, 22, 5, 5221–5236. https://doi.org/10.2166/ws.2022.172
Najafzadeh, M., Zahiri, A., 2015. Neuro-fuzzy GMDH-based evolutionary algorithms to predict flow discharge in straight compound channels. J. Hydrol. Eng., 20, 12, 04015035.
Otero, F.E., Freitas, A.A., Johnson, C.G., 2016. Multi-expression programming: An extended description and a new implementation. Swarm and Evolutionary Computation, 30, 1–16.
Parsaie, A., Yonesi, H., Najafian, S., 2017. Prediction of flow discharge in compound open channels using adaptive neuro fuzzy inference system method. Flow Meas. Instrum., 54, 288–297.
Parsaie, A., Yonesi, H.A., Najafian, S., 2015. Predictive modeling of discharge in compound open channel by support vector machine technique. Model. Earth Syst. Environ., 1, 1–2. DOI: 10.1007/s40808-015-0002-9
Patel, V.C., 1965. Calibration of the Preston tube and limitations on its use in pressure gradients. J. Fluid Mech., 231, 85–208.
Patra, K.C., Kar, S.K., 2000. Flow interaction of meandering river with floodplains. J. Hydraul. Eng., 126, 8, 593–604.
Patra, K.C., 1999. Flow interaction of meandering river with flood plains. PhD Thesis. Indian Institute of Technology Kharagpur, Kharagpur, India.
Patra, K.C., Kar, S.K., Bhattacharya, A.K., 2004. Flow and velocity distribution in meandering compound channels. J. Hydraul. Eng., 130, 5, 398–411.
Pradhan, A., Khatua, K.K., 2017. Gene expression programming to predict Manning's n in meandering flows. Can. J. Civ. Eng., 45, 4, 304–313.
Proust, S., Rivière, N., Bousmar, D., Paquier, A., Zech, Y., 2006. Flow in the compound channel with abrupt floodplain contraction. J. Hydraul. Eng., 132, 9, 958–970.
Rezaei, B., Knight, D.W., 2009. Application of the Shiono and Knight Method in the compound channel with nonprismatic floodplains. J. Hydraul. Res., 47, 6, 716–726.
Rezaei, B., Knight, D.W., 2011. Overbank flow in compound channels with nonprismatic floodplains. J. Hydraul. Eng., 137, 815–824.
Rezaei, B., 2006. Overbank Flow in Compound Channels with Prismatic and Nonprismatic Floodplains. PhD Thesis. University of Birmingham, Birmingham, UK.
Sellin, R.H.J., 1964. A laboratory investigation into the interaction between flow in the channel of a river and that of its flood plain. LaHouille Blanche, 7, 793–801.
Sharma, S., Ahmad, S., Gupta, A., 2012. Support vector regression for financial time series forecasting: A comparative study with artificial neural network. Expert Systems with Applications, 39, 3, 3509–3518.
Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and Computing, 14, 3, 199–222.
Stephenson, D., Kolovopoulos, P., 1990. Effects of momentum transfer in compound channels. J. Hydraul. Eng., 116, 12, 1512–1522.
Wang, J., Sui, J., Guo, L., Karney, B.W., Jüpner, R., 2010. Forecast of water level and ice jam thickness using the back propagation neural network and support vector machine methods. International Journal of Environmental Science & Technology, 7, 2, 215–224. https://doi.org/10.1007/bf03326131
Yang, S-Q., Lim, S-Y., 2005. Boundary shear stress distributions in trapezoidal channels. J. Hydraul. Res., 43, 1, 98–102.
Yeh, A.G.-O., Huang, C.-H., Li, X., 2010. A support vector machine with a quantum-behaved particle swarm optimization approach for predicting transportation mode choice. Expert Systems with Applications, 37, 12, 7922-7929.
Yonesi, H.A., Omid, M.H., Ayyoubzadeh, S.A., 2013. The hydraulics of flow in nonprismatic compound channels. J. Civ. Eng. Urbanism, 3, 6, 342–356. https://doi.org/10.1061/(ASCE)0733-9429(2000)126:4(299)
Yonesi, H.A., Parsaie, A., Arshia, A., Shamsi, Z., 2022. Discharge modeling in compound channels with nonprismatic floodplains using GMDH and MARS models. Water Supply, 22, 4, 4400–4421. https://doi.org/10.2166/ws.2022.058
Zahiri, A., Azamathulla, H.M., 2014. Comparison between linear genetic programming and M5 tree models to predict flow discharge in compound channels. Neural Comput. Appl., 24, 2, 413–420.

Received 23 November 2023
Accepted 10 January 2024

Uncorrected Proof

Water Practice & Technology

Check for updates

© 2024 The Authors

Water Practice & Technology Vol 00 No 0, 1 doi: 10.2166/wpt.2024.117

Prediction of the flow resistance in non-prismatic compound channels

Vijay Kaushik [a,*], Bandita Naik[b], Munendra Kumar[a] and Vijay K. Minocha[a]

[a] Department of Civil Engineering, Delhi Technological University, Delhi 110042, India
[b] Department of Civil Engineering, Methodist College of Engineering, Hyderabad 500001, India
*Corresponding author. E-mail: vijaykaushik_2k20phdce01@dtu.ac.in

VK, 0000-0002-7270-4520

ABSTRACT

Achieving an accurate estimation of the flow resistance in open channel flows is crucial for resolving several critical engineering difficulties. In instances when there is excessive flow on both banks of a river, it results in the breach of the primary channel, leading to the discharge of water into the adjacent floodplain. The alteration of floodplain geometry occurs as a consequence of agricultural and developmental practises, leading to the emergence of compound channels that exhibit converging, diverging, or skewed characteristics throughout the course of the flow. The efficacy of conventional equations in accurately forecasting flow resistance is limited due to their heavy reliance on empirical approaches. As a result of this phenomenon, there persists a significant need for methodologies that possess both novelty and precision. The objective of this work is to use the support vector machine (SVM) technique for the estimation of the Manning's roughness coefficient in a compound channel with converging floodplains. Statistical indicators are used to validate the constructed models in the experimental investigation, enabling the assessment of their performance and efficacy. The findings indicate a significant correlation between the Manning's roughness coefficient predicted by SVM and both experimental data and prior research outcomes.

Key words: geometric and flow parameters, Manning's roughness coefficient, non-prismatic compound channel, statistical analysis, support vector machine

HIGHLIGHTS

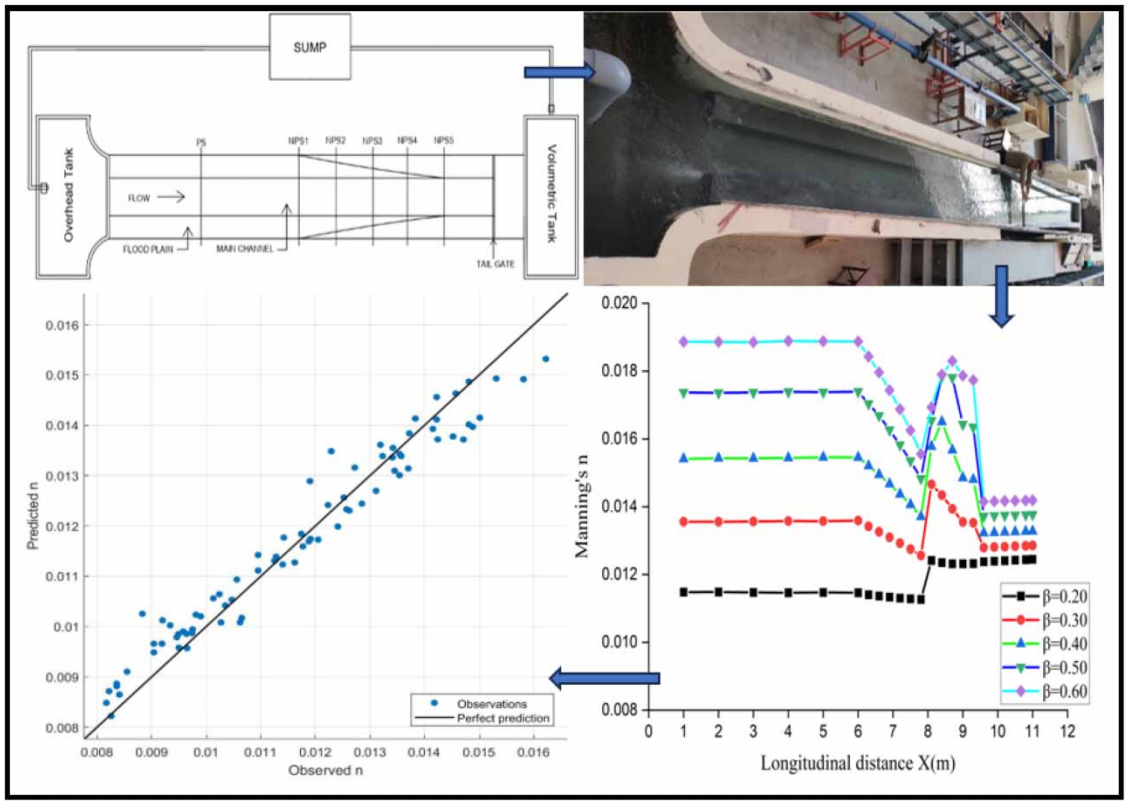
- Flow resistance is crucial across various engineering contexts, such as planning irrigation systems, managing drainage networks, and constructing flood control structures.
- This study delved into assessing flow resistance in non-prismatic compound channels by employing support vector machines.
- The results indicate a strong alignment between the predicted Manning's roughness coefficient and the actual observed values.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Downloaded from http://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2024.117/1416749/wpt2024117.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

GRAPHICAL ABSTRACT



NOTATIONS

The following symbols are used in this paper:

B	total width of compound channel
b	width of the main channel
h	height of the main channel
H	flow depth
α	width ratio (B/b)
β	relative flow depth [$(H–h)/H$]
δ	aspect ratio (b/h)
θ	converging angle
X_r	relative distance (x/L)
L	converging length
x	distance between two consecutive sections
S_o	longitudinal bed slope
S_e	energy slope
Q_r	discharge ratio (Q/Q_b)
Q	discharge at any depth
Q_b	bankfull discharge
F_r	Froude number
R_e	Reynolds number
A	flow area
P	wetted perimeter
R	hydraulic radius (A/P)
n	Manning's roughness coefficient
C	Chezy's roughness coefficient
f	Darcy–Weisbach resistance coefficient
a	actual values
p	predicted values
$\bar{a}$	mean of actual values
$\bar{p}$	mean of predicted values
N	number of datasets

Downloaded from http://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2024.117/1416749/wpt2024117.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

k number of variables
y/D ratio of water depth to culvert diameter

ABBREVIATIONS

SVM	support vector machine
GEP	gene expression programming
ANN	artificial neural network
ANFIS	adaptive neuro-fuzzy inference system
MARS	multivariate adaptive regression splines
MLP	multilayer perceptron
GMDH	group method of data handling
MLPNN	multilayer perceptron neural networks
NF-GMDH	neuro-fuzzy group method of data handling
CCNPF	compound open channels with non-prismatic floodplains
PS	prismatic section
NPS	non-prismatic section
ADV	acoustic Doppler velocimeter
CM	Cox method
EBM	Einstein and Banks method
LM	Lotter method
SVR	support vector regression
RBF	radial basis function
BPNN	back-propagation neural network
GPR	Gaussian process regression
HDPE	high density polyethylene
R^2	coefficient of determination
MSE	mean squared error
RMSE	root mean squared error
MAE	mean absolute error
MAPE	mean absolute percentage error
SI	scatter index
E	coefficient of efficiency
AIC	Akaike information criterion

INTRODUCTION

The consideration of flow resistance is of paramount importance in compound open channels, as it has a direct influence on the hydraulic behavior and overall efficiency of the channel system. Flow resistance is the phenomenon that arises as water moves inside a channel, resulting in the counteracting force against its motion. This resistance induces changes in velocity and causes a dissipation of energy. The comprehension and precise estimation of flow resistance are of utmost importance in a range of engineering applications, including the design and administration of irrigation systems, drainage networks, and flood control structures. The impact of flow resistance on water conveyance capacity is a significant aspect to consider in compound open channels. The velocity of water is substantially influenced by flow resistance, which therefore impacts the total flow rate that can be sustained by the channel. Engineers have the ability to enhance the efficiency of water distribution for irrigation or flood channeling by designing channels that optimize water conveyance via precise quantification of flow resistance. The significance of flow resistance is crucial in the determination of water surface profiles. The presence of flow resistance induces energy losses that have a direct impact on both the water surface slope and the distribution of water depths inside the channel. The precise forecasting of water surface profiles has significant importance in the fields of hydraulic design, floodplain management, and flood control operations. The comprehension of flow resistance is crucial for engineers as it enables them to develop models and simulations to analyze channel behavior. This, in turn, facilitates the accurate prediction of floods and the development of efficient strategies for flood mitigation planning. Several researchers have conducted studies on the flow, velocity, and shear stress distribution on compound sections (Kar 1977; Knight & Demetriou 1983; Myers 1987; Bhattacharya 1995; Patra 1999; Patra & Kar 2000; Patra *et al.* 2004; Yang & Lim 2005; Khatua *et al.* 2011a, 2011b; Mohanty 2013). Traditional channel division techniques are used for the assessment of discharge capacity in compound channels. It is postulated that the compound section is delineated by hypothetical interface planes located at the confluence of the main channel and the floodplain. The absence of shear

Downloaded from http://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2024.117/1416749/wpt2024117.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

stress in these regions suggests a lack of momentum transmission. The momentum transfer phenomena were first examined by Sellin (1964) via laboratory research. Extensive investigations were conducted in laboratory flumes to study both prismatic and meandering compound channel designs. Nevertheless, a comparison between the compound section data of prismatic compound channels and non-prismatic compound channels revealed significant discrepancies in the estimation of shear force carried by floodplains. These errors can be attributed to the omission of additional mass and momentum transfer, as elucidated by Bousmar & Zech (2002), Bousmar *et al.* (2004), Rezaei (2006), and Rezaei & Knight (2009). The additional momentum exchange associated with this phenomenon must be considered when doing flow modeling for non-prismatic compound channels. Yang *et al.* (2005) carried out the estimation of the Manning and Darcy–Weisbach resistance coefficients in composite waterway channels. The experimental results suggest that the Darcy–Weisbach resistance coefficient for a compound channel is dependent on the Reynolds number. Bjerklie *et al.* (2005) developed equations to estimate inbank river discharge, which are constructed based on theoretical considerations and calibrated and confirmed using a comprehensive experimental database, rely on the Manning's and Chezy equations. The study conducted by Proust *et al.* (2006) investigated the phenomenon of asymmetric geometry, specifically focusing on its enhanced convergence rate. An augmented convergence angle of 22° leads to heightened levels of mass transfer and head loss. Cao *et al.* (2006) introduces novel methodologies for the computation of flow resistance and momentum flux in compound open channels. Yang *et al.* (2007) demonstrated that the methodologies used to investigate the resistance characteristics of inbank and overbank flows in compound channel consisting of a rough main channel and rough floodplains resulted in substantial errors in evaluating the composite roughness. The flow behavior of skewed, two-stage converging and diverging channels was investigated by Chlebek *et al.* (2010). The researchers noted an elevation in head losses resulting from the transfer of mass and momentum, as well as the homogeneity of velocity on shrinking floodplains. Additionally, they discovered an augmented velocity gradient on the expanding floodplains. The observed variations in flow distribution may be attributed to alterations in the flow force between the main channel and floodplains within different subsections. Rezaei & Knight (2011) conducted a study on compound channels with floodplains that are non-prismatic. The authors examined the depth-averaged velocity, local velocity distributions, and boundary shear stress distributions at different convergence angles. The analysis of depth-averaged velocities provides insights into the impact of contractions on velocity distributions. Notably, these contractions lead to an augmentation of velocity in close proximity to the walls of the main channel. This effect is particularly pronounced in the latter portion of the convergence reach. Additionally, for cases where the relative depth is high, the lateral flow entering the main channel also contributes to these velocity variations. The study conducted by Yonesi *et al.* (2013) examines the influence of roughness in floodplains on the occurrence of overbank flow in compound channels with non-prismatic floodplains. The rate of change in velocity, as shown by the gradient, exhibited a rise in magnitude as the angle of divergence saw an increase. The magnitude of shear stress increases with the augmentation of surface roughness on the floodplain. Naik *et al.* (2017) developed novel mathematical equations to predict the spatial variation of boundary shear stress in a composite channel that includes a convergent floodplain. The effective calculation of relationships for stage–discharge with decreased width ratios in compound channels with converging floodplains has been facilitated by the use of developed mathematical expressions. Das & Khatua (2018) conducted experiments on a composite stream that had symmetric geometry. This stream had sections that expanded and narrowed, resulting in zones of flood with varying depths of relative flow. The experiments were conducted to examine the resistance properties of the flow in terms of different roughness coefficients such as C, n, and f.

Creating models to estimate Manning's roughness coefficient using various methods complicates the analysis of interdependencies among components. These models become intricate and time-consuming, leading to a significant reduction in the time required for experiments and labor-intensive calculations. SVM, GEP, ANN, ANFIS, and the M5 decision tree models are commonly used for predicting flow in composite channels (Parsaie *et al.* 2017). Parsaie & Haghiabi (2017) employed analytical methods, specifically ANN and MARS, to forecast the discharge capacity of compound open channels, revealing that the MARS model outperformed the MLP model in accuracy. Mohanta *et al.* (2018) used the GMDH-NN technique to estimate Manning's n in meandering compound channels. The evaluation of GMDH-NN performance is conducted using two distinct machine learning approaches, namely SVR and MARS, using a range of statistical methodologies. The results demonstrate that the GMDH-NN model accurately predicts Manning's n in comparison to the MARS and SVR models. Khuntia *et al.* (2018) developed an ANN model using key characteristics to predict boundary shear stress distribution in straight compound channels, showcasing the effectiveness of BPNN models across diverse parameter spaces. Das

Downloaded from http://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2024.117/1416749/wpt2024117.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

et al. (2020) utilized ANN and ANFIS techniques to predict discharge in compound channels with converging and diverging sections, with the ANFIS model demonstrating superior performance over the ANN model. Naik *et al.* (2022) proposed a unique equation using GEP to estimate water surface profile in converging compound channels using nondimensional variables. Yonesi *et al.* (2022) utilized soft computation models such as MARS and GMDH to model discharge in compound open channels with non-prismatic floodplains (CCNPF), showing sufficient accuracy in flow discharge estimation. The effective parameters of the Manning's n in HDPE culverts were investigated by Matin (2022) by the utilization of intelligence methodologies, including GPR and SVM. The results of the sensitivity analysis indicate that the parameters y/D and Q play a crucial role in predicting the Manning coefficient for HDPE culvert models. Bijanvand *et al.* (2023) employed soft computing models like MLPNN, GMDH, NF-GMDH, and SVM to predict water surface elevation in compound channels with converging and diverging floodplains, where SVM model demonstrated superior performance in testing. Kaushik & Kumar (2023a) utilized ANN, SVM, and GEP to forecast water surface profile in compound channels with converging floodplains, with the ANN model displaying high accuracy. Kaushik & Kumar (2023b) formulated a novel equation for compound channels with converging floodplains using GEP, estimating the boundary shear force conveyed by floodplains. Kaushik & Kumar (2023c) utilized SVM to approximate water surface profile in compound channels featuring narrowing floodplains, showing a high degree of concordance with empirical data and previous research findings.

Historically, research has been undertaken on the flow characteristics in compound channels with non-prismatic floodplains. However, there has been little focus on the flow resistance in non-prismatic compound channel with converging floodplains using creative methodologies. The current investigation focused on a compound channel including smooth converging floodplains. To estimate the flow resistance in terms of Manning's roughness coefficient using both nondimensional geometric and hydraulic parameters, a supervised learning approach known as SVM was used. The efficacy of the suggested models was evaluated using theoretical methodologies and established approaches using statistical analysis.

MATERIALS AND METHODS

Experimental setup

The tests were conducted in the Hydraulics laboratory, which is situated under the Department of Civil Engineering at Delhi Technological University. The trials were carried out inside a masonry flume with dimensions of 12 m in length, 1.0 m in width, and 0.8 m in depth. In this study, a compound cross-section was created using brick masonry, with a main channel width of 0.5 m and a depth of 0.25 m (Figure 1). Figure 1 provides a comprehensive depiction of the geometric attributes associated with a two-stage channel. The construction of the converging portion of the channel included the use of brick masonry, which was employed to achieve a converging angle denoted as $\theta = 4°$. The floodplains were made smooth with the help of trowel finish. The compound channel consists of a prismatic section (PS) of 6 m in length, followed by a non-prismatic section (NPS)

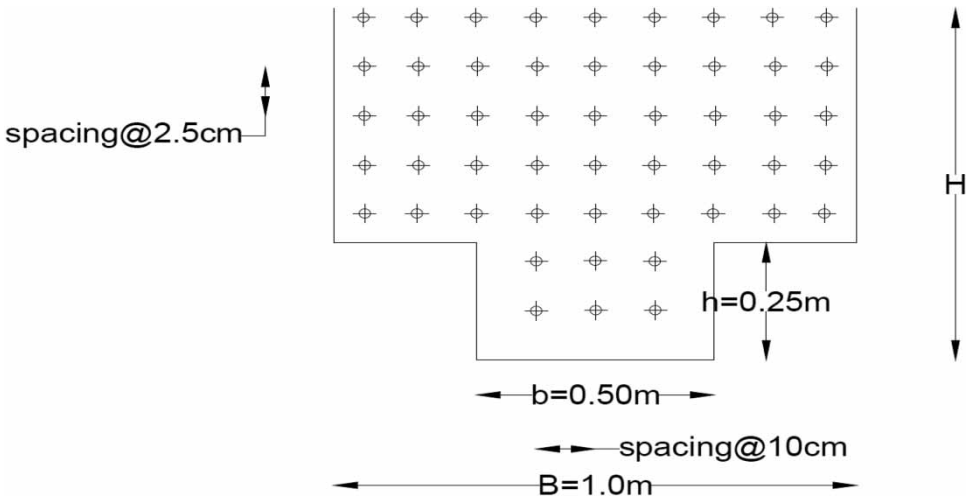


Figure 1 | Cross-section of compound channel.

Downloaded from http://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2024.117/1416749/wpt2024117.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

spanning 3.6 m, with the remaining portion being the downstream section. The experiments were conducted at six relative flow depths ranging from 0.10 to 0.60. Flow properties, including the stage-discharge relationship, water surface profile, velocity distribution, and shear stress distribution, were measured for each discharge in both the PS and different NPS, as seen in Figure 2. The parts denoted as NPS1 and NPS5 represent the first and final segments of the converging component, respectively. On the other hand, NPS3 corresponds to the central region of the converging section. The NPS2 serves as an intermediary segment situated between NPS1 and NPS3. NPS4 may be regarded as the intermediary segment situated between NPS3 and NPS5.

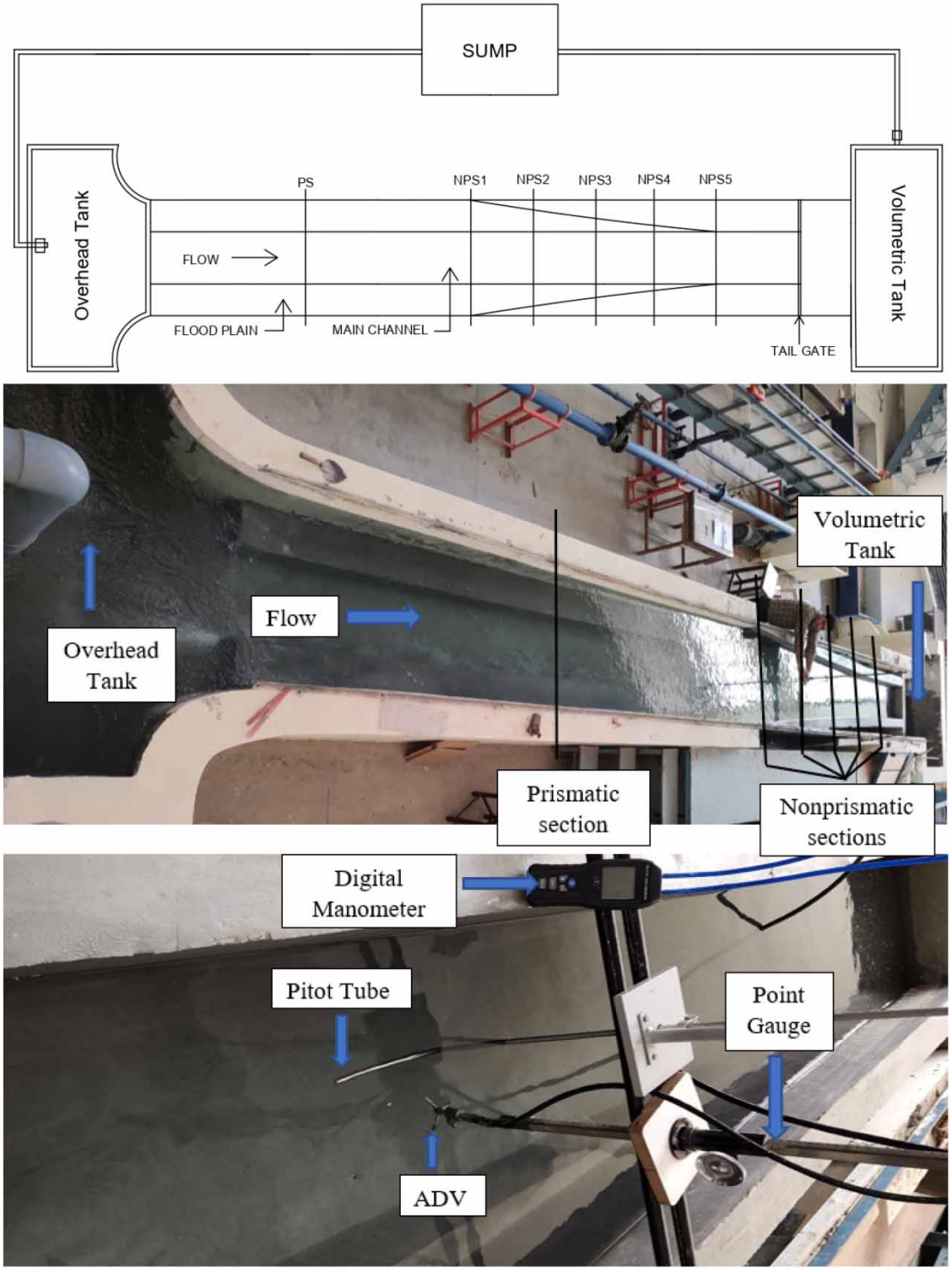


Figure 2 | Experimental setup.

Downloaded from http://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2024.117/1416749/wpt2024117.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

The subcritical flow regime was seen under various circumstances in the two-stage channel, which had a longitudinal bed slope of 0.001. The estimation of Manning's n value was conducted based on data obtained from both in-bank and overbank flows in the floodplains and main channel. The water supply in the experimental channel is sourced from an underground sump and transferred to an above tank by the system. The water originating from the channel is collected inside a volumetric tank that is equipped with a v-notch. The v-notch was appropriately adjusted in order to accurately quantify the discharge emanating from the experimental channel. Subsequently, it facilitates the return of the fluid to the sump situated underneath. Figure 2 depicts the experimental configuration and the assortment of apparatus used in the investigation. A tailgate was constructed at the downstream end of the flume to regulate the water surface profile and enforce a predetermined flow depth inside that section of the flume. A precision point gauge with a precision of 0.1 mm was used to measure the profile of the water surface at distances of 1.0 and 0.3 m in the PS and NPS, respectively. The researchers used an ADV to measure the mean velocity of the cross-section and the three-dimensional velocity distributions around the wetted perimeter. The measurements were taken at vertical intervals of 2.5 cm and horizontal intervals of 10 cm, as seen in the grid format presented in Figure 1. The data obtained via the use of the ADV were processed and refined using the Horizon ADV program. The measurement of lateral distributions of boundary shear stress was conducted using a Preston tube with an outer diameter of 5 mm at the identical portions where the velocity distributions were examined. The pressure differential was measured using a digital manometer. Subsequently, the calibration equations proposed by Patel (1965) were used to compute the shear stress values.

Model development for Manning's *n*

The lateral variation of channel section roughness is observed along the wetted perimeter in a river that is experiencing flooding. The determination of the conveyance capacity of a compound channel necessitates the consideration of the composite roughness. Despite the varied structures and forms of equations, they can be categorized into distinct groups based on different criteria. These groups include the CM (Cox 1973) (as represented by Equation (1)), the EBM (Einstein & Banks 1950) (as represented by Equation (2)), and the LM (Lotter 1933) represented by Equation (3)). The Equation (4) was formulated by Das & Khatua (2018) through the utilization of multivariate analysis, with consideration of geometric and flow parameters, for a compound channel featuring converging floodplains. The composite roughness equations are

$$n = \frac{1}{A}\sum_{i=1}^{i=N} A_i n_i \tag{1}$$

$$n = \left(\frac{1}{P}\sum_{i=1}^{i=N} P_i\, n_i^{3/2}\right)^{2/3} \tag{2}$$

$$n = PR^{5/3}\left(\sum_{i=1}^{i=N}\frac{P_i\, R_i^{5/3}}{n_i}\right)^{-1} \tag{3}$$

$$n = \{0.096\{1 - 0.0131e^{0.211(\alpha)} - 0.068e^{0.3(\beta)} - 0.082(\theta)^{0.046} - 0.235e^{0.101(x)} - 0.791(R_e)^{-0.214} + ln\,(\delta^{0.0114} S_0^{0.073} F_r^{-0.167})\} \tag{4}$$

Composite roughness for compound channel with converging floodplains is calculated by the following expression:

$$n = \frac{AR^{2/3}\sqrt{S_e}}{Q} \tag{5}$$

The determination of Manning's n values is subject to various factors, among which are the width ratio, relative depth ratio, aspect ratio, converging angle, relative distance, longitudinal bed slope, energy slope, discharge ratio, Froude's number, and Reynold's number. Several parameters are utilized to facilitate the application of the model equation to various compound channels. The details of the dataset used in the present investigation are given in Table 1.

The requisite equation in a dimensionless form can be formulated as follows:

$$n = F(\alpha,\ \beta,\ \delta,\ \theta,\ X_r,\ S_o, S_e,\ Q_r,\ F_r,\ R_e) \tag{6}$$

Downloaded from http://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2024.117/1416749/wpt2024117.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

Table 1 | Statistical characteristics of the dataset used

Range	α	β	δ	θ	X_r	S_o	S_e	Q_r	F_r	R_e	n
Minimum	0.84	0.053	2	4	0	0.001	0.00003	0.98	0.05	76912.63	0.0081
Maximum	2.00	0.600	2	4	1.00	0.001	0.00012	2.43	0.18	229804.35	0.0162
Average	1.48	0.349	2	4	0.50	0.001	0.00008	1.58	0.12	135006.63	0.0116
Median	1.50	0.358	2	4	0.50	0.001	0.00009	1.48	0.12	130987.07	0.0117
Standard deviation	0.34	0.175	0	0	0.31	0	0.00003	0.47	0.04	34764.67	0.0021

SVM model

SVM is a highly effective and extensively used machine learning technique employed for both classification and regression purposes. This approach has notable efficacy in scenarios when the data exhibits non-linear separability, indicating that the classes cannot be distinctly partitioned by a linear or hyperplane boundary. The SVM algorithm is designed to identify the ideal hyperplane that effectively distinguishes different classes by maximizing the margin. The margin refers to the distance between the hyperplane and the closest data points from each class. SVR algorithm is a commonly employed machine learning technique utilized for the purpose of regression analysis. The method in question operates on analogous principles to those of SVMs utilized for classification purposes, albeit with the objective of forecasting a continuous output variable on the basis of input variables. The fundamental concept underlying SVR is to identify a function that can effectively map input variables to output variables, with minimal error, while also satisfying a predetermined margin. The process of attaining this objective involves the conversion of the input variables into a space of higher dimensions through the application of a kernel function. Subsequently, the regression line that optimizes the margin while maintaining the error within a predetermined tolerance level is determined (Drucker *et al.* 1997). SVR exhibits a notable benefit in its capacity to effectively manage non-linear associations between input and output variables. The attainment of this objective is facilitated through the utilization of kernel functions, including but not limited to polynomial, RBF, and sigmoid functions. These functions enable the capture of intricate relationships between variables (Smola & Schölkopf 2004). SVR has been effectively utilized in diverse domains, including finance (Sharma *et al.* 2012), transportation (Yeh *et al.* 2010), and medical imaging (Borges *et al.* 2016). It is of utmost significance to meticulously choose and adjust the kernel function and other associated parameters to attain optimal performance. One common method for transforming a linear classifier into a non-linear classifier involves the utilization of a non-linear function to map the input space x onto a feature space F. An alternative approach involves the utilization of a non-linear function for the purpose of mapping. The separating function in space f can be expressed as (Kaushik & Kumar 2023a):

$$f(x) = w^T \varphi(x) + b \tag{7}$$

Within the subspace, the statistical methodology denoted by the algebraic function $f(x, w)$ can be articulated as follows:

$$w = \sum_{i=1}^{n} \alpha_i x_i \tag{8}$$

$$f(x, w) = \sum_{j=1}^{n} \alpha_i x_i \varphi_i(x) + b \tag{9}$$

$$f(x) = \sum_{i=1}^{n} \alpha_i x_i^T x + b \tag{10}$$

In the feature space, F, this expression takes the form:

$$f(x) = \sum_{i=1}^{n} \alpha_i \varphi(x_i)^T \varphi(x) + b \qquad 0 \le \alpha_i \le C \tag{11}$$

$$k(x,\ x') = \varphi(x)^T \varphi(x') \tag{12}$$

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x, x_i) + b \tag{13}$$

Downloaded from http://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2024.117/1416749/wpt2024117.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

Conversely, there exists a diverse range of supplementary kernel functions that can be employed in various settings to enhance performance.

Linear kernel: $k(x_i, x_j) = x_i^T x_j$
Polynomial kernel: $k(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \quad \gamma > 0$
RBF kernel: $k(x_i, x_j) = exp(-\gamma \|x_i - x_j\|)^2, \quad \gamma > 0$
Sigmoid kernel: $k(x_i, x_j) = tanh(\gamma x_i^T x_j + r), \quad \gamma > 0$

The SVM model employs parameters *C*, *γ*, *r*, and *d*. The *C* parameter governs the balance between optimizing the margin and minimizing the classification error on the training dataset. A lesser value of *C* leads to a broader margin, which may let some instances to be misclassified. Conversely, a bigger value of *C* aims to minimize misclassification by sacrificing margin width. The parameter *γ* is unique to the RBF kernel. It regulates the impact of specific training samples. A decrease in the gamma value results in a more refined decision boundary, characterized by a higher degree of smoothness. Conversely, an increase in the gamma value introduces more fluctuations and irregularities to the border, potentially leading to overfitting. The parameter denoted as '*d*' is exclusive to polynomial kernels. The parameter determines the order of the polynomial function used for the transformation. The attainment of advanced academic degrees has the potential to result in decision boundaries of greater complexity. However, it is important to acknowledge that this pursuit may also elevate the likelihood of overfitting. The efficacy of the SVM model is contingent upon the caliber of the selected settings for said parameters, in addition to the kernel parameters. The selection of values for the parameters *C, γ, and r* plays a crucial role in regulating the complexity of the regression model. However, the process of determining the optimal parameters is further complicated by the intricate nature of their selection. Kernel functions are utilized to reduce the dimensionality of the input space and facilitate classification. The refinement of the SVM model is achieved through its exposure to data sets, a process that is analogous to the refinement of other neural network models. The flow resistance in terms of Manning's roughness coefficient (*n*) of compound channel with converging floodplains was evaluated utilizing a reliable and high-quality experimental dataset. Prior to building the SVM model, the data are partitioned into training and testing sets using a 70:30 ratio in MATLAB R (2019). Table 2 illustrates the various models that were developed to predict roughness coefficient, utilizing distinct kernel functions and kernel scales while maintaining a consistent relationship and proportion of training and testing datasets.

Table 2 | Different models developed using SVR

Model	Preset	Kernel function	Kernel scale	Prediction speed (obs/s)	Training time (s)
M1	Linear SVM	Linear	Automatic	330	17.48
M2	Quadratic SVM	Quadratic	Automatic	370	16.30
M3	Cubic SVM	Cubic	Automatic	360	16.07
M4	Fine Gaussian SVM	Gaussian	0.79	360	15.85
M5	Medium Gaussian SVM	Gaussian	3.2	380	15.58

Statistical measures

Various error analysis approaches were used to evaluate the accuracy of the machine learning models. The aforementioned metrics included the coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), scatter index (SI), coefficient of efficiency (*E*), and Akaike Information Criterion (AIC). The relevant equations were used to compute these measures (Kaushik & Kumar 2023a).

$$R^2 = \frac{\sum_{i=1}^{N} (a_i - \bar{a})^2 (p_i - \bar{p})^2}{\sum_{i=1}^{N} (a_i - \bar{a})^2 \sum_{i=1}^{N} (p_i - \bar{p})^2} \tag{14}$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (p_i - a_i)^2 \tag{15}$$

Downloaded from http://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2024.117/1416749/wpt2024117.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - a_i)^2} \tag{16}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|p_i - a_i| \tag{17}$$

$$\text{MAPE}\ (\%) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{|p_i - a_i|}{a_i} \times 100\right) \tag{18}$$

$$SI = \sqrt{\frac{\sum_{i=1}^{N}[(p_i - \bar{p}) - (a_i - \bar{a})]^2}{\sum_{i=1}^{N} a_i^2}} \tag{19}$$

$$E = 1 - \frac{\sum_{i=1}^{N}(p_i - a_i)^2}{\sum_{i=1}^{N}(a_i - \bar{a})^2} \tag{20}$$

$$AIC = N\log\left(\frac{1}{N}\sum_{i=1}^{N}(p_i - a_i)^2\right) + 2k \tag{21}$$

RESULTS AND DISCUSSION

Figure 3 depicts the longitudinal distance-dependent variation of Manning's roughness coefficient (n) in a compound channel with smooth converging floodplains for different relative flow depths. The decrease in hydraulic radius along the longitudinal distance results in a corresponding reduction in the n values within the PS, extending up to the midpoint of the converging portion. Following the midpoint of the converging section, there is a

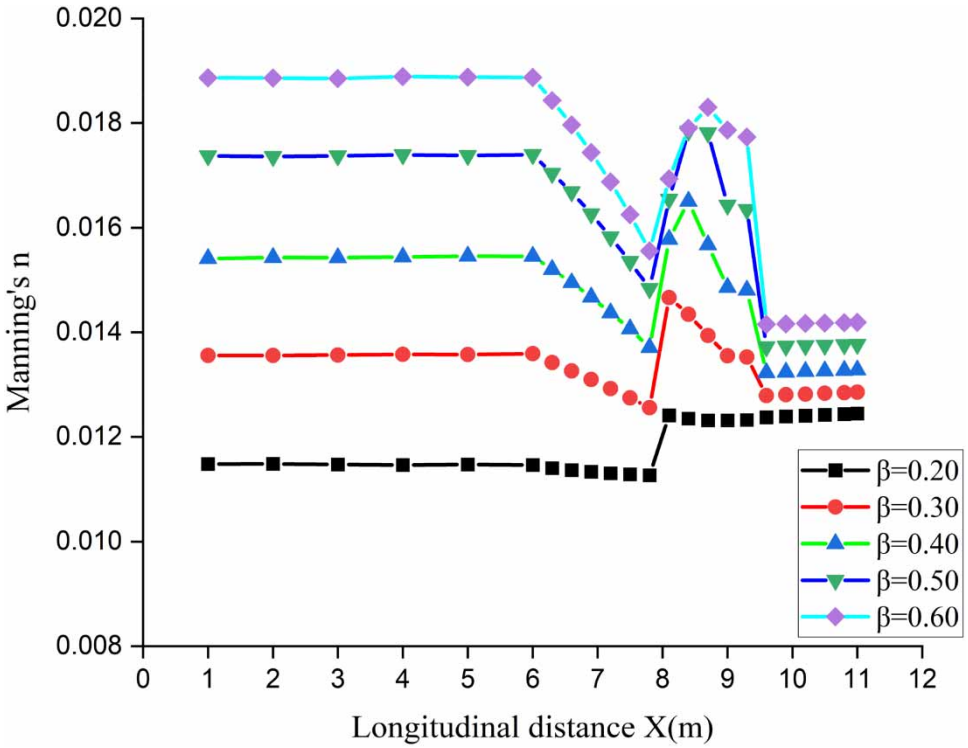


Figure 3 | Variation of Manning's n along the longitudinal distance of the non-prismatic compound channel.

Downloaded from http://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2024.117/1416749/wpt2024117.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

notable increase in the numerical values of n due to the convergent nature of the channel's geometry. The observed outcomes are attributed to a significant rise in the energy gradient relative to the hydraulic radius. Subsequently, the downstream section experiences a decrease in n values due to the reduction in hydraulic radius and energy slope as the flow progresses further along its length. Similar fluctuation patterns are observed in higher relative flow depths as well as in lower relative flow depths. The study reveals that a significant increase in roughness occurs primarily in the converging regions as the relative depth escalates from 0.20 to 0.60. This phenomenon can be attributed to the transfer of momentum from the main channel to the floodplains. The efficacy of the SVM methodology in generating precise forecasts is evidenced by the scatter diagrams that employ diverse models to juxtapose the projected and actual n values, as depicted in Figure 4. The SVM model exhibits

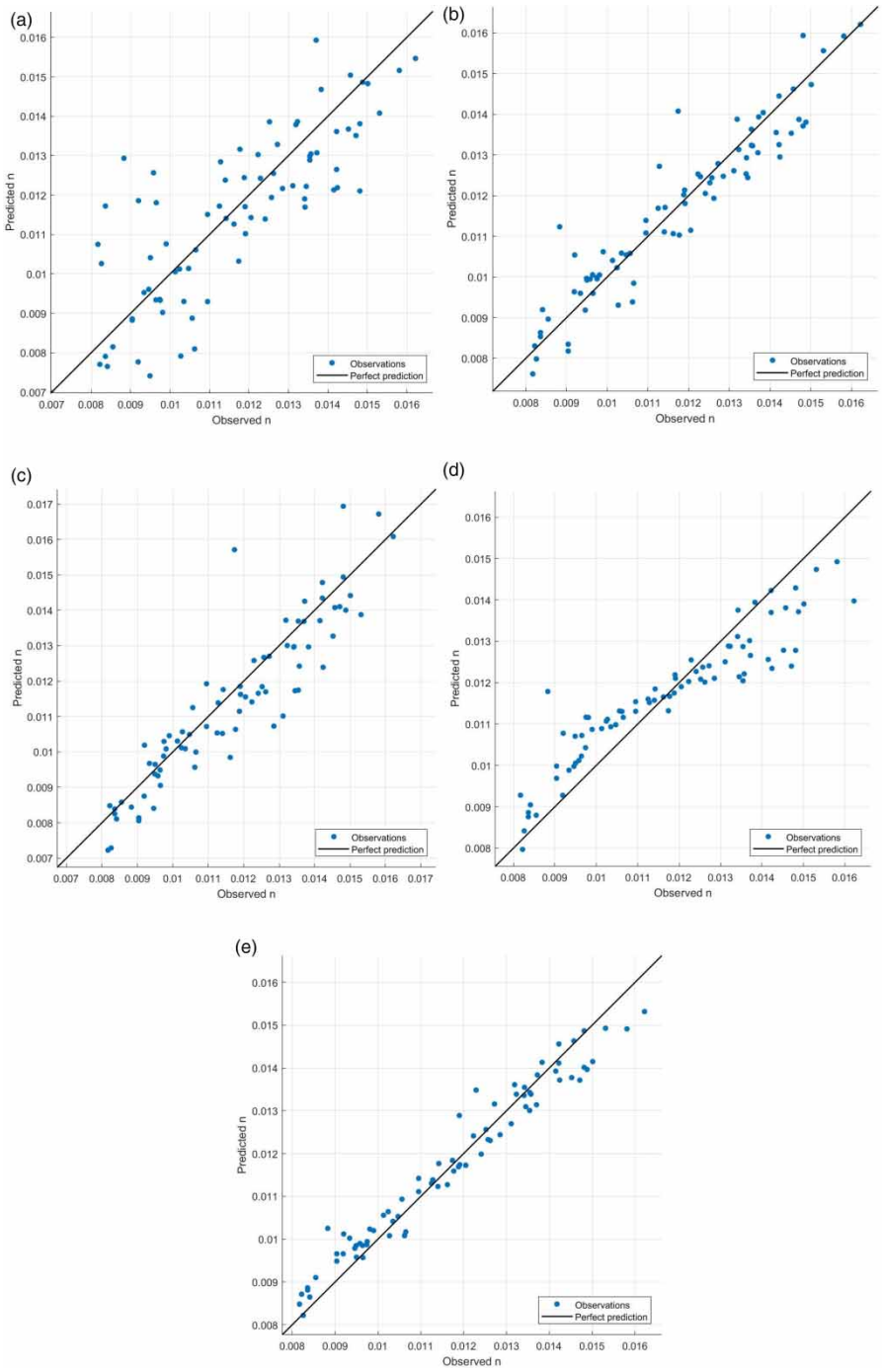


Figure 4 | Scatter plot to predict Manning's roughness coefficient (*n*) for (a) model 1; (b) model 2; (c) model 3; (d) model 4; and (e) model 5.

Downloaded from http://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2024.117/1416749/wpt2024117.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

significant potential, as evidenced by the notable indication of the values' proximity to the line representing exceptional concurrence. The SVM predicted values that are dispersed over the best-fitting line exhibit a superior level of concurrence with the empirical data, resulting in elevated R^2 values. According to the data presented in Figure 4(e), the SVM model five that was created using a 70:30 ratio of data in training and testing yielded predicted values of Manning's n that were more closely aligned with the best fitted line than other developed models. This suggests that the SVM model five is more accurate in its predictions.

In order to assess the effectiveness of the SVM models, statistical metrics including R^2, MSE, RMSE, MAE, MAPE, SI, E, and AIC were employed, and their corresponding numerical values are displayed in Table 3. The SVM model five has demonstrated superior predictive performance in estimating the Manning's roughness coefficient of composite waterways with converging floodplains, when compared to alternative approaches such as Das & Khatua (2018), as well as established theoretical methods including the Lotter method for composite roughness computation. This is evident from the highest R^2 value, lowest MSE value, lowest RMSE value, lowest MAE value, and lowest MAPE value. The SI serves as a standardized measure for the quantification of inaccuracy. A lower SI value is indicative of a higher level of model performance. The SVM model five has a significantly reduced SI value of 0.033 in both the training and testing stages, indicating a minimum level of error. The efficiency measure, represented as E, is assigned a value of 1 when there is a complete correspondence between the model and the observed data. The interval of E extends from negative infinity to 1. The SVM model five has an E value of 0.95 for both the training and testing datasets. The AIC is a widely used criterion in statistical model selection. It is used to choose the best suitable model by evaluating the statistical likelihood function. The model that has the lowest AIC value is regarded as the most optimum model. The SVM model five has the lowest AIC score of −510.754 when compared to the other models proposed. The statistical study of several indices indicates that SVM model five exhibits superior performance compared to other models and established methodologies in forecasting the Manning's roughness coefficient in compound channels with narrowing floodplains.

Table 3 | Statistical analysis of predicted n by various approaches

Statistical parameters	SVM model 1	SVM model 2	SVM model 3	SVM model 4	SVM model 5	Das & Khatua Method (2018)	Lotter method (1933)
R^2	0.62	0.90	0.82	0.82	0.95	0.80	0.78
MSE	1.82×10^{-6}	4.84×10^{-7}	8.82×10^{-7}	8.67×10^{-7}	2.33×10^{-7}	0.0019	0.0025
RMSE	0.00135	0.000696	0.000939	0.000931	0.000483	0.043	0.05
MAE	0.00104	0.000516	0.000676	0.000727	0.000386	0.002	0.045
MAPE	5.168	2.917	4.309	3.728	2.405	10.921	15.813
SI	0.059	0.042	0.061	0.044	0.033	0.088	0.117
E	0.62	0.90	0.82	0.82	0.95	0.70	0.78
AIC	−469.873	−505.382	−472.319	−493.169	−510.754	−250.614	−105.291

CONCLUSIONS

The study employs SVM to estimate the Manning's roughness coefficient (n) in non-prismatic compound channel with smooth narrowing floodplains. The SVM models are developed using high-quality experimental datasets with varying geometrical and flow variables. The study indicates that the proposed model is influenced by a range of features, such as the width ratio, relative depth of flow, aspect ratio, angle of convergence, relative longitudinal distance, longitudinal bed slope, energy slope, discharge ratio, Froude number, and Reynolds number. In the prismatic portion, the Manning roughness coefficient decreases, and in the NPS, it decreases up to the middle. After then, a precipitous increase in the roughness value is seen. The present study aims to examine the correlation between the Manning's roughness coefficient and the nondimensional geometric and flow variables of a composite waterway with narrowing floodplains. It has been discovered that there is a connection that is not linear between each of the parameters. In comparison to previously developed techniques, the newly created model demonstrates superior performance in terms of statistical parameters for a variety of

Downloaded from http://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2024.117/1416749/wpt2024117.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

datasets. In line with the evaluation criteria, the SVM model five could fairly forecast the Manning's coefficient in compound channels with converging floodplains. This was shown by the fact that the technique was successful. Since it had the best R^2 and the lowest MSE, RMSE, MAE, MAPE, SI, and AIC, the SVM model developed using Gaussian kernel function demonstrated superior performance in the prediction of Manning's n in non-prismatic compound channels. The models proposed in this paper possess practical implications for non-prismatic rivers, such as the River Main in Northern Ireland, the Brahmaputra River in India, and other analogous river systems. The results of this research will be valuable for informing the design of flood control and diversion structures, leading to a reduction in both economic and human losses.

ACKNOWLEDGEMENTS

The authors express their gratitude for the help received from the Department of Civil Engineering at Delhi Technological University in Delhi, India.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Bhattacharya, A. K. 1995 *Mathematical Model of Flow in A Meandering Channel*. PhD Thesis, Indian Institute of Technology Kharagpur, India.

Bjerklie, D. M., Dingman, S. L. & Bolster, C. H. 2005 Comparison of constitutive flow resistance equations based on the Manning and Chezy equations applied to natural rivers. *Journal of Water Resources Research* **41**, 11. https://doi.org/10.1029/2004wr003776.

Borges, L. M., Mekitarian Filho, E. R. & Paiva, A. C. 2016 Support vector regression applied to magnetic resonance imaging: An approach to predicting hepatic iron concentration. *Journal of Digital Imaging* **29**(1), 70–77.

Bousmar, D. & Zech, Y. 2002 Periodical turbulent structures in compound channels. In: *River Flow International Conference on Fluvial Hydraulics*. Louvain-la-Neuve, Belgium, pp. 177–185.

Bousmar, D., Wilkin, N., Jacquemart, J. H. & Zech, Y. 2004 Overbank flow in symmetrically narrowing floodplains. *Journal of Hydraulic Engineering* **130**(4), 305–312.

Cao, Z., Meng, J., Pender, G. & Wallis, S. 2006 Flow resistance and momentum flux in compound open channels. *Journal of Hydraulic Engineering* **132**(12), 1272–1282. https://doi.org/10.1061/(asce)0733-9429(2006)132:12(1272).

Chlebek, J., Bousmar, D., Knight, D. W. & Sterling, M. A. 2010 Comparison of overbank flow conditions in skewed and converging/diverging channels. In: *River Flows International Conference*, pp. 503–511.

Cox, R. G. 1973 *Effective hydraulic roughness for channels having bed roughness different from bank roughness*. Miscellaneous paper H-73-2. U.S. Army Corps of Engineers Waterways Experiment Station, Vicksburg, MS.

Das, B. S. & Khatua, K. K. 2018 Flow resistance in a compound channel with diverging and converging floodplains. *Journal of Hydraulic Engineering* **144**, 8. https://doi.org/10.1061/(asce)hy.1943-7900.0001496.

Das, B. S., Devi, K., Khuntia, J. R. & Khatua, K. K. 2020 Discharge estimation in converging and diverging compound open channels by using adaptive neuro-fuzzy inference system. *Canadian Journal of Civil Engineering* **47**(12), 1327–1344. https://doi.org/10.1139/cjce-2018-0038.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. J. & Vapnik, V. 1997 Support vector regression machines. *Advances in Neural Information Processing Systems* **9**, 155–161.

Einstein, H. A. & Banks, R. B. 1950 Fluid resistance of composite roughness. *Eos, Transactions American Geophysical Union* **31**(4), 603–610. https://doi.org/10.1029/TR031i004p00603.

Kar, S. K. 1977 *A Study of Distribution of Boundary Shear in Meander Channel with and Without Floodplain and River Floodplain Interaction*. *PhD Thesis*, Indian Institute of Technology Kharagpur, India.

Kaushik, V. & Kumar, M. 2023a Assessment of water surface profile in nonprismatic compound channels using machine learning techniques. *Water Supply* **23**(1), 356–378. https://doi.org/10.2166/ws.2022.430.

Kaushik, V. & Kumar, M. 2023b Sustainable gene expression programming model for shear stress prediction in nonprismatic compound channels. *Sustainable Energy Technologies and Assessments* **57**, 103229. https://doi.org/10.1016/j.seta.2023.103229.

Kaushik, V. & Kumar, M. 2023c Water surface profile prediction in non-prismatic compound channel using support vector machine (SVM). *AI in Civil Engineering* **2**, 6. https://doi.org/10.1007/s43503-023-00015-1.

Khatua, K. K., Patra, K. C. & Mohanty, P. K. 2011a Stage-discharge prediction for straight and smooth compound channels with wide floodplains. *Journal of Hydraulic Engineering* **138**(1), 93–99.

Downloaded from http://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2024.117/1416749/wpt2024117.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

Khatua, K. K., Patra, K. C. & Nayak, P. 2011b Meandering effect for evaluation of roughness coefficients in open channel flow. *WIT Transactions on Ecology and the Environment* **146**, 213–224.

Khuntia, J. R., Devi, K. & Khatua, K. K. 2018 Boundary shear stress distribution in straight compound channel flow using artificial neural network. *Journal of Hydrologic Engineering* **23**(5), 04018014. doi:10.1061/(asce)he.1943-5584.0001651.

Knight, D. W. & Demetriou, J. D. 1983 Flood plain and main channel flow interaction. *Journal of Hydraulic Engineering* **109**(8), 1073–1092.

Lotter, G. K. 1933 Considerations on hydraulic design of channels with the different roughness of walls. *Transactions, All-Union Scientific Research Institute of Hydraulic Engineering* **9**, 238–241.

Matin, N. G. 2022 Evaluation of effective parameters of Manning roughness coefficients in HDPE culverts via kernel-based approaches. *Journal of Hydroinformatics* **24**(6), 1111–1126. https://doi.org/10.2166/hydro.2022.003.

MATLAB R 2019 [Computer Software]. MathWorks, Natick, MA.

Mohanta, A., Patra, K. & Sahoo, B. 2018 Anticipate Manning's coefficient in meandering compound channels. *Hydrology* **5**(3), 47. https://doi.org/10.3390/hydrology5030047.

Mohanty, P. K. 2013 *Flow Analysis of Compound Channels with Wide Floodplains*. PhD Thesis, National Institute of Technology Rourkela, Odisha, India.

Myers, W. R. C. 1987 Velocity and discharge in compound channels. *Journal of Hydraulic Engineering* **113**(6), 753–766.

Naik, B., Khatua, K. K., Wright, N. G. & Sleigh, A. 2017 Stage-Discharge prediction for converging compound channels with narrow floodplains. *Journal of Irrigation and Drainage Engineering* **143**, 8. https://doi.org/10.1061/(asce)ir.1943-4774.0001184.

Naik, B., Kaushik, V. & Kumar, M. 2022 Water surface profile in converging compound channel using gene expression programming. *Water Supply* **22**(5), 5221–5236. https://doi.org/10.2166/ws.2022.172.

Parsaie, A. & Haghiabi, A. H. 2017 Mathematical expression of discharge capacity of compound open channels using MARS technique. *Journal of Earth System Science* https://doi.org/10.1007/s12040-017-0807-1.

Parsaie, A., Yonesi, H. & Najafian, S. 2017 Prediction of flow discharge in compound open channels using adaptive neuro fuzzy inference system method. *Flow Measurement and Instrumentation* **54**, 288–297.

Patel, V. C. 1965 Calibration of the Preston tube and limitations on its use in pressure gradients. *Journal of Fluid Mechanics* **231**, 85–208.

Patra, K. C. 1999 *Flow Interaction of Meandering River with Flood Plains*. PhD Thesis, Indian Institute of Technology Kharagpur, India.

Patra, K. C. & Kar, S. K. 2000 Flow interaction of meandering river with floodplains. *Journal of Hydraulic Engineering* **126**(8), 593–604.

Patra, K. C., Kar, S. K. & Bhattacharya, A. K. 2004 Flow and velocity distribution in meandering compound channels. *Journal of Hydraulic Engineering* **130**(5), 398–411.

Proust, S., Rivière, N., Bousmar, D., Paquier, A. & Zech, Y. 2006 Flow in the compound channel with abrupt floodplain contraction. *Journal of Hydraulic Engineering* **132**(9), 958–970.

Rezaei, B. 2006 *Overbank Flow in Compound Channels with Prismatic and Nonprismatic Floodplains*. Ph.D. Thesis, University of Birmingham, Birmingham, UK.

Rezaei, B. & Knight, D. W. 2009 Application of the Shiono and Knight method in the compound channel with nonprismatic floodplains. *Journal of Hydraulic Research* **47**(6), 716–726.

Rezaei, B. & Knight, D. W. 2011 Overbank flow in compound channels with nonprismatic floodplains. *Journal of Hydraulic Engineering* **137**(8), 815–824. https://doi.org/10.1061/(asce)hy.1943-7900.0000368.

Sellin, R. H. J. 1964 A laboratory investigation into the interaction between flow in the channel of a river and that of its flood plain. *LaHouille Blanche* **7**, 793–801.

Sharma, S., Ahmad, S. & Gupta, A. 2012 Support vector regression for financial time series forecasting: A comparative study with artificial neural network. *Expert Systems with Applications* **39**(3), 3509–3518.

Smola, A. J. & Schölkopf, B. 2004 A tutorial on support vector regression. *Statistics and Computing* **14**(3), 199–222.

Yang, S.-Q. & Lim, S.-Y. 2005 Boundary shear stress distributions in trapezoidal channels. *Journal of Hydraulic Research* **43**(1), 98–102.

Yang, K., Cao, S. & Liu, X. 2005 Study on resistance coefficient in compound channels. *Acta Mechanica Sinica* **21**, 353–361.

Yang, K., Cao, S. & Liu, X. 2007 Flow resistance and its prediction methods in compound channels. *Acta Mechanica Sinica* **23**, 23–31.

Yeh, A. G.-O., Huang, C.-H. & Li, X. 2010 A support vector machine with a quantum-behaved particle swarm optimization approach for predicting transportation mode choice. *Expert Systems with Applications* **37**(12), 7922–7929.

Yonesi, H. A., Omid, M. H. & Ayyoubzadeh, S. A. 2013 The hydraulics of flow in nonprismatic compound channels. *Journal of Civil Engineering and Urbanism* **3**(6), 342–356.

Yonesi, H. A., Parsaie, A., Arshia, A. & Shamsi, Z. 2022 Discharge modeling in compound channels with nonprismatic floodplains using GMDH and MARS models. *Water Supply* **22**(4), 4400–4421. https://doi.org/10.2166/ws.2022.058.

First received 27 November 2023; accepted in revised form 24 April 2024. Available online 6 May 2024

Downloaded from http://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2024.117/1416749/wpt2024117.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE) | 979-8-3503-8328-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICWITE59797.2024.10503360

Remote Sensing and Machine Learning Fusion: A Robust Framework for Land Use and Land Cover Change Detection

Kriti Tripathi
School of Geoinformatics, R.S.A.C
Dr. APJ Abdul Kalam University Lucknow, India
tripathikriti1905@gmail.com

Abhishek Verma
Department of Information Technology
Delhi Technological University New Delhi, India
Abhishekcms08@gmail.com

Abstract— **Determining land use and land cover (LULC) changes in an area is crucial before proceeding with any developmental or restoration plan. Traditionally, field surveys were done to mark significant changes in an area, often compared with toposheets of the region. However, this task was time-consuming and required a substantial human workforce and resources. With the advent of satellite sensor technology, the task of field surveys was minimised by a considerable percentage. The satellite imageries, capable of providing high-resolution and historical data, were analysed for land use/cover changes. The results of this study show a reduction in water bodies by a percentage of nearly 41% in a span of 28 years. Rapid industrialisation and illegal residency have contributed to an increase of 33% in built-up. Barren/fallow land shows a growth of nearly 79%, agricultural practices remain dominant in the region, whereas forest decreases 4%. Despite the time-saving capabilities, satellite imagery requires substantial processing. With the advancements in computer science and Geographical Information Systems (GIS), classification techniques came into the image, and LULC studies gave further refined and near-accurate results. The machine learning techniques, i.e. Maximum Likelihood Classification (MLC), have been used in this study to recognise its benefits in the timely analysis of a problem statement of LULC and will help review and monitor the study area.**

Keywords: Supervised classification, LULC, Hindon River basin, GIS, Maximum Likelihood Classification, ML

I. INTRODUCTION

A global scarcity of well-preserved watersheds leads to an imbalanced ecology. Riverine systems support biodiversity and serve as a crucial supply of freshwater. Nevertheless, the deterioration of watersheds and the current predicament can be attributed to unsustainable activities. The research area encompasses the Hindon River basin, which stretches over several unauthorised settlements close to the river, where the residents dispose of solid waste and sewage straight into the river. It is imperative to designate these settlements and develop a strategy for rehabilitating the inhabitants[1].

To achieve this objective, satellite imagery may be utilised to delineate the floodplain and watershed of the river. Subsequently, image analysis can be conducted to identify the distinct land use categories and track their temporal variations. Furthermore, analysing elements such as rainfall patterns, groundwater levels, and agricultural practices contributing to these changes is possible. Comprehending the connection between land use and land cover (LULC) and water resources is crucial for efficiently managing watersheds, formulating policies, considering future LULC development, and safeguarding freshwater [2]. In the monsoon season of 2023, the Hindon River basin had copious precipitation. The Hindon, often arid and devoid of life throughout the non-perennial period, experiences a resurgence of activity during the monsoon season. Due to copious rainfall this year, the river in Ghaziabad and Noida districts experienced a significant increase in water volume, resulting in the inundation of surrounding encroached areas and subsequent flooding of dwellings. As a result, the occupants were compelled to evacuate. The villages of Chajjarsi, Chotpur, Yusufpur, Kakrala, Mubarakpur, and Shafipur have witnessed the construction of several illegal residences near the river[3].

Using categorisation strategies in Land Use and Land Cover (LULC) research may assist in comprehending land and water resources more effectively, facilitating their suitable administration. Remote sensing imagery can provide a comprehensive and detailed perspective of a region, allowing for a broad overview and the ability to analyse changes over time. [4].

Artificial Intelligence (AI) and Machine Learning (ML) are robust technologies that significantly enhance the precision of such forms of analysis. Classification is a machine-learning approach that groups comparable information based on shared properties. A classification model is utilised to forecast the category of a dataset. Image classification algorithms are employed in GIS to categorise an image. Each pixel is regarded as a unit of data. Classes are allocated according to the user's requirements. Supervised classification refers to the scenario where the user provides classes or training datasets to the model before categorising the dataset. On the other hand, unsupervised classification refers to when the model first classifies the data into distinguished classes, and then the user contributes names to those classes[5].

Another form of image classification approach utilised in remote sensing is object-based categorisation. The first two strategies are widely used, whereas Object-based

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:55:15 UTC from IEEE Xplore. Restrictions apply.

categorisation is particularly beneficial for high-resolution data.[6]

A. Study Area

The river Hindan, a notable tributary of the river Yamuna, flows through the six districts of western Uttar Pradesh: Saharanpur, Muzaffarnagar, Baghpat, Meerut, Ghaziabad, and Gautambudh Nagar. The river holds considerable historical and cultural significance and serves as the primary water source for the primarily rural inhabitants of western Uttar Pradesh. The Hindan River is located in the western region of the northern Indian state of Uttar Pradesh. Its basin extends between coordinates 77°49'51.42"E and 30°16'47.85"N to 77°29'43.5588"E and 28°24'32.2416"N. [7]

II. Related Work

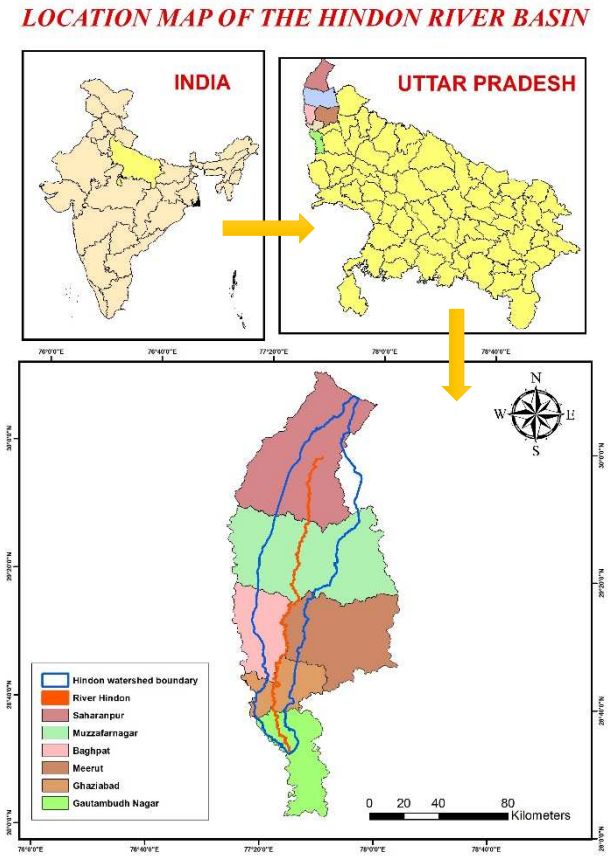


Fig. 1 Study Area

This study did an extensive literature review to detect any existing gaps. Additionally, it hoped to utilise remotely sensed data for mapping and monitoring changes in land use and land cover (LULC). The study also sought to examine the impact of these changes on the Hindon River. Furthermore, the study wanted to determine the percentage of LULC classes that have changed over the years.

A. Study on Remote Sensing Data Classification using Statistical Techniques.[8]

This study did a comprehensive literature analysis to identify any existing gaps. It then utilised remotely sensed data to accurately map and monitor land use and land cover (LULC) changes and their impact on the Hindon River. Additionally, the study quantified the proportion of LULC class changes throughout the years.

The mentioned study examined the statistical methodologies used for categorising data. Below are detailed explanations of the Minimum distance, Maximum likelihood, and Mahalnobis distance algorithms. The Maximum Likelihood is a widely used supervised classification approach in remote sensing image data. The algorithms have demonstrated favourable outcomes. Nevertheless, the correctness of the algorithm is significantly contingent upon the calibre of the generated data. These image classification techniques are effective for classifying satellite images. This work can be enhanced by combining many algorithms, which might improve the result's accuracy. Among the three methods used, Maximum likelihood classification had the highest accuracy of 92%, followed by Minimal distance with 90% and Mahalanobis distance with 86% accuracy.

B. Analysis of Supervised Maximum Likelihood Classification using Remote Sensing Image [8]

The MLC algorithm relies on Bayes' categorisation. The technology of MLC is quite sturdy. The categorisation was performed on a Landsat ETM+ remote sensing image from 2011. Ground truthing and validation have been conducted at 20 sites, including the entirety of the research region. The study also reveals that utilising MLC reduces the likelihood of misclassification. MLC assumes a normal data distribution; however, this may not always be true. Therefore, it is essential to pick the training data carefully. Additionally, the study revealed a direct correlation between the increase in image size and the corresponding increase in processing cost. The study also includes an accuracy evaluation; the algorithm attained an overall accuracy rate of 93.75%.

III. Proposed Approach

The suggested technique entails acquiring toposheets for the designated research region and satellite imageries, specifically from 2020. The initial phase is demarcating the watershed border of the river by utilising data from the Shuttle Radar Topographic Mission (SRTM) Digital Elevation Model (DEM). The SRTM provides DEM data with a spatial resolution of 30 meters. The watershed delineation was performed using the toposheet as the foundation map. The watershed covers an expanse of 5744 square kilometres. The extended configuration of the basin facilitates a gradual flow of water, resulting in a delayed discharge of all the water from the basin into the exit, which is the river Yamuna. Once the watershed has been delineated, our next objective is to conduct a classification of the satellite images of the research region[9]. The photos are from the Landsat 8 mission[10], with a temporal resolution of 16 days and a spatial resolution of 30 meters. The obtained images are timestamped as 16/05/2020.

The Hindon River spans a length of around 350 kilometres and predominantly experiences aridity. The river is heavily impacted by the disposal of solid waste, industrial effluents, and sewage discharge from residential and industrial areas along its banks. The primary tributaries include the Kali, Krishna, and Dhamola. The imagery was subjected to supervised classification to quantify the area covered by five land use classes[11]. These classes were identified and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:55:15 UTC from IEEE Xplore. Restrictions apply.

compared, including river/waterbody, forest, barren/fallow, built-up, and agricultural land/. The classification is performed using the maximum likelihood classification (MLC) method. The MLC is a widely utilised technique for categorising pixel data. The spectral pattern detected in the data for each pixel is used for data classification. The numerical value of each pixel is determined based on its spectral response. The bands utilized in the construction of a colour infrared mosaic for Landsat images are band 5 (near infrared), band 4 (red), and band 3 (green. Subsequently, these bands are allocated red, green, and blue to create a composite image with false colours. In a supervised technique, the user furnishes a training dataset for each class, which is then recorded as a signature file. Each class is allocated three to four distinct regions in the images corresponding to a specific land cover type. Each pixel in the image is assigned a specific grey value using three bands or channels. The grey values are subsequently matched to the signature of each land cover type in the training dataset. The particular pixel is allocated after being compared.

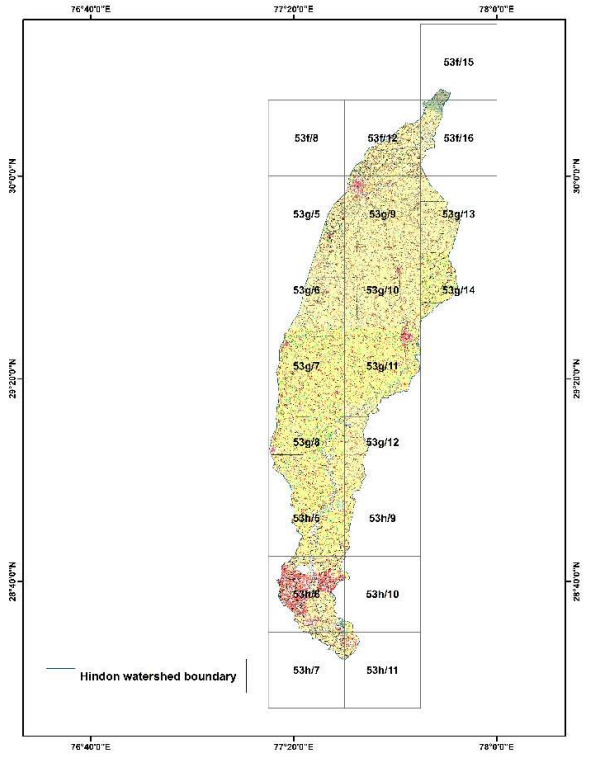


Fig. 2 Toposheets for Hindon basin

A. Satellite imagery used

Satellite imagery from the Landsat 8 mission has been obtained from earthexplore.usgs.gov. Landsat 8 was deployed on February 11, 2013, and has two sensors: the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). Landsat 8, capable of capturing images in several wavelengths, offers imagery in 9 distinct bands. The bands used to create an infrared colour composite are 5 (near-infrared), 4 (red), and 3 (green). The bands possess a spatial resolution of 30 meters, and the images were obtained during the pre-monsoonal period in May 2020[12].

B. Toposheet used

Toposheets have been obtained from the official website of the Survey of India, surveyofIndia.gov.in, and they are on a scale of 1:50000. The whole research region has been covered by a total of 18 toposheets that have been procured, cropped, georeferenced and a mosaic is obtained[13].

C. Evaluation Techniques

After obtaining the satellite images, they undergo preprocessing for subsequent analysis. The topographic sheets are georeferenced and combined to create a mosaic. The desired bands of the images are then selected and assigned false colours. The study concluded that analysing land use and cover for only one year would not provide significant benefits. Therefore, it is necessary to investigate the temporal changes in the region. The toposheets are utilised for this specific objective. The land use categories were measured to facilitate a more accurate comparison study. MLC has been used to achieve an accuracy of around 92%[8]. The MLC tool, accessible in the QGIS software, has been recognised as a highly efficient classification method.

III. METHODOLOGY

The suggested technique entails acquiring toposheets for the designated research region and satellite imageries specifically from the year 2020. The initial phase is demarcating the watershed border of the river by utilizing data from the Shuttle Radar Topographic Mission (SRTM) Digital Elevation Model (DEM)and flow diagram is well illustrated in fig 3. The SRTM provides DEM data with a spatial resolution of 30 meters. The watershed delineation was performed using the toposheet as the foundation map. The watershed covers an area of 5744 square kilometers.

The extended configuration of the basin facilitates the gradual flow of water, resulting in a delayed discharge of all the water from the basin into the exit, namely River Yamuna. Once the watershed has been delineated, our next objective is

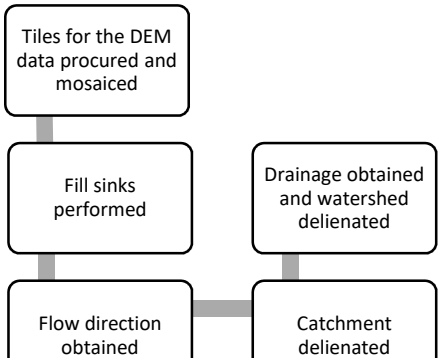


Fig. 3 Flow diagram for generation of drainage map

to conduct classification of the satellite images in the research region. The images are from the Landsat 8 mission, which has a temporal resolution of 16 days and a spatial resolution of 30 meters. The obtained images are from the date 16/05/2020. The Hindon river spans a length of around 350 kilometers and predominantly experiences aridity. The river is heavily impacted by the disposal of solid waste, industrial effluents, and sewage from residential and industrial areas located along its banks. The primary tributaries of the river include the Kali River, Krishni River, and Dhamola River. The

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:55:15 UTC from IEEE Xplore. Restrictions apply.

imagery has been classified into five distinct classes: river/waterbody, forest/plantation, barren/fallow, built up, and agricultural land/plantation.

The classification is performed using the maximum likelihood classification (MLC) method and workflow is well explained in fig 4. The MLC is a widely utilized technique for categorizing pixel data. The spectral pattern detected in the data for each pixel is utilized for data classification. The numerical value of each pixel is determined based on its spectral response. The bands utilized in the construction of a color infrared mosaic for Landsat images are band 5 (near infrared), band 4 (red), and band 3 (green). Subsequently, these bands are allocated the colors red, green, and blue, respectively, in order to create a composite image with false colors. Under a supervised method, the user furnishes a training dataset for each class, which is then recorded as a signature file. Each class is allocated three to four distinct regions in the images that correspond to a certain land cover type. Each pixel in the picture is assigned a specific gray value using three bands or channels. Subsequently, the grayscale values are compared to the signature of each land cover type in the training dataset. Upon conducting a comparison, the specific pixel is allocated to its corresponding class.

All the pixels are categorized uniformly. To categorize data, QGIS 3.28.2 has been utilized. The signature files are specified and stored in the gsg format. The maximum likelihood supervised classification method is used to classify overlapping signatures. Within the framework of MLC, the pixel is allocated to the class that possesses the highest likelihood. The MLC algorithm assumes that the likelihood of a pixel belonging to a class is equal for all classes, and that the input bands follow a normal distribution. The classification process involves computing the Spectral Euclidean distance between the target pixel's vector and the average vector for each signature class. The spectral Euclidean distance refers to the distance between two multiband raster, where the spectra of each pixel are considered as vectors [9] The algorithm is time-consuming and frequently results in misclassification of pixels that should remain unclassified. Ground truthing is a method used to confirm or validate anything.

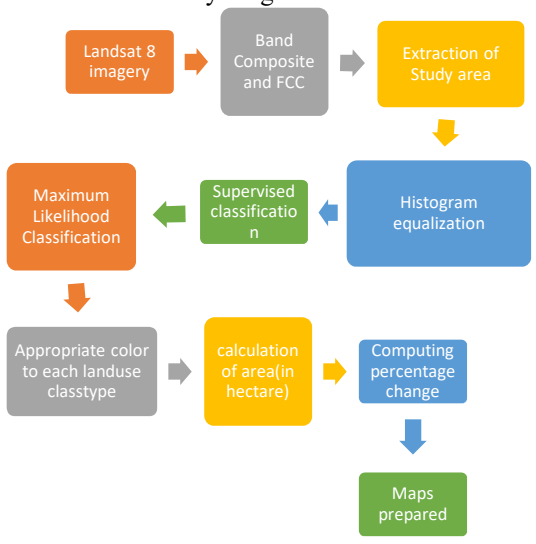


Fig. 3 Workflow for performing Maximum Likelihood Classification

IV. Result & Discussion

After the image pixels have been classified into the five classes of land cover type, each class is assigned an appropriate colour, and a map is prepared that depicts the LULC in the study area. Each class's area (in hectares) in the basin is calculated using mathematical functions.

For determining the temporal change in each class, a LULC map was prepared based on the toposheets of the study area.

Table 1 Percentage change of classes in Hindon Basin

Classes	1972	2020	Percentage change
	Area in Hectares		
Water	5204.5	3040.6	41.5%decrease
Forest	6230.8	5977.6	4% decrease
Barren	3042.9	5470.8	79.78%increase
Built up	42508.6	56931.3	33.9%increase
Agricultural	518572.9	502479.6	3.1% decrease

Subsequently another map was prepared based on satellite imagery.

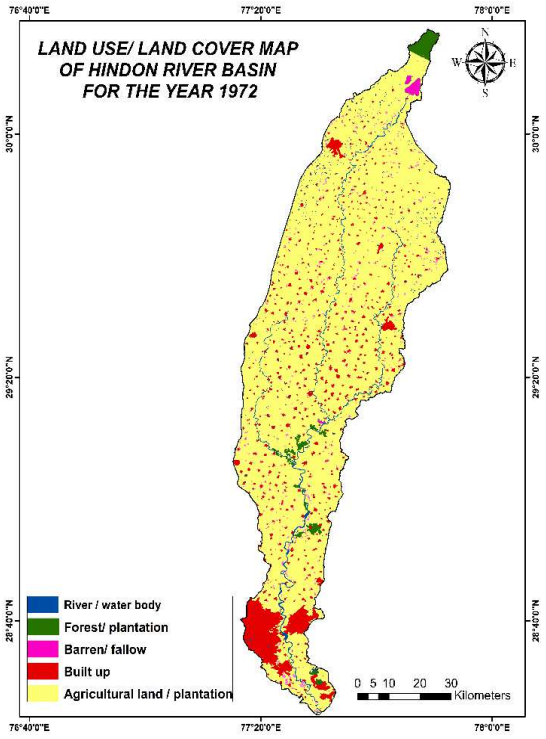


Fig. 4 LULC 1972

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:55:15 UTC from IEEE Xplore. Restrictions apply.

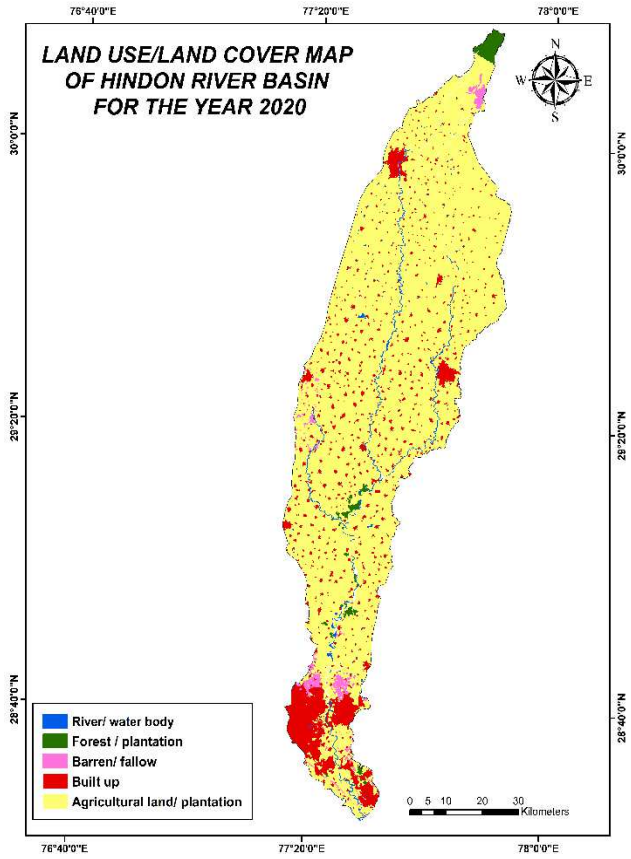


Fig. 5 LULC for 2020

A total of 18 toposheets have been downloaded and mosaiced from which the study area has been extracted. Likewise, each class's area is calculated. The difference between the values of the two data gives the percentage change in the area of each landcover type (see Table 1)

This result shows the land use and cover change over twenty-eight years. The river/water body has seen a 41.577% decrease, which indicates the deteriorating condition of the hydrological systems of the area. Agricultural practices remain prevalent in the basin. Plantation and forests see a four per cent decrease, corresponding to ignorant forestation practices. Built-up has increased by approximately 34%, given the present scenario where the basin has many industrial setups, especially in the highly populated regions of Ghaziabad and Noida. Visually, the maps prepared after classification show the population cluster in the basin and how much of the floodplain is illegally encroached upon.

V. Conclusion and Future Scope

In the wake of significant land use changes observed over the past 28 years in the study area, future endeavors should focus on implementing a continuous monitoring system for a more nuanced understanding of evolving patterns. Integrating advanced remote sensing technologies and machine learning algorithms can enhance the precision of land cover assessments. A comprehensive environmental impact assessment is imperative to discern the consequences of these changes on biodiversity and ecosystem health, guiding targeted conservation efforts. Moreover, fostering community engagement and integrating findings into regional planning will be vital for sustainable land use practices. The study's conclusion underscores the urgency of informed policy decisions to address environmental challenges and promote a balanced approach to land use in the region. Furthermore, these machine learning methodologies have achieved exceptional levels of accuracy in many categorization assignments. These models are versatile and may be used in different real-life settings as future endeavors, including sentiment categorization and hate speech detection[14], [15], [18], [19]. Furthermore, it is crucial to enhance the resilience of these models against adversarial perturbations to achieve robust generalizations in adversarial settings[16], [17].

REFERENCES

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

[2] M. J. Mashala, T. Dube, B. T. Mudereri, K. K. Ayisi, and M. R. Ramudzuli, "A Systematic Review on Advancements in Remote Sensing for Assessing and Monitoring Land Use and Land Cover Changes Impacts on Surface Water Resources in Semi-Arid Tropical Environments," *Remote Sens.*, vol. 15, no. 16, Art. no. 16, Jan. 2023, doi: 10.3390/rs15163926.

[3] "Effect of Seasonal Variations and Impact Assessment of COVID-19 Lockdown on Water Quality Status of Hindon River, Ghaziabad | Water Resources." Accessed: Dec. 10, 2023. [Online]. Available: https://link.springer.com/article/10.1134/S0097807823600018

[4] N. Tyagi and S. Sahoo, "Assessing the status of changing regimes of water bodies in Gorakhpur District, Uttar Pradesh, India," *Environ. Monit. Assess.*, vol. 194, no. 2, p. 67, Jan. 2022, doi: 10.1007/s10661-021-09630-w.

[5] E. A. Alshari, M. B. Abdulkareem, and B. W. Gawali, "Classification of land use/land cover using artificial intelligence (ANN-RF)," *Front. Artif. Intell.*, vol. 5, 2023, Accessed: Dec. 10, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2022.964279

[6] J. Wang, W. Li, Y. Wang, R. Tao, and Q. Du, "Representation-Enhanced Status Replay Network for Multisource Remote-Sensing Image Classification," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–13, 2023, doi: 10.1109/TNNLS.2023.3286422.

[7] A. Saha and P. Singh, "Drainage Morphometric Analysis and Water Resource Management of Hindon River Basin, using Earth Observation Data Sets," *Imp. J. Interdiscip. Res. IJIR*, vol. 3, pp. 2051–2057, May 2017.

[8] P. Sisodia, V. Tiwari, and A. Kumar, "Analysis of Supervised Maximum Likelihood Classification for remote sensing image," *Int. Conf. Recent Adv. Innov. Eng. ICRAIE 2014*, Sep. 2014, doi: 10.1109/ICRAIE.2014.6909319.

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

[10] "Landsat 8 | Landsat Science." Accessed: Dec. 10, 2023. [Online]. Available: https://landsat.gsfc.nasa.gov/satellites/landsat-8/

[11] M. E. D. Chaves, M. C. A. Picoli, and I. D. Sanches, "Recent Applications of Landsat 8/OLI and Sentinel-2/MSI for Land Use and Land Cover Mapping: A Systematic Review," *Remote Sens.*, vol. 12, no. 18, Art. no. 18, Jan. 2020, doi: 10.3390/rs12183062.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:55:15 UTC from IEEE Xplore. Restrictions apply.

[12] Z. Yu *et al.*, "Selection of Landsat 8 OLI Band Combinations for Land Use and Land Cover Classification," in *2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, Jul. 2019, pp. 1–5. doi: 10.1109/Agro-Geoinformatics.2019.8820595.

[13] S. Sarma and T. Saikia, "Prioritization of Sub-watersheds in Khanapara–Bornihat Area of Assam–Meghalaya (India) Based on Land Use and Slope Analysis Using Remote Sensing and GIS," *J. Indian Soc. Remote Sens.*, vol. 40, no. 3, pp. 435–446, Sep. 2012, doi: 10.1007/s12524-011-0163-6.

[14] A. Yadav and D. K. Vishwakarma, "MRT-Net: Auto-adaptive weighting of manipulation residuals and texture clues for face manipulation detection," *Expert Syst. Appl.*, vol. 232, p. 120898, Dec. 2023, doi: 10.1016/j.eswa.2023.120898.

[15] A. Yadav and D. K. Vishwakarma, "Deep learning algorithms for person re-identification: sate-of-the-art and research challenges," *Multimed. Tools Appl.*, Aug. 2023, doi: 10.1007/s11042-023-16286-w.

[16] A. Bajaj and D. Kumar Vishwakarma, "Evading text based emotion detection mechanism via adversarial attacks," *Neurocomputing*, vol. 558, p. 126787, Nov. 2023, doi: 10.1016/j.neucom.2023.126787.

[17] A. Bajaj and D. K. Vishwakarma, "HOMOCHAR: A novel adversarial attack framework for exposing the vulnerability of text based neural sentiment classifiers," *Eng. Appl. Artif. Intell.*, vol. 126, p. 106815, Nov. 2023, doi: 10.1016/j.engappai.2023.106815.

[18] A. Pandey and D. K. Vishwakarma, "Progress, achievements, and challenges in multimodal sentiment analysis using deep learning: A survey," *Appl. Soft Comput.*, vol. 152, p. 111206, Feb. 2024, doi: 10.1016/j.asoc.2023.111206.

[19] A. Pandey and D. K. Vishwakarma, "VABDC-Net: A framework for Visual-Caption Sentiment Recognition via spatio-depth visual attention and bi-directional caption processing," *Knowl.-Based Syst.*, vol. 269, p. 110515, Jun. 2023, doi: 10.1016/j.knosys.2023.110515.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 06:55:15 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully edi
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2024.3404413

Rethinking Latency-Aware DNN Design with GPU Tail Effect Analysis

Fuxun Yu, Zirui Xu, Longfei Shangguan, *Member, IEEE,* Di Wang, Dimitrios Stamoulis, Rishi Madhok, Nikolaos Karianakis, Ang Li, *Member, IEEE,* ChenChen Liu, *Member, IEEE,* Yiran Chen, *Fellow, IEEE,* and Xiang Chen, *Member, IEEE,*

***Abstract*—As the size of Deep Neural Networks (DNNs) continues to grow, their runtime latency also scales. While model pruning and Neural Architecture Search (NAS) can effectively reduce the computation workload, their effectiveness fails to consistently translate into runtime latency reduction. In this paper, we identify the root cause behind the mismatch between workload reduction and latency reduction is *GPU tail effect* — a classic system issue caused by resource under-utilization in the last processing wave of the GPU. We conduct detailed DNN workload characterization and demonstrate the prevalence of GPU tail effect across different DNN architectures, and meanwhile reveal that the unique deep structure and the light-weight layer workload of DNNs exacerbate the tail effect for DNN inference. We then propose a tail-awareness design space enhancement and DNN optimization algorithm to optimize existing NAS and pruning designs and achieve better runtime latency and model accuracy performance. Extensive experiments show 11%-27% latency reduction over SOTA DNN pruning and NAS methods.**

I. Introduction

Deep Neural Networks (DNNs) have achieved remarkable progress on cognitive applications [3], [16], [6] and have drawn attentions from both industry and academia on deploying DNN inference service on the cloud, edge and mobile devices [19], [39]. However, the superior performance of DNNs is largely built upon growing model volume and structure complexity. Consequently, how to run these bulky DNNs on accelerators efficiently becomes the key to meet the ever-growing user experience, *e.g.*, DNN model inference *latency*.

To relieve the computation cost and improve runtime performance, many research efforts have been made in different design perspective and implementation levels. These works include model-level optimization [11], [9], [10], compiler-level computing optimization [25], [35], [1], [30], and architecture-level optimizations [28], [27], [26]. Among these works, model-level design solutions are usually the most flexible and straightforward way, since such methods can reduce model volume and computation workload like FLOPs, which can directly translate to latency reduction. However, recent hardware-aware works questioned the effectiveness of these optimization works by revealing a widespread concept, *i.e.*, the lower amount of FLOPS does not equal to small latency during actual deployment on certain hardwares [38], [36], which underlines the importance of understanding of the hardware execution mechanisms for DNN design optimizations.

In this work, we re-examine the effectiveness of model design works for latency optimization on the general processing units (GPUs). By delving into the DNN to GPU deployment and execution mechanisms, we reveal a series of findings that current model design fails to recognize and thus lead to sub-optimal latency optimization performance. Figure 1 illustrates such an ineffective latency optimization case with a structural filter pruning example. Specifically, two layer configurations (config#1, #2) are compared, in which the config#2 conducts filter pruning and reduces partial workloads of this layer. Surprisingly, although config#2 reduces the filters, it ends up having the exact same runtime latency as config#1. In fact, as we keep pruning filters, Figure 1 (c) shows that the execution time exhibits an interesting non-linear *latency staircase* phenomenon. Why does filter pruning not always lead to latency reduction? In this paper, We identify the root cause of this phenomenon as the lack of awareness of the GPU *tail effect* — a classic parallel system issue that causes the resource under-utilization in the last processing cycle [20].

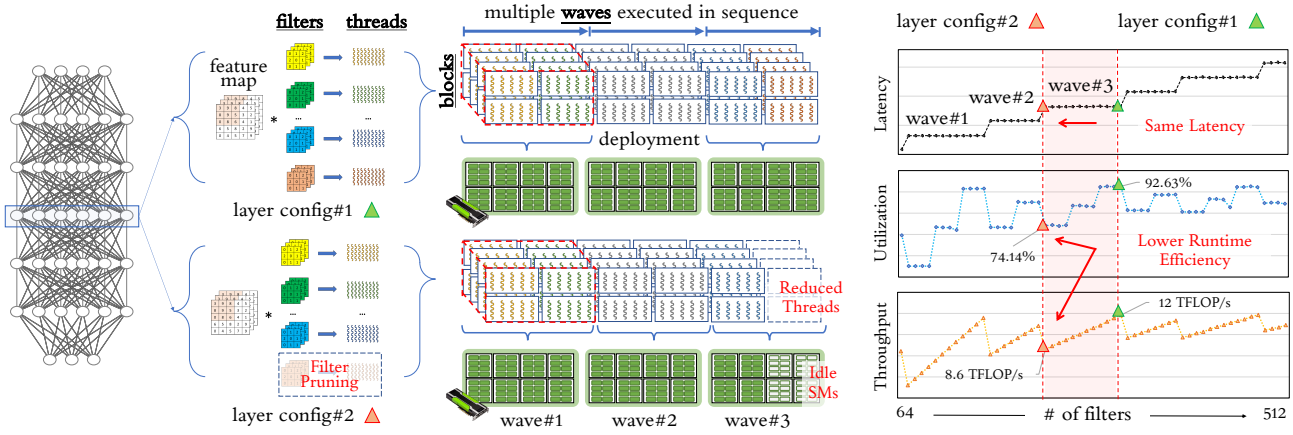
In particular, during execution on GPUs, a DNN layer's computation workload will be translated into massive parallel threads and allocated to GPU for processing on CUDA cores (Figure 1 (a)). However, due to the resource limit (*e.g.*, number of cores, maximal concurrency), the excessive parallel threads would be partitioned into multiple *waves* (Figure 1 (b)). Threads in each wave run concurrently while these multiple waves execute sequentially. Due to such an execution mechanism, although layer config#2 reduces the number of threads in the last wave, it takes the same runtime latency as config#1. Figure 1 (c) verifies the above execution analysis by showing a *latency staircase* pattern. Comparing layer config#2 to config#1, pruning certain amounts of filters brings no latency gain, but rather significantly lower down the GPU utilization and throughput due to the last partial wave.

Motivated by such an observation, **we conduct comprehensive benchmarking analysis and reveal the systematic mechanism** that causes such inefficiency in DNN inference: ***(i)*** Due to massive CUDA core parallelism, GPU presents a large *hardware-level compute granularity*. This is one major factor that causes the non-linear DNN workload to GPU runtime latency translation, and lead to the *latency staircase*

Fuxun Yu, Di Wang, Dimitrios Stamoulis, Rishi Madhok, Nikolaos Karianakis, are with Microsoft (contact email: fuxunyu@microsoft.com).
Zirui Xu is with CVS Helalth.
Longfei Shangguan is with University of Pittsburgh.
Ang Li is with University of Maryland, College Park.
ChenChen Liu is with University of Maryland, Baltimore County.
Yiran Chen is with Duke University.
Xiang Chen is with George Mason University and Peking University.
Manuscript received 2023; revised May, 2024.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:37:56 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 Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully edi
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2024.3404413



(a) DNN Design Optimization (b) GPU Execution with Tail Effect (c) GPU Inefficiencies due to Latency Staircase

Fig. 1. An Example of Ineffective Latency Optimization by Filter Pruning. Specifically, by analyzing the GPU execution mechanism, we identify that software-level DNN optimization algorithms such as pruning and NAS algorithms fail to consider the native GPU runtime and thus can easily cause *GPU tail effect* that under-utilize the GPU during runtime, which makes these methods ineffective. Therefore, we propose a tail-aware software-hardware DNN co-optimization method to achieve ultimate DNN optimization effectiveness and efficiency.

phenomenon; ***(ii)*** Most current works are non-aware of such a GPU compute granularity and their DNN optimization can easily lead to the runtime *tail effect*; ***(iii)*** Furthermore, due to common DNN's deep layer structure, such *tail effect accumulates and amplifies with hundreds of layers*, resulting in significant overall resource under-utilization.

Based on such tail effect analysis, **we rethink the current DNN design optimization and propose following insights**: ***(i)*** For DNN structure design, the observation of latency staircase implies *DNN architecture design should take GPU compute granularity into consideration*, which would help enable more efficient hardware utilization for model optimization, *e.g.*, help defining the search space in NAS solutions; ***(ii)*** For DNN structure optimization, the tail effect implies that there are *free runtime resources for potential accuracy boost*, *e.g.*, by fulfilling each layer's tail wave of workload to utilize idle SMs but without increasing the latency; ***(iii)*** For complex DNNs with hundreds of layers, accumulated tail effects within these layers implies great chances for *non-trivial latency reduction* while maintaining accuracy.

Performing the design insights above could guide both DNN pre-design and DNN post-optimizations. We demonstrate our optimization approaches regarding two common design-level solutions, neural architecture search and structural pruning: ***(i)*** For NAS solutions, we enhance the current design space with expected GPU granularity awareness. Such a GPU granularity-aware configuration discretizes previous continuous layer width design space, thus greatly reducing the search candidates and the searching complexity. Meanwhile, as the design space already avoids the tail effect, the DNN structure could also achieve better GPU runtime efficiency and thus better latency performance. ***(ii)*** For structure post-optimization like pruning, we can also leverage the GPU-aware layer design space to fine-tune the sub-optimal pruned models. Specifically, we adopt layer growing and layer pruning two operations to utilize the free resources and to eliminate the GPU tail effect. Featured with such flexibility, our structural post-optimization can achieve both accuracy lifting and latency reduction objectives, which is one major difference from previous works.

Through comprehensive evaluations, we demonstrated that our DNN design and optimization methods have exceptional latency performance and model generality with optimal accuracy. Across benchmarks, our design achieves 11%–27% latency reduction and 2.5%–4.0% accuracy improvement over state-of-the-art (SOTA) DNN structural pruning and NAS techniques.

The rest of this paper is organized as follows: Section §II provides the research preliminaries for this work. Section §III presents the observed DNN latency issue with the underlying GPU computing granularity; Section §IV then rethinks current DNN design pitfalls regarding latency-awareness; and Section §V proposes effective latency-aware DNN design insights and demonstrates different design and optimization approaches, following comprehensive evaluations in Section §VI; Section §VIII concludes this work.

II. Background and Related Work

This section introduces the background for the primary research focus of this paper. A GPU workflow overview is first given and then the state-of-the-art DNN computing latency optimization works are reviewed. Specifically, their *pros and cons* are cross-compared to highlight this work's optimization perspective and motivation.

A. GPU Computing Preliminary

GPU Architecture. We take NVIDIA's GPU architecture as the primary investigation object in this work, which also applies to other GPU products. The GPU architecture is featured with an array of *Streaming Multiprocessors* (SMs), each of which is formed by a group of CUDA Cores. For example, NVIDIA Titan-V GPU [28] has 80 SMs and can process 5120 threads

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:37:56 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 Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully edi
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2024.3404413

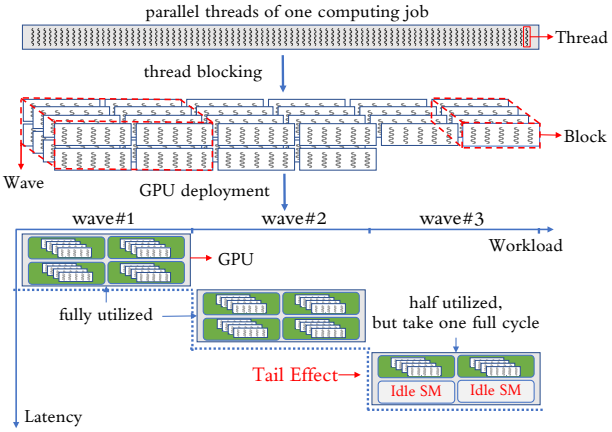


Fig. 2. GPU Execution Model. A computing workload will be split into massive threads and then split into Blocks and Waves for parallel processing. When the final wave comes unfulfilled, the GPU tail effect happens that under-utilizes the GPU resources.

simultaneously. Such a physical parallelism with massive computing units also defines the distinct GPU execution model.

GPU Execution Model. As shown in Figure 2, when a multi-thread job is deployed to a GPU, it will be first parsed into multiple logical workload partitions — ***Blocks*** with identical thread capacity. Depending on the physical capacity of the SM, the number of deployed thread blocks per SM varies.

For a large computing job, whose total workload exceeds the overall SM capacity in a GPU, the GPU needs to take multiple processing cycles. In such as case, the blocks are further clustered into logical ***Waves*** for sequential deployment. Specifically, wave describes the workload that can be concurrently processed per GPU cycle, which equals to the number of SMs $\times$ the number of blocks per SM. And the ***GPU Latency*** can be roughly estimated by the overall number of waves $\times$ the time per GPU computing cycle.

GPU Tail Effect. In practical deployment, the job workload usually will not be perfectly clustered into complete waves; in other words, the last wave could contain fewer blocks, as shown Figure 1. Nevertheless, such an incomplete tail wave will still consume a full GPU cycle, due to the GPU concurrent execution mechanisms as mentioned above, resulting in under-utilized SMs and sub-optimal runtime efficiency, namely, the GPU ***Tail Effect.***

The tail effect generally exists in parallel systems [20] and is considered a trivial issue with conventional GPU workload (*e.g.*, streaming graphic rendering), since the impact of the incomplete tail wave tail can be amortized with dozens of total wave amounts. *However, when it comes to DNN computing, particular DNN design characteristics (*e.g.*, the deep layers) will significantly exacerbate the negative impact of the tail effect with considerable latency issues.*

B. Related Works on Latency Optimization

Software Perspective — DNN Design. In early methods, the volume of the DNN model or FLOPs used to be considered as a linear factor of the workload and therefore the computing latency. Thus, early latency optimization works from the DNN design perspective are through either lightweight structure designs, such as neural architecture search (NAS) [12], [34]; or post-design model architecture tuning, such as structural DNN pruning [11], [9], [14].

However, recent works show the workload measured in FLOPs fail to consistently translate into latency due to different hardware execution mechanisms [38]. Therefore, ***latency-aware*** DNN design is expected to consider the actual hardware runtime [38]. Some recent works leverage off-line profiling of DNN models for latency estimation, such as latency-aware NAS works [10], [2].

Hardware Perspective — GPU Compiling. Currently there are also compiler optimization works that specifically target at optimizing the GPU kernel computing efficiency, such as TensorFlow-XLA and TensorRT [35]. These works tuned the low-level kernel computing primitives (*e.g.*, tiling, unrolling) for each DNN model and each target computing hardware, and thus could achieve optimal hardware efficiency, such as maximal SM utilization, memory throughput. Recently, many ML-based auto-tuning frameworks are proposed, which leverage machine learning to search scheduling factors automatically, such as TVM, TASO [1], [13], *etc.*.

It is worthy to note that these mentioned software and hardware optimizations are current mainstream, but not all. Beyond GPU compiler-level, there are also architecture-level works, such as GPU resource management and scheduling [27, 28], memory system optimizations [15], etc. However, due to their optimization cost and method generality issues, they will not be addressed in this work.

C. Pros and Cons of Different Perspectives

Latency optimizations from the software and hardware perspectives have their respective advantages and disadvantages.

From the DNN design perspective, as mentioned above, recent latency-aware DNN designs are more direct, which optimizes the profiled latency metric already, instead of indirect metrics like FLOPs [38]. However, most latency-aware optimizations treat the underlying runtime mechanism as a black box and only utilize rough look-up-table (LUT) based latency modeling [2], [36]. Although being able to approximate the end-to-end DNN inference latency, they reveal limited understanding of how varied structures perform differently on the hardware. Therefore, the searching performance can be easily limited with an improper design space configuration or insufficient sampling density[1].

Compared to that, hardware-perspective optimizations mainly focus on GPU computing efficiency and provide complementary computing speedup.

Take the TensorRT or TVM compiler for GPU as an example. In compiler-level, re-compiling the hardware-specific kernels for pruned layers can renew the shape-specific optimization (e.g., rescheduling the thread blocking/tiling settings), thus

[1]For example, most NAS works [36], [34] heuristically set or hand-pick layer width search space such as {16, 32, 64, 128, 192}, which can miss optimal layer settings in terms of GPU runtime efficiency.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:37:56 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 Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully edit
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2024.3404413

reducing/removing the tail effect and achieving the optimal efficiency. However, due to the practical cost issue, the current limitation still exists even in the most popular CuDNN/TensorRT lib since TensorRT only develops a static kernel base with around 100+ kernel optimization templates [22], [21] in order to support all conv shapes and many of them are also legacy ones for GEMM-based MatMul kernels. Although such kernel templates take heavy hardware engineering tuning efforts, they are still heavily tuned towards and thus restricted to the most common layer shapes (like ResNet50 layer shapes, EfficientNet layer shapes), which cannot fit various pruning operations with different shapes. To remediate this inefficiency, pytorch and TensorRT also have a "cudnn_benchmark" function that will exhaustively test all kernel shapes' runtime in advance and rank the relative performance with these kernel templates. However, a lot of different pruning/NAS methods' customized configuration cannot find optimal matched templates and thus tail effect problems still exist in such static kernel bases.

D. Our Design Perspective and Motivation

Targeting on optimization efficiency and generality, our work mainly rethink the software-level approaches, *i.e.*, targeting at the DNN model design. But instead of treating GPU as a black box, we incorporate its computing mechanism awareness into the DNN design by diving into the DNN model-to-GPU workload mapping during compiling for an explicit GPU computing granularity analysis, and therefore guide our latency-aware DNN model optimization. Our later proposed solution and evaluation also reveal an interesting finding: DNN designers could effectively alleviate the GPU computing inefficiency (*i.e.*, tail effect) directly from the early design stages, delivering optimal computing performance with low costs, and meanwhile achieving expected generality and feasibility.

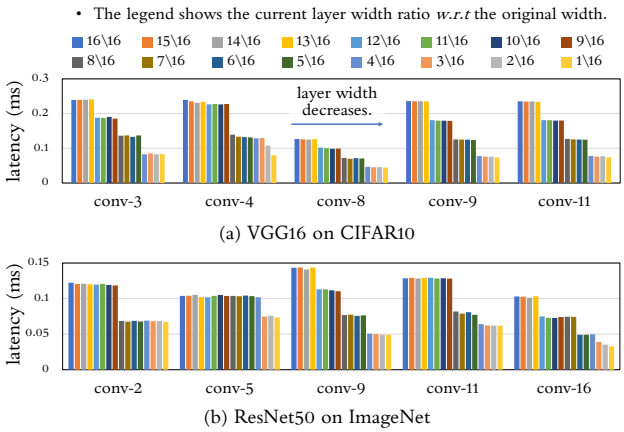
III. Revealing the GPU Computing Granularity for Runtime Latency

In this work, we first dive into the DNN latency analysis to reveal the underlying GPU computing mechanisms. A particular DNN latency issue is identified through detailed observation as the *latency staircase* phenomenon, which triggers our rethinking of the latency-aware DNN model design.

A. DNN Latency Staircase Phenomenon

We profile the inference latency of two representative DNN model settings on an NVIDIA Titan-V GPU[2]: (a) VGG16 on CIFAR10 dataset; (b) ResNet50 on ImageNet dataset. For each DNN model, we investigate the latency impact from different model volumes by gradually pruning the layer width step by step and measure the runtime latency in each layer width setting. All experiments are conducted using CUDA 10.2 and CuDNN 7.6.5 backend.

Through the analysis, we identify a distinct DNN runtime issue, *i.e.*, the ***latency staircase*** phenomenon. Specifically, Figure 3 illustrates the latency of five randomly selected



(a) VGG16 on CIFAR10

(b) ResNet50 on ImageNet

Fig. 3. DNN Runtime Latency Profiling.

DNN layers (denoted by conv-N) of VGG16 and ResNet50, respectively. For each layer's dataset, although the layer volume is gradually shrunk by filter pruning, runtime latency doesn't reduce linearly but demonstrates a "staircase" trend, are changes occurring only after a certain plateau period. Taking the left-upper corner dataset of as an example (conv-3 on VGG16), we observe that the runtime latency does not change as the layer width shrinks gradually to 13/16 (4th yellow bar from the left) *w.r.t* its original size (1st blue bar). However, the latency suddenly drops as the layer width shrinks from 13/16 further to 12/16, after which stays the latency still again until the layer width drops to 9/16 of its original size. And such a staircase dropping pattern repeats through the whole latency profiling.

Such results reveal several findings in DNN latency optimization: ***(i)*** *the layer workload reduction by conventional structural pruning cannot directly translate to GPUs' runtime latency optimization;* ***(ii)*** *such an issue is prevalent across DNN models and layer configurations;* ***(iii)*** *DNN latency optimization requires a deeper investigation of the hardware computing mechanism than a pure software perspective.*

B. Reflection of GPU Computing Granularity

Through dedicated analysis, we find that the above latency staircase is a reflection of specific GPU computing granularity due to the huge-parallelism execution mechanism. Specifically, we dive into fundamental GPU computing primitives (*e.g.*, number of GPU thread blocks B and waves W) using NVIDIA Nsight Compute [24], and reveal the DNN deployment and GPU computing mechanism.

Latency Staircase Analysis. Figure 4 shows the GPU computing changes when we vary a layer's width from 64 to 512 filters [3]. As we can see, the number of thread blocks B grows with increasing number of filters F (top figure), which is expected. Correspondingly, the number of waves also increase. However, due to the granularity mismatch (*i.e.*, one filter's workload doesn't equal to one block, and not equal to one wave.), the number of wave increase in a discrete way (step-by-step), forming the staircase growing trend and with a larger granularity (middle and bottom figure).

[2]We will show in later experiments that such phenomenon is general for both high-capacity and low-capacity GPUs like Jetson Nano.

[3]Specifically, the filter shape and input feature map sizes are set to 3×3 and 64×64×512. And we keep the batch size to 1 throughout the experiment.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:37:56 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 Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully edi
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2024.3404413

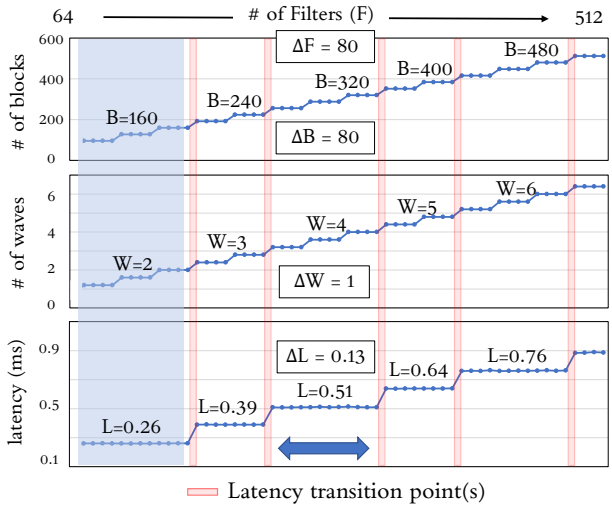


Fig. 4. GPU Computing Analysis *w.r.t.* Blocks, Waves, and Latency with Increasing DNN Layer Size.

As highlighted by the blue box in Figure 4, the number of waves W grows gradually from 1 to 2 with increasing workload. However, the latency stays still until the workload grows to above 2 waves, and then GPU consumes a new processing cycle to process the new wave of workload. We thus see a jump in the runtime latency as highlighted in the red box, *i.e.*, the latency transition point.

C. Latency Formulation with GPU Granularity

Similar staircase patterns continuously emerge with growing workloads, which reflects a GPU-specific compute granularity. We can thus model the latency staircase based on the above analysis. For a given DNN layer, its GPU runtime latency L can be represented by:

$$L = \Delta L \times \lceil \#W \rceil, \qquad \text{where } \#W = \frac{B}{S}, \tag{1}$$

where ΔL represents the latency for the GPU to finish one wave of workload, $\#W$ is the number of waves and B is the number of blocks that the overall workload being translated to. S is processing capacity of the GPU, *e.g.*, the maximal number of blocks that can be concurrently executed. $\lceil \cdot \rceil$ is the rounding-up operation that returns the least integer greater than or equal to the input (*e.g.*, $\lceil 2.1 \rceil = 3$), which reveals the impact of ***tail effect*** on the runtime latency, that is, one underutilized tail wave still leads to the latency of one full computing cycle for each DNN layer.

Eq. 1 demonstrates that the DNN execution latency on GPUs is highly dependent on a GPU-specific compute granularity, i.e., the wave W. Such a granularity is not only a hardware-level concept to describe the GPU capacity, but also an important factor for latency-aware DNN design. Without being aware of this, DNN structure design can easily generate granularity-unmatched models, thus resulting in sub-optimal GPU efficiency and DNN inference latency.

IV. Rethinking the DNN Design Pitfalls with GPU Tail Effects

The previous observation corresponds to the ***GPU tail effect*** mentioned in §II-A. As the DNN structure is usually composed of massive layers, the latency staircase and the tail effect can thus be prevalent across most DNN layers. Whereas, unlike ordinary GPU computing tasks where a single tail effect can be amortized, DNN inference with deep sequential layers incurs massive accumulated tail effects. And even worse, each tail wave may occupy a larger proportion of runtime since each layer has relatively lightweight workload.

In this section, we reveal the latency impact of the tail effect, and rethink the DNN design with GPU hardware-awareness.

A. Excessive Tail Effects within DNN Design

We characterize the workload of mainstreaming DNN models and reveal the excessive tail effects' existence in DNN structures due to two factors: ***(i)*** *Massively sequential layers*: while DNN models come in a wide variety of shapes and sizes, they all share the similar deep structure containing massive fragmented runtime units as layers. The tail-induced inefficiency in each layer thus accumulates with the feed-forward inference; ***(ii)*** *Limited per-layer workload*: While the entire DNN workload can be large, the workload of each layer is light-weight and thus requires much fewer waves to finish. amplifying the per-layer tail wave's negative impact.

Massively Sequential Layers. Figure 5 illustrates the workload of each layer (termed as Layer workload) in three different DNN models. To facilitate the presentation, we also plot the actual GPU workload in the figure, both of which are quantified by the number of waves. The gap between the layer workload and the actual GPU workload reflects the GPU under-utilization ratio. We observe these three models all contain more than 50 layers, and more importantly, the GPU under-utilization can easily occur in almost every layer of these models. During inference, such per-layer GPU under-utilization accumulates with the model execution, and thus severely undermines the GPU runtime efficiency. Beyond such lightweight models, higher under-utilization may appear in even deeper DNN models such as ResNet101 [7].

Limited per-Layer Workload. We further measure the number of waves incurred by each layer's workload in different batch size settings, and show their cumulative distribution function (CDF) in Figure 6. Here, we can observe that the majority of layers incur less than five waves for all three different models. For ResNet50, we can see over 80% of layers only introduces less than three waves from batch size=1∼4. Similarly, over 80% of layers in both MobileNet and MNasNet requires less than three GPU waves when the batch size is relatively small (*i.e.*, B=1). The maximum 80-percentile per-layer workload with batch size=4 only reaches 5-6 waves, which is far limited compared to the aforementioned same-sized GPGPU workload (30-80 waves). Due to the limited workload per layer, under-utilization in the tail wave causes more significant resource under-utilization (*e.g.*, up to 33% in a layer with three waves).

It is worthy to note that, the GPU tail effect can potentially become increasingly prominent with more light-weight DNN design (such as in the edge domain), while GPUs are becoming increasingly powerful with higher capacity even for the edge GPUs like NVIDIA Jetson [23]. *Therefore, without dedicated*

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:37:56 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 Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully edit
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2024.3404413

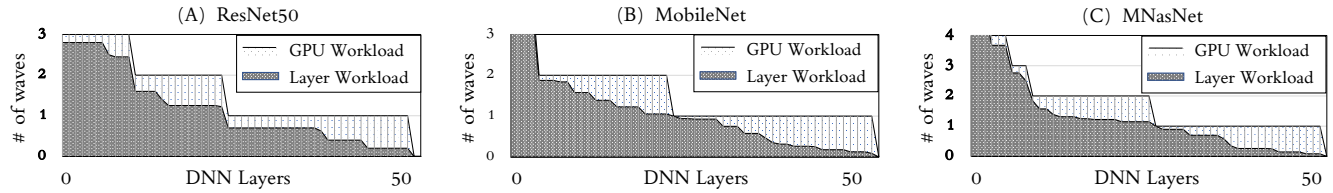


Fig. 5. The offset between the runtime GPU workload and the layer workload illustrates the GPU under-utilization caused by the tail effect (BS=1). Such under-utilized happens in each layer and thus accumulates to a non-negligible amount.

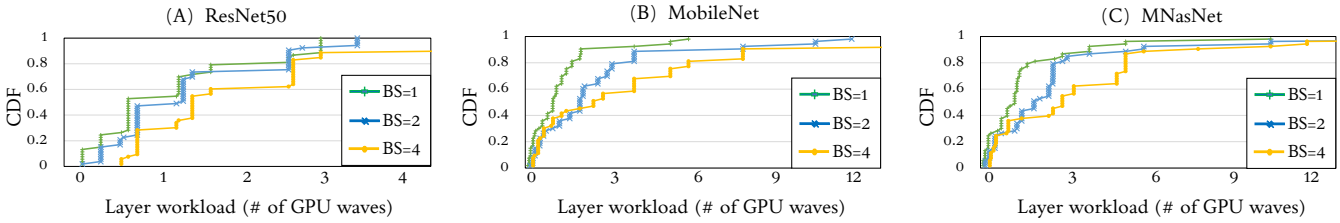


Fig. 6. The CDF of the layer workload shows that most (>90%) DNN layers consume less than 3∼9 GPU waves with varied batch sizes (BS=1∼4). In such cases, the relative influence of tail effect (affecting the last wave) is amplified.

optimization, such DNN latency sub-optimality can be exacerbated with the increased tail effects caused by mismatched GPU and DNN design granularity.

B. Pitfalls in DNN Optimization

We then demonstrate that conventional DNN optimizations such as filter pruning and architecture search (NAS) can even exacerbate the tail effect, leading to ineffective DNN optimization pitfalls.

In the beginning of Figure 1 (c), we have shown a real case of GPU runtime efficiency variation when layer workload is not matching the GPU granularity. Specifically, as the latency exhibiting a staircase pattern, the two GPU runtime metrics (GPU utilization and throughput) also show significant oscillation. The highlighted red box in Figure 1 (c) annotates a concrete example: the GPU utilization and throughput peak at the local maxima (92.63% and 12.0 TLOPS/s respectively) when the workload fully occupies the current wave, while decreases to the local minima for the lower-bound workload (74.14% and 8.6 TFLOP/s). Caused by ineffective optimization, the GPU runtime efficiency gap can reach up to 1.4×.

Such "ineffective pruning" pitfalls, that is, blindly pruning a DNN layer can lead to small runtime latency reduction but rather cause sub-optimal GPU utilization and throughput, can also be common in general pruning algorithms. For example, by reproducing the model pruning result using one of the recent SOTA pruning work [14], Figure 7 shows the runtime latency and throughput information of two resulted DNN layers. The performance result achieved by [14] is highlighted in orange. As we can observe that, without such tail awareness, two pruned layers did not achieve the optimal GPU throughput, but rather leading to the *ineffective pruning*. Especially in the first case that pruning 40% of filters doesn't lead to latency reduction, the gap between the throughput achieved by [14]'s configuration and the optimal one can reach up to 1.5×.

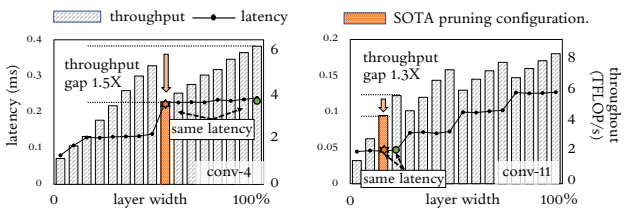


Fig. 7. The "ineffective pruning" pitfall. Without being aware of tail effect, the shown filter pruning configuration is sub-optimal, *i.e.*, 1.3×-1.5× less GPU throughput under the same latency. The DNN model is VGG16 [32] on CIFAR10 dataset as per the official open-sourced repository [15].

V. Approaching Latency-Aware DNN Design with GPU Awareness

In this section, we summarize the lessons learned from our analysis and characterization to guide DNN design on GPUs. We then demonstrate that by eliminating the GPU tail effect, we can enhance current DNN optimizations with better GPU awareness and achieve better performance.

A. DNN Design with Latency-Awareness

By understanding the GPU granularity-based latency modeling and the inefficiency caused by the tail effect, we can illustrate a series of DNN design insights:

• *Latency-aware DNN optimization on GPUs should consider a new basic unit, i.e., the GPU computing granularity (§III-B).* Previously, many DNN design and latency optimization works consider filters as the primary structural unit. However, our latency staircase analysis shows that, as one filter's workload is usually much smaller than the GPU computing granularity, workload measured by filters fails to translate to the latency consistently. By understanding this, latency-oriented DNN design could improve both their latency modeling effectiveness and design efficiency. For example, given a targeted GPU, each DNN layer's design only needs to consider a discrete design space, *i.e.*, with a step size of a certain number of filters that matches the GPU granularity.

• *For existing DNN structures, the existence of tail effects is prevailing, which exposes many opportunities for both accuracy and latency performance improvements*. The reason of prevalence is that, as the GPU platform (and other factors like compilers) can vary, the unified DNN structure design

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:37:56 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 Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully edi
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2024.3404413

can hardly avoid tail effects in all different settings. In such a case, tail effect accumulation in DNNs can incur huge inefficiency, meanwhile exposing many optimizations potentials. Specifically, the GPU resource under-utilization enables us to increase DNN workload and utilize the "free" tail resources[4] to boost accuracy while without affecting the runtime latency. In addition, as tail effect can emerge in many layers, by simultaneously pruning certain tail waves and filling up other ones, we can also obtain certain latency gains while maintain the overall DNN capacity and accuracy.

The above design insights can guide various DNN design optimization approaches. In this work, we mainly include approaches: GPU-aware NAS design space enhancement (§V-B) and lightweight DNN structure fine-tuning (§V-C).

B. GPU-aware NAS Design Space Enhancing

The first optimization we can apply is to enhance the design space of NAS methods with the proposed GPU awareness.

As discussed in related work (§II-C), previous NAS works mainly heuristically choose layer width in the design space, which may have sub-optimal GPU runtime performance. Therefore, one direct way of utilizing our GPU granularity-aware modeling is to find the optimal layer configurations and integrate into the NAS design space. The benefits are twofold: ***(i)*** our granularity-based modeling discretizes a continuous layer width design space and reduces the amount of search candidates for each layer, reducing the searching complexity; ***(ii)*** meanwhile, as each configuration already avoids the tail effect, our design space brings better GPU runtime performance without the tail effect related inefficiency.

Profiling-Guided Identification. We take a profiling strategy, *i.e.*, profiling the GPU latency staircase offline to identify the optimal design space for each layer.

Based on the layer runtime modeling in Eq. 1, the tail effect diminishes when DNN layer workload matches the GPU compute granularity, *i.e.*, the last wave of workload fully occupies the GPU. In such case, the GPU achieves optimal utilization, and the throughput also reaches local maximum in the range of current wave. Therefore, we leverage the GPU throughput T to identify the existence of tail effect in each layer and further detect its optimal layer configurations. Taking a retrospect of Figure 1 (c), we can see the throughput always peaks its local maximal (denoted by the green triangle) at the optimal layer size setting without tail effect and thus can serve as a reliable clue to track the optimal configurations. Specifically, we obtain the optimal configuration $C_i^\star$ for each layer i based on the following rule:

$$C_i^\star = \arg\max_m (T_{i,m}) \tag{2}$$

where $T_{i,m}$ denotes GPU throughput of layer i in the mth layer width configuration. We then use Eq. 2 to identify optimal configurations in different layer width intervals.

Overhead Analysis. Table I shows the offline profiling overhead on different DNN models. For each layer, we sample 16 different layer width configurations (from 1/16 to 16/16 with respect to the default layer size) and record their throughput. We observe the profiling overhead stays in the range of one to three minutes, which is generally acceptable considering the NAS overall search time.

[4]Here we mainly denote latency-wise "free". The power consumption may still show differences, which is out of scope of this work.

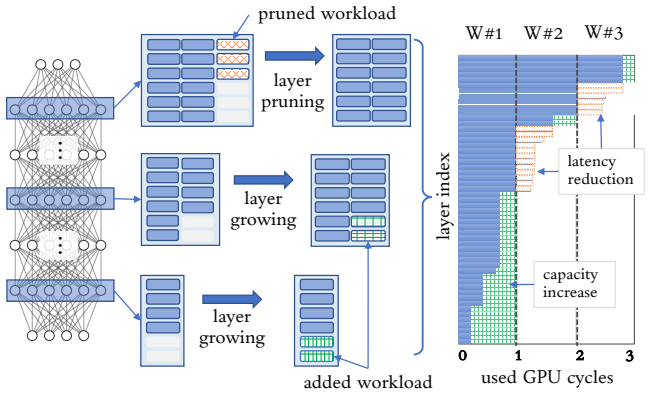


Fig. 8. We leverage two operations (*i.e.*, pruning and grafting) to eliminate the tail effect. Such two operations provide complementary latency gain and accuracy gain and facilitate the DNN accuracy-latency optimization.

C. Lightweight DNN Structure Fine-Tuning

Besides the end-to-end DNN design apporach of NAS, we can also conduct lightweight DNN structure fine-tuning to obtain the optimal pruned models in structural pruning methods.

Structure pruning without GPU awareness can generate pruned models with excessive tail effects (§IV-B). Therefore, we can leverage ***pruning*** **and** ***grafting*** two operations to eliminate the tail effect. An overview of our method is shown in Figure 8. Specifically, *layer pruning* reduces the number of filters to remove the remaining workload on the last wave and thus can effectively harvest latency gain by an entire GPU wave processing cycle. *Layer grafting* instead increases the number of filters in other laters to fill up the GPU capacity in the last wave, which can compensate the accuracy loss of layer pruning without sacrificing the latency (*i.e.*, accuracy gain). These two operations provide us flexible opportunities to match the GPU granularity and thus eliminate the tail effect. Meanwhile, they also provides DNN optimization flexibility towards two optimization targets: latency reduction and accuracy lifting, which is one major difference from traditional pruning works.

Here one challenge remains, *i.e.*, how to select two operations between layer pruning and grafting across multiple layers regarding different optimization objectives? We propose

TABLE I
OFFLINE PROFILING OVERHEAD.

	Model	BS=32	BS=64	BS=128
CIFAR (32x32x3)	VGG16	2m22s	2m23s	2m25s
	ResNet56	1m27s	1m30s	1m40s
	Model	**BS=1**	**BS=2**	**BS=4**
ImageNet (224x224x3)	VGG16	2m14s	2m15s	2m18s
	ResNet50	2m50s	2m51s	2m56s

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:37:56 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 Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully edi
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2024.3404413

Algorithm 1 Lightweight DNN Structure Fine-tuning.

1: **Input:** Model's initial layer configs $r[l]$, latency $L[l]$ and throughput profiling $T[l]$.
2: **Output:** New config R_{new}.
3: Identify design space C_l for each layer l by Eq. 2.
4: **for** layers $i = 1$ **to** l **do**
5: Get LG_i, PG_i estimation through profiling results.
6: Sort the layer index list [1,2...i...l] by LG_i or PG_i.
7: **while** layer index list is not empty **do**
8: Pop out layer j with argmax$_j$ $LG[l]$.
9: Prune layer j with one step in design space C_l.
10: **while** $\Sigma^l PG(R_{new}) \notin (-\tau, \tau)$ **do**
11: Pop out layer k with argmin$_k$ $LG[l]$.
12: Grow layer k with one step in design space C_l.
13: Get runtime latency evaluation L_{new} of config R_{new}.
14: **if** L_{new} achieves the target latency **then**
15: Train and evaluate the model accuracy.
16: **else**
17: Set τ *= 2 and repeat the algo. from line 9.
18: **Return** Fine-tuned config R_{new}.

a lightweight fine-tuning algorithm to flexibly balance the corresponding operations to reach different objectives.

Optimization Objectives Formulation. We first define the DNN inference accuracy and runtime latency optimization objectives. We define the latency gain LG to indicate the latency being saved, and parameter gain PG (amounts of parameters been deleted or added) to indicate the model capacity as a guidance for retaining accuracy. The two metrics could be estimated by the change in different layer widths R_i and their latency profiling L_i.

Accuracy-Oriented Optimization aims to improve the model accuracy without extra latency overhead. Therefore, it can be formulated as maximizing the parameter gain while constraining latency gain:

$$\text{Maximize} \quad \sum_i^l PG_i, \quad \text{s.t.} \quad \sum_i^l LG_i \geq 0. \tag{3}$$

Latency-Oriented Optimization aims to reduce the runtime latency without accuracy drop, *i.e.*, maximizing latency gain while maintaining parameter gain in a tolerable range (τ):

$$\text{Maximize} \quad \sum_i^l LG_i, \quad \text{s.t.} \quad \sum_i^l PG_i \in (-\tau, \tau). \tag{4}$$

Algorithm Walk-through. Without loss of generality, we take the latency-oriented optimization as an illustrate example to explain the algorithm. The accuracy-oriented optimization follows the same principle by simply switching the optimization objectives as defined in Eq. 3–4.

Step 1. Layer Candidates Identification (Line 3): Given a DNN model structure, we first identify the optimal layer configurations, *i.e.*, design space C_i by Eq. 2.

Step 2. Intra-Layer Performance Gain Estimation (Line 4–6): For each layer, we get the latency and parameter gain estimation assuming conducting layer pruning and layer growing operations by one step in the design space C_i. These performance estimation will guide the inter-layer adjustment.

Step 3. Inter-Layer Adjustment Strategy (Line 7–13): To achieve latency optimization, we maintains a bi-directional queue of layer indexes, and greedily prune layers with maximal LG_j (max latency gain) and minimal LG_k for growing (balanced capacity) . Same procedure applies to accuracy optimization by switching the objective/constrain in Line-9, 11.

Step 4. Model Structure Determination (Line 14–18): Finally, we return the new configuration if it reaches the targeted latency. Otherwise, we can adjust the constraints, *e.g.*, with larger parameter gain tolerance ($\tau* = 2$), to allow more aggressive layer pruning for latency reduction.

Algorithm Discussion. Note that different from traditional pruning-from-scratch method, our method is designed as an orthogonal pruning "post-calibration" method on top of well-searched pruning/NAS configurations, which could be easily combined with any global pruning method that finds optimal accuracy-aware pruning configurations. In fact, our pruning and grafting operations usually yiled minor pruning size calibration within one staircase as shown in Figure 8, for example, adjusting from pruning 40 filters (an accuracy-aware best pruning configuration but NOT latency-aware) to 48 filters (both latency- and accuracy-aware best setting) in a 512-filter layer. Such a practice incurs minimal accuracy change to the optimal pruning configurations. Meanwhile, with the small filter number changes, we adopt the simple but effective guidance metrics that are the parameter gain (PG) and latency gain (LG). Even though the PG indicator is not mathematically proven to be related to accuracy modeling like traditional from-scratch pruning method, it is more designed as a lower-bound indicator that our pruning post-calibration method is not hurting the model capacity (e.g., the parameter gain remains largely positive/non-negative), thus are shown to achieve fairly stable accuracy improvement and latency reduction, which we will show later in our experiments.

VI. Experimental Evaluation

In this section, we evaluate our latency-aware DNN design methodology on a series of common benchmarks. Specifically, we demonstrate our performance gain in different approaches.

Experimental Setup. We conduct experiments using the common software stack including CUDA 10.2 and CuDNN 7.6.5. For GPU hardwares, we select four GPUs from high-end (Ampere 3080Ti, Volta Titan-V, Pascal P6000) to embedded ones (Jetson Nano). The detailed specifications of these GPUs are shown in Table II, including number of SMs and CUDA cores, as well as the peak FP32 performance.

We apply our method into two common model optimization approaches, *i.e.*, *structural filter pruning* and *NAS*. For pruning, we compare with several state-of-the-art pruning methods including HRank [14], SoftPruning [8], Network Slimming [18] and Dependency Graph [5]. For NAS, we apply our model optimization method onto the EfficientNet series of structures [34], which are one type of the SOTA efficient model architectures. For model performance evaluation, we compare the model's accuracy and runtime latency. Without specific mentioning, the batch sizes for latency evaluation on CIFAR10

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:37:56 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 Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully edi
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2024.3404413

TABLE II
THE EVALUATED GPUS AND SPECIFICATIONS.

GPU Name	Archicture	#SMs	#Cores	Peak FLOP/s
3080Ti	Ampere	80	10240	29.8T
Titan-V	Volta	80	5120	14.9T
Quadro P6000	Pascal	30	3840	12.0T
Jetson Nano	Maxwell	1	128	0.24T

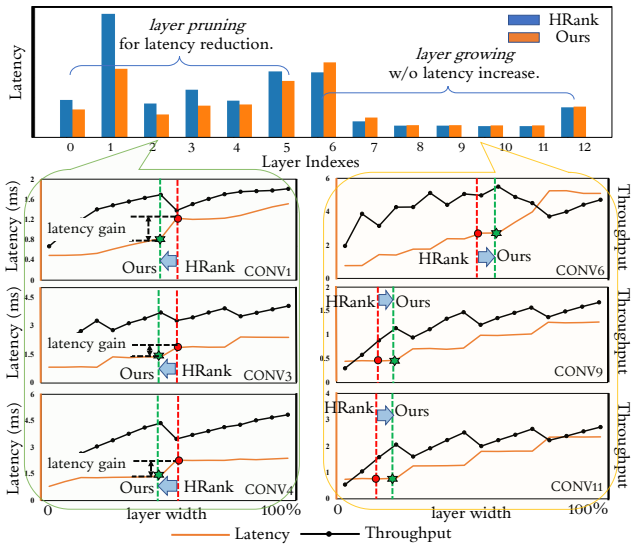


Fig. 9. Our DNN optimization illustration. We conduct layer pruning in Conv1-5 to get latency reduction, and layer growing in Conv6-12 to recover model capacity but without obvious latency increment. During this, we eliminate tail effects across layers and get latency reduction.

and ImageNet are set to 128 and 1 to observe the latency variations. Meanwhile, we also collect the GPU throughput (FLOP/s) information as a indicator to demonstrate the GPU runtime efficiency comparison.

A. Evaluation on the Pruning Approach

A Proof-of-Concept Case Study. As a proof-of-concept, we first apply our optimization to a prevalent DNN benchmark model VGG16 and compare the results with one SOTA pruning competitor HRank [14] to illustrate our algorithm mechanism. The result is shown in Figure 9.

Our results (orange bar) consistently outperform HRank (blue bar) on run-time latency reduction, as shown in Layer 0∼5. We further showcase the latency and throughput of three randomly selected layers for both layer pruning and growing operations (bottom figure). As shown in CONV-1,3,4, we successfully remove one tail latency by conducting layer pruning operation. At the same time, the GPU throughput achieves its local maximal. The layer growing actions, on the other hand, brings little to no latency overhead, as shown in CONV-6,9,11. Taking a further scrutiny on their latency results, we can see that the layer width is increased by a proper amount such that the layer's runtime latency stays similar with HRank. Overall, we can achieve **17.7%** latency reduction compared to HRank with only **0.2%** accuracy drop (93.1%→92.9%), as shown in the first two rows in Table III.

TABLE III
LATENCY OPTIMIZATION AND GPU THROUGHPUT ANALYSIS (TITAN V).

	Method	Params	#FLOPs	FLOP/s	Acc.%	Time (ns)
VGG16 (CIFAR10)	HRank [14]	1.90M	67.0M	2.41T	**93.1**	2.50E6
	Ours	2.85M	104.1M	**3.90T**	92.9	**2.05E6 (-17.7%)**
ResNet56 (CIFAR10)	HRank [14]	0.48M	65.9M	0.83T	93.6	4.20E6
	Ours-1	0.50M	79.1M	**1.00T**	**93.8**	**3.72E6 (-11.3%)**
	Ours-2	0.50M	75.1M	**0.95T**	93.5	**3.51E6 (-16.3%)**
ResNet56 (CIFAR10)	SOFT-1 [8]	0.53M	68.8M	0.88T	93.1	4.08E6
	Ours-1	0.43M	71.4M	**0.91T**	**93.2**	**3.52E6 (-13.7%)**
	SOFT-2 [8]	0.45M	53.1M	0.68T	92.3	3.64E6
	Ours-2	0.43M	66.0M	**0.79T**	**92.3**	**3.01E6 (-17.3%)**

Pruning Accuracy and GPU Throughput Analysis. We then apply our model optimization method in more pruning benchmarks and analyse both the latency and GPU throughput. The result comparisons are shown in Table III. Note that, SOFT method [8] has two pruning configurations (denoted by SOFT-1 and SOFT-2) with varied pruning degrees, which we compare separately. The latency evaluation is conducted on Titan-V.

Latency Reduction: As Table III presents, by optimizing the model structure configurations, our method achieves consistently lower latency than baseline methods (**11.3%-17.7%** latency reduction in all settings). Different from previous work that focuses on reducing workload to reduce latency, our method's latency reduction mainly comes from the higher GPU runtime efficiency, *e.g.*, GPU throughput improvement.

GPU Throughput Improvement: As Table III shows, our method shows consistently higher GPU throughput than the baselines, *i.e.*, in the (*FLOP/s*) column. For VGG16, our optimized model achieves 3.90 TFLOP/s, **1.6**× throughput than the baseline method, 2.41 TFLOP/s. For ResNet56, our models also achieve **1.2**× throughput than HRank (1.00 vs. 0.83 TFLOP/s) and SOFT-2 (0.79 vs. 0.68 TFLOP/s). Such throughput enhancement indicates the GPU runtime efficiency improvement, and shows the advantages of our configuration optimization by tail effect elimination.

Accuracy Maintenance: In addition to lower latency, our method also maintains the model accuracy change within negligible ranges (±0.2%). This is benefited by constraining the parameter gain (PG) in the algorithm. As (*Params*) column in Table III shows, our optimized models have maintained the amount of parameters either in positive range or very small negative range. By doing so, we ensure the models' capacity to be well-maintained, thus keeping the accuracy during the optimization.

Fine-Grained Comparison with SOTA Benchmarks. To further illustrate the advantages of our methods, we conduct a fine-grained analysis by applying our method on top of two state-of-the-art methods: *Network Slimming (SLIM)* [18] and *Dependency Graph (DEPG)* [5].

Specifically, given an upper bound pruning ratio for one network (e.g., 70%), we enumerate all pruning configurations below the upper bound with a specific pruning ratio step size (e.g, 5%). For each pruning ratio, we run the baseline

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:37:56 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 Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully edi
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2024.3404413

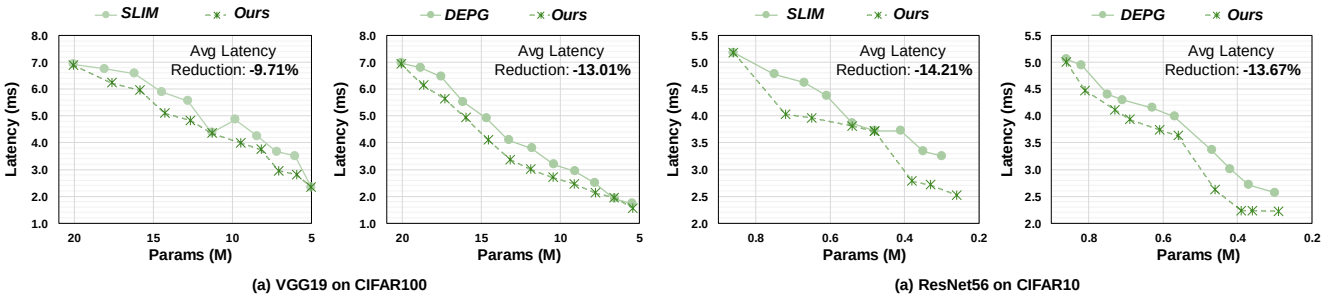


Fig. 10. We applied our fine-tuning algorithm (*Ours*) onto each pruning configuration generated by *SLIM* [18] and *DEPG* [5]. Experiments show our optimization achieves consistent latency reduction (on average 10% to 14%) on top of the baseline. Latency measured on 3080Ti with batch size = 128.

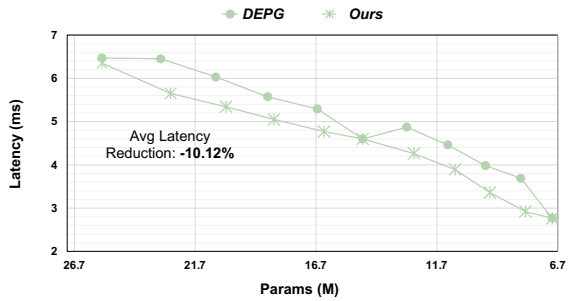


Fig. 11. We applied our fine-tuning algorithm (*Ours*) onto each pruning configuration generated by *DEPG* [5] (ResNet50 backbone on ImageNet). Experiments show our pruning optimization achieves consistent latency reduction (on average 10.12%) on top of the baseline methods. Latency measured on 3080Ti with batch size = 8.

algorithms (*SLIM*, and *DEPG*) to find the accuracy-optimized pruning configurations. Then our layer pruning and grafting Algorithm 1 is applied on top to finetune the baseline pruning configuration to remove the GPU tail effect, and thus generated a both accuracy- and latency-optimized configuration. We then compare the latency vs parameter amount plot to compare each of the baseline and our optimized results.

Figure 10 shows the detailed latency curve comparison. Latency are measured on 3080Ti GPU with batch size 128 for both CIFAR10 and CIFAR100. As we can see, most pruning configurations generated by tail-unaware baselines (*SLIM*, and *DEPG*) run with longer latencies, indicating they are far from optimal in terms of actual GPU runtime efficiency. By contrast, by slightly optimizing each layer's configuration to remove the tail effect, our fine-tuned model architecture could easily save 10%-14% average latency for nearly all pruning configurations.

To further demonstrate that our pruning and grafting algorithm only has minimal accuracy impact, we compare the model re-training performance between the baselines and our optimized model configurations. Here for simplicity, we only compare the model pruning settings that are reported in the official codebase benchmarks[5]. The results are shown in Table. IV. As we can see, ours algorithms yileds minimal accuracy variation (-0.15% to +0.02%) but achieves consistently 10% to 17.7% latency reduction.

The above experiments demonstrate the major benefits of our algorithm: ***orthogonal design***, that is, our pruning and grafting algorithm could be combined with various of different SOTA pruning algorithm logics (e.g., local-based ranking [14], [8], [18] or global-based ranking mechanisms [5], different l1-norm or other importance criteria, etc). When the SOTA algorithms generate an accuracy-oriented optimal configuration, further applying our algorithm on top will then generate a both accuracy-optimal and GPU efficiency-optimal configuration, thus achieving consistently better accuracy-latency trade-offs.

[5]https://github.com/VainF/Torch-Pruning/tree/master/benchmarks

TABLE IV
ACCURACY / LATENCY COMPARISON WITH SOTA BENCHMARKS.

		Org. Params	Params	Accuracy	Latency (ns)
VGG19 (CIFAR100)	SLIM [18]	20.09M	1.27M	67.75%	1.03E+06
	SLIM+Ours	20.09M	1.25M	67.64%	**9.2E+05 (-10.7%)**
	DEPG [5]	20.09M	1.18M	70.39%	1.01E+06
	DEPG+Ours	20.09M	1.23M	70.24%	**9.1E+05 (-9.9%)**
ResNet56 (CIFAR10)	SLIM [18]	0.86M	0.41M	93.29%	3.72E+06
	SLIM+Ours	0.86M	0.41M	93.31%	**3.46E+06 (-6.99%)**
	DEPG [5]	0.86M	0.37M	93.64%	2.71E+06
	DEPG+Ours	0.86M	0.36M	93.57%	**2.23E+06 (-17.7%)**
ResNet50 (IMGNET)	DEPG [5]	25.56M	12.74M	75.83%	4.87E+06
	DEPG+Ours	25.56M	12.53M	75.78%	**4.23E+06 (-13.1%)**

TABLE V
ACCURACY OPTIMIZATION ON EFFICIENTNETS.

	Method	Acc.%	GPU:Titan-V		GPU:P6000	
			Time(ms)	FLOP/s	Time(ms)	FLOP/s
EfficientNet (ImageNet)	B0	77.52	12.6	61.9G	13.8	56.5G
	Ours	**81.49 (+3.97)**	12.7	**414.2G**	14.1	**373.0G**
EfficientNet (ImageNet)	B1	79.38	17.8	78.7G	19.8	70.7G
	Ours	**82.08 (+2.7)**	18.0	**442.7G**	19.9	**396.0G**
EfficientNet (ImageNet)	B2	80.18	18.2	109.9G	19.9	100.5G
	Ours	**82.72 (+2.54)**	18.4	**547.3G**	20.4	**488.2G**

B. Evaluation on the NAS Approach

In this part, we apply our optimization method to further optimize the NAS network's performance on GPUs. Specifically, we optimize the EfficientNet [34] series of model structures to achieve better accuracy latency trade-offs. The evaluation results are shown in Table V. The latency evaluation is conducted on Titan-V and P6000 GPUs with batch size = 1 to simulate the real-time performance.

Accuracy Maximization: As Table V shows, with the similar runtime latency, our optimized EfficientNet model could achieve much better accuracy (**+3.97%**, **+2.7%**, **+2.54%**, respectively) on the challenging ImageNet dataset.

Better Latency Accuracy Trade-offs: We composedly compare the model accuracy-latency performance with original Efficient-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:37:56 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 Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully edi
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2024.3404413

Nets [34], and several major types of DNN structures including ResNets, ResNeXts, InceptionNet, and Dense Net. The results are shown in Fig. 12. From the *accuracy* perspective, our optimized B0 to B3 models achieve the highest accuracies than most baselines under the same latency. From the *latency* perspective, our model B2's latency is **1.5**× less than the models which achieve the same accuracy level, *e.g.*, EfficientNet-B3 and B4. Therefore, both perspectives demonstrate the effectiveness of our method in achieving better accuracy latency trade-offs on GPUs.

Optimization Explanation: Here we explain our optimization mechanisms for the EfficientNet series of networks. When conducting EfficientNet inference on GPUs, we observe that most of layers in these models consume less than one GPU wave of workload, *i.e.*, highly under-utilizing the GPUs. The GPU throughput is only 61.9 - 109.9 GFLOP/s for B0 to B2 as shown in Table V. Such under-utilized GPU capacity allows us to conduct layer growing to reach one full wave for each layer without latency increment, thus bringing higher model capacity and model accuracy.

For implementation, EfficientNet series of networks have three scaling dimensions in their initial design: layer width (w), input resolution (r) and network depth (d) [34]. Out of the three, we balanced scale up both (w) and (r) to increase the network capacity. The depth (d) dimension, however, is kept same with the baseline model, as network layers are run sequentially and depth scaling would inevitably incur latency increase. By width and resolution growing, the optimized models have higher accuracy as well as improved GPU throughput, *e.g.*, 414.2 - 547.3 GFLOP/s, while maintaining the similar latency with baselines, as shown in Table V.

C. Generalizability across GPUs

We further extend the generalizability evaluation to different high-end GPU, P6000 and embedded GPU, Jetson Nano. The detailed GPU specifications can be found in Table II.

The overall results are shown in Table VI and Table VII. On the P6000 GPU (Table VI), our method could achieve **9.0%** to **27.2%** latency reduction than SOTA methods while maintaining similar accuracy for VGG16 and ResNet56, demonstrating the generality across different high-end GPU architectures. For Jetson Nano GPU (Table VII), although it has relatively smaller computing capacity, our method still achieves **13.3%** to **20.5%** latency reduction, demonstrating the generalizability for both high-end and embedded GPU platforms. The above results demonstrate the effectiveness and generalizability of our DNN optimization algorithm. Without any GPU-specific assumptions, our profiling-guided design could be applied to a spectrum of GPUs to enhance the DNNs' accuracy-latency trade-offs.

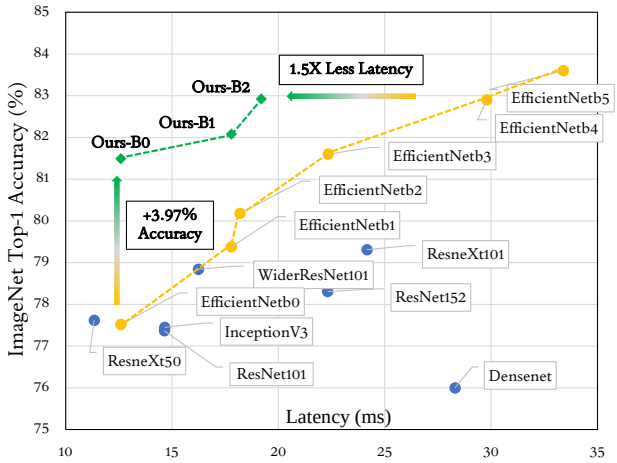


Fig. 12. ImageNet accuracy and latency comparisons. Compared to EfficientNets (yellow) and other models, our GPU-aware optimized models achieve better accuracy latency trade-offs. Evaluated platform: NVIDIA Titan-V GPU.

TABLE VI
GENERALIZABILITY EVALUATION ON PASCAL P6000 GPU.

	Method	Params	FLOP/s	Acc.%	Time (ns)
VGG16	HRank	1.90M	1.68T	**93.1**	4.15E6
	Ours	2.04M	**1.86T**	92.9	**3.02E6 (-27.2%)**
ResNet56	HRank	0.48M	0.84T	93.6	5.84E6
	Ours	0.50M	**1.00T**	**93.7**	**5.32E6 (-9.0%)**

TABLE VII
GENERALIZABILITY EVALUATION ON JETSON NANO GPU.

	Method	Params	FLOP/s	Acc.%	Time (ms)
VGG16	HRank	1.90M	35.5G	**93.1**	60.5
	Ours	2.04M	**39.4G**	92.9	**48.1 (-20.5%)**
ResNet56	HRank-1	0.48M	25.6G	93.6	82.1
	Ours	0.50M	**28.8G**	**93.7**	**68.0 (-17.1%)**
	HRank-2	0.24M	16.7G	92.3	67.2
	Ours	0.39M	**24.3G**	**92.5**	**58.1 (-13.3%)**

D. Generalizability in DNN Architectures

In this part, we conduct generalizability evaluation of DNN latency staircase in terms of different DNN hyper-parameters, including batch sizes, input resolutions and filter shapes.

Batch Size Variation: Fig. 13 shows the latency *w.r.t.* varied batch size from 1 to 256 for two CONV layers of VGG16 on ImageNet resolution. The latency staircase consistently exists for all layers with different batch sizes. Specifically, layers with larger batch size incurs a longer processing cycle for a wave, leading to a higher latency staircase.

Number of Filters Variation: Fig. 13 also shows that CONV-5 has more levels of staircase compared to CONV-3. The reason is that CONV-5 has more filters (128) than CONV-3 (64). Since the number of filters determines the number of thread blocks, more sequential GPU waves are needed with a larger number of filters, leading to more levels of staircase.

Input Resolution Variation: Similar latency evaluation is also conducted w.r.t varied input resolution from 128^2 to 1024^2 for CONV-3 and CONV-5 in Fig. 14. The latency staircase still consistently shows in both low to high resolutions.

Filter Shape Variation: In addition, we evaluate layers with different filter shapes from 1x1, 3x3, 5x5 dense filters to 3x3 depthwise filter, while keep the maximum layer workload for

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:37:56 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 Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully edi
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2024.3404413

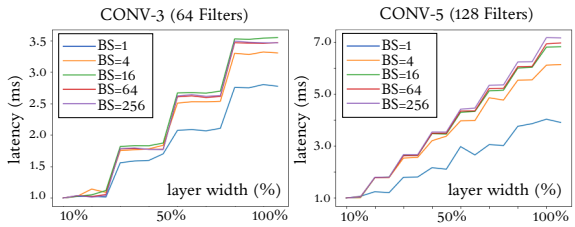


Fig. 13. Latency staircase *w.r.t* varied batch size (1 to 256 in sub-figure) and number of filters (64 to 128).

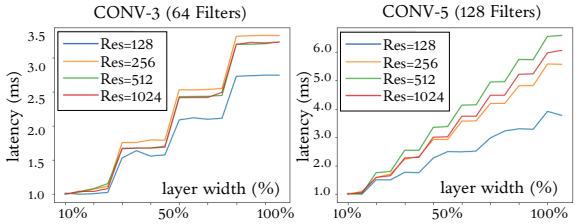


Fig. 14. Latency Staircase *w.r.t* varied input resolution.

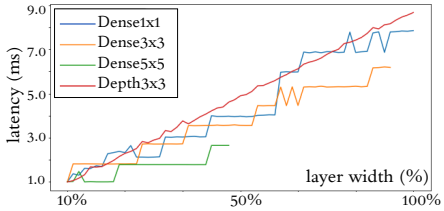


Fig. 15. Latency staircase *w.r.t* varied filter shape.

each type to be the same[6]. Fig. 15 shows the generalizability of latency staircase w.r.t. 1x1, 3x3 and 5x5 dense filters. One exception here is the depthwise 3x3 filter, whose latency results demonstrate a relatively linear pattern. The reason is that the depthwise filter is more light-weight compared to dense filters. Each GPU processing cycle for one wave of depthwise workload is thus smaller, leading to no dramatic upward edge in the latency curve. However, as depthwise filters are usually combined with 1x1 dense filters to compose depth-width layers, *e.g.*, in MobileNets [31], MnasNets [33], the latency staircase still exists for such models.

VII. Limitations and Future Works

A. Future Works for Concurrent Kernel Analysis

Our current work adopts a ***single-model*** execution assumption as stated in NN-meter [40], i.e., *the assumption that kernels run sequentially on each device, even for those ones without dataflow dependency.* This is true for most edge GPUs with limited capacity [40]. Meanwhile, we have observed the same sequential execution patterns hold for our tested architectures (e.g., VGGs, ResNets, EfficientNets) and tested server-level GPUs (e.g., Ampere 3080Ti, Volta Titan-V, Pascal P6000).

However, with the continuous CUDA feature upgrading, we do see recently some new CUDA features (such as multi-stream, multi-process service MPS [27], multi-instance MIG [26]) enable ***multiple models*** and ***concurrent kernel execution*** on the same GPU since the Ampere architecture. This will lead to complex cross-effects between several kernels such that when multiple kernels share the sane resources, the tail effect's impact can be different from current single-kernel execution. However, in the current stage, a practical limitation that we have been facing is that, up-to-now, the available CUDA profiling toolsets (NVIDIA NSight/Compute [24]) still cannot force two kernels to accurately align to co-run, thus it can be hard to analyze their concurrency states, nor profile two arbitrary kernels' concurrent execution time accurately. Multi-round end-to-end multi-model co-running and profiling can be a potential way to remediate such limitations but also require more statistical analysis.

B. Limitations for Vision Transformers

Our current work mainly focuses on convolutional neural networks including traditional VGG and ResNets, as well as the NAS-based networks like EfficientNets, which cover most of the basic convolution components for modern CNNs. Recently, vision transformer (ViT) architectures are introduced [17], [4]. ViT architecture features with a new meta-layer type, i.e., multi-head self-attention (MHSA) layer, and a new type of transformer blocks composed MHSA layers and feed-forward fully connected (FC) layers.

With the specific transformer block structure and MHSA layers, the CNN-oriented channel pruning mechanisms for CNNs are no longer suitable. Instead, ViT pruning method design and the latency analysis faces multiple major challenges that are still awaiting research and analysis: (1) ViT pruning can happen within more dimensions, e.g., image patch embedding size (E), number of head (H), query & query size (QK), value size (V) and FC layer hidden dimension (M). (2) Pruning different ViT dimensions will require corresponding dimension matching across inter-dependent layer shapes [37], [29], for example, image patch embedding size (E) will also impact all query sizes and FC layer's output sizes. With compound dimension pruning challenges and different attention computation mechanisms, our current tail effect analysis is not directly applicable onto the ViT-based architecture pruning, as shown in Figure 16. However, the fundamental parallel computing mechanism and the wave-based granularity of GPU

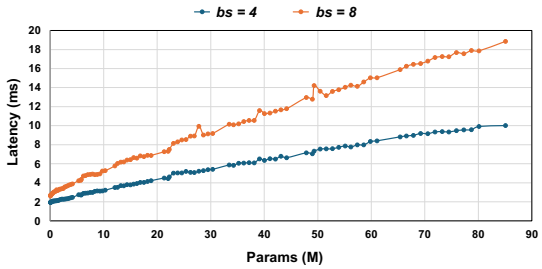


Fig. 16. We applied fine-grained pruning on VIT-Base. As we can observe, the model latency tends to be linearly correlates to the ratio of pruned num of heads, indicating the limitation of the current tail effect analysis on MHSA regarding the multi-step multi-head computation patterns, correlated FFA computation changes, etc. Latency measured on 3080Ti with batch size = 8/4.

[6]The difference between dense and depthwise filter is the dense kernel is of shape $n \times n \times 512$, while the depth-wise kernel is only $n \times n \times 1$. The latter one is mainly used in light-weight models like MobileNets [31].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:37:56 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 Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully edi
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2024.3404413

architecture still remains. Therefore, we leave the compound tail-effect analysis on ViTs as future work.

VIII. Conclusion

In this work, we revisit the GPU tail effect, a classic parallel system issue, and shows it can cause significant GPU under-utilization and lower the GPU throughput for DNN inference. Detailed DNN workload characterization demonstrates the prevalence of GPU tail effect across different DNN architectures is due to the unique deep structure and the light-weight layer workload, which exacerbates the tail effect in DNN inference. We then demonstrate two user cases with simple yet effective solutions to eliminate the tail effect and enhance the DNN run-time efficiency. With effectiveness, feasibility, and generalizability well proved, our optimization methods show outstanding DNN latency accuracy trade-offs, *e.g.*, 11%–27% latency reduction over state-of-the-art (SOTA) DNN pruning and NAS approaches.

References

[1] Tianqi Chen, Thierry Moreau, Ziheng Jiang, Lianmin Zheng, Eddie Yan, Haichen Shen, Meghan Cowan, Leyuan Wang, Yuwei Hu, Luis Ceze, et al. {TVM}: An automated end-to-end optimizing compiler for deep learning. In *13th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 18)*, pages 578–594, 2018.

[2] Xiaoliang Dai, Peizhao Zhang, Bichen Wu, Hongxu Yin, Fei Sun, Yanghan Wang, Marat Dukhan, Yunqing Hu, Yiming Wu, Yangqing Jia, et al. Chamnet: Towards efficient network design through platform-aware model adaptation. In *Proceedings of the IEEE Conference on computer vision and pattern recognition*, pages 11398–11407, 2019.

[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[4] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[5] Gongfan Fang, Xinyin Ma, Mingli Song, Michael Bi Mi, and Xinchao Wang. Depgraph: Towards any structural pruning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16091–16101, 2023.

[6] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. IEEE, 2013.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[8] Yang He, Guoliang Kang, Xuanyi Dong, Yanwei Fu, and Yi Yang. Soft filter pruning for accelerating deep convolutional neural networks. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 2234–2240, 2018.

[9] Yang He, Ping Liu, Ziwei Wang, Zhilan Hu, and Yi Yang. Filter pruning via geometric median for deep convolutional neural networks acceleration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4340–4349, 2019.

[10] Yihui He, Ji Lin, Zhijian Liu, Hanrui Wang, Li-Jia Li, and Song Han. Amc: Automl for model compression and acceleration on mobile devices. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 784–800, 2018.

[11] Yihui He, Xiangyu Zhang, and Jian Sun. Channel pruning for accelerating very deep neural networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1389–1397, 2017.

[12] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[13] Zhihao Jia, Oded Padon, James Thomas, Todd Warszawski, Matei Zaharia, and Alex Aiken. Taso: optimizing deep learning computation with automatic generation of graph substitutions. In *Proceedings of the 27th ACM Symposium on Operating Systems Principles*, pages 47–62, 2019.

[14] Mingbao Lin, Rongrong Ji, Yan Wang, Yichen Zhang, Baochang Zhang, Yonghong Tian, and Ling Shao. Hrank: Filter pruning using high-rank feature map. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1529–1538, 2020.

[15] Mingbao Lin, Rongrong Ji, Yan Wang, Yichen Zhang, Baochang Zhang, Yonghong Tian, and Ling Shao. Hrankplus github repo, 2020. https://github.com/lmbxmu/HRankPlus.

[16] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[17] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[18] Zhuang Liu, Jianguo Li, Zhiqiang Shen, Gao Huang, Shoumeng Yan, and Changshui Zhang. Learning efficient convolutional networks through network slimming. In *Proceedings of the IEEE international conference on computer vision*, pages 2736–2744, 2017.

[19] microsoft. Deep learning inference service at microsoft, 2020. https://www.usenix.org/system/files/opml19papers-soifer.pdf.

[20] NVIDIA. Cuda pro tips (page 13), 2020. https://on-demand.gputechconf.com/gtc/2012/presentations/S0514-GTC2012-GPU-Performance-Analysis.pdf.

[21] NVIDIA. Cudnn kernel invoking logic, 2020. https://docs.nvidia.com/deeplearning/cudnn/api/index.html-cudnnGetConvolutionForwardAlgorithm_v7.

[22] NVIDIA. Dnn compiling stack and kernel base, 2020. https://developer.download.nvidia.com/video/gputechconf/gtc/2020//s21685-cuDNN-v8-New-Advances-in-Deep-Learning-Acceleration-APIs-Optimizations-and-How-to-Tackle-the-Future-Challenges-in-Hardware-and-Software.pdf.

[23] NVIDIA. Jetson devices, 2020. https://www.nvidia.com/en-us/autonomous-machines/jetson-store/.

[24] Nvidia. Nsight compute — nvidia, 2020. https://developer.nvidia.com/nsight-compute.

[25] NVIDIA. Nvidia cudnn documentation — kernel selection heuristics, 2020. https://docs.nvidia.com/deeplearning/cudnn/developer-guide/index.htmltroubleshooting.

[26] NVIDIA. Nvidia multi instance gpu (mig), 2020. https://docs.nvidia.com/datacenter/tesla/mig-user-guide/.

[27] NVIDIA. Nvidia multi process service (mps), 2020. https://docs.nvidia.com/deploy/pdf/cuda_multi_process_service_overview.pdf.

[28] NVIDIA. Nvidia titan v — volta architecture, 2020. https://www.nvidia.com/en-us/titan/titan-v/.

[29] Lorenzo Papa, Paolo Russo, Irene Amerini, and Luping Zhou. A survey on efficient vision transformers: algorithms, techniques, and performance benchmarking. *arXiv preprint arXiv:2309.02031*, 2023.

[30] Jonathan Ragan-Kelley, Connelly Barnes, Andrew Adams, Sylvain Paris, Frédo Durand, and Saman Amarasinghe. Halide: a language and compiler for optimizing parallelism, locality, and recomputation in image processing pipelines. *Acm Sigplan Notices*, 48(6):519–530, 2013.

[31] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[32] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[33] Mingxing Tan, Bo Chen, Ruoming Pang, Vijay Vasudevan, Mark Sandler, Andrew Howard, and Quoc V Le. Mnasnet: Platform-aware neural architecture search for mobile. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2820–2828, 2019.

[34] Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*, 2019.

[35] Han Vanholder. Efficient inference with tensorrt, 2016.

[36] Bichen Wu, Xiaoliang Dai, Peizhao Zhang, Yanghan Wang, Fei Sun, Yiming Wu, Yuandong Tian, Peter Vajda, Yangqing Jia, and Kurt Keutzer. Fbnet: Hardware-aware efficient convnet design via differentiable neural architecture search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10734–10742, 2019.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:37:56 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 Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully edi
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2024.3404413

[37] Huanrui Yang, Hongxu Yin, Maying Shen, Pavlo Molchanov, Hai Li, and Jan Kautz. Global vision transformer pruning with hessian-aware saliency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18547–18557, 2023.
[38] Tien-Ju Yang, Andrew Howard, Bo Chen, Xiao Zhang, Alec Go, Mark Sandler, Vivienne Sze, and Hartwig Adam. Netadapt: Platform-aware neural network adaptation for mobile applications. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 285–300, 2018.
[39] Peifeng Yu and Mosharaf Chowdhury. Salus: Fine-grained gpu sharing primitives for deep learning applications. *arXiv preprint arXiv:1902.04610*, 2019.
[40] Li Lyna Zhang, Shihao Han, Jianyu Wei, Ningxin Zheng, Ting Cao, Yuqing Yang, and Yunxin Liu. Nn-meter: Towards accurate latency prediction of deep-learning model inference on diverse edge devices. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, pages 81–93, 2021.



Fuxun Yu received the B.S. degree from Harbin Institute of Technology, Harbin, China, in 2017. He received his Ph.D. degree in the Department of Electrical and Computer Engineering at George Mason University under the supervision of Prof. Xiang Chen. His current research interests include high performance deep neural network computing, full-stack DNN computing optimization, deep learning security, interpretability and explainability of deep learning.



Zirui Xu received the B.S. and M.S. degree from Beijing Jiaotong University, Beijing, China, in 2014 and in 2017, respectively. He received his Ph.D. degree with the Department of Electrical and Computer Engineering at George Mason University under the supervision of Prof. Xiang Chen. His current research interests include high performance mobile computing system, neural network model optimization, and mobile intelligent application robustness and security.



Longfei Shangguan is an assistant professor in the Computer Science Department. Before joining Pitt, he was a senior researcher at Microsoft, Redmond. Dr. Shangguan's research interests are in all aspects of IoT systems: from building novel IoT applications, solving security issues, all the way down to optimizing the network stack and designing low-power IoT hardware. His research has been recognized by a Google Research Scholar Award, MobiCom Best Paper Runner-up Award, Ubicomp Distinguished Paper Award, and an AIoTsys Young Scientist Award.



Di Wang is currently a Principal Research Manager with Microsoft. He earned his B.E. in Computer Science and Technology from Zhejiang University (2005), M.S. in Computer Systems Engineering from Technical University of Denmark (2008), and Ph.D. in Computer Science and Engineering from The Pennsylvania State University (2014). His research spans the areas of artificial intelligence, computer systems, computer architecture, and energy efficient system design and management. He has authored more than 40 peer-reviewed papers. He received 5 best paper awards and 2 best paper nominations.



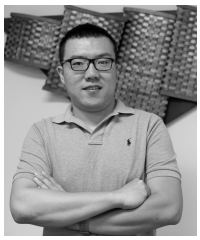
Dimitrios Stamoulis is currently a Principal Research Manager at Microsoft AI, working on Neural Architecture Search (NAS) for object detection applications. He received his PhD in Electrical and Computer Engineering from Carnegie Mellon University under the supervision of Diana Marculescu. Previously, He received a MEng in ECE from McGill University, Montreal, and a Diploma in ECE from NTUA, Athens.



Rishi Madhok is currently a Senior Applied Science Manager at Microsoft and leads high-impact engineering and applied AI projects. He received his M.S. in Computer Science from Carnegie Mellon University. Before that, he received his B.S. degree from Delhi Technological University. His current research interests lie in object detection, semantic segmentation, and tracking models from aerial/satellite imagery.



Nikolaos Karianakis is currently a Principal Research Manager in Mixed Reality at Microsoft and leads high-impact engineering and applied AI projects. He received his Ph.D. in Computer Science from UCLA at the Vision Lab under the supervision of Stefano Soatto. His research spans the broader areas of Computer Vision and Machine Learning. His current focus is Computer Vision and Large Language Models for Aerial and Horizontal Imaging.



Ang Li is a tenure-track assistant professor in the Department of Electrical and Computer Engineering at University of Maryland College Park. He obtained Ph.D. from Duke University under the supervision of Professor Yiran Chen. His research interests lie in the intersection of machine learning and edge computing, with a focus on building large-scale networked and trustworthy intelligent systems to solve practical problems in a collaborative, scalable, secure, and ubiquitous manner.



ChenChen Liu received the M.S. degree from Peking University, Beijing, China, in 2013, and the Ph.D. degree from the ECE Department at the University of Pittsburgh, Pittsburgh, USA, in 2017. She is currently an Assistant Professor in the Department of Computer Science and Electrical Engineering at University of Maryland, Baltimore County. Her current research interests include brain-inspired computing system, machine learning, and emerging nonvolatile memory technologies.



Yiran Chen received B.S. (1998) and M.S. (2001) from Tsinghua University and Ph.D. (2005) from Purdue University. After five years in the industry, he joined the University of Pittsburgh in 2010 as Assistant Professor and was promoted to Associate Professor with tenure in 2014, holding Bicentennial Alumni Faculty Fellow. He is now the John Cocke Distinguished Professor of Electrical and Computer Engineering at Duke University and serving as the director of the NSF AI Institute for Edge Computing Leveraging the Next-generation Networks (Athena), the NSF Industry-University Cooperative Research Center (IUCRC) for Alternative Sustainable and Intelligent Computing (ASIC), and the co-director of Duke Center for Computational Evolutionary Intelligence (DCEI).



Xiang Chen is a Tenure-Track Associated Professor of the Department of Computer Science and Technology, the School of Computer Science, Peking University. He achieved his MS. and Ph.D. degrees with a major in Computer Engineering under the supervision of Dr. Yiran Chen at the University of Pittsburgh and finished the bachelor's degree at Northeastern University in China in 2010. After his Ph.D. graduation in 2016, he directly joined George Mason University, where he led 7 National Science Foundation (NSF) projects and achieved the NSF CAREER Award. In 2023, he joined Peking University. His research works focus on mobile and distributed computing systems, high performance intelligence computing, and related Edge, IoT, and CPS applications.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 09:37:56 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.

Neural Computing and Applications
https://doi.org/10.1007/s00521-024-09987-2

RETRACTION NOTE

Retraction Note: *CanarDeep*: a hybrid deep neural model with mixed fusion for rumour detection in social data streams

Deepak Kumar Jain[1] · Akshi Kumar[2] · Akshat Shrivastava[3]

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

Retraction Note to: Neural Computing and Applications (2021) 34:15129–15140
https://doi.org/10.1007/s00521-021-06743-8

The Editor-in-Chief and the publisher have retracted this article. The article was submitted to be part of a guest-edited issue. An investigation by the publisher found a number of articles, including this one, with a number of concerns, including but not limited to compromised editorial handling and peer review process, inappropriate or irrelevant references or not being in scope of the journal or guest-edited issue. Based on the investigation's findings the Editor-in-Chief therefore no longer has confidence in the results and conclusions of this article.

Akshi Kumar has stated on behalf of all authors that they disagree with this retraction.

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

The original article can be found online at https://doi.org/10.1007/s00521-021-06743-8.

✉ Akshi Kumar
akshi.kumar@nsut.ac.in

[1] Key Laboratory of Intelligent Air-Ground Cooperative Control for Universities in Chongqing, College of Automation, Chongqing University of Posts and Telecommunications, Chongqing, China

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

[3] Department of Computer Science & Engineering, Delhi Technological University, Delhi, India

Published online: 20 May 2024

R-Peak Detection Using Wavelet Scattering Transform for Pre-Term Infant ECG Dataset

Shikha Singhal
Department of Electronics & Communication Engineering
Delhi Technological University
Delhi, India
ggshikha12@gmail.com

Manjeet Kumar
Department of Electronics & Communication Engineering
Delhi Technological University
Delhi, India
manjeetchhillar@gmail.com

2024 2nd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT) | 979-8-3503-7284-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/DICCT61038.2024.10532950

***Abstract*—Heart-related disease is one of the pre-eminent origins of mortality across the globe. Early and accurate detection of heart-related disease is essential to avoid an uncertain mishap. Peak detection is a pivotal step for the analysis of different spectral signals. The dislocation of peaks disrupts the detection accuracy of a heart-related disease. This study proposes a method for accurate R-peak detection of ECG signal. The acute R-peak detection of ECG signal is achieved using wavelet scattering transform. It eliminates the artifacts associated to signal with the help of inbuilt filters, proves reliable as it identifies the R-peaks at correct location. The wavelet scattering transform generates rational, instructive, and invariable translation representation of signals. The efficiency of the proposed method for R-peak detection is evaluated on pre-term infant ECG dataset available at physionet. The attained average accuracy and specificity for R-peak detection is 99.23% and 99.86% respectively. The proposed method outperforms the state-of-the-art methods namely continuous wavelet transform (CWT) and thresholding technique.**

Keywords— Electrocardiogram, R-peak detection, pre-term infant, wavelet scattering transform.

I. INTRODUCTION

A precise detection of R-peaks is necessary for recognition of cardiovascular diseases (CVDs) in electrocardiogram (ECG) signals. The R-peak identification is the major feature in the diagnosis of different cardiac pathologies. An electrocardiogram serves as a non-invasive tool used to detect the irregularities of heart, including the contractions in ventricular and atrial area, and supraventricular premature beats [1]. The advent of wearable ECG monitors underscores the significance of examining resilient automated R-peak detection in single-lead ECG signals. Although, several methods of R-peak detection have been proposed, it is still an important issue in terms of noise effects, alternative rhythms etc. The R-peak detection is the origin of arrhythmia recognition. Signal processing-based applications comprises of data collection, preprocessing, feature extraction, and classification models [2]. The characterization of the signal is achieved with time-varying electrical depictions that may generate from various sources. A typical ECG waveform includes the P wave, QRS complex, and T wave. The QRS complex represents the electrical depolarization of the muscles in the ventricular area of the heart [3]-[4]. The HRV assessment are generally influenced by noise level, including artifacts present in the raw signal that makes the R-peak detection complex. Noise sources [5] consists of power-line interference, baseline wandering, and electromyographic noise that affects the ECG signal in various aspects. It has been observed that R-peak detection is basically split into threshold methods, wavelet analysis method, and deep learning methods. The traditional thresholding method includes filtering the signal, increase the QRS complex, and detect the R-peak position with threshold decision. A classic threshold algorithm for R-peak detection is introduced in [6], utilizing the bandpass filter to denoise the ECG signal, then overall steps involving the first-order difference, and squaring to mention the QRS complex components. The R-peak detection using fast Fourier transform with denoising ECG signal is discussed in [7], achieved the sensitivity of 99.98%. To enhance the adaptive ability of peak detection, a weighted continuous wavelet transform is introduced in [8] that generates spectral peak characteristics in low-scale regions. In [9], a combined algorithm of biorthogonal wavelet transforms and run-length encoding is introduced for QRS complex detection achieved a low data error rate of 2×10^{-3}. Several other techniques of peak detection are illustrated in [10]-[12]. A brown exponential smoothing model [13] is implemented for an adaptive R-peak detection procedure, achieved precision and recall of 99.6% and 99.7% respectively. The challenge of noise interruption is removed with this implementation and proves well suited for noisy signals. Based on the noise suppression and peak detection, few decomposition techniques are introduced [14]-[15]. The objective of this study is to detect the R-peak of the pre-term individuals in terms of spectral analysis of the peaks illustrated using scattering transform. The purpose of the study is discussed as follows.

1) The wavelet scattering transform (WST) is proposed for R-peak detection of ECG signals.
2) Examining the wavelet analysis with scattering coefficients generated through WST.
3) The proposed method is employed on pre-term infant ECG database.

This manuscript is framed as follows. In Section I, the introduction and the related work corresponding to R-peak detection of pre-term infant ECG signal is discussed. The detailed information of proposed methodology and the database employed is exhibited in Section II. Section III outline the steps involved in implementing the proposed methodology. In Section IV, the results and discussions are presented. The conclusion is discussed in Section V

II. METHODOLOGY

A. Wavelet Scattering Transform

The WST generates systematic, revealing, and translation-invariant presentation of ECG signals. It exhibits resilience to disruptions and preserves class discriminability, making it highly effective for feature extraction [16]-[17]. The method consists of three-layered iterative signal transformation involving wavelet convolution, modulation, and denoising. The features are extracted using monotonous wavelet decomposition, local averaging, and complex modulus operations. Specifically, the high and low frequency components of respective signal are detected by the wavelet filter and the scaling function respectively. At each

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 05:52:04 UTC from IEEE Xplore. Restrictions apply.

decomposition level, the characteristics associated with low frequency are improved. The balanced frequency characteristics are determined through local averaging following an assessment of the modulus of high-frequency signal coefficients. Through local averaging, the high-frequency data might be lost that can be restored utilizing the high-frequency wavelet transform [18]. Ultimately, after arranging the signal through various scattering routes, the stability is sustained between invariability and bias.

Assume the inspection of ECG signal is presented by $x(t)$. The low pass filter, $\phi_f(t)$ reflects the local translation invariant specifications of x at given scale T. The grouped wavelet indices $\mathbb{W}\mathbb{i}$, shows an octave frequency resolution q. The high pass filter bank is given by $\{\psi_f\}_{f\in\mathbb{W}\mathbb{i}}$ evaluated by widen the wavelet ψ. The convolution $(x * \phi_f(t))$ develops a confined translation invariant characteristic of x, exhibiting the information loss of high frequency. Those high frequencies can be retrieved through wavelet modulus transform and discussed in following steps.

Step1: $|W_1|x = \{C_o x(t), |x * \phi_{f_1}(t)|\}_{f_1\in\mathbb{W}1}$

The first order scattering coefficients after evaluating the mean of wavelet modulus coefficients with ϕ_f is

Step2: $C_1 x(t) = \{|x * \psi_{f_1}| * \phi_f(t)\}_{f_1\in\mathbb{W}1}$

The information can be recovered by calculating mean, where low frequency unit is shown by $C_1 x(t)$ of the high frequency coefficients $|x * \psi_{f_1}|$, can be released by

Step3: $|W_2||x * \psi_{f_1}| = \{C_1 x(t), ||x * \psi_{f_1}| * \psi_{f_1}(t)|\}_{f_2\in\mathbb{W}2}$

Second order scattering coefficients is defined as

Step4: $C_2 x(t) = \{||x * \psi_{f_1}| * \psi_{f_2}| * \phi_f(t)\}_{f_i\in\mathbb{W}i}, i = 1,2$

Reiterating the process, the wavelet modulus convolution is given as

Step5: $V_m x(t) = \{|| |x * \psi_{f_1}| * \dots | * \psi_{f_m}|\}_{f_i\in\mathbb{W}i}, i = 1,2,\dots,m$

By averaging of $V_m y(t)$ with ϕ_f, m-th order scattering coefficients is calculated

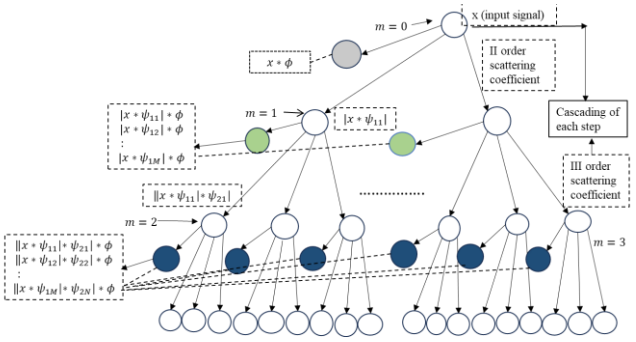


Fig1. The structure of wavelet scattering transform

Step7: $C_m x(t) = \{|| |x * \psi_{f_1}| * \dots | * \psi_{f_m}| * \phi_f(t)\}_{f_i\in\mathbb{W}i}$, $i = 1,2,\dots,m$

The scattering matrix is formed by the sum of scattering coefficients of all orders to verify the input signal characteristics, where the maximum decomposition order is presented by n, $Cx(t) = \{C_m x(t)\}_{0\le m\le n'}$

The wavelet scattering noticed the small variations in time span and dimensions of non-stationary signal which are tough to evaluate but reflects the heart condition [19]. Thus, WST is imposed to produce the robust arrangement of ECG which reduces the differences in one category of arrhythmia during the discriminability maintenance between different categories of arrhythmia. The implementation of WST for feature extraction is introduced in [20]. The formation of WST with scattering order at different layers is structured in Fig 1.

B. Dataset

The data of ten preterm infant is considered with postconceptional age of $29\frac{3}{7}$ to $34\frac{2}{7}$ weeks (average: $31\frac{1}{7}$ weeks) and weights of 843 gm to 2100 gm (average: 1468 gm) [21], available at physionet which is freely accessible [22]. A three-lead ECG signal was recorded for $20-70$ hrs per infant of sampling frequency 500 Hz. A compound ECG signal was registered at 250Hz for infant 1 and 5 in the absence of an ECG channel. The remaining recordings are performed at 500Hz. The whole information of pre-term infant database is summarized in Table1.

Table 1. Summary of pre-term infant ECG database at physionet

Infants	Age (weeks)	Weight (kg)	Duration (hrs)	Mean [SD]	Sampling freq
Inf_1	29 3/7	1.20	45.6	155	250
Inf_2	30 5/7	1.76	43.8	131	500
Inf_3	30 5/7	1.71	43.7	131	500
Inf_4	30 1/7	0.84	46.8	167	500
Inf_5	32 2/7	1.67	48.8	143	250
Inf_6	30 1/7	1.14	48.6	137	500
Inf_7	30 1/7	1.11	20.3	162	500
Inf_8	32 3/7	2.10	24.6	141	500
Inf_9	30 4/7	1.23	70.3	150	500
Inf_10	34 2/7	1.90	45.1	156	500

III. PROPOSED METHOD

Initially, the ECG signal is extracted from pre-term infant database consisting information of ten infants for different age groups. The extracted signal is normalized for precise and acute processing, to keep the different levels of signal in an appropriate range. The wavelet scattering transform (WST) is implemented on the normalized data, generating coherent, informative, and translation-invariant representation of ECG signals. The benefit of implementing the WST is to utilize the in-built filters. Thus, an extra step of preprocessing is removed by employing this technique. The technique works on both low and high frequency signal at different levels with scattering coefficients. The 2D time-frequency scalogram is generated using WST, represents the amplitude and location of peaks. The energy intensity can be estimated through the formation of peaks, which also separates the noise signal from the original signal. The scattering matrix is generated after combining all scattering coefficients obtained at each level from WST. The matrix obtained after combination is investigated through average of samples at different scales.

An automatic selection of scales corresponding to maximum energy intensity is performed. Thus, the maximum wavelet analysis at a selected scale is achieved that helps to notice the R-peak of the ECG signal. The overall flow involved in R-peak detection of ECG signal utilizing WST is presented in Fig. 2, consists of data acquisition, preprocessing using WST, and R-peak detection using wavelet analysis.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 05:52:04 UTC from IEEE Xplore. Restrictions apply.

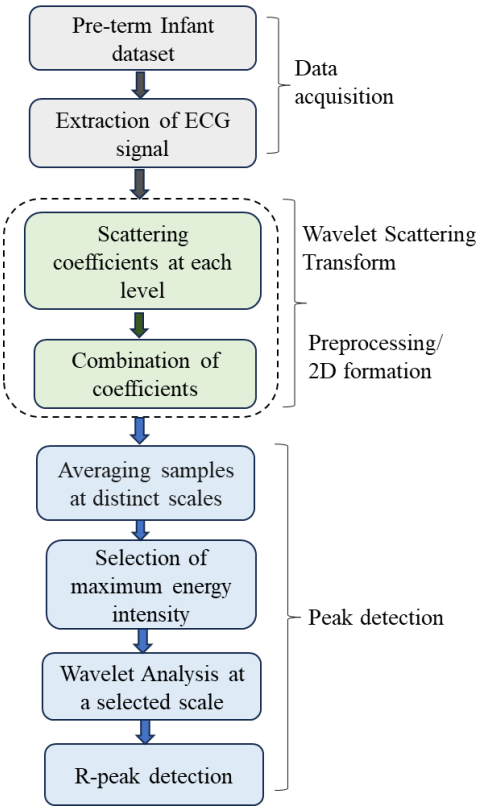


Fig 2. The proposed methodology for peak detection using WST

IV. RESULTS AND DISCUSSIONS

The simulations are executed on MATLAB 2021a using GPU: NVIDIA GEFORCE RTX 3070, processor: 11th Gen Intel(R) Core (TM) i7-11370H@3.30GHz, and memory of 128 GB with software: Win 10 64-bit operating system.

Initially, the ECG records are fetched from pre-term infant database for R-peak detection. The input signal, Inf_1 consisting 2500 samples is normalized an applied to the WST. The scalogram generated from WST conveys the spectral analysis of the respective signal. It represents the energy intensity of the specified peaks through which the R-peak detection of ECG signal is achieved. From WST, a resultant scattering coefficient matrix is produced which is analyzed at different samples. Finally, the averaging of these sample value reflects the selected time-scale. Thereafter, localized the R-peaks in the given range of samples. The results obtained through implementation of WST is illustrated in Fig. 3, represents the normalized ECG signal, the obtained 2D scalogram representation, the combined form of scattering coefficients, and the detected R-peaks of ECG signal with signal, noise, and wavelet analysis levels using blue, black, and green color respectively. The implementation of WST is made upon Inf_3 and the results obtained are discussed in Fig. 4. In Fig. 4, it is seen that the input signal of ten peaks of sample length 2300 is processed and after wavelet analysis, the number of R-peaks detected reduced to nine, shows the dislocation of peak to other scale. A random selection of samples is made and correspondingly, the detection accuracy of R-peaks is calculated using WST implementation, discussed in Table 2. The evaluation parameters named as accuracy and specificity are calculated and compared with the existing techniques for R-peak detection.

Accuracy (Acc.): It may be defined as the correctly identification of R-peaks to the total number of peaks in the ECG signal.

$$Acc. = \left(\frac{tp + tn}{tp + tn + fn + fp}\right) \times 100$$

Specificity (Spec.): It is defined as the ratio of the true negatives to all negative results obtained. It specifies how correctly the model identify the occurrences of negative class.

$$Spec. = \left(\frac{tn}{tn + fp}\right) \times 100$$

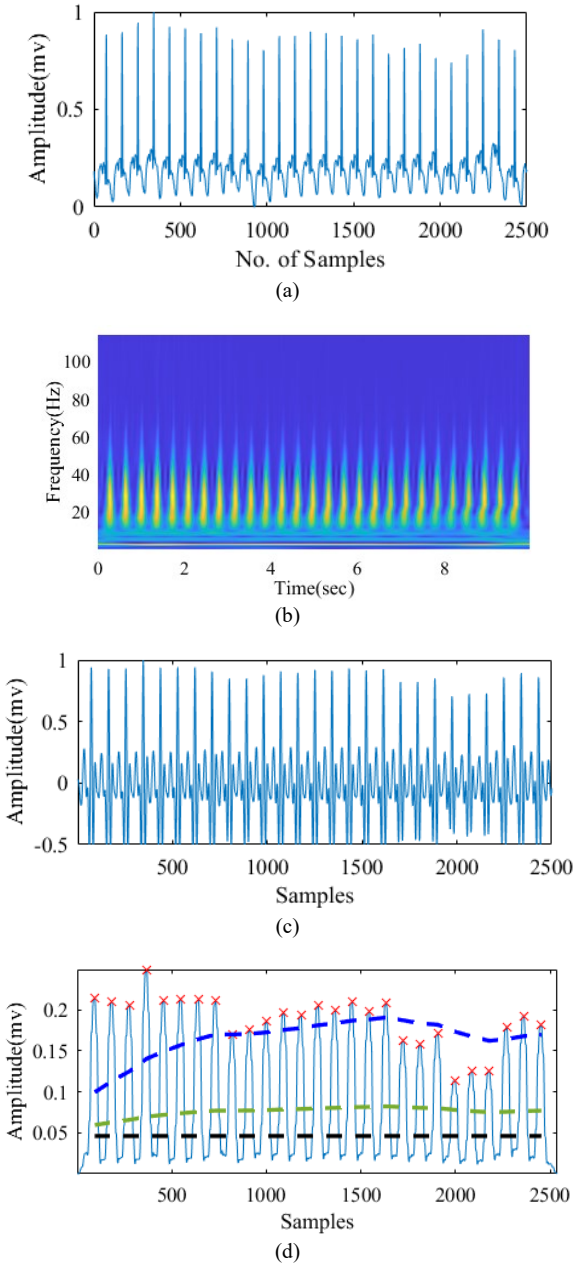


Fig 3. (a) ECG signal of infant *Inf_1* (b) Scalogram generated through WST (c) result obtained through different combinations of WST coefficients (d) R-peak detection using maximum wavelet analysis: black shows the noise level, green reflects the wavelet analysis, blue shows the signal level.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 05:52:04 UTC from IEEE Xplore. Restrictions apply.

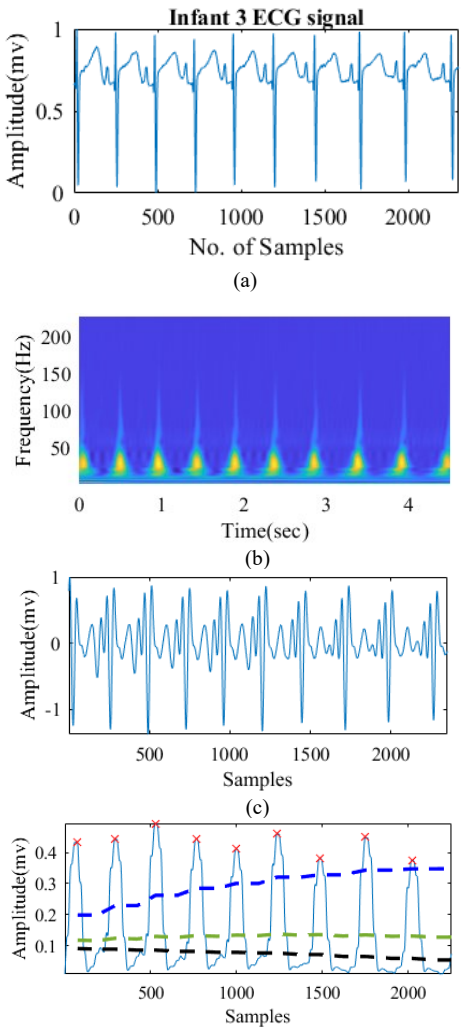


Fig 4. (a) ECG signal of infant *Inf_3* (b) Scalogram obtained through WST (c) Result obtained through different combinations of WST coefficients (d) R-peak detection using maximum wavelet analysis.

Table 2. The summary of detection accuracy at different samples of infants.

Infants	Samples	Total no. of Peaks given	Detected Peaks using WST	Detection accuracy (%)
Inf_1	3600	27	27	100
Inf_2	2500	23	23	100
Inf_2_17	4000	28	28	100
Inf_3	2300	10	9	90.0
Inf_4	1800	15	15	100
Inf_5	1300	12	12	100
Inf_5_9	3200	24	24	100
Inf_6	1000	8	8	100
Inf_7	1000	9	9	100
Inf_8	1200	11	11	100
Inf_8_11	1400	12	12	100
Inf_9	1700	16	16	100
Inf_10	2200	21	21	100

The proposed method is compared with existing methods to assess performance in terms of accuracy and specificity. Other methods such as continuous wavelet transform performed well but somehow lags in computational time which is an important parameter while handling a diseased subject. Few existing methods are compared with proposed method discussed in Table 3.

Table 3. Summary of comparing the proposed method with existing methods for R-peak detection using ECG signals.

Author	Database	Method	Performance (%)
Das et al [23]	BIH-DB, PTB-DB, Fantasia DB	Wavelet Transform (scale 4)	Spec.:99.98
E.Celik et al [24]	MIT-BIH DB	CWT + U-Net3+ model	Acc.: 98
Peng et al [25]	MIT-BIH DB CPSC2019	ST-Res U-net	Sens.: 99.76,90.01
Proposed method	**Pre-term Infant ECG**	**WST**	**Acc.:99.23 Spec.: 99.86**

DNN: Deep Neural Network; BIH-DB: Boston's Beth Israel Hospital Database; PTB-DB: Physikalisch Technical Bundesanstalt Database

The R-peak detection on the same database (pre-term infant database) is not available. Hence, discussed the R-peak detection with other databases in Table 3.

V. CONCLUSION & FUTURE SCOPE

The R-peak detection of ECG signal is a crucial step in diagnosing an arrhythmia disease. The dislocation of peaks from their original position disrupts the heart status, correspondingly decrease the detection accuracy of the respective disease. In this manuscript, executed an R-peak detection algorithm based on the wavelet scattering transform. The core of this framework is to detect the peak status on the basis of sample and scale group. The proposed method deals with the coefficients generated though WST, where wavelet analysis is utilized to detect the level of the obtained peaks. The verification of detected R-peaks of ECG signal with the original number of peaks is performed using a combined form of scattering coefficients, which is being averaged over different samples. The proposed methodology was tested on the preterm infant dataset and achieved an average accuracy and specificity of 99.23%, 99.86%. The computed average computational time of proposed method is 13 seconds. In future, peak detection can be processed utilizing the WST on real-time dataset.

REFERENCES

[1] M. U. Zahid *et al*., "Robust R-Peak Detection in Low-Quality Holter ECGs Using 1D Convolutional Neural Network," in IEEE Transactions on Biomedical Engineering, vol. 69, no. 1, pp. 119-128, Jan. 2022.

[2] Shikha Singhal, and Manjeet Kumar, "A Systematic Review on Artificial Intelligence-Based Techniques for Diagnosis of Cardiovascular Arrhythmia Diseases: Challenges and Opportunities", Archives of Computational Methods in Engineering (Springer), Vol. 30, pp. 865–888, March 2023.

[3] Banerjee S, Mitra M, "Application of cross wavelet trans - form for ECG pattern analysis and classifcation", IEEE Trans Instrum Meas vol. 63 no. 2, pp. 326–333. 2013.

[4] S. Singhal and M. Kumar, "Cardiovascular Diseases Classification Using High- Resolution Superlet Transform on ECG and PCG Signals," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, pp. 1-5, 2023.

[5] Tripathi, P.M., Kumar, A., Komaragiri, R., " A Review on Computational Methods for Denoising and Detecting ECG Signals to Detect Cardiovascular Diseases" Arch Computat Methods Eng, vol. **29**, pp. 1875–1914, 2022.

[6] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," IEEE Trans. Biomed. Eng., vol. BME-32, no. 3, pp. 230–236, Mar. 1985.

[7] Kumar, A., Ranganatham, R., Komaragiri, R., "Efficient QRS complex detection algorithm based on Fast Fourier Transform" Biomed. Eng. Lett. Vol. 9, pp. 145–151, 2019.

[8] Y. Zhou, J. Ma, F. Li, B. Chen, T. Xian and X. Wei, "An Improved Algorithm for Peak Detection Based on Weighted Continuous Wavelet Transform," in *IEEE Access*, vol. 10, pp. 118779-118788, 2022.

[9] Kumar A, Kumar M, Komaragiri R, "Design of a biorthogonal wavelet transform based R-peak detection and data compression scheme for

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 05:52:04 UTC from IEEE Xplore. Restrictions apply.

implantable cardiac pacemaker systems", J Med Syst., vol. 42, no 6, pp. 102, 2010.
[10] Kumar A, Komaragiri R, Kumar M., "From pacemaker to wearable: techniques for ECG detection systems", J Med Syst, vol. 42, no. 2, pp. 34, 2018;.
[11] Kumar A, Komaragiri R, Kumar M. "Heart rate monitoring and therapeutic devices: a wavelet transform based approach for the modeling and classification of congestive heart failure", ISA transactions, vol. 79, pp. 239-250, 2018.
[12] Kumar A, Komaragiri R, Kumar M., "Design of wavelet transform based electrocardiogram monitoring system" ISA transactions, vol 80, pp. 381-398, 2018.
[13] Z. Hao, X. Zhang and Z. Lai, "Adaptive R-Peak Detection Algorithm Based on Brown Exponential Smoothing Model," in IEEE Access, vol. 10, pp. 114355-114363, 2022.
[14] S. Singhal and M. Kumar, "GSMD-SRST: Group Sparse Mode Decomposition and Superlet Transform Based Technique for Multi-Level Classification of Cardiac Arrhythmia," in IEEE Sensors Journal, vol. 24, no. 6, pp. 8160-8169, 2024.
[15] V. N. V, H. Cao and L. Peyrodie, "Variational Mode Decomposition-Based Simultaneous R Peak Detection and Noise Suppression for Automatic ECG Analysis," in IEEE Sensors Journal, vol. 23, no. 8, pp. 8703-8713, 15 April15, 2023.
[16] S Nahak, A Pathak, G Saha, "Fragment-level classification of ECG arrhythmia using wavelet scattering transform", Expert Systems with Applications, vol. 224, Article ID 120019, 2023,.
[17] Z Liu, G Yao, Q Zhang, J Zhang, X Zeng, "Wavelet Scattering Transform for ECG Beat Classification", Computational and Mathematical Methods in Medicine, vol. 2020, Article ID 3215681, 11 pages, 2020.
[18] S. Mollai, "Recursive interferometric representations", 18th European Signal Processing Conference, Aalborg, Denmark, 2010, pp. 716-720, 2010
[19] Marzog, H. A., and Abd, H. J. Machine Learning ECG Classification Using Wavelet Scattering of Feature Extraction. Applied Computational Intelligence and Soft Computing, pp. 1–8, 2022.
[20] S. Singhal and M. Kumar, "SPTDMD-WST: Arrhythmia Classification by Extracting Features from Spatiotemporal Modes Generated through Dynamic Mode Decomposition Using Wavelet Scattering Transform," in Biomedical Sugnal Processing and Control, vol. 92, 2024, 105983.
[21] A. H. Gee, R. Barbieri, D. Paydarfar and P. Indic, "Predicting Bradycardia in Preterm Infants Using Point Process Analysis of Heart Rate," in IEEE Transactions on Biomedical Engineering, vol. 64, no. 9, pp. 2300-2308, Sept. 2017.
[22] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh. Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and Physionet: Components of a New Research Resource for Complex Physiologic Signals. Circulation Electronic, vol. 101, no. 23: pp. e215-e220, 2000.
[23] S. Das, S. Mukherjee, S. Chatterjee and H. K. Chatterjee, "Noise elimination and ECG R peak detection using wavelet transform," *2016* IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, pp. 1-5, 2016.
[24] E. Çelik *et al*., "DSD-R: Deep Learning Based Segmentation for the Detection of R peaks in ECG Signals," *2022 Medical Technologies Congress (TIPTEKNO)*, Antalya, Turkey, 2022, pp. 1-4.
[25] X. Peng, H. Zhu, X. Zhou, C. Pan and Z. Ke, "ECG Signals Segmentation Using Deep Spatiotemporal Feature Fusion U-Net for QRS Complexes and R-Peak Detection," in IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1-12, 2023, Art no. 2506212.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2024 at 05:52:04 UTC from IEEE Xplore. Restrictions apply.



RUHate-MM: Identification of Hate Speech and Targets using Multimodal Data from Russia-Ukraine Crisis

Surendrabikram Thapa
Virginia Tech
Blacksburg, Virginia, USA
surendrabikram@vt.edu

Farhan Ahmad Jafri
Jamia Millia Islamia
New Delhi, Delhi, India
farhanjafri88888@gmail.com

Kritesh Rauniyar
Delhi Technological University
New Delhi, Delhi, India
rauniyark11@gmail.com

Mehwish Nasim
University of Western Australia
Flinders University
Australia
mehwish.nasim@uwa.edu.au

Usman Naseem
Macquarie University
Sydney, New South Wales, Australia
usman.naseem@mq.edu.au

ABSTRACT

During the conflict between Ukraine and Russia, hate speech targeted toward specific groups was widespread on different social media platforms. With most social platforms allowing multimodal content, the use of multimodal content to express hate speech is widespread on the Internet. Although there has been considerable research in detecting hate speech within unimodal content, the investigation into multimodal content remains insufficient. The limited availability of annotated multimodal datasets further restricts our ability to explore new methods to interpret and identify hate speech and its targets. The availability of annotated datasets for hate speech detection during political events, such as invasions, are even limited. To fill this gap, we introduce a comprehensive multimodal dataset consisting of 20,675 posts related to the Russia-Ukraine crisis, which were manually annotated as either 'Hate Speech' or 'No Hate Speech'. Additionally, we categorize the hate speech data into three targets: 'Individual', 'Organization', and 'Community'. Our benchmarked evaluations show that there is still room for improvement in accurately identifying hate speech and its targets. We hope that the availability of this dataset and the evaluations performed on it will encourage the development of new methods for identifying hate speech and its targets during political events like invasions and wars. The dataset and resources are made available at https://github.com/Farhan-jafri/Russia-Ukraine.

CCS CONCEPTS

• **Information systems** → **World Wide Web**; *Information retrieval*; • **Computing methodologies** → **Natural language processing**.

KEYWORDS

Russia-Ukraine Crisis, Multimodal Data, Content Moderation, Hate Speech

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WWW '24 Companion, May 13–17, 2024, Singapore, Singapore

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0172-6/24/05
https://doi.org/10.1145/3589335.3651973

ACM Reference Format:
Surendrabikram Thapa, Farhan Ahmad Jafri, Kritesh Rauniyar, Mehwish Nasim, and Usman Naseem. 2024. RUHate-MM: Identification of Hate Speech and Targets using Multimodal Data from Russia-Ukraine Crisis. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24 Companion), May 13–17, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3589335.3651973

1 INTRODUCTION

Social media platforms have revolutionized the way people communicate and express their opinions. This has been a double-edged sword, as while social media platforms have strengthened unique opinions and promoted freedom of speech, they have also served as a fertile ground for cyberbullying, hate speech, and offensive language [22]. The unmatched broadcasting ability of social media platforms has made the dissemination of false information or hate speech a major social concern. As a result, several tools are being developed to help users identify harmful or deceptive content, including everything from sexualization and pornography to hate speech and disinformation [14]. Most of the research is focused on developing sophisticated machine learning algorithms to automatically detect and remove harmful content from their platforms. These algorithms use a variety of features, such as linguistic cues, social context, and user behavior, to identify and flag potentially harmful content [13]. In addition to automated tools, social media platforms are also developing policies and guidelines to govern user behavior and content creation Bhandari et al. [3]. These policies and guidelines are designed to promote responsible and respectful communication on social media platforms, while also protecting users from harmful content. The worldwide prevalence and growing significance of tweets have led to increasing studies on tweets. Several studies have been conducted to investigate the use of hate speech and other forms of offensive language on Twitter, with the aim of identifying strategies to curb their prevalence. Most of the research in tweets today focuses on a few key categories, such as ethnicity, sex, and religion. There are also several types of assaults where the target is mocked and dehumanized [20] that need a good focus of research.

On 24 February 2022, Russia launched a full-scale invasion of Ukraine by land, sea, and air, leading to a polarized response from the international community, with some supporting the invasion and others opposing it [38]. This led to widespread condemnation of



(a) No Hate **(b) Hate** **(c) Tar: Individual** **(d) Tar: Organization** **(e) Tar: Community**

Figure 1: Examples of tweets during Russia-Ukraine Crisis: *Task A*: Hate speech detection - Figure (a - 'No Hate' and b - 'Hate') and *Task B*: Target (Tar) detection - Figure (c - 'Individual', d - 'Organization', and e - 'Community')

the war and sanctions imposed on Russia. As a result, social media became a hub of activity as people expressed their opinions on the humanitarian and economic crisis caused by the invasion. Despite respectful discourse and discussion, there was also a significant amount of hateful content (as shown in Figure 1) targeting various groups of people [5].

Hate speech can have severe consequences for society, and the microblogging and social media platforms put in a significant effort to manage it, mostly through human moderators [21]. However, in political events such as invasions, the volume of hate speech can become overwhelming for human moderators to flag and regulate effectively [3, 31]. Therefore, there is a pressing need for an automated system to identify and manage such content [37]. Thus, to tackle this challenge, we propose a new dataset and benchmark them with multiple state-of-the-art algorithms. Our main contributions are as follows:

- We create and release a new large-scale dataset called **RUHate-MM**, which contains 20,675 manually annotated tweets to identify hate speech and their targets during the Russia and Ukraine crisis.
- We conduct a preliminary analysis of the data. We have set benchmarks with several state-of-the-art textual, visual, and multimodal models.
- Our experiments analysis shows that to address issues of hate speech, an integration of multimodal inputs is important.

The structure of our manuscript is as follows. Section 2 offers an overview of related work. In Section 3, we introduce the dataset, along with annotation guidelines, and present a comprehensive statistical analysis. Section 4 provides detailed information on the experimental methodology. Moving on to Section 5, we present the results and conduct a thorough analysis. Finally, Section 6 serves as the conclusion of the study.

2 RELATED WORKS

In previous studies, different aspects of mischief, such as offensiveness, false news, and hateful tweets, have been investigated separately [22]. Waseem and Hovy [36] annotated 16K tweets-dataset to classify a given tweet as sexist, racist, or neither. The aspect-based annotation is useful for understanding the theme of hate speech. Fortuna et al. [8] introduced a dataset comprising 5,668 tweets that were meticulously annotated across 81 categories of hate speech in the Portuguese language. Concurrently, Pereira-Kohatsu et al. [24] curated a dataset consisting of 6,000 Spanish tweets specifically addressing hate speech. Additionally, they provided an unlabeled corpus containing a substantial 2 million tweets. Recognizing the profound impact of political events on societal perspectives, it becomes imperative to construct datasets that encapsulate a pertinent political context. Kumar and Pranesh [15] addressed this need by presenting TweetBLM, a dataset intricately linked to the Black Lives Matter (BLM) movement. This dataset underwent manual annotation for the identification of hate speech. In a similar vein, Grimminger and Klinger [9] contributed to the landscape by introducing a dataset focused on the 2020 US elections. Their dataset comprised 3,000 tweets, each categorized based on its political stance toward a candidate. Moreover, the tweets were further classified into offensive and non-offensive categories, providing a nuanced understanding of the political discourse surrounding the election.

These studies hold a high importance in studying and understanding hate speech, bullying, and different related tasks. However, these studies focus on unimodal textual data. It is evident that textual data alone cannot provide the information a post wants to convey [2]. Nowadays, most social networks support multimodal content, and therefore it has become extremely important to tackle multimodal hate speech. Kiela et al. [14] proposed multimodal data for detecting hate speech in multimodal memes. Apart from telling if a meme is harmful or not, it is extremely important to identify who is being targeted in the memes. Annotating targets in addition to hate speech can provide valuable insights and help in understanding the nature and scope of hate speech. Sharma et al. [28] developed a meme dataset that contained 3,552 memes about US politics. The annotations of the memes in the dataset include information on whether or not the meme is harmful and the specific group, organization, community, or general public that it targets. The authors also propose DISARM, a multimodal neural architecture for detecting harmful targeting in memes. Similarly, Pramanick et al. [25] presented HarMeme, which is the dataset of 3,544 memes related to the COVID-19 pandemic. The annotation was done for the harm level as well as targets.

Despite the importance of political events, limited work has been done to address hate speech during these events such as the

Russia-Ukraine conflict. Hasan et al. [10] assembled a dataset comprising 10,861 Bengali comments discussing the Russia-Ukraine crisis on YouTube news channels. These comments were systematically categorized into three groups: 'Pro-Ukraine,' 'Pro Russia,' and 'Neutral.' In a parallel effort, Toraman et al. [33] meticulously curated a dataset featuring 5,284 English and 5,064 Turkish tweets. These tweets centered around contemporary issues such as the Russia-Ukraine war, COVID-19, and Refugees, with a specific focus on analyzing the spread of misinformation within the tweeted content. These efforts are important efforts in understanding Russia-Ukraine war sentiments. However, these datasets contain text-only information. The current social media users tend to use image along with text in their tweets. Such posts can propagate hate more easily as they can be more visually appealing than text-only tweets. To address this problem, Thapa et al. [32] presented a data set related to the detection of hate speech during the Russia-Ukraine crisis. They annotated 5,680 text-image pairs related to the crisis which were collected from Twitter. However, the authors only address the problem of hate speech detection but detection of intended targets of hate speech is also an important task which has received limited research interest. Apart from this, authors only annotate around 5,000 tweets which is a small dataset.

The significance of political events, coupled with the harmful consequences of hate speech, underscores the need for the development of hate speech detection tools for such contexts. Hate speech detection and target detection during political events such as the Russia-Ukraine war are necessary to understand the dynamics of hate speech, protect vulnerable groups, develop targeted interventions, and understand and assess the impact of hate speech on society. Furthermore, such analyses hold significant potential for enhancing the understanding of hate speech within the context of wargames, where the dynamics of conflict and the role of information warfare are increasingly complex. Table 1 shows a detailed comparison of previous work in the literature. As seen in the table, our study fills the gap of studying multimodal hate speech and related targets during political events such as the Russia-Ukraine war. To this end, we investigate hateful posts related to the Russia-Ukraine crisis and address two main tasks: (i) Task A: Detecting hateful posts – determining if a post is hateful or not. (ii) Task B: Identification of targets of hateful posts – identifying whether a given hateful post targets an individual, an organization, or a community.

3 DATASET

3.1 Data Collection

The development of the events started when Russia started to attack Ukraine on 22 February 2022. Since we were interested in the tweets around this event, we started to crawl tweets from 22 February 2022 to 28 March 2022 using Twitter API[1]. We collected the tweets with certain list of keywords namely *ukraine*, *putin*, *russia*, *zelensky*, *kyiv*, *kiev*, *kremlin*, *ukrainian*, *nato*, *russian*, *soviet*, *moscow*, *kharkiv*, and *donbas*. The tweets for keywords *kharkiv*, and *donbas* were collected from 1 March 2022 whereas, for all other keywords, tweets were collected starting from 22 February 2022. The tweets revolving around the Russia-Ukraine crisis had these keywords mentioned very frequently. The keywords *kharkiv*, and *donbas* were collected from 1 March 2022 as the crisis in those regions started later than 22 February 2023. To filter the data and ensure we collect appropriate data for our study, We included tweets in the dataset that contained media and were written in English. We eliminated tweets that had media in the form of videos or animations. Similarly, If the annotators do not find the tweet to be relevant, it was labeled as 'Non-Informative,' and such tweets were later dropped. The criteria for filtering tweets are given in section 3.2. The resulting dataset comprises 20,675 labeled tweets, each containing an image and text along with annotations. All tweets had unique tweet IDs and hence there are no duplicate data.

[1]https://developer.twitter.com/en/docs/twitter-api

3.2 Selection of image-text tweets

The tweet (image-text) selection procedure was quite straightforward. The dataset contained a collection of images that included aspects of hate, such as making derogatory comments about specific individuals, targeting a company, and raising issues relevant to the country. The dataset also contained cartoon photos that made fun of politicians, nations, and organizations. Images that were just text-oriented or blurry were not taken into consideration. We were focusing on the English language so the images that majorly included text in other languages such as Russian or Ukrainian were discarded. More concretely, the criteria for dropping data are as follows:

- The tweet contains non-textual elements such as videos, gifs, or audio recordings which do not provide any relevant information about the Russia-Ukraine conflict. We have only considered image as a media.
- The tweet does not have any context. The image and text do not contain any relevant information about the Russia-Ukraine conflict.
- The tweet has a language other than English. Since the tweets revolve around the Russia-Ukraine conflict, there were tweets in Russian and Ukrainian language as well.
- The tweet has a blurry image from which no relevant information can be gained. Such tweets are discarded on a case-by-case basis.
- The tweet contains only a link to an external website or a picture that does not provide any relevant information about the Russia-Ukraine conflict.
- The tweet has more than one image. For the purpose of creating this multimodal dataset, we take tweets with only one image as media to create a text-image pair from a given tweet.

Figure 2 shows some of the images that are filtered. In Figure 2 (a), it can be seen that the text is in the Russian language. Similarly, Figure 2 (b) has a blurred image, and hence no information can be obtained from the tweet. We hence discard such tweets.

3.3 Data Annotation:

Annotation of data was done in two folds, first, as Task A, we annotated the data for binary classes viz. 'Hate' vs 'Non-Hate'. In the second fold, as task B, we annotated the hate speech for target

Table 1: Summary of related datasets in the literature

Work	Year	Data Source	Multimodal	Sub-classes	Size	Context
Waseem and Hovy [36]	2016	Twitter	✗	✗	16,000	✗
Pereira-Kohatsu et al. [24]	2019	Twitter	✗	✗	6,000	✗
Kumar and Pranesh [15]	2021	Twitter	✗	✗	9,165	BLM Movement
Kiela et al. [14]	2020	Self-generated	✓	✗	8,000	✗
Hasan et al. [10]	2023	YouTube	✗	✗	10,861	Russia-Ukraine Crisis
Toraman et al. [33]	2022	Twitter	✗	✗	10,348	Russia-Ukraine Crisis
Grimminger and Klinger [9]	2021	Twitter	✗	✓	3,000	2020 U.S. Election
Thapa et al. [32]	2022	Twitter	✓	✗	5,680	Russia-Ukraine Crisis
Sharma et al. [28]	2022	Online sources	✓	✓	3,552	U.S. Politics
Pramanick et al. [25]	2021	Online sources	✓	✓	3,544	COVID-19
RuHateMM (Ours)	2024	Twitter	✓	✓	20,675	Russia-Ukraine War



(a) Non-English Language (b) Blurred Image (c) Blurred Image (d) No useful Information

Figure 2: Examples of discarded tweet images

labels. For the target labels, we annotated data into three labels viz. 'Community', 'Organization', and 'Individual'. As the name suggests, the class 'Hate' pertains to tweets that are offensive and contains hateful content such as personal attacks, homophobic abuse, racial abuse, or attack on the minority. The 'Non-Hate' class has tweets that report the events objectively and have no offensive or hateful content. Further details on annotation are discussed in Section 3.4. Similarly, for the annotation of targets, we defined targets as the following:

- **Community**: A community is a group of individuals who share common interests, beliefs, or characteristics and interact with one another. It can be defined as a social unit that shares a sense of identity, purpose, and values.
- **Organization**: An organization is a structured group of individuals, created to achieve a specific goal or set of goals. Examples of organizations are 'Republican Party', 'NATO', 'United Nations', etc.
- **Individual**: It refers to a person as an autonomous entity. In the context of our dataset, some of the most frequently noted individuals are 'Putin', 'Trump', 'Biden', etc.

Our team of annotators consisted of four individuals, both male and female, with diverse educational qualifications, including undergraduate, MS, and Ph.D. degrees, as well as researchers with experience in NLP and data collection. All the annotators had at least 10 years of formal English education and were familiar with Russia-Ukraine crisis. This diversity in qualifications helped to ensure clear instructions and maintain a high standard of annotations. The use of diverse annotators is an important aspect of data annotation in order to minimize potential biases [34]. In our study, the diversity in the annotator's background was intended to mitigate any potential bias in the annotation process. The manual annotation process was time-consuming and required significant effort from the annotators, but it helped to ensure a high level of accuracy in the annotations. The use of clear guidelines and regular quality checks were also employed to maintain consistency and reliability in the annotation process.

Three-Phase Annotation: Given the difficulty of labeling tweets that have both text and images, we implemented a 3-phase annotation process as explained in Section 3.4.1.

3.4 Annotation Schema

Clear-cut annotations are crucial in order to ensure that the dataset is labeled consistently and accurately. This is important as the results of any analysis or model development will be based on the labeled data. If the annotations are inconsistent or inaccurate, it can

lead to inaccurate or unreliable results. Clear-cut annotations also ensure that the dataset is representative of the underlying phenomenon being studied and that any conclusions drawn from the data are valid. Thus, we follow a 3-phase annotation. To quantitatively assess the inter-annotator agreement, we have used Cohen's Kappa (κ) (for agreement between two annotators) and Fleiss' Kappa (for agreement between multiple annotators) as our inter-rater agreement measure. We collectively refer to them as Kappa in further sections.

3.4.1 ***3-Phase Annotation***. For annotation of the data, an instruction set was created. The instructions were revised iteratively until all the annotators were clear about the instructions. To ensure that the annotation instructions are clear, we follow a three-phase annotation. The first phase involved a pilot run to make sure everyone understood the annotation. The second phase ensured that the revised instructions after the first phase were clear enough. The third phase eliminated the conflicts in the annotation.

- **Pilot Run:** As the first phase of the annotations, we run a pilot annotation for 50 tweets to ensure that everyone understood the annotation instructions. This is important as the task of labeling tweets can be challenging and it is important that everyone has the same understanding of what constitutes hate speech. There was some confusion among the annotators. Some annotators demanded annotation to be clear and we revised the instructions. The annotation instructions were then revised to address all the confusion.
- **Revised Instructions:** The second phase of annotation of 200 tweets by all four annotators was done to ensure that the instructions revised after the first stage were clear enough. During this phase, the annotators were given the revised instructions and asked to annotate the tweets. This phase was important to confirm that the revised instructions were clear and that the annotators were able to consistently identify hate speech.
- **Conflict Resolution:** The third phase was a group discussion of conflicts in the second phase of annotation, during which the annotators discussed any discrepancies in their annotations and reached a consensus. This phase was important as it helped to resolve any disagreements and ensure that all the tweets were labeled consistently. The group discussion also helped to make the instructions more apparent and provided an opportunity to identify any further ambiguities or inconsistencies in the instructions.

  The annotation guidelines is given in detail in Appendix A.1.

3.5 Inter-Annotator Agreement and Statistics:

To calculate the inter-annotator agreement, we have used Kappa (κ). Kappa is a statistical measure of inter-rater agreement for categorical items. It is a more robust measure of inter-rater agreement than simple percent agreement calculation, as it takes into account the agreement that can be expected by chance [35]. It is used to measure the level of agreement between two or more annotators on a categorical item. The coefficient ranges from -1 to 1, with values close to 1 indicating strong agreement and values close to -1 indicating strong disagreement. A value of 0 represents chance agreement. It's a widely used measure in social science and medical research. For our annotations, the inter-annotator agreement, Fleiss' Kappa for the **Task A** i.e. 2-class annotation of 'Hate' vs 'Non-Hate' ($\kappa_{2-Class}$) is 0.74. Similarly, Fleiss' Kappa for **Task B** i.e. 3-class target annotation ($\kappa_{3-Class}$) is 0.69. The value of Cohen's Kappa among different annotators for different stages of annotation is given in Table 2.

Table 2: Cohen's Kappa (κ) for annotation during different Phases by four annotators

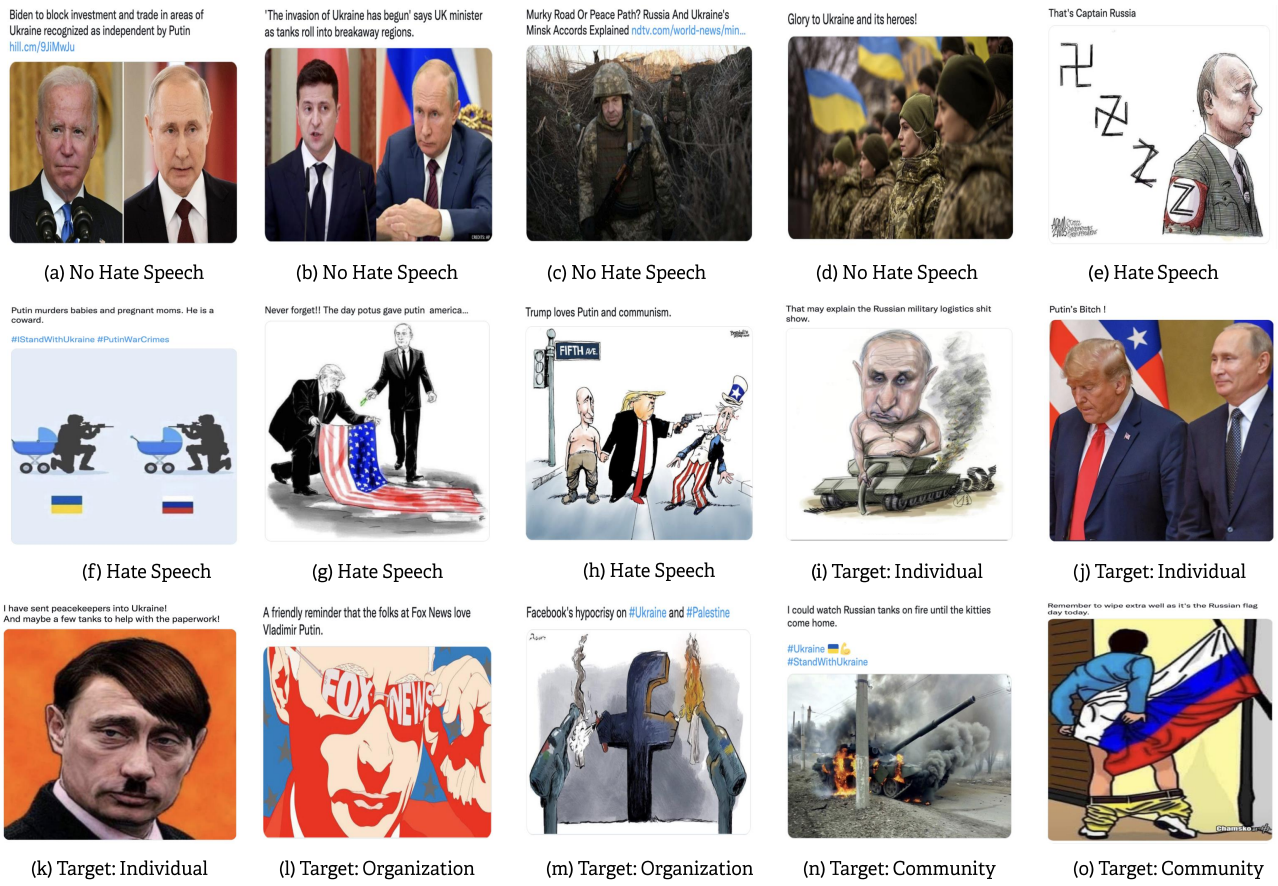
Phase	Annotators	$\kappa_{2-class}$	$\kappa_{3-class}$
Pilot Annotation	α_1 and α_2	0.49	0.53
	α_1 and α_3	0.50	0.51
	α_1 and α_4	0.55	0.56
	α_2 and α_3	0.47	0.52
	α_2 and α_4	0.57	0.60
	α_3 and α_4	0.50	0.49
Final Annotation	α_1 and α_2	0.76	0.74
	α_1 and α_3	0.78	0.76
	α_1 and α_4	0.79	0.76
	α_2 and α_3	0.80	0.73
	α_2 and α_4	0.73	0.71
	α_3 and α_4	0.81	0.73

3.6 Dataset Statistics

Our new multi-modal dataset, RUHate-MM includes 20,675 tweets, with 4,222 (20.33%) tweets being labeled as 'hate speech' label whereas 16,543 (79.67%) tweets are labeled as 'no hate' label (Table 3). The dataset statistics represent a true distribution in a real-world scenario where many posts are neutral, and only some are related to hate speech. The hate speech tweets were further annotated into three broad categories of targets. The targets are individuals, organizations, and communities. Among 4,222 hate speech tweets, 2,402 targeted individuals, 918 of them targeted organizations whereas the remaining 902 posts targeted specific communities. The statistics of data are shown in table 3. It also provides the average number of characters and words per tweet for each of the categories. These statistics can be used to gain an understanding of the overall composition and size of the dataset. The values in parentheses are calculated after preprocessing of text, which includes cleaning and normalizing the text data. Further details of text preprocessing can be found in Section 4.3. It can further be observed that the average character counts for hate and non-hate speech is somewhat distinctive. Figure 1 shows examples of annotated tweets whereas Figure 3 shows some more examples of targeted hate speech.

3.7 Exploratory Data Analysis

Table 4 presents the top 10 most frequent words in each class of the dataset along with their TF-IDF scores. TF-IDF (term frequency-inverse document frequency) is a statistical measure that is used to



(a) No Hate Speech (b) No Hate Speech (c) No Hate Speech (d) No Hate Speech (e) Hate Speech

(f) Hate Speech (g) Hate Speech (h) Hate Speech (i) Target: Individual (j) Target: Individual

(k) Target: Individual (l) Target: Organization (m) Target: Organization (n) Target: Community (o) Target: Community

Figure 3: A few examples from our dataset. The tweets are annotated as hate and non-hate for *Task A*. Hate tweets are further annotated into targets (Tar) as an individual, organization, and community for *Task B*

Table 3: Dataset Statistics for RUHate-MM. The values in parentheses in average characters per tweet (Avg. Char) and average words per tweet (Avg. words) are calculated after preprocessing of text.

Problem	Labels	Tweets	Avg. Char	Avg. words
Hate Speech	Hate	4,222	54.56 (46.46)	9.94 (9.63)
	Non-Hate	16,543	56.14 (51.63)	10.22 (10.12)
Targets	Individual	2,402	53.80 (45.62)	9.87 (9.57)
	Organization	918	56.85 (48.31)	10.14 (9.84)
	Community	902	54.23 (46.79)	9.92 (9.58)

evaluate the importance of a word in a document or a collection of documents. The fundamental principle behind TF-IDF involves assigning weights to words based on their frequency within a document and their rarity across the entire corpus. The TF-IDF score for a word in a document is calculated as the product of its term frequency (TF) and inverse document frequency (IDF) scores. Words with higher TF-IDF scores are deemed more significant and relevant to the document in which they appear. This significance arises from their frequency within the document and their scarcity across the broader collection of documents.

Analyzing the presented table 4, it becomes evident that certain prominent personalities, such as Putin, Trump, Biden, etc., are consistently targeted more frequently in individual categories. TF-IDF aids in highlighting these key terms by emphasizing their importance within specific documents while considering their rarity across the entire dataset. This approach enables a more nuanced understanding of the distinctive features and themes associated with each class, shedding light on the salient words that contribute significantly to the characterization of each category.

4 EXPERIMENTAL SETUP

In this section, we describe our experimental setup and baselines.

4.1 Experimental Settings

We used the pre-trained models for each baseline and used the F1-score as the evaluation metric. We train all the models using Tensorflow and Pytorch on a Tesla T4 or Tesla V100 GPU, with 32

Table 4: Top-10 most frequent words in each class. The TF-IDF scores are given for each words.

All Posts		Hate Speech Posts		HS: Target Individual		HS: Target Organization		HS: Target Community	
Words	TF-IDF	Words	TF-IDF	Words	TF-IDF	Words	TF-IDF	Words	TF-IDF
ukraine	0.3463	putin	0.3274	putin	0.4524	ukraine	0.2284	ukraine	0.2997
russia	0.1659	ukraine	0.2025	ukraine	0.1449	putin	0.1684	russia	0.2161
putin	0.1629	russia	0.1458	russia	0.1140	russia	0.1531	russian	0.1883
russian	0.1211	russian	0.1320	russian	0.0931	russian	0.1529	putin	0.1030
ukrainian	0.0812	trump	0.0488	trump	0.0746	nato	0.1027	ukrainian	0.0694
war	0.0505	war	0.0459	biden	0.0433	ukrainian	0.0572	war	0.0587
kyiv	0.0333	ukrainian	0.0429	war	0.0406	nazis	0.0554	fuck	0.0333
people	0.0245	nato	0.0305	like	0.0259	war	0.0524	people	0.0323
says	0.0241	biden	0.0287	ukrainian	0.0242	nazi	0.0438	just	0.0238
biden	0.0216	just	0.0241	fuck	0.0234	nestle	0.0191	embassy	0.0186

GB dedicated memory. For the unimodal text models, we import all the pre-trained models of transformers from the hugging-face library. Similarly, for the visual models, we import all the pre-trained models from the PyTorch Image Models library (timm) [23]. All the models we experimented with used Adam optimizer [4]. The hyper-parameters that can be supplemental to reproduce the experiments are given in Table 5.

4.2 Baselines

We established baselines using various techniques, including both unimodal and multimodal methods.

Unimodal Models: We used the following unimodal methods:

- **Textual Unimodal:** For textual models, we used Bidirectional Encoder Representations from Transformers (BERT) [7], DistilBERT (a distilled version of BERT) [27], optimized variant of BERT, i.e., RoBERTa [18] and Albert (A Lite BERT for Self-supervised Learning of Language Representations) [16].
- **Visual Unimodal:** For the image-based unimodal baseline methods, we used 4 pretrained methods viz. DenseNet [12], Visformer [6], Improved Multiscale Vision Transformers for Classification and Detection (MVITV2) [17] and VGG19 [29].

Multimodal Models: We employed 3 multimodal models that have been widely used in earlier research involving hate speech classification. The first is a combination of ResNet [11] and BERT, where we first trained ResNet and BERT on text and image data respectively, and then merged their representations using a linear layer. We also implemented the state-of-the-art model CLIP (Contrastive Language-Image Pre-training) in our study, which has shown remarkable performance in a wide range of vision-and-language tasks [26]. We also utilized ViLBERT (Visual-Linguistic BERT) in our study, which is a pre-trained transformer-based model that is specifically designed to handle vision-and-language tasks [19].

4.3 Text Preprocessing

Text preprocessing is important in any Natural Language Processing (NLP) task [1, 30]. In order to prepare the tweet text for further analysis, a preprocessing step was conducted to remove various non-alphanumeric elements, including but not limited to special characters, hyperlinks, mentions, and emojis. Special characters, such as punctuation marks and other symbols, were removed as they can add noise to the data and potentially impact the accuracy of subsequent analysis. Hyperlinks and mentions, which reference external web pages or other users respectively, were removed as they are not relevant to the content of the tweet itself and can also introduce noise into the data. Emojis, while commonly used in social media, were also removed as they are not part of the standard character set used in most natural language processing techniques, and thus could cause errors or inaccuracies in downstream analysis. Overall, the preprocessing step serves to clean and standardize the text data for further analysis, while ensuring that only relevant and meaningful content is retained.

5 RESULTS AND ANALYSIS

In this section, we report the performance of various models on both tasks viz. hate speech classification and target classification.

5.1 Unimodal Baseline Results

Table 6 summarizes the performance of different unimodal algorithms on our dataset in terms of $F1_{HateSpeech}$ (f1-score in hate speech detection) and $F1_{Target}$ (f1-score in target identification). The results for the unimodal textual models (BERT, DistilBERT, ROBERTa, and Albert) demonstrate competitive performance in hate speech detection, with F1 scores ranging from 0.749 to 0.798. Among these, BERT exhibits the highest F1 score, indicating its effectiveness in identifying hate speech instances.

In target identification, the unimodal textual models also showcase notable performance, with $F1_{Target}$ scores ranging from 0.644 to 0.679. BERT again emerges as the leading model in target identification.

For unimodal visual models (DenseNet-161, Visformer_small, MVITV2_base, and VGG19), the results in hate speech detection (ranging from 0.758 to 0.774) and target identification (ranging from 0.598 to 0.628) demonstrate a competitive performance across the board. DenseNet-161 achieves the highest F1-score in both hate speech detection and target identification among the visual models.

Table 5: Implementation Details of the Experiments

Modality	Models	Batch Size	Epochs	Learning Rate	Parameters	Image Encoder	Text Encoder
Textual	BERT	16	3	5×10^{-5}	110M	-	bert-base-uncased
	DistilBERT	16	3	5×10^{-5}	67M	-	distilbert-base-uncased
	RoBERTa	16	3	5×10^{-5}	125M	-	roberta-base
	ALBERT	16	3	5×10^{-5}	12M	-	albert-base-v2
Visual	DenseNet-161	16	5	10^{-5}	26.5M	densenet161	-
	Visformer_small	16	5	10^{-5}	39.5M	visformer_small	-
	MVITV2_base	16	5	10^{-5}	50.7M	mvitv2_base	-
	VGG19	16	5	10^{-5}	139.6M	vgg19	-
Multimodal	ResNet + BERT	16	5	10^{-3}	172M	ResNet-152	Bert-base-uncased
	CLIP	4	6	10^{-3}	63M	ViT-Large-Patch14	
	ViLBERT-CC	16	5	10^{-3}	112M	FasterRCNN	Bert-base-uncased

5.2 Multimodal Baseline Results

The performance of multimodal algorithms on our dataset is presented in Table 6, focusing on $F1_{HateSpeech}$ and $F1_{Target}$. The multimodal models, combining ResNet with BERT, CLIP, and ViLBERT-CC, exhibit enhanced performance compared to unimodal models.

In hate speech detection, the multimodal models achieve higher $F1_{HateSpeech}$ scores compared to their unimodal counterparts. Particularly, ViLBERT-CC stands out with the highest $F1_{HateSpeech}$ score of 0.848, indicating its superior ability in identifying hate speech instances when leveraging both textual and visual modalities.

Similarly, in target identification, multimodal models outperform unimodal models, emphasizing the effectiveness of leveraging both textual and visual information. ViLBERT-CC leads in $F1_{Target}$ with a score of 0.741, demonstrating its proficiency in identifying diverse targets within hate speech instances.

These results underscore the advantages of multimodal approaches, showcasing improved performance in hate speech detection and target identification compared to unimodal models on our dataset. The combination of textual and visual information proves to be a promising avenue for advancing the field of hate speech detection during politically charged events.

Table 6: Performance of different unimodal and multimodal algorithms on our dataset

Modality	Model	$F1_{HateSpeech}$	$F1_{Target}$
Unimodal Textual	BERT	0.798	0.679
	DistilBert	0.787	0.668
	ROBERTa	0.780	0.668
	Albert	0.749	0.644
Unimodal Visual	DenseNet-161	0.768	0.628
	Visformer_small	0.762	0.611
	MVITV2_base	0.774	0.625
	VGG19	0.758	0.598
Multimodal	ResNet + BERT	0.806	0.700
	CLIP	0.826	0.719
	ViLBERT-CC	0.848	0.741

5.3 Analysis

Multimodal models, leveraging both textual and visual information, consistently outperform unimodal models in hate speech detection and target identification. ViLBERT-CC particularly shines, showcasing superior performance and emphasizing the synergistic benefits of combining both modalities. This suggests a promising avenue for future research, highlighting the potential of multimodal approaches in enhancing the discernment of hate speech instances and the identification of diverse targets during complex socio-political events.

The benchmarked evaluations reveal areas for improvement in accurately identifying hate speech and its targets, indicating the intricate nature of the dataset. This opens avenues for refining existing models and developing novel approaches to address the unique challenges presented by hate speech during geopolitical crises. The release of our multimodal dataset contributes a valuable resource to the research community, fostering further exploration and innovation in hate speech detection. The dataset's annotations, particularly the categorization of hate speech targets, offer a nuanced understanding of the manifestations of hate speech during the Russia-Ukraine crisis.

6 CONCLUSION

This paper introduces a new multi-modal dataset for identifying hateful content on social media, consisting of 20,675 text-image pairs collected from Twitter, labeled for hate speech and their targets. The experimental analysis of the presented dataset has shown that understanding both text and image modalities is crucial for detecting hateful content. In future work, we plan to develop new multi-modal models specifically for hate-speech detection, with the goal of gaining a deeper understanding of the relationship between text and images. Another potential area of research could be expanding the dataset to include more languages and different types of social media platforms. Additionally, it would be interesting to explore the hate speech detection possibilities for more specific or nuanced targets.

REFERENCES

[1] Surabhi Adhikari, Surendrabikram Thapa, Usman Naseem, Priyanka Singh, Huan Huo, Gnana Bharathy, and Mukesh Prasad. 2022. Exploiting linguistic information from Nepali transcripts for early detection of Alzheimer's disease using natural language processing and machine learning techniques. *International Journal of Human-Computer Studies* 160 (2022), 102761.

[2] Firoj Alam, Stefano Cresci, Tanmoy Chakraborty, Fabrizio Silvestri, Dimiter Dimitrov, Giovanni Da San Martino, Shaden Shaar, Hamed Firooz, and Preslav Nakov. 2022. A Survey on Multimodal Disinformation Detection. In *Proceedings of the 29th International Conference on Computational Linguistics*. 6625–6643.

[3] Aashish Bhandari, Siddhant B Shah, Surendrabikram Thapa, Usman Naseem, and Mehwish Nasim. 2023. CrisisHateMM: Multimodal Analysis of Directed and Undirected Hate Speech in Text-Embedded Images From Russia-Ukraine Conflict. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1993–2002.

[4] Sebastian Bock, Josef Goppold, and Martin Weiß. 2018. An improvement of the convergence proof of the ADAM-Optimizer. *arXiv preprint arXiv:1804.10587* (2018).

[5] Jean-Christophe Boucher. 2022. Disinformation and Russia-Ukrainian War on Canadian Social Media. *The School of Public Policy Publications* 15, 1 (2022).

[6] Zhengsu Chen, Lingxi Xie, Jianwei Niu, Xuefeng Liu, Longhui Wei, and Qi Tian. 2021. Visformer: The vision-friendly transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 589–598.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[8] Paula Fortuna, Joao Rocha da Silva, Leo Wanner, Sérgio Nunes, et al. 2019. A hierarchically-labeled portuguese hate speech dataset. In *Proceedings of the third workshop on abusive language online*. 94–104.

[9] Lara Grimminger and Roman Klinger. 2021. Hate Towards the Political Opponent: A Twitter Corpus Study of the 2020 US Elections on the Basis of Offensive Speech and Stance Detection. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. 171–180.

[10] Mahmud Hasan, Labiba Islam, Ismat Jahan, Sabrina Mannan Meem, and Rashedur M Rahman. 2023. Natural Language Processing and Sentiment Analysis on Bangla Social Media Comments on Russia–Ukraine War Using Transformers. *Vietnam Journal of Computer Science* (2023), 1–28.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[12] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.

[13] Farhan Ahmad Jafri, Mohammad Aman Siddiqui, Surendrabikram Thapa, Kritesh Rauniyar, Usman Naseem, and Imran Razzak. 2023. Uncovering Political Hate Speech During Indian Election Campaign: A New Low-Resource Dataset and Baselines. (2023).

[14] Douwe Kiela, Hamed Firooz, Aravind Mohan, Vedanuj Goswami, Amanpreet Singh, Pratik Ringshia, and Davide Testuggine. 2020. The hateful memes challenge: Detecting hate speech in multimodal memes. *Advances in Neural Information Processing Systems* 33 (2020), 2611–2624.

[15] Sumit Kumar and Raj Ratn Pranesh. 2021. Tweetblm: A hate speech dataset and analysis of black lives matter-related microblogs on twitter. *arXiv preprint arXiv:2108.12521* (2021).

[16] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942* (2019).

[17] Yanghao Li, Chao-Yuan Wu, Haoqi Fan, Karttikeya Mangalam, Bo Xiong, Jitendra Malik, and Christoph Feichtenhofer. 2022. MViTv2: Improved Multiscale Vision Transformers for Classification and Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4804–4814.

[18] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).

[19] Jiasen Lu, Dhruv Batra, Devi Parikh, and Stefan Lee. 2019. Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. *Advances in neural information processing systems* 32 (2019).

[20] Lambert Mathias, Shaoliang Nie, Aida Mostafazadeh Davani, Douwe Kiela, Vinodkumar Prabhakaran, Bertie Vidgen, and Zeerak Waseem. 2021. Findings of the WOAH 5 shared task on fine grained hateful memes detection. In *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*. 201–206.

[21] Sünje Paasch-Colberg and Christian Strippel. 2022. "The Boundaries are Blurry…": How Comment Moderators in Germany See and Respond to Hate Comments. *Journalism Studies* 23, 2 (2022), 224–244.

[22] Anil Singh Parihar, Surendrabikram Thapa, and Sushruti Mishra. 2021. Hate Speech Detection Using Natural Language Processing: Applications and Challenges. In *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 1302–1308.

[23] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32 (2019).

[24] Juan Carlos Pereira-Kohatsu, Lara Quijano-Sánchez, Federico Liberatore, and Miguel Camacho-Collados. 2019. Detecting and monitoring hate speech in Twitter. *Sensors* 19, 21 (2019), 4654.

[25] Shraman Pramanick, Dimitar Dimitrov, Rituparna Mukherjee, Shivam Sharma, Md Akhtar, Preslav Nakov, Tanmoy Chakraborty, et al. 2021. Detecting harmful memes and their targets. *arXiv preprint arXiv:2110.00413* (2021).

[26] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*. PMLR, 8748–8763.

[27] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108* (2019).

[28] Shivam Sharma, Md Akhtar, Preslav Nakov, Tanmoy Chakraborty, et al. 2022. DISARM: Detecting the victims targeted by harmful memes. *arXiv preprint arXiv:2205.05738* (2022).

[29] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[30] Surendrabikram Thapa, Surabhi Adhikari, Usman Naseem, Priyanka Singh, Gnana Bharathy, and Mukesh Prasad. 2020. Detecting Alzheimer's disease by exploiting linguistic information from Nepali transcript. In *Neural Information Processing: 27th International Conference, ICONIP 2020, Bangkok, Thailand, November 18–22, 2020, Proceedings, Part IV 27*. Springer, 176–184.

[31] Surendrabikram Thapa, Farhan Jafri, Ali Hürriyetoğlu, Francielle Vargas, Roy Ka-Wei Lee, and Usman Naseem. 2023. Multimodal hate speech event detection-shared task 4, case 2023. In *Proceedings of the 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text*. 151–159.

[32] Surendrabikram Thapa, Aditya Shah, Farhan Ahmad Jafri, Usman Naseem, and Imran Razzak. 2022. A multi-modal dataset for hate speech detection on social media: Case-study of russia-ukraine conflict. In *CASE 2022-5th Workshop on Challenges and Applications of Automated Extraction of Socio-Political Events from Text, Proceedings of the Workshop*. Association for Computational Linguistics.

[33] Cagri Toraman, Oguzhan Ozcelik, Furkan Şahinuç, and Fazli Can. 2022. Not Good Times for Lies: Misinformation Detection on the Russia-Ukraine War, COVID-19, and Refugees. *arXiv preprint arXiv:2210.05401* (2022).

[34] Francielle Vargas, Isabelle Carvalho, Fabiana Rodrigues de Góes, Thiago Pardo, and Fabrício Benevenuto. 2022. HateBR: A Large Expert Annotated Corpus of Brazilian Instagram Comments for Offensive Language and Hate Speech Detection. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. 7174–7183.

[35] Matthijs J Warrens. 2015. Five ways to look at Cohen's kappa. *Journal of Psychology & Psychotherapy* 5, 4 (2015), 1.

[36] Zeerak Waseem and Dirk Hovy. 2016. Hateful symbols or hateful people? predictive features for hate speech detection on twitter. In *Proceedings of the NAACL student research workshop*. 88–93.

[37] P William, Ritik Gade, Rup esh Chaudhari, AB Pawar, and MA Jawale. 2022. Machine Learning based Automatic Hate Speech Recognition System. In *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*. IEEE, 315–318.

[38] Xi-Yin Zhou, Gang Lu, Zhicheng Xu, Xiaoqing Yan, Soon-Thiam Khu, Junfeng Yang, and Jian Zhao. 2023. Influence of Russia-Ukraine War on the Global Energy and Food Security. *Resources, Conservation and Recycling* 188 (2023), 106657.

A APPENDIX

A.1 Annotation Guidelines

A.1.1 Hate Speech. A post (text or image or both) that contains hateful content such as a personal attack, homophobic abuse, racial abuse, or attack on minorities.

- **Targeted language:** Hate speech during political events often targets specific groups based on their political beliefs or affiliations. This can include language that demeans, degrades, or dehumanizes a particular political group or individual.

- **Hostility and aggression:** Hate speech during political events often expresses hostility or aggression towards a particular political group or individual. This can include language that promotes or glorifies violence or hatred against a particular political group or individual.
- **Use of memes and images:** Hate speech during political events often uses memes and images to disseminate harmful messages that are intended to demean, degrade or dehumanize a particular political group or individual.

Further, it is important to note that sarcasm and political satire can be used to express hate speech, and they can be difficult to identify. Sarcasm and satire can be used to mask hate speech, making it more subtle and harder to detect. Sarcasm can be used to express hate speech in a way that is less obvious and less likely to be flagged as hate speech. Satire can also be used to express hate speech in a way that is intended to be humorous or satirical but can still be harmful. Annotators were trained to identify sarcasm and satire in tweets and understand the context in which it is used. They were also trained to differentiate between sarcasm and satire which is intended to be humorous or satirical and sarcasm and satire which is used to express hate speech. Annotation guidelines included clear examples of sarcasm and satire and how they can be used to express hate speech.

A.1.2 No Hate Speech. A post (text or image or both) reports the events or others' opinions objectively and contains no offensive or hateful content. To make guidelines clear, the following were discussed as the significant characteristics of non-hate speech.

- **Constructive criticism:** Non-hate speech during political events often includes constructive criticism of political figures, policies or parties. It can also include criticism of political events and happenings.
- **Factual and informative:** Non-hate speech on Twitter during political events often includes factual and informative content, it can be news, updates, and analysis of the political events.
- **Respectful and civil:** Non-hate speech during political events is respectful and civil in nature, it doesn't use derogatory language or hate speech symbols.
- **Lack of hostility:** Non-hate speech during political events does not express hostility or aggression towards a particular political group or individual.
- **Lack of misinformation or fake news:** Non-hate speech during political events does not spread misinformation or fake news, it is based on facts and credible sources.
- **Lack of targeting specific group:** Non-hate speech during political events does not target specific groups of people based on their political beliefs or affiliations.
- **Lack of sarcasm and/or political satire intended for hate:** Non-hate speech during political events does not use sarcasm or political satire to express hate speech.

A.1.3 Conflict Resolution Between Organization and Community Targets. Organization refers to a group of people who come together for a specific purpose, such as a business, non-profit, or government agency. An organization can have a clear leadership structure, a specific goal or mission, and a defined membership. Community refers to a group of people who share a common bond, such as geographic location, culture, or interest. A community can be more loosely defined than an organization, and it may not have a clear leadership structure or defined membership.

The guidelines for delineating organization and community are given below:

Community Targets:

- The tweet references a specific group of people based on their shared characteristics, such as race, ethnicity, national origin, religion, sexual orientation, gender identity, or disability with an intention of demeaning.
- The tweet references a specific geographic location or culture or specific interest or shared activity as the basis for the targeted group with an intention of demeaning it.

Organization Targets:

- The tweet specifically mentions the name of an organization or a business.
- The tweet references a specific goal or mission of the organization. The tweet references a specific set of actions or decisions made by the organization.
- The tweet references a specific leadership structure or hierarchy within the organization.

The annotation guidelines were exhaustive and the annotators regularly communicated the problems in annotations to each other. Some resolutions in annotation were made through meetings and annotation sessions. Annotators were able to distinguish between tweets that target an organization and tweets that target a community by analyzing the language used in the tweets and understanding the context in which the tweets were written.

A.2 Limitations

In this paper, we present a large-scale multimodal dataset for hate speech detection and target identification. We also present baselines for detecting hate speech and identifying targets using this dataset. However, there are several limitations to our work that should be acknowledged. First, our dataset is limited to tweets from a specific time period surrounding a political event, and may not be representative of hate speech in other contexts. Additionally, our dataset is based on tweets from a single microblogging platform, and it is not clear how well our approach would generalize to other platforms or modalities. Second, our annotation scheme for targets is based on broad categories (Individuals, Organizations, and Communities), and may not capture more specific or nuanced targets. Furthermore, the annotation process is subjective, and different annotators may have different opinions on whether certain tweets should be considered hate speech or not. Third, the baselines we provide are based on a limited set of features, and it is possible that other features or architectures could lead to improved performance. Finally, it's important to note that hate speech detection and target identification technologies can raise ethical concerns, such as the potential for misuse, bias, and invasion of privacy. These ethical concerns should be considered and addressed in the development and deployment of such technology.

Wireless Personal Communications
https://doi.org/10.1007/s11277-024-11115-1

Scalable and Cost-Efficient PoA Consensus-Based Blockchain Solution for Vaccination Record Management

Neetu Sharma[1] · **Rajesh Rohilla**[1]

Accepted: 17 April 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
The integrity of vaccination records is essential for protecting the public from deadly diseases. Currently, vaccination records are scattered across various health organizations' custody in a centralized manner, making them vulnerable to invalid alteration or even complete loss. The benefits of blockchain include immutability, integrity, availability, and transparency, which ensure that trustworthy vaccination records are easily accessible. This research aims to create a scalable and cost-effective proof of authority (PoA) consensus-based blockchain solution for vaccination record management. It can be beneficial to ensure that accurate records are available for scheduled vaccinations, medication, migration, education, or employment. Scalability, limited functionality, high execution costs, resource consumption, and store-house costs are all issues with current blockchain-based vaccination systems. By verifying the hash of vaccination certificates on the Ethereum blockchain, the suggested system preserves the accuracy of vaccination records. Vaccination records are stored off-chain through the interplanetary le system (IPFS) based on patient preference for scalability, availability, and store-house cost savings. Furthermore, the Proof of Work (PoW) consensus engine used by the current blockchain-based systems is extremely energy-intensive. The PoA consensus process employed by the suggested system was economical. Through the remix IDE, smart contracts with extra features and access controls are successfully implemented. The performance of the contracts is assessed by examining the execution costs of various transaction sizes. We presented a security analysis for detecting vulnerabilities using the Mythril tool. Additionally, we assessed the data availability using patient preference and IPFS-network parameters for optimal results.

Keywords Blockchain · Vaccination · IPFS · Scalable · Ethereum · Smart Contract

✉ Neetu Sharma
neetusharma85@gmail.com; neetu_2k19phdec25@dtu.ac.in

Rajesh Rohilla
rajesh@dce.ac.in

[1] Delhi Technological University, Delhi, New Delhi 110042, India

Published online: 07 May 2024

1 Introduction

Vaccinations are a basic necessity for ensuring public protection against infectious diseases. They need to be mandated by the government [1] as a crucial solution for protection against communicable diseases [2], saving lives and economies during pandemics [3]. Vaccines stimulate the production of antibodies, safeguarding individuals from contagious illnesses [4]. This secured and immunized public contributes to broader herd immunity by preventing disease transmission to those who are not fully vaccinated.

Vaccination can also serve as an immunity passport for medical consultations, migration, education, employment, etc. [5], safeguarding public health. The digitization of vaccination records has addressed the practical challenge of patients maintaining hard copies obtained from various sources. However, it has introduced a new challenge of securing sensitive health information. Mismanagement of vaccination data can lead to delayed or erroneous treatment, impacting public health. Integrating various healthcare sector members who operate independently is a significant challenge. Traditional healthcare systems maintain vaccination records in centralized systems susceptible to single point failures and various attacks like DDoS, ransomware, etc. The lack of security, integrity, and interoperability in centralized systems leads to record blocking and breaches, eroding public trust in healthcare providers. Innovative strategies are needed to track and validate vaccination records by all authorized members and securely share data to combat pandemics and threats, as evidenced by the importance of reliable systems highlighted by COVID-19 [6]. Distributed technologies, such as blockchain, have emerged for data handling across multiple nodes [7], finding success in the business community, particularly in cryptocurrency and token markets [8]. Blockchain, when integrated with artificial intelligence, has the potential to create an innovative business environment [9], benefiting drug supply chains and drug discovery processes [10, 11]. Unlike centralized systems, blockchain ensures distributed control, preventing data alteration [12, 13]. Optimizing storage pressure is essential to leveraging blockchain's benefits in securing sensitive vaccination records and data from smart devices [14]. Additionally, governance policies are necessary for sustainable blockchain implementation [15].

1.1 Motivation

One of the most significant issues facing the healthcare industry is health data breaches [16], wherein hackers use stolen data for fraud. Over the past ten years, a significant number of incidents involving healthcare data breaches have occurred [17]. Protecting patient information is crucial for improving health.

Blockchain presents a promising solution to challenges in existing centralized systems, yet its application in vaccination management remains immature. Few COVID-specific prototype systems exist, with limited smart contract functionality, scalability, high execution costs, and storage cost issues. Additionally, existing blockchain-based systems employ a PoW consensus engine that consumes excessive energy. Establishing a proper information security policy to govern data shared with stakeholders is imperative, achieved by incorporating access controls as security provisions within the contract. This work focuses on PoA-based blockchain for lifelong patient vaccination record management, aiming to enhance smart contract functionalities to encompass the entire vaccination process at minimal execution fees. To address scalability, availability, and storage costs, we opt to store only the vaccination record hash in the Ethereum blockchain, with actual data stored in off-chain storage based on patient preference.

1.2 Contributions

We propose a decentralized model for managing patients' lifetime vaccination records. Utilizing the Ethereum blockchain enhances decentralization and network security, with the vaccination certificate hash stored to ensure integrity. A smart contract with additional functionality and access control checks is created at a low execution cost. This contract also handles complaints against health organizations to ensure smooth governance of the vaccination process on the blockchain network. Storage costs are optimized based on patient preference and IPFS-network parameters. Our key contributions include:

- Proposing a scalable and cost-efficient proof-of-authority (PoA) consensus-based blockchain solution for managing patients' lifetime vaccination records.
- Storing vaccination records off-chain in IPFS-based on patient preference for scalability, availability, and reduced storage costs.
- Developing a vaccination smart contract with additional functionality, successfully deployed over Remix IDE, with performance evaluated through analysis of execution costs with different transaction sizes.
- Utilizing the PoA consensus engine to reduce energy waste in mining vaccination data simultaneously by all network nodes and to expedite validation.
- Evaluating data availability based on IPFS-network parameters and patient preference, along with analyzing network parameters to optimize storage costs.
- Conducting a comparative analysis demonstrating that the proposed blockchain-based scheme outperforms existing schemes across various performance metrics.

1.3 Organization

The rest of the paper proceeds as follows: Sect. 2 introduces blockchain technology and its consensus mechanisms. Section 3 presents existing work related to blockchain-based data management systems, followed by Sect. 4, which discusses preliminaries and possible solutions for blockchain scalability. Section 5 provides an overview of the proposed work, including architecture, information flow, smart contracts, and patient preferences regarding data availability. Experimental results and discussion are offered in Sect. 6, analyzing data availability, execution cost, energy efficiency, and discussing comparative analysis in terms of different metrics. Finally, Sect. 7 concludes the paper.

2 Blockchain Technology

A blockchain is a decentralized computer network based on cryptography [18], with blockchain nodes serving to store copies of transactions and maintain the network [19]. Its primary use is secure transaction recording and sharing across multiple nodes, fostering trust. Satoshi Nakamoto founded blockchain technology [20, 21], with the bitcoin blockchain being its first application in 2008 for cryptocurrency transactions [22]. Following bitcoin's success, researchers have developed various blockchain-based frameworks for cryptocurrencies [23, 24], extending its applications beyond cryptocurrency to reinforce smart home security [25, 26], manage evidence in digital forensic investigations [27], copyright management [28], transparent voting systems [29], and more. Blockchain's advantages, investigated widely in literature, include facilitating fair exchange [30] and addressing password

forgetting issues [31]. The increasing demand for blockchain emphasises the importance of accountability and confidentiality [32], as it prevents intruder manipulation, granting access rights only to authorized users [33]. In blockchain networks, transactions are recorded within blocks added after the consensus process, with databases always online and transparent, accessible to permitted members with internet facilities. Furthermore, blockchain eliminates intermediaries, making processes cheaper and more streamlined.

The technology's basic features include security, integrity, immutability, transparency, and transaction availability [34], enabling verification without third-party involvement and ensuring data immutability [35]. It employs asymmetric cryptography for data block security [36], with each block dependent on its previous block to complete the encryption process. Transactions undergo trust verification through broadcasting, consensus, and auditing before being attached to blocks, which are timestamped and shared across each node, ensuring decentralization and lifetime data availability. Transactions are visible to all authorized participants, preserving transparency, integrity, and trust, with records traceable to their origin. Consensus mechanisms verify transaction data, eliminating the risk of duplicate records, while smart contracts facilitate automatic transactions based on pre-set conditions.

Figure 1 illustrates a blockchain transaction, where a digitally signed transaction request is disseminated to all network nodes through the gossip protocol [37].

Blockchain employs a public–private key pair for transaction signing [38], where a private key encrypts transactions and only corresponding public keys can decrypt them. Blocks are encrypted using hashes, with each block hash dependent on the previous block hash, current block data, and nonce [39]. Mining, the process of generating a valid hash for new block creation, involves complex math to find a suitable nonce and obtain an acceptable hash to meet difficulty requirements. Any change to block data necessitates re-mining subsequent blocks. Once consensus among validator nodes is reached, a new block is added to the chain, and the transaction is executed.

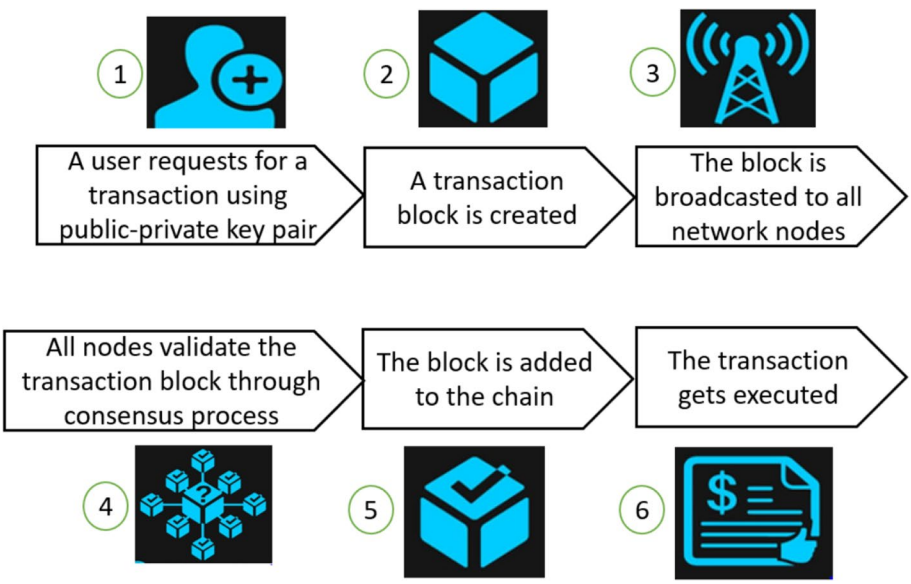


Fig.1 Blockchain transaction flow

2.1 Types of Blockchain

There are three types of blockchains: public, private, and consortium [40, 41]. A public blockchain is non-restrictive, allowing anyone to join as an authorized node to perform, verify, and record transactions. For example, bitcoin operates on a public blockchain, where individuals worldwide can become nodes. A private blockchain, on the other hand, is restrictive, with only authorized users permitted to join. The advantage of a private blockchain lies in its high transaction speed, while its weaknesses include lower security and transparency. A consortium blockchain, meanwhile, operates as a permissioned blockchain, with multiple organizations collaborating to run the blockchain platform.

2.2 Consensus in Blockchain

In the blockchain network, mathematical algorithms validate transactions and transaction blocks through consensus mechanisms, enabling communal maintenance of a sustainable ledger among members who may not trust each other. The three fundamental consensus engines are PoW, PoA, and Proof of Stake (PoS), each with its own benefits, costs, and trade-offs, with selection often contingent on application requirements. PoW, requiring complex mathematical algorithms and substantial computing power to confirm transaction blocks, offers high decentralization and security but consumes significant energy [42]. PoS mitigates energy consumption issues present in PoW, boasting low block creation time and energy usage, lower security and a monetary stake requirement for validating new blocks [43].

PoA, an enhanced version of the PoS consensus protocol, achieves consensus through authorized validators, with block sealing power limited to a set number of nodes. Validators' identities are at stake, as any faulty behavior tarnishes their reputation. In PoA consensus, with n validators, a maximum of b nodes where $b < n$ can be Byzantine to reach a consensus. PoA's cost-efficiency, throughput, and scalability make it a viable alternative to address PoW and PoS limitations, particularly suitable for applications requiring user trust in validators. The advantages of the PoA consensus protocol include: reduced energy consumption compared to other consensus schemes due to absence of block sealing competition; uniform transaction processing without dynamic transaction fees; suitability for decentralized application development; enhanced efficiency and throughput relative to other consensus models owing to its centralized reputation-based approach; and publicly shared identities of sealing nodes preventing fraudulent activities in the blockchain network, resulting in improved processing speed and reduced chances of network forks.

2.3 Comparison of PoA With PoW and PoS

PoW, a computationally intensive consensus protocol, results in slow block creation speed and fosters network forks due to block mining competition among nodes, complicating the consensus process. In contrast, PoA is energy-efficient and scalable, owing to the limited number of miner nodes in the network. PoA facilitates rapid development of decentralized networks and diminishes potential forks and network attacks. PoS, an energy-saving consensus scheme, allows nodes with higher stakes to become block sealers by locking their stakes in the network. Conversely, PoA operates as an identity-based consensus scheme for sealing new blocks, enhancing network scalability and cost-efficiency [44].

3 Related Work

This section addresses the work related to blockchain-based health data management systems, emphasizing the critical importance of medical data for healthcare [45]. The merging of digital identities from various medical records presents a significant challenge [46], with security, privacy, and authorization emerging as key concerns for real-world applications [47]. Security services are essential to ensure the secure capture, inventory, handling, transmission, and monitoring of healthcare records on the network. Blockchain has piqued the interest of investors across various sectors, including healthcare [48], offering a privacy-preserving method for the secure exchange of medical and government data [49]. While blockchain-based record management holds promise for enhancing the existing healthcare system [50], scalability remains a major research challenge [51].

Several prototypes and approaches for blockchain-based vaccination process security have been developed [1–5]. However, these prototypes lack comprehensive analysis, security, cost-efficiency, functionality to cover the entire vaccination process, and scalability. Additionally, decentralized applications aimed at supporting record-keeping for COVID-19 test reports and vaccination certificates suffer from scalability and functionality issues [5]. Previous studies have addressed security risks in cloud computing environments [52] and showcased secure sharing of Electronic Health Records (EHRs) using blockchain [53]. Moreover, applications of blockchain in information management have been reviewed [54], and proposals for cloud-supported electronic health (eHealth) systems leveraging blockchain for medical record management have been made [55].

Overall, the discussed implementations and studies emphasise the role of blockchain in ensuring the security, immutability, transparency, trust, and accessibility of recorded data, suggesting that blockchain could be the most effective tool for vaccination record management.

4 Preliminaries

4.1 Ethereum

Ethereum, founded by Vitalik Buterin in December 2013, is an open-source platform for building blockchain-based distributed applications, renowned for its execution of smart contracts [56, 57]. Smart contracts, computer programs stored on a distributed ledger, automate transactions when specific conditions are met, significantly expanding the scope of blockchain applications. The Ethereum Virtual Machine (EVM) enables machines to execute smart contracts compiled into EVM bytecode. These contracts are deployed onto the Ethereum network through contract creation transactions. Ethereum transactions comprise the following components:

- From: A 20-byte address of the transaction sender.
- To: A 20-byte address of the transaction recipient.
- Value: The amount (in wei) sent from sender to recipient.
- Data: Transaction inputs.
- Gas: A unit measuring the fee required for transaction execution.
- Gas Price: The amount of ether per unit of gas the sender is willing to pay for transaction execution.
- Gas Limit: The maximum gas set by the sender for the transaction.

4.2 Blockchain Scalability

Currently, blockchain faces scalability issues that limit its usability across various applications. Scalability refers to the limited capability of a highly distributed network to handle a large number of transactions and update ledgers within a short period of time. Transaction per second (TPS) can be calculated using the mathematical formula:

$$\text{Transactions per second} = \text{Transactions Per Block} / \text{Block Generation Time} \quad (1)$$

The above equation highlights that to increase transaction speed, the number of transactions per block should be high and the block generation time should be low. However, block generation time is typically fixed by blockchain platforms to maintain network stability. The number of transactions per block can be calculated using the formula:

$$\text{Transactions per block} = \text{Block Size} / \text{Average Transaction Size} \quad (2)$$

This equation demonstrates that increasing transactions per block can be achieved by either increasing the block size or decreasing the average transaction size. However, increasing block sizes may disrupt the harmony of the blockchain network by splitting the community. To expedite transaction validation, transactions can be partitioned and distributed across the network through sharding.

Traditional cryptocurrency-based Proof of Work (PoW) is an energy-inefficient consensus engine [58, 59]. Hence, throughput can be enhanced by adopting a more efficient consensus algorithm, such as switching from PoW to Proof of Authority (PoA). PoA boosts transaction speed, reduces energy waste by validators, and mitigates the risk of 51% attacks.

Blockchain systems often struggle to achieve scalability, decentralization, and security simultaneously. While Ethereum blockchains are secure and decentralized, they exhibit slow transaction speeds. On-chain scaling, which involves increasing block size and optimizing consensus mechanisms, may compromise decentralization and security. Therefore, improving the storage efficiency of smart contracts is crucial to enhance overall throughput [60], possibly achieved through the creation of lightweight blockchains [61].

4.3 Interplanatery File System

IPFS is a peer-to-peer file-sharing system designed for off-chain storage, enabling Off-Chain scaling without reliance on a blockchain network. It functions as a protocol for secure peer-to-peer data storage, encrypting all data files stored on IPFS with cryptographically generated hash values. This approach diminishes the computational overhead associated with mining large data files. IPFS empowers users to store and share data files within the network using content addresses rather than location addresses.

5 Proposed Work

5.1 Overview

In today's world, frequent relocations often result in individuals missing one or more vaccination records, either due to misplaced physical copies or poorly managed online records.

When such records are needed for scheduled vaccinations, illness management, migration, education, or employment purposes, individuals typically attempt to retrieve them from the relevant health organizations. However, current health organizations often maintain data in a centralized manner, leaving it vulnerable to alteration or loss due to cyber-attacks. The process of transferring vaccination data from hospitals often involves multiple hospital-to-patient communications, leading to delays that can be costly during times of illness. Moreover, individuals receive various vaccines throughout their lifetime, including Hepatitis B, Rotavirus, Haemophilus Influenzae type b, Poliovirus, Diphtheria, Tetanus, Pertussis, Pneumococcal, Meningococcal, Influenza, Measles, Mumps, Rubella, Varicella, Hepatitis A, Human Papillomavirus, Covaxin, Covishield, etc. The scattered nature of vaccination records complicates their management and retrieval. Centralizing different hospitals onto a single system is not a viable solution, as the database would be under the control of a single administrator and susceptible to tampering or loss. Therefore, there is a pressing need for the health ministry to initiate a decentralized system for maintaining immutable and trustworthy vaccination records accessible to any requesting entity throughout an individual's life.

Figure 2 illustrates the typical and proposed usecase scenario.

When a person moves out of town or country and urgently requires their past vaccination record for reasons such as illness, scheduled vaccinations, or applying for education or employment, the traditional process involves several steps:

1. The person, lacking some records, contacts the previous health organization where vaccinations were administered.
2. The health organization first authenticates and validates the person's request.
3. Subsequently, the organization takes time to issue a duplicate copy of the vaccination certificate to the person, with no guarantee of its integrity.
4. The person then submits the received record to the local health organization, academic institution, or employer, often resulting in delayed submission.

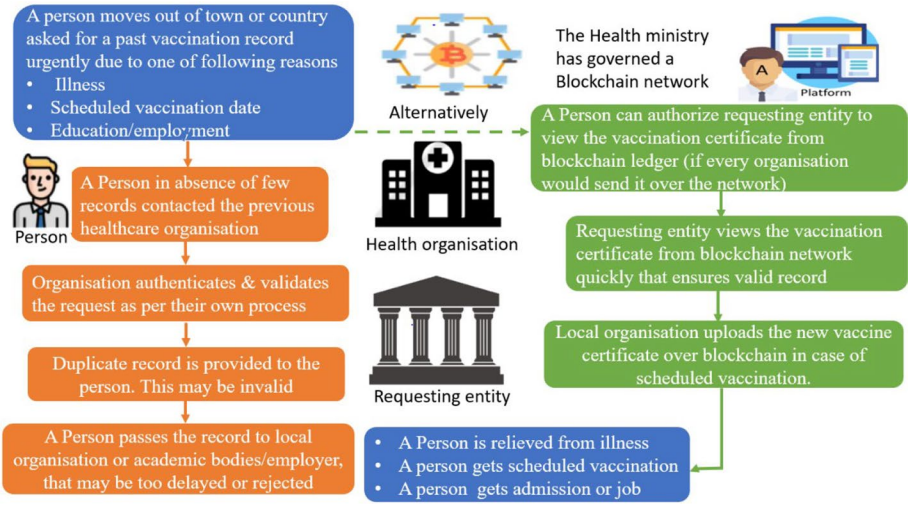


Fig.2 Typical and proposed usecase scenarios

In contrast, the proposed scenario streamlines the process:

1. The person can authorize the viewing of their vaccination record over a blockchain network by the requesting entities. This could become feasible if the health ministry mandates the uploading of vaccination records onto the blockchain network for all health organizations.
2. The requesting entity swiftly accesses the authenticated vaccination records from the blockchain network.
3. By verifying the valid vaccination certificate, the local organization can promptly provide the necessary treatment or vaccinations.
4. Additionally, the local organization uploads the vaccination record onto the blockchain network.
5. As a result, the person can easily address their illness, receive vaccinations, gain admission, or secure employment without encountering any obstacles.

5.2 System Architecture

A blockchain-based vaccination system enables users to update and retrieve vaccination records from any location and at any time. The objective is to guarantee the availability of accurate records for scheduled vaccinations or in cases of illness, thereby safeguarding patients against diseases. The proposed model for vaccination record management operates without the need for a central authority, yet it ensures accessibility of records to all requesting entities. The architecture of the proposed system, depicted in Fig. 3, consists of the following layers:

- Users layer comprises the Health Ministry, health organizations, patients, and requesting entities that initiate transactions on the blockchain network. The Health Ministry serves as the administrator, overseeing the entire network. Patients can authorize requesting entities to view their vaccination certificate using their patient ID and vaccination ID.
- The application layer serves as the frontend of the application. In the proposed model, this user interface facilitates interaction with the blockchain-IPFS network, enabling users to upload and retrieve transactions. Upon completion of a transaction, the sender receives a transaction receipt containing transaction details through this interface.
- The interface layer consists of APIs used to communicate with the blockchain and obtain required results, such as data retrieval or addition.
- The smart contract layer constitutes the logic of the network, ensuring adherence to all rules and regulations governing the blockchain network.
- The Decentralized Network and Storage layer comprise the Ethereum blockchain and IPFS. The Ethereum blockchain forms the foundation of the blockchain network, consisting of blocks (ledgers) that hold transaction data. The Ethereum blockchain network functions to broadcast uploaded transactions to all peer nodes, create new transaction blocks, mine transaction blocks, validate blocks by all authorized nodes, and append blocks into the ledger. The PoA consensus mechanism is utilized for mining purposes. Transaction hashes are stored in the blockchain, while complete transactions are stored on off-chain storage. Each transaction sender runs both a blockchain node and an IPFS node, with the API server serving as middleware. When publishing off-chain data, this middleware stores the original data in IPFS, then creates a blockchain transaction containing that data's hash. The middleware extracts the hash from the blockchain, then uses it to fetch the complete content from IPFS. The local IPFS node automatically

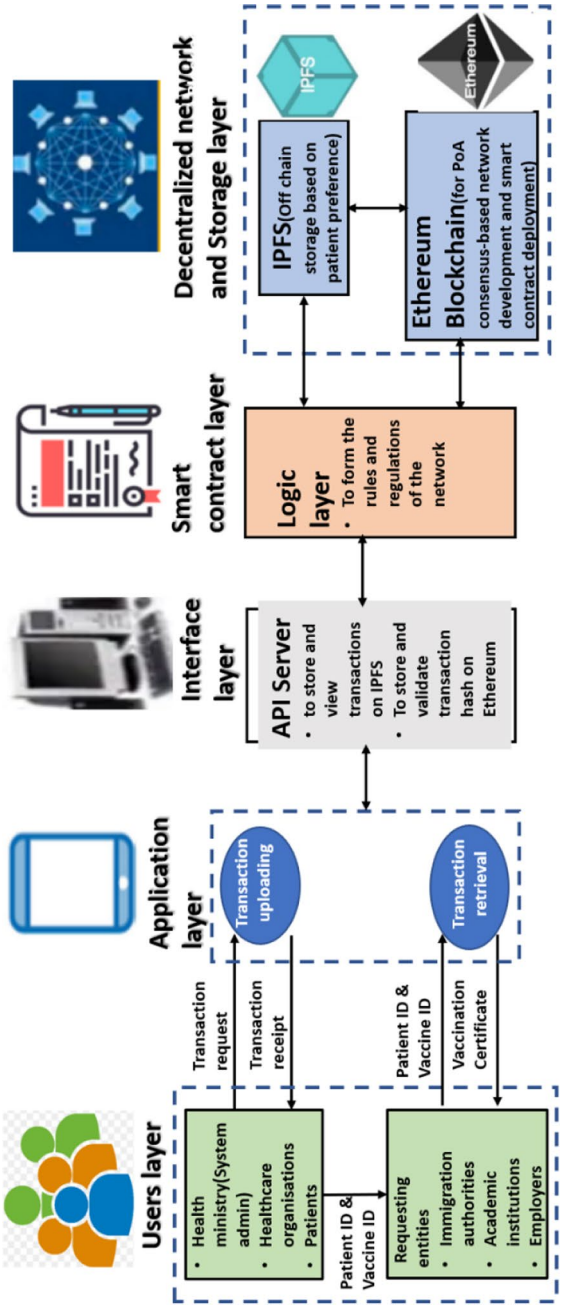


Fig. 3 Architecture of proposed decentralised vaccination system

verifies the retrieved content against the hash to ensure it hasn't been altered. IPFS is employed with the blockchain for scalability, and the patient preference feature is utilized to reduce storage costs.

5.3 Information Flow

The use-case diagram for all parties involved in the proposed system is depicted in Fig. 4. Initially, the Health Ministry adds health organizations with vaccination facilities. Patients can invoke registration requests on the network. Health organizations can register new patients in response to registration requests. Patients can view all health organization details, including addresses, available vaccines, and date-time slots. After reviewing organization details, patients add a vaccination request to any suitable health organization to schedule an appointment.

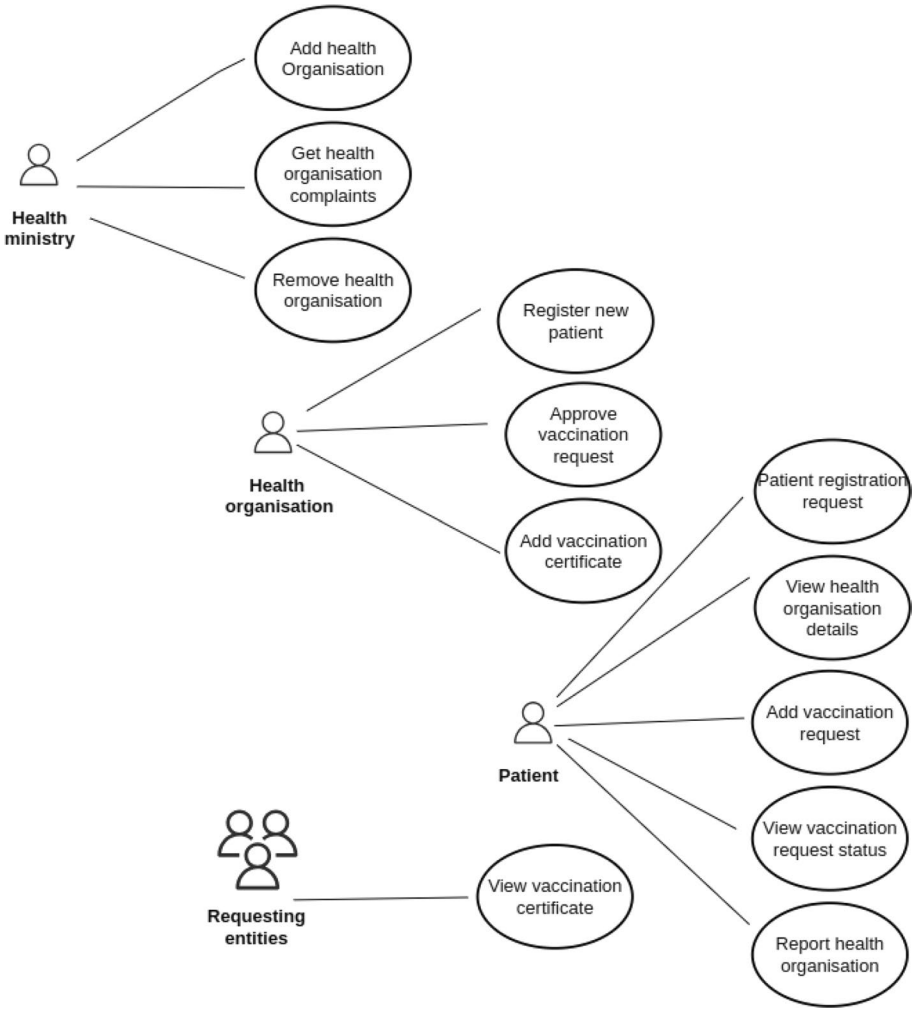


Fig. 4 Use-case diagram of proposed vaccination system

The health organization accepts the vaccination request by adding an approved status. Patients can then view the transaction request status and visit the health organization physically for vaccination. After the vaccination process, the health organization adds the vaccination certificate, which is accessible to any authorized requesting entity. Patients can report complaints against health organizations in cases of issues during the vaccination process. The Health Ministry can fetch reported complaints against any organization and remove health organizations if complaints exceed a predefined value. The Ethereum blockchain includes a smart contract for various functionalities.

Algorithm 1 Smart contract functionality of health ministry.

```
function addHealthOrganisation(Organisation ID, Address, Registration number, Available Vaccines slots)
    if msg.sender == Health ministry then
        if Organisation ID does not exist then
            Add health organisation to the Health organisation asset
        else
            Invalid transaction
        end if
    end if
end function
function getHealthOrganisationComplaints(Organisation ID)
    if Organisation ID exist then
        return Complaints reported against health organisation
    else
        Invalid Organisation ID
    end if
end function
function removeHealthOrganisation(Organisation ID)
    if Organisation ID exist then
        if Complaint reported > 1/3(Vaccination count of organisation) then
            remove health organisation
        else
            Invalid Organisation ID
        end if
    end if
end function
```

Figure 5 presents the workflow diagram, and algorithm 1 illustrates the smart contract functionality of the Health Ministry. Only the Health Ministry has the authority to perform the following functions:

- addHealthOrganisation(): Used to add health organizations, containing organization ID, address, registration number, and available vaccines with date and time slots.
- getHealthOrganisationComplaints(): Retrieves complaints reported against health organizations, containing organization ID.
- removeHealthOrganisation(): Removes health organizations if complaints are reported by more than one-third of vaccinated patients, containing the organization ID field.

Algorithm 2 outlines the smart contract functionality, and Fig. 6 presents the workflow diagram of a health organization. These organizations possess access control to invoke the following functions:

- registerNewPatient(): This function enables the health organization to register a new patient. Any health organization can initiate the registration of a new patient. It includes fields for patient ID, Aadhaar number, age, and data availability choice.
- approveVaccinationRequest(): This function is utilized to endorse vaccination requests. The health organization adds the "approved status" to the request, encompassing patient ID, date-time, requested vaccination, and status.
- 3. addVaccinationCertificate(): This function facilitates the addition of a vaccination certificate and increments the vaccination count. It includes fields for organization ID, vaccination ID, patient ID, and vaccination certificate.

Algorithm 2 Smart contract functionality of health organisation.

```
function registerNewPatient(Patient Name, Aadhaar Number, Age, Data Availability choice)
    if Patient ID does not exist then
        Add patient to the Registered patient asset
    else
        Invalid transaction
    end if
end function
function ApproveVaccinationRequest(patient name, Date Time, Requested vaccination, status)
    if Patient ID exist then
        Approve vaccination request by adding "approved" status
    else
        Invalid transaction
    end if
end function
function AddVaccinationCertificate(Organisation ID, Vaccination ID, patient ID, Vaccination certificate)
    if patient ID and vaccination ID exist then
        add vaccination certificate and increment vaccination count of organisation
    else
        Invalid transaction
    end if
end function
```

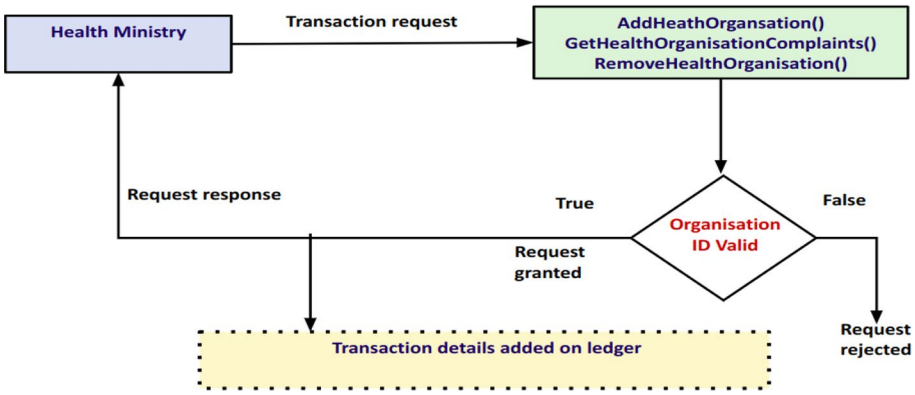


Fig. 5 Workflow diagram for health ministry functions

Figure 7 presents the workflow diagram, and algorithm 3 illustrates the smart contract functionality of a patient. Patients are the only ones with access control to perform the following actions:

Algorithm 3 Smart contract functionality of patient.

```
function patientRegistrationRequest(Patient ID, Aadhaar Number, Age, Data Availability choice)
    if Patient ID does not exist then
        Add patient to the registration request asset
    else
        Invalid transaction
    end if
end function
function ViewhealthorganisationOrganisation ID)
    if Organisation ID exist then
        return Health organisation details
    else
        Invalid Organisation ID
    end if
end function
function addVaccinationRequest(patient ID, Date Time, Requested vaccination, Organisation ID)
    if patient ID exist then
        add vaccination request to the vaccination request asset
    else
        Invalid transaction
    end if
end function
function viewVaccinationRequestStatus(Patient ID)
    if Patient ID exist then
        return Vaccination request status
    else
        Invalid Patient ID
    end if
end function
function ReportHealthOrganisation(Organisation ID)
    if Organisation ID exist then
        Increment complaint reported
    else
        Invalid transaction
    end if
end function
```

- patientRegistrationRequest(): This function allows the patient (or the parent in the case of a child patient) to submit a registration request. It includes fields for patient ID, Aadhaar number, age, and data availability choice.
- viewHealthOrganisationDetails(): This function is utilized to access health organization details, containing the organization ID field.
- addVaccinationRequest(): This function facilitates the submission of a vaccination request, encompassing fields for patient ID, date-time, requested vaccination, and organization ID.
- ViewHealthOrganisationStatus(): This function enables the patient to view the status of vaccination requests, containing the organization ID field.

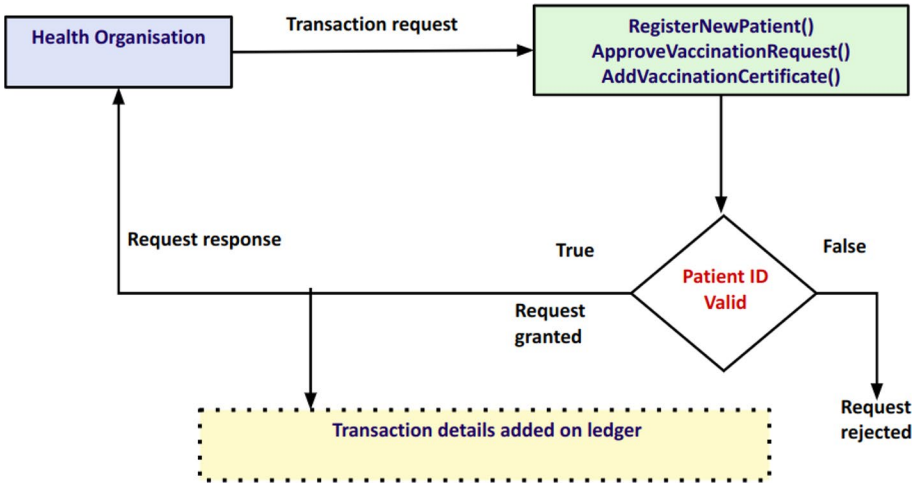


Fig. 6 Workflow diagram for health organisation functions

- reportHealthOrganisation(): This function is utilized to lodge a complaint against a health organization by incrementing the reported complaints. It includes the organization ID field.

Figure 8 presents the workflow diagram and algorithm 4 outlines the smart contract functionality of a requesting entity. Any requesting entity with a valid patient ID and vaccination ID can invoke the following function:

- View Vaccination Certificate(): This function allows any requesting entity to view the vaccination certificate. It includes fields for patient ID and vaccination ID.

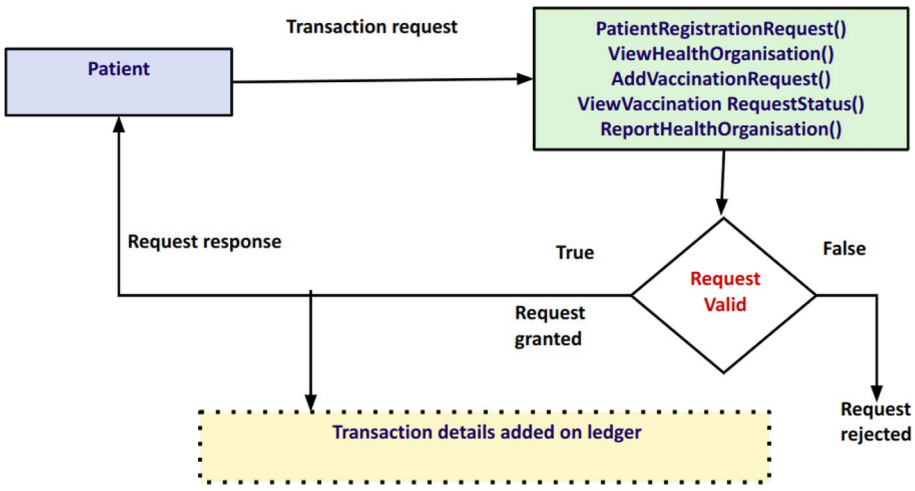


Fig. 7 Workflow diagram for patient functions

Algorithm 4 Smart contract functionality of requesting entity..

```
1: function viewVaccinationCertificate(Patient ID, vaccination ID)
2:     if Patient ID and vaccination ID exist then
3:         return Vaccination certificate
4:     else
5:         Invalid Patient ID and vaccination ID
6:     end if
7: end function
```

5.4 Patient Preferences of Data Availability

In the proposed system, patients specify their preferences for vaccination data availability as HIGH, MODERATE, or LOW in the transaction request. Based on these preferences, the system stores the data at the desired number of nodes in off-chain storage. The HIGH data availability preference may increase the probability of data availability at the cost of a larger storage size, while the low data availability preference may reduce the storage size at the cost of the risk of data unavailability. Therefore, we have devised a mechanism to assess the probability of data availability with specified patient preferences. In this section, Table 1 lists the notations used in the paper and their descriptions.

In this mechanism, the involved network parameters of the off-chain storage IPFS are denoted as TN, TDN, TIAN, and IADN. TN represents the total number of nodes, TDN represents the total number of data nodes, TIAN represents the total number of inactive nodes, and IADN represents the number of inactive-data nodes. Initially, we calculate the possible ways to obtain inactive data nodes from the total inactive nodes, active data nodes from all active nodes, and total data nodes from all nodes. This yields the probability of data unavailability as:

$$^{\mathbf{TIAN}}\mathbf{C}_{\mathbf{IADN}} * ^{\mathbf{TN-TIAN}}\mathbf{C}_{\mathbf{TDN-IADN}} / ^{\mathbf{TN}}\mathbf{C}_{\mathbf{TDN}} \tag{3}$$

Thus, the probability of data availability can be calculated as:

$$\mathbf{P} = \mathbf{1} - [(^{\mathbf{TIAN}}\mathbf{C}_{\mathbf{IADN}} * ^{\mathbf{TN-TIAN}}\mathbf{C}_{\mathbf{TDN-IADN}}) / ^{\mathbf{TN}}\mathbf{C}_{\mathbf{TDN}}] \tag{4}$$

For example, if TN = 100, TDN = 90, IAN = 10, and IADN = 0, this gives a probability of data availability (P) equal to 1.

6 Experimental Results & Discussion

To assess the performance of the proposed framework, we conducted experiments using an Intel Core i5, 8th generation CPU running on the Ubuntu 64-bit operating system with 12.00 GB of RAM. We utilized the Ethereum Geth environment to create the blockchain network and developed the smart contract in the Solidity programming

language. The smart contract of the proposed design was deployed and tested on the Remix IDE. The set of APIs to communicate with the blockchain was written in Node.js, while the algorithms to compute the IPFS-network parameters and probability of data availability were implemented in JavaScript.

6.1 Computation of IPFS-Network Parameters

In this section, we calculate the IPFS-network parameters, as shown in Algorithm 5. Initially, the data availability preferences of patients are retrieved from the ledger. This algorithm performs the following functions:

- getTotalDataNodes(): This function takes the patient preference of data availability and the total number of nodes as input and returns the total number of data nodes as output. If the input preference is high, the total number of data nodes would be 90% of all nodes. If the input preference is moderate, the total number of data nodes would be 75% of all nodes. If the input preference is low, the total number of data nodes would be 50% of all nodes.
- getTotalInactiveNodes(): This function takes the total number of nodes and the percentage of inactive nodes in all nodes as input and returns the total number of inactive nodes as output.
- getInactiveDataNodes(): This function takes the total number of nodes and the percentage of inactive data nodes in all nodes as input and returns the total number of inactive nodes as output.

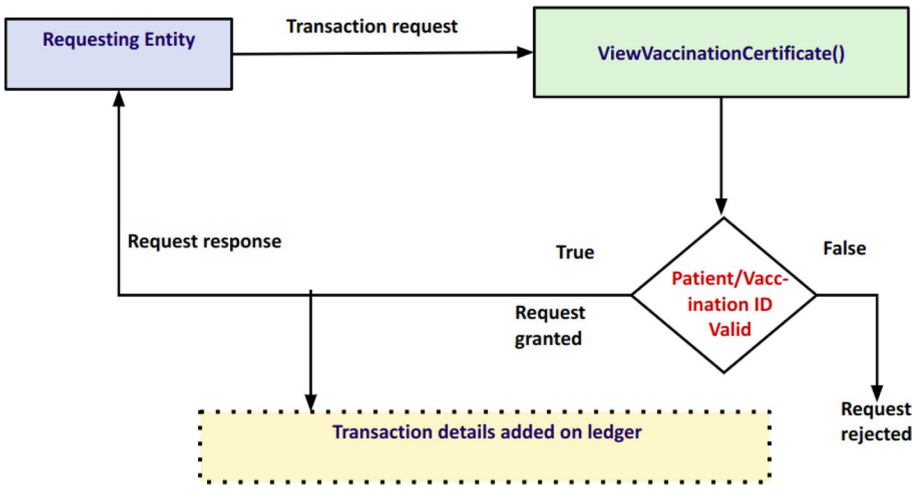


Fig. 8 Workflow diagram for requesting entity function

Algorithm 5 Computation of IPES-network parameters.

```
INPUT:Patient preferences PPDA, Total number of nodes TN, Total number of inactive
nodes in % TIAN%, Number of inactive-data nodes in % IADN%.
OUTPUT:Total number of data nodes TDN, Total number of inactive nodes TIAN,
Number of inactive-data nodes IADN
function getTotalDataNodes(PPDA, TN)
    if PPDA == "HIGH" then
        TDN% = 90
        TDN= TDN%*TN/100
        return TDN
    else if PPDA == "MODERATE" then
        TDN% = 75
        TDN= TDN%*TN/100
        return TDN
    else if PPDA == "LOW" then
        TDN% = 50
        TDN= TDN%*TN/100
        return TDN
    else
        TDN% = 50
        TDN= TDN%*TN/100
        return TDN
    end if
end function
function getTotalInactiveNodes(TN,TIAN%)
    TIAN= TIAN%*TN/100
    return TIAN
end function
function getInactiveDataNodes(TN,IADN%)
    IADN= IADN%*TN/100
    return IADN
end function
```

6.2 Computation of Probability of Data Availability

In this section, we compute the probabilities of data availability using Algorithm 6.

Table 1 Used notations with their descriptions

Symbol	Description
TN	Total nodes
TDN	Total data nodes
TIAN	Total inactive nodes
IADN	Inactive data nodes
(TN,TDN,TIAN,IADN)	IPFS-network parameters
Q	The potential methods for determining the total number of inactive nodes
R	Approaches that could be used to determine the number of active-data nodes from the total active nodes
S	Potential methods to obtain the total number of data nodes from total node
P	Data availability probability

Algorithm 6 Computation of probability of data availability.

```
1: INPUT: Set of IPFS-network parameters(TN, TDN, TIAN, IADN)
2: OUTPUT: Probability of data availability P
3: function getProbabilityofDataAvailability(TN, TDN, TIAN, IADN)
4:     Q= combination(TIAN,IADN)
5:     R= combination((TN-TIAN), (TDN-IADN))
6:     S = combination(TN, TDN)
7:     P = (Q*R)/S
8:     return P
9: end function
```

This algorithm performs the following functions:

- getProbabilityofDataAvailability(): This function takes the set of IPFS-network parameters (TN, TDN, IAN, IADN) as input and returns the probability of data availability as output. This function determines the total number of data nodes, the total number of inactive nodes, the number of inactive data nodes, possible ways to obtain inactive data nodes from total inactive nodes, active data nodes from total active nodes, and total data nodes from total nodes, and finally the probability of data availability.

6.3 Data Availability Analysis

Patients have specified their data availability preferences as HIGH, MODERATE, and LOW. Based on these preferences, the system stores the data at the desired number of network nodes in off-chain storage. If the preference is high, data is shared among 90% of the total number of network nodes. If the preference is moderate, data is maintained at 75% of the total number of network nodes. If the preference is low, data is maintained at 50% of the total number of network nodes.

The probability of data availability can be calculated using Eq. 4 for different patient preferences. We have calculated the probability of data availability for each patient preference by considering the total number of nodes, the total number of inactive nodes (TIAN%) as a percentage of all nodes, and the number of inactive-data nodes (IADN%) as a percentage of all nodes. In our investigation, TN is kept at 100, while TDN can be 90%, 75%, or 50%, according to patient preference. The IAN values are taken as 10%, 20%, and 30%, and the IADN values are taken as 0%, 1%, 2%, 3%, 4%, and 5%. Initially, we investigated the probability of data availability for different patient preferences at 10% inactive nodes and 0–5% inactive-data nodes. Figure 9 illustrates the data availability probability with different user preferences at 10% inactive nodes.

Next, we explored the probability of data availability with 20% inactive nodes and 0–5% inactive data nodes. Figure 10 illustrates the data availability probability with various user preferences at 20% inactive nodes.

Finally, we investigated the probability of data availability for different patient preferences with 30% inactive nodes and 0–5% inactive-data nodes. Figure 11 illustrates the data availability probability with different user preferences at 30% inactive nodes.

The obtained results indicate that relatively better data availability is achieved when IADN ranges from 0 to 3%. Specifically, this lies between 0.9999 and 1.0000 at 10% inactive nodes, between 0.9933 and 0.9999 at 20% inactive nodes, and between 0.9932 and 0.9999 for high data availability choice. Similarly, for moderate data availability choice,

optimal results are obtained at IADN = 0–3%, with probabilities ranging from 0.9935 to 0.9999 at 10% inactive nodes, from 0.9864 to 0.9929 at 20% inactive nodes, and from 0.9848 to 0.9919 at 30% inactive nodes. Likewise, for low data availability choice, optimal results are again achieved at IADN = 0–3%, with probabilities ranging from 0.9934 to 0.9869 at 10% inactive nodes, from 0.8553 to 0.9721 at 20% inactive nodes, and from 0.8427 to 0.9613 at 30% inactive nodes. The probability of data availability exceeds 90% when the inactive-data nodes constitute up to 3% of all nodes at 10% down nodes. Thus, we conclude that the optimal probability of data availability can be achieved at low storage cost by ensuring that the number of inactive data nodes does not exceed 3% of the total network nodes.

6.4 Execution Cost Analysis

We developed the proposed Ethereum smart contract in the Solidity programming language. The smart contract of the proposed design was compiled and deployed in the Remix environment for testing. Figures 12 and 13 illustrates the Remix IDE interface representing the smart contract deployment read and write transactions.

We presented the logs triggered upon executing the smart contract function on the Remix ID environment in Fig. 14 and the execution cost analysis for different transaction sizes in Fig. 15. The maximum execution cost in the contract transaction is 4,291,404 gas, while all transaction functions have low costs. Additionally, during the testing phase, we endeavored to decrease execution fees by implementing various checks in the smart contract. These checks revert invalid transactions to avoid unnecessary execution fees resulting from the wastage of computational power. To enhance scalability and reduce costs, vaccination records are stored off-chain through IPFS based on patient preference. Furthermore, execution costs can be further reduced by consolidating multiple transactions into a single transaction.

6.5 Energy Efficiency Analysis

For the development of the blockchain network, we established 1 boot node and 4 peer nodes representing the users (Health Ministry, health organization, patient, requesting entity) in the Ethereum Geth environment. The boot node's role is to connect all the peers. The PoA consensus mechanism is employed for mining in the proposed system. Using the Puppeth tool, only one admin node representing the Health Ministry has been authorized to mine blocks. Since the mining process is not performed by all peers, it reduces energy wastage. Figure 11 illustrates the block sealing and block importing process.

It is evident from this figure that node1 can successfully seal a new block by committing new mining work. Since only node1 has the authority to seal a new block, when node2 or any other node attempts to commit new mining work, block sealing fails with the error message "Unauthorised signer." In our scheme, only one node out of four nodes consumes energy in mining a new block. Hence, our proposed PoA consensus-based system reduces energy consumption in sealing new blocks by a factor of 4 compared to the same system designed using the PoW consensus engine.

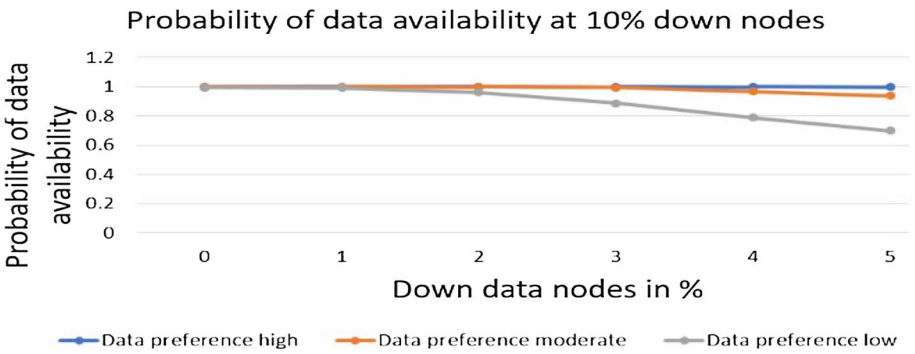


Fig. 9 Analysis of data probability at 10% inactive nodes

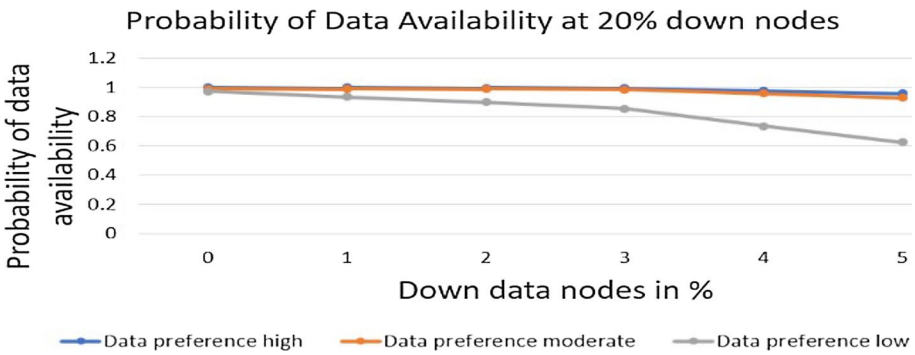


Fig. 10 Analysis of data probability at 20% inactive nodes

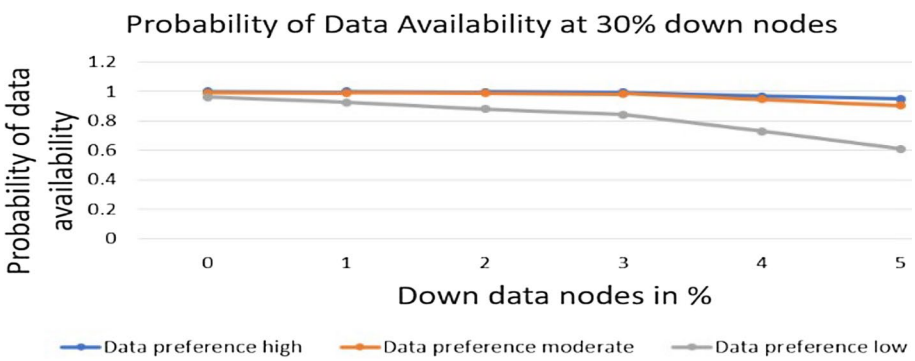


Fig. 11 Analysis of data probability at 30% inactive nodes

6.6 Comparative Analysis

A comparison of the proposed blockchain-based vaccination system with related work in terms of different metrics is presented in Table 2. Limited work related to blockchain-based

vaccination systems exists, with most authors focusing on COVID vaccination records and COVID vaccine supply schemes.

Limited work related to blockchain-based vaccination systems exists, with most authors focusing on COVID vaccination records and COVID vaccine supply schemes. In this work, we propose a blockchain-based scalable and optimal vaccination record management model that differs from prior blockchain-based models in the literature. It is not limited to COVID-19 vaccination but aims to create vaccination records for a patient's entire life, from childhood to death. Some authors have implemented their systems on the Hyperledger platform, while others have used the Ethereum platform. Our proposed system is deployed over the Ethereum private network. Furthermore, the smart contract functionalities used in our proposed system cover the entire vaccination process, which was missing in existing papers. Additionally, during the smart contract testing phase, we attempted to reduce execution fees by introducing various checks. These checks revert invalid transactions to avoid unnecessary execution fees due to the wastage of computational power. Execution costs were not evaluated in most existing works and were not analyzed for reduction. Our proposed system is energy-efficient compared to existing systems as it is based on the PoA consensus mechanism. Scalability is a major issue in the adoption of blockchain technology, and most existing works have not addressed it. In our proposed work, we attempt to provide an optimal solution. Off-chain storage, IPFS, has the potential to improve scalability, which was not widely used in existing works but is utilized in our proposed scheme. We introduced a patient preference option in our proposed system to make it cost-effective. Moreover, we analyzed the IPFS-network parameters to achieve optimal data availability at a low storage cost.

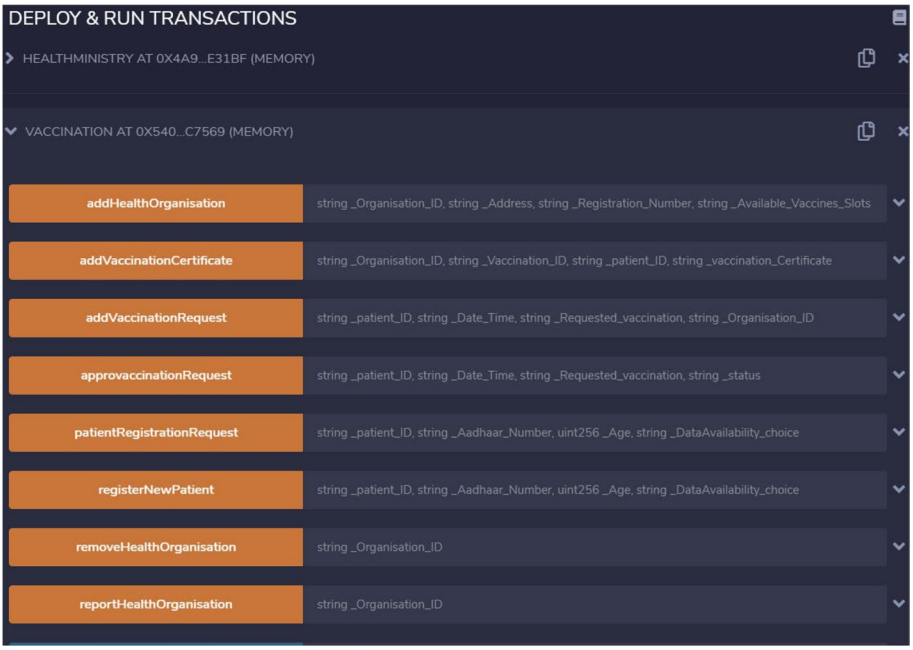


Fig. 12 Result of smart contract deployment write transactions

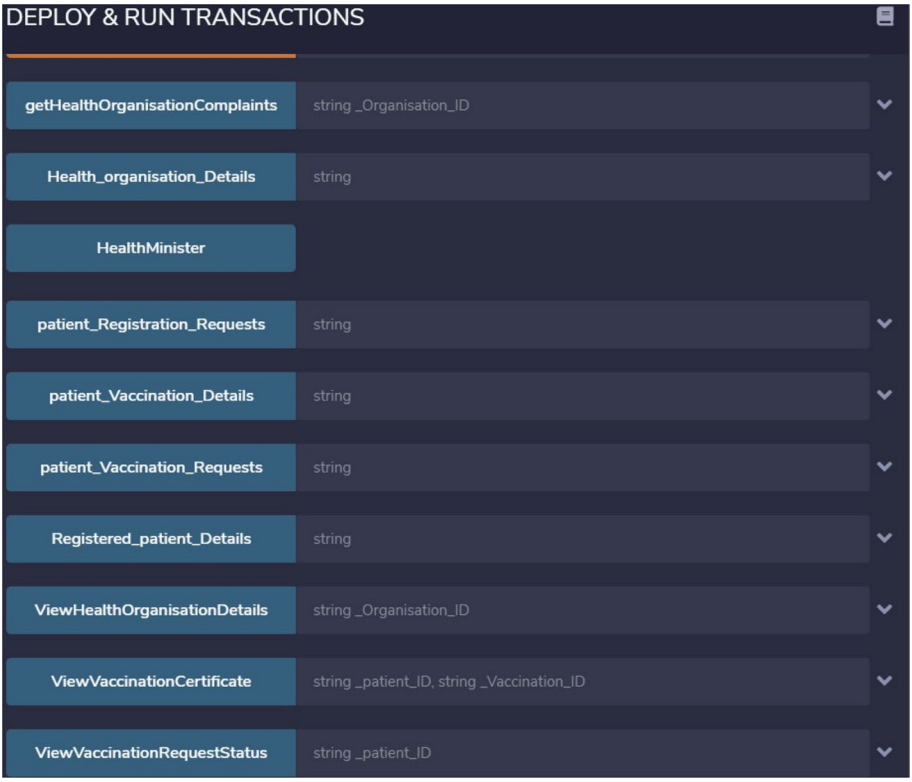


Fig. 13 Result of smart contract deployment read transactions



Fig. 14 Logs triggered upon executing smart contract function on Remix IDE

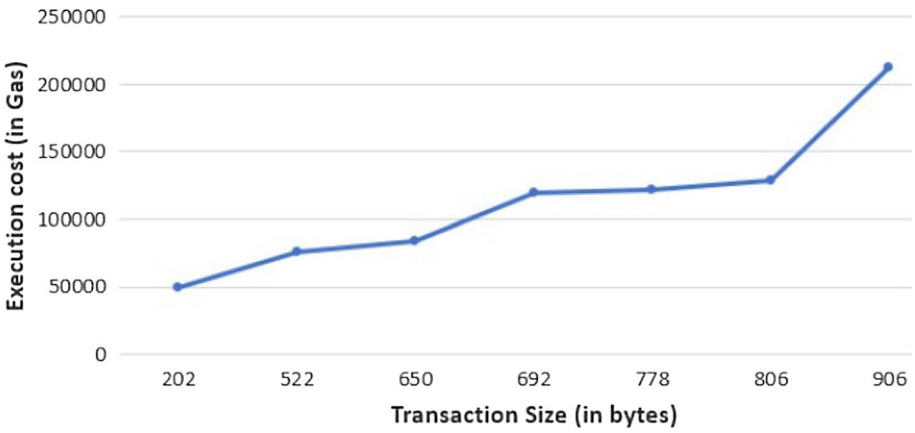


Fig. 15 Execution cost analysis for different transaction sizes

6.7 Security Analysis

The possible vulnerabilities or attacks that an intruder could perform to steal data are:

- DDOS: A malicious attempt to stop a targeted server, service, or network from operating normally by flooding it with traffic from several sources is known as a Distributed Denial of Service (DDoS) attack.
- Ransomware: A form of malware attack where the attacker encrypts and locks the victim's critical files and data, then demands money in exchange for the key to unlock and decrypt the information.
- Man-in-the-Middle (MITM) Attack: An attacker can obtain records by listening in on user communications with the system, changing messages, or taking the identity of a real user.
- Vulnerabilities in Smart Contracts: Although smart contracts are unchangeable once they are implemented, they can have flaws like reentrancy bugs or integer overflow/underflow vulnerabilities that could be used by hackers to alter data or take off funds.
- Phishing Attacks: Cybercriminals may utilize phishing tactics to deceive users into disclosing their private keys or login credentials. This information could then be exploited to obtain unauthorized access to the system.

Leveraging the benefits of blockchain, including immutability, integrity, availability, and transparency, the research has aimed to establish a scalable and cost-effective proof of authority (PoA) consensus-based blockchain solution for vaccination record management. Such a system holds promise for ensuring the availability of accurate records for scheduled vaccinations, medication, migration, education, or employment purposes. The immutable and decentralized nature of the proposed blockchain-based system prevents data theft. Because the recorded data is encrypted, dispersed among several nodes, and tamper-proof, it is very difficult for adversaries to access or change the data. Furthermore, safe data sharing is ensured by cryptographic methods like digital signatures and public-key encryption. The integrity and authenticity of vaccination records are confirmed by the Ethereum blockchain's verification of vaccination certificate hashes (Fig. 16).

Table 2 Comparison of the proposed vaccination scheme with prior works

Qualitative metrics	[1]	[2]	[3]	[4]	[5]	Proposed work
Ethereum based	×	×	✓	✓	✓	✓
Covid-specific	✓	×	×	×	×	✓
Smart contract	✓	✓	✓	✓	×	✓
Energy efficient	×	×	×	×	✓	✓
Cost analysis	×	×	✓	×	×	✓
Cost efficient	×	×	×	×	×	✓
Scalability	×	×	✓	×	×	✓
Data availability analysis	×	×	×	×	×	✓

×: Unavailable, ✓: Available

The security level of the proposed system against online data theft is relatively high due to various features that are listed below:

- The Ethereum blockchain's verification of vaccination certificate hashes protects records from unauthorized changes and tampering by ensuring their authenticity and integrity.
- By decentralizing data storage and lowering the possibility of single-point vulnerabilities, the off-chain storage of vaccination records via IPFS in accordance with patient preferences adds an additional layer of security.
- The system uses a Proof of Authority (PoA) consensus process, which is less vulnerable to attacks and more secure than the energy-intensive Proof of Work (PoW) consensus mechanism.
- Remix IDE-implemented smart contracts with extra features and access controls provide fine-grained control over transaction authorization and data access, reducing the possibility of unauthorized data access or manipulation.
- Frequent execution cost assessment and smart contract performance evaluation aid in the detection of anomalies or suspicious activity, enabling the early identification and mitigation of possible security threats.
- Network parameter analysis, including IPFS-network parameters and patient preferences, minimizes storage costs, minimizes attack surfaces, and minimises possibility to data theft.

We have adopted various strategies to mitigate potential vulnerabilities and enhance the security of the proposed smart contracts. We have analysed the smart contract to identify, validate and address all vulnerabilities. We have fixed all input data to prevent vulnerabilities such as integer overflow/underflow, buffer overflow, and other types of injection attacks. We have used standard libraries and implemented mechanisms such as function modifiers, access control, and permission checks to ensure that only authorized parties can invoke critical functions and perform state-changing operations. We have made the smart contract simple, modular and avoided unnecessary functionality that could introduce potential security risks.

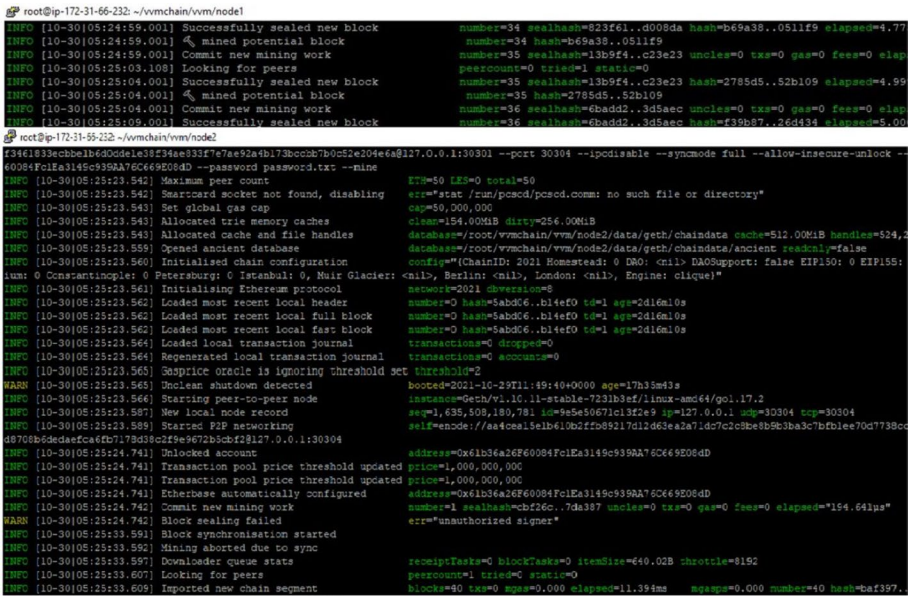


Fig. 16 New block sealing at node 1 and node 2

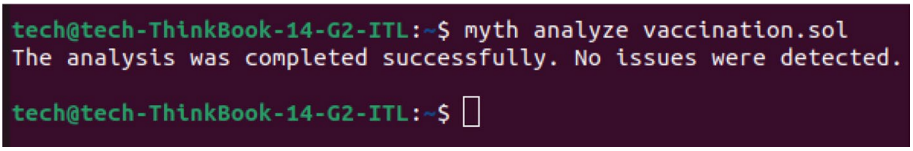


Fig. 17 Analysis report of smart contract vulnerabilities utilizing the mythril tool

We have performed the security analysis for potential vulnerabilities within the smart contracts of our proposed vaccination record management solution. Our model ensures trust, data integrity, availability, non-repudiation, and resilience against cyber-attacks. Although the Ethereum blockchain offers robustness, it remains susceptible to security threats such as network attacks. Network attackers exploit communication weaknesses through DDoS and Sybil attacks. MyThril is a tool renowned for its adeptness in identifying smart contract vulnerabilities like reentrancy bugs or integer overflow/underflow. Our analysis relies on threat analysis, dynamic symbolic execution, and control flow verification to scrutinize EVM bytecode for flaws. The audit of our smart contract yielded reassuring results, as depicted in Fig. 17. This analysis affirms the credibility of our system by providing concrete evidence of its security and demonstrating its robustness and effectiveness in managing vaccination records securely and efficiently.

7 Conclusions

This study proposes a PoA consensus-based blockchain solution for vaccination record management, which differs from prior blockchain-based models in the literature. It is not limited to COVID-19 vaccination records but aims to maintain vaccination records throughout a patient's life. This system enables healthcare organizations to upload vaccination records and allows requesting entities to easily view trustworthy vaccination records at any time. Hence, the proposed system will be beneficial in protecting the public from diseases and in situations where vaccination records are needed for non-medical purposes such as education or employment. Existing prototype systems suffer from limited smart contract functionality, scalability issues, high execution costs, and storage cost problems. Additionally, prior systems have used a PoW consensus engine, which is very energy-consuming, whereas our proposed system is based on PoA. The smart contract functionalities used in our proposed system cover the entire vaccination process with various checks to reduce execution fees. For scalability, availability, and storage cost reduction, vaccination records are stored off-chain through IPFS based on patient preference. All functionalities of the developed smart contract were successfully tested using the Remix IDE environment. We evaluated the performance of the deployed smart contract by analyzing execution costs with different transaction sizes. Moreover, we evaluated data availability based on IPFS-network parameters and patient preference. Furthermore, we analyzed network parameters to achieve optimal data availability at a low storage cost. Comparative analysis shows that our proposed blockchain-based vaccination system outperforms existing systems in terms of different metrics. Further, we presented a security analysis for detecting vulnerabilities using the Mythril tool and proved that our proposed smart contract is threat free.

Author Contributions The complete manuscript was written by Neetu Sharma. The first draft of the manuscript included the conception, design, and methodology of the work, which was reviewed several times. Then the design was implemented to produce the final manuscript. Rajesh Rohilla reviewed and approved the final manuscript.

Funding No sources of funding were received for the preparation of this article.

Availability of Data and Material Not applicable to this article as no datasets were analyzed during the current study.

Declarations

Conflict of interest Neetu Sharma and Rajesh Rohilla declare that they have no conflicts of interest.

Code Availability The authors built all code, and the software application that supports the execution of the code is openly available.

References

1. Fiquaro, M. A., Zahilah, R., Othman, S. H., Arshad, M. M., Saad, S. M. S. (2021).Vaccination system using blockchain technology: A prototype development, in: 2021 3rd International Cyber Resilience Conference (CRC), IEEE, pp. 16

2. Yong, B., Shen, J., Liu, X., Li, F., Chen, H., & Zhou, Q. (2020). An intelligent blockchain-based system for safe vaccine supply and supervision. *International Journal of Information Management, 52*, 102024.
3. Deka, S. K., Goswami, S., Anand, A. (2020). A blockchain based technique for storing vaccination records, in: 2020 IEEE Bombay Section Signature Conference (IBSSC), IEEE, pp. 135139
4. Carniel, A., Leme, F., de Melo Bezerra, J., Hirata, C. M. (2021). A blockchain approach to support vaccination process in a country. In *ICEIS(1)* (pp. 343–350). https://doi.org/10.5220/0010520003430350
5. Eisenstadt, M., Ramachandran, M., Chowdhury, N., Third, A., & Domingue, J. (2020). Covid-19 antibody test/vaccination certication: there's an app for that. *IEEE Open Journal of Engineering in Medicine and Biology, 1*(148), 155.
6. Zhu, P., Hu, J., Zhang, Y., & Li, X. (2021). Enhancing traceability of infectious diseases: a blockchain-based approach. *Information Processing & Management, 58*(4), 102570. https://doi.org/10.1016/j.ipm.2021.102570
7. Cangir, O. F., Cankur, O., & Ozsoy, A. (2021). A taxonomy for blockchain based distributed storage technologies. *Information Processing & Management, 58*(5), 102627. https://doi.org/10.1016/j.ipm.2021.102627
8. Zheng, W., Zheng, Z., Dai, H.-N., Chen, X., & Zheng, P. (2021). Xblock-eos: Extracting and exploring blockchain data from eosio. *In-formation Processing & Management, 58*(3), 102477. https://doi.org/10.1016/j.ipm.2020.102477
9. Wang, Z., Li, M., Lu, J., & Cheng, X. (2022). Business innovation based on articial intelligence and blockchain technology. *Information Processing & Management, 59*(1), 102759. https://doi.org/10.1016/j.ipm.2021.102759
10. Sharma, N., & Rohilla, R. (2023). A novel hyperledger blockchain-enabled decentralized application for drug discovery chain management. *Computers & Industrial Engineering*. https://doi.org/10.1016/j.cie.2023.109501
11. Sharma, N., & Rohilla, R. (2023). A multilevel authentication-based blockchain powered medicine anti-counterfeiting for reliable IoT supply chain management. *The Journal of Supercomputing*. https://doi.org/10.1007/s11227-023-05654-w
12. Sai, A. R., Buckley, J., Fitzgerald, B., & Gear, A. L. (2021). Taxonomy of centralization in public blockchain systems: A systematic literature review. *Information Processing & Management, 58*(4), 102584. https://doi.org/10.1016/j.ipm.2021.102584
13. Sharma, N., & Rohilla, R. (2022). Blockchain based electronic health record management system for data integrity. In *Proceedings of International Conference on Computational Intelligence* (pp. 289297), Springer. https://doi.org/10.1007/978-981-16-3802-2_24
14. Li, C., Zhang, J., Yang, X., & Youlong, L. (2021). Lightweight blockchain consensus mechanism and storage optimization for resource-constrained iot devices. *Information Processing & Management, 58*(4), 102602. https://doi.org/10.1016/j.ipm.2021.102602
15. Dursun, T., & Üstündağ, B. B. (2021). A novel framework for policy based on-chain governance of blockchain networks. *Information Processing & Management, 58*(4), 102556. https://doi.org/10.1016/j.ipm.2021.102556
16. Zaabar, B., Cheikhrouhou, O., Jamil, F., Ammi, M., & Abid, M. (2021). Healthblock: A secure blockchain-based healthcare data management system. *Computer Networks, 200*, 108500. https://doi.org/10.1016/j.comnet.2021.108500
17. Pandey, A. K., Khan, A. I., Abushark, Y. B., Alam, M. M., Agrawal, A., Kumar, R., & Khan, R. A. (2020). Key issues in healthcare data integrity: Analysis and recommendations. *IEEE Access, 8*, 4061240628.
18. Patil, P., Sangeetha, M., & Bhaskar, V. (2021). Blockchain for iot access control, security and privacy: A review. *Wireless Personal Communications, 117*(3), 18151834.
19. Wang, X., Chen, Y., & Zhang, Q. (2021). Incentivizing cooperative relay in utxo-based blockchain network. *Computer Networks, 185*, 107631. https://doi.org/10.1016/j.comnet.2020.107631
20. Alam, S., Shuaib, M., Khan, W. Z., Garg, S., Kaddoum, G., Hossain, M. S., & Zikria, Y. B. (2021). Blockchain-based initiatives: Current state and challenges. *Computer Networks, 198*, 108395. https://doi.org/10.1016/j.comnet.2021.108395
21. Patil, P., Sangeetha, M., & Bhaskar, V. (2022). A consortium blockchain based overseas fund transfer system. *Wireless Personal Communications, 122*(2), 13671389.
22. Platt, M., & McBurney, P. (2021). Sybil attacks on identity-augmented proof-of-stake. *Computer Networks, 199*, 108424. https://doi.org/10.1016/j.comnet.2021.108424

23. Min, X., Kong, L., Li, Q., Liu, Y., Zhang, B., Zhao, Y., Xiao, Z., & Guo, B. (2022). Blockchain-native mechanism supporting the circulation of complex physical assets. *Computer Networks, 202*, 108588. https://doi.org/10.1016/j.comnet.2021.108588
24. Bai, S., Yang, G., Rong, C., Liu, G., & Dai, H. (2020). Qhse: An ecient privacy-preserving scheme for blockchain-based transactions. *Future Generation Computer Systems, 112*, 930944.
25. Ren, Y., Leng, Y., Qi, J., Sharma, P. K., Wang, J., Almakhadmeh, Z., & Tolba, A. (2020). Multiple cloud storage mechanism based on blockchain in smart homes. *Future Generation Computer Systems, 115*, 304–313.
26. Ammi, M., Alarabi, S., & Benkhelifa, E. (2021). Customized blockchain-based architecture for secure smart home for lightweight iot. *Information Processing & Management, 58*(3), 102482. https://doi.org/10.1016/j.ipm.2020.102482
27. Li, M., Lal, C., Conti, M., & Hu, D. (2020). Lechain: A blockchain-based lawful evidence management scheme for digital forensics. *Future Generation Computer Systems., 115*, 406–420.
28. Jing, N., Liu, Q., & Sugumaran, V. (2021). A blockchain-based code copyright management system. *Information Processing & Management, 58*(3), 102518. https://doi.org/10.1016/j.ipm.2021.102518
29. Yang, X., Yi, X., Nepal, S., Kelarev, A., & Han, F. (2020). Blockchain voting: Publicly veriable online voting protocol without trusted tallying authorities. *Future Generation Computer Systems, 112*, 859874.
30. Delgado-Mohatar, O., Sierra-Cámara, J. M., & Anguiano, E. (2020). Blockchain-based semi-autonomous ransomware. *Future Generation Computer Systems., 112*, 589–603.
31. Huang, X., & Zhang, Y. (2020). Indistinguishability and unextractablility of password-based authentication in blockchain. *Future Generation Computer Systems., 112*, 561–566.
32. Chow, S. S., Choo, K.-K.R., & Han, J. (2020). Editorial for accountability and privacy issues in blockchain and cryptocurrency. *Future Generation Computer Systems, 114*, 647–648.
33. Bonnah, E., & Shiguang, J. (2020). Decchain: A decentralized security approach in edge computing based on blockchain. *Future Generation Computer Systems, 113*, 363379.
34. Khan, A. A., Khan, M. M., Khan, K. M., Arshad, J., & Ahmad, F. (2021). A blockchain-based decentralized machine learning framework for collaborative intrusion detection within uavs. *Computer Networks, 196*, 108217. https://doi.org/10.1016/j.comnet.2021.108217
35. Wu, Y., Wang, Z., Ma, Y., & Leung, V. C. (2021). Deep reinforcement learning for blockchain in industrial iot: A survey. *Computer Networks, 191*, 108004. https://doi.org/10.1016/j.comnet.2021.108004
36. Sharmila, A. H., & Jaisankar, N. (2021). Edge intelligent agent assisted hybrid hierarchical blockchain for continuous healthcare monitoring & recommendation system in 5g wban-iot. *Computer Networks, 200*, 108508. https://doi.org/10.1016/j.comnet.2021.108508
37. Arnold, R., & Longley, D. (2021). Continuity: A deterministic byzantine fault tolerant asynchronous consensus algorithm. *Computer Networks, 199*, 108431. https://doi.org/10.1016/j.comnet.2021.108431
38. da Silva Rodrigues, C. K. (2021). Analyzing blockchain integrated architectures for effective handling of iot-ecosystem transactions. *Computer Networks, 201*, 108610. https://doi.org/10.1016/j.comnet.2021.108610
39. Sharma, N., & Rohilla, R. (2020). Blockchain based approach for managing medical practitioner record: A secured design. In *International advanced computing conference* (pp. 7382), Springer. https://doi.org/10.1007/978-981-16-0404-1_6
40. Al-Marridi, A. Z., Mohamed, A., & Erbad, A. (2021). Reinforcement learning approaches for ecient and secure blockchain-powered smart health systems. *Computer Networks, 197*, 108279. https://doi.org/10.1016/j.comnet.2021.108279
41. Pandey, P., & Litoriya, R. (2021). Securing e-health networks from counterfeit medicine penetration using blockchain. *Wireless Personal Communications, 117*(1), 725.
42. Liu, Y., & Xu, G. (2021). Fixed degree of decentralization dpos consensus mechanism in blockchain based on adjacency vote and the average fuzziness of vague value. *Computer Networks, 199*, 108432. https://doi.org/10.1016/j.comnet.2021.108432
43. Liu, Y., Lan, Y., Li, B., Miao, C., & Tian, Z. (2021). Proof of learning (pole): Empowering neural network training with consensus building on blockchains. *Computer Networks, 201*, 108594. https://doi.org/10.1016/j.comnet.2021.108594
44. Joshi, S. (2021). Feasibility of proof of authority as a consensus protocol model, arXiv preprint arXiv:2109.02480
45. Farouk, A., Alahmadi, A., Ghose, S., & Mashatan, A. (2020). Blockchain platform for industrial healthcare: Vision and future opportunities. *Computer Communications, 154*, 223–235.

46. Houtan, B., Had, A. S., & Makrakis, D. (2020). A survey on blockchain-based self-sovereign patient identity in healthcare. *IEEE Access, 8*, 90478–90494.
47. Yuen, T. H. (2020). Pachain: Private, authenticated & auditable consortium blockchain and its implementation. *Future Generation Computer Systems, 112*, 913–929.
48. Chukwu, E., & Garg, L. (2020). A systematic review of blockchain in healthcare: Frameworks, prototypes, and implementations. *IEEE Access, 8*, 21196–21214.
49. Piao, C., Hao, Y., Yan, J., & Jiang, X. (2021). Privacy preserving in blockchain-based government data sharing: A service-on-chain (soc) approach. *Information Processing & Management, 58*(5), 102651. https://doi.org/10.1016/j.ipm.2021.102651
50. Tanwar, S., Parekh, K., & Evans, R. (2020). Blockchain-based electronic healthcare record system for healthcare 4.0 applications. *Journal of Information Security and Applications., 50*, 102407.
51. Sanka, A. I., & Cheung, R. C. (2021). A systematic review of blockchain scalability: Issues, solutions, analysis and future research. *Journal of Network and Computer Applications, 195*, 103232. https://doi.org/10.1016/j.jnca.2021.103232
52. Li, C., Liang, S., Zhang, J., Wang, Q. E., & Luo, Y. (2022). Blockchain-based data trading in edge-cloud computing environment. *Information Processing & Management, 59*(1), 102786. https://doi.org/10.1016/j.ipm.2021.102786
53. Tang, X., Guo, C., Choo, K.-K.R., Liu, Y., & Li, L. (2021). A secure and trustworthy medical record sharing scheme based on searchable encryption and blockchain. *Computer Networks, 200*, 108540. https://doi.org/10.1016/j.comnet.2021.108540
54. Berdik, D., Otoum, S., Schmidt, N., Porter, D., & Jararweh, Y. (2021). A survey on blockchain for information systems management and security. *Information Processing & Management, 58*(1), 102397. https://doi.org/10.1016/j.ipm.2020.102397
55. Zhang, G., Yang, Z., & Liu, W. (2021). Blockchain-based privacy preserving e-health system for healthcare data in cloud. *Computer Networks*. https://doi.org/10.1016/j.comnet.2021.108586
56. Hewa, T., Ylianttila, M., & Liyanage, M. (2021). Survey on blockchain based smart contracts: Applications, opportunities and challenges. *Journal of Network and Computer Applications, 177*, 102857. https://doi.org/10.1016/j.jnca.2020.102857
57. Sookhak, M., Jabbarpour, M. R., Safa, N. S., & Yu, F. R. (2021). Blockchain and smart contract for access control in healthcare: A survey, issues and challenges, and open issues. *Journal of Network and Computer Applications, 178*, 102950. https://doi.org/10.1016/j.jnca.2020.102950
58. Haouari, M., Mhiri, M., El-Masri, M., & Al-Ya, K. (2022). A novel proof of useful work for a blockchain storing transportation transactions. *Information Processing & Management, 59*(1), 102749. https://doi.org/10.1016/j.ipm.2021.102749
59. Song, H., Zhu, N., Xue, R., He, J., Zhang, K., & Wang, J. (2021). Proof-of-contribution consensus mechanism for blockchain and its application in intellectual property protection. *Information Processing & Management, 58*(3), 102507. https://doi.org/10.1016/j.ipm.2021.102507
60. Spataru, A. L., Pungila, C.-P., & Radovancovici, M. (2021). A high-performance native approach to adaptive blockchain smart-contract transmission and execution. *Information Processing & Management, 58*(4), 102561. https://doi.org/10.1016/j.ipm.2021.102561
61. Ekanayake, O. A., & Halgamuge, M. N. (2021). Lightweight blockchain framework using enhanced master-slave blockchain paradigm: Fair rewarding mechanism using reward accuracy model. *Information Processing & Management, 58*(3), 102523. https://doi.org/10.1016/j.ipm.2021.102523

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.



Neetu Sharma Neetu Sharma is currently a full time PhD Scholar at the Delhi Technological University, India. She is PG certified in blockchain technology from IIT Bangalore. Earlier, she worked as an assistant Professor for 10 years. She received his MTech and BE degrees in electronics and communication engineering from RGPV, Madhya Pradesh technical university, India in 2010 and 2007 respectively. Her research interests include blockchain technology, IOT, Machine learning, Deep learning, Image Processing and VLSI. She is gold medallist in MTech and has received Srijin award for good teaching. She has co-guided 10 + MTech thesis and has 38 research papers publications



Rajesh Rohilla Rajesh Rohilla is currently a professor in Delhi Technological university, India. He received his PhD, MTech and BE degree in electronics and communication engineering. His research interests include computer vision, machine learning, embedded system, robotics and digital design

Semi-Empirical DFT Based Investigation of Electronic and Quantum Transport Properties of Novel GS-AGNR (N) FET

Anshul, *Student Member, IEEE*, and Rishu Chaujar, *Senior Member, IEEE*

***Abstract*—In this article, the electronic and quantum transport properties for the bulk configuration of armchair graphene nanoribbons (AGNRs) with varied number of carbon atoms along AGNR width (N) are investigated. The semi-empirical (SE) Density Functional Theory (DFT) approach is used to calculate the band structure, density of states (DOS), and transmission spectrum for the bulk configuration of AGNR. Further, the AGNRs are used in channel material to analyze the performance of field-effect transistors with Gate Stack (GS) architecture. The result shows that the bandgap value is higher for AGNR (N = 4) with a value of 1.98 eV compared to another bulk configuration of AGNRs. In addition to this, AGNR (N = 4) also shows an improved transmission spectrum. Moreover, the transmission spectrum at varied input voltages and projected local density of states (PLDOS) are also analyzed to study the performance of the proposed devices. The parameters mentioned above give a unique idea for evaluating the performance in terms of resonance peaks and electronic structure for device configurations. The off current (I_{off}) is remarkably reduced, and the switching ratio (I_{on}/I_{off}) is significantly improved in GS-AGNR (N = 4) FET compared with other device configurations. Owing to the enhanced switching, this paper highlights GS-AGNR (N = 4) FET as a suitable candidate for low-power applications such as low-power sensors, wireless communication, and medical devices.**

***Index Terms*—Atomic-scale, density functional theory (DFT) analysis, GS-AGNR (N) FET, quantum transport properties.**

I. Introduction

TO INCREASE the functionality of electronic gadgets; the new technology tries to shrink more components on integrated chips (IC). The co-founder of Intel, Gordon Moore, in 1965 predicted that the number of electronic devices would double every 18 months, i.e., the size of the devices would be reduced [1]. The scaling down in ICs will bring significant hurdles in the electronic industry. The majority of electronic devices are built using silicon technology, which has been widely used in the semiconductor industry. According to the International Roadmap for Devices and Systems (IRDS), transistor channel lengths will reach 5 nm by 2021 [2]. However, as the physical limits of silicon transistors are being approached, there is a growing demand for alternative materials and device architectures that can offer superior performance, energy efficiency, and scalability [3]. Graphene nanoribbon field-effect transistors (GNRFETs) have emerged as potential candidates for revolutionizing IC technology in this context [4]. Due to its outstanding electrical properties, graphene, a 2D carbon allotrope, has drawn much attention [5]. This unique material possesses amazing qualities, including outstanding electrical conductivity, excellent mechanical strength, great thermal conductivity, and gas impermeability [6]. As graphene possesses exceptional electrical characteristics, the absence of bandgap limits its employment in electronics. However, it is possible to create a bandgap in graphene through quantum confinement to make graphene nanoribbons (GNRs) [7]. The generally utilized edge configurations for GNRs are the zigzag and armchair edges [8]. Zigzag GNRs (ZGNR) shows a near-zero bandgap and exhibit metallic nature. In the AGNR category, there are three groups, N = 3p, N = 3p + 1, and N = 3p + 2, where N is the number of carbon atoms along AGNR width and p is a positive integer [9]. The 3p+1 and 3p groups in AGNR exhibit a direct energy gap and are semiconducting in nature, while the 3p+2 group is metallic in nature [10]. Semiconducting AGNRs are ideal materials for channels in GNRFET [10]. The materials with a high-κ dielectric constant are needed for the transistor gate oxide in order to upgrade the performance of nanoscale transistors and prevent Short-Channel Effects (SCEs) [11]. While choosing a high-κ dielectric material, a trade-off must be made between the high-κ dielectric and an appropriate bandgap for a transistor [10], [12]. There is an adverse relationship between the bandgap and dielectric constant. Implementing a high-κ dielectric layer on the AGNR channel can generate undesirably strong fringing fields at the source and drain contacts, which can cause various issues like bandgap degradation, threshold voltage instability, and mobility degradation [13]. The issues above, like bandgap degradation, threshold voltage instability, and mobility degradation, can be overcome using gate stack (GS) architecture [14]. In GS architecture, a combination of two dielectric layers has been used. Specifically, a low-dielectric layer of SiO_2 (κ = 3.9) has been sandwiched between the high-κ dielectric layer and the GNR channel. The narrow SiO_2 layer in GS stabilizes the high-κ dielectric ($LaAlO_3$) layer and AGNR substrate without

Manuscript received 7 December 2023; revised 16 March 2024; accepted 23 April 2024. Date of publication 30 April 2024; date of current version 7 May 2024. This work was supported by the Department of Applied Physics, Delhi Technological University, New Delhi. The review of this article was arranged by Associate Editor K. Namura. *(Corresponding author: Rishu Chaujar.)*

The authors are with the Microelectronics Research Lab, Applied Physics Department, Delhi Technological University (formerly the Delhi College of Engineering), New Delhi 110042, India (e-mail: anshul_2k20phdap502@dtu.ac.in; chaujar.rishu@dtu.ac.in).

Digital Object Identifier 10.1109/TNANO.2024.3394547

1536-125X © 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 May 09,2024 at 07:10:44 UTC from IEEE Xplore. Restrictions apply.

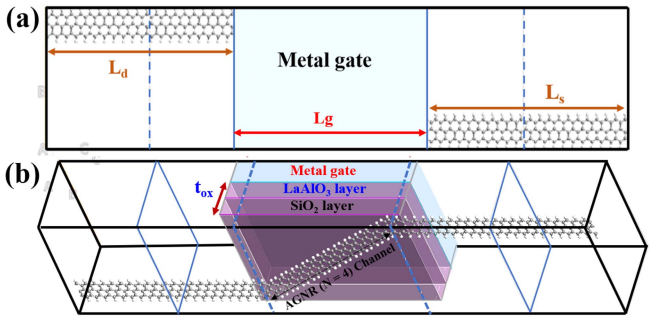


Fig. 1. (a) Front view and (b) rotated view of the proposed device taken from the ATK Simulator.

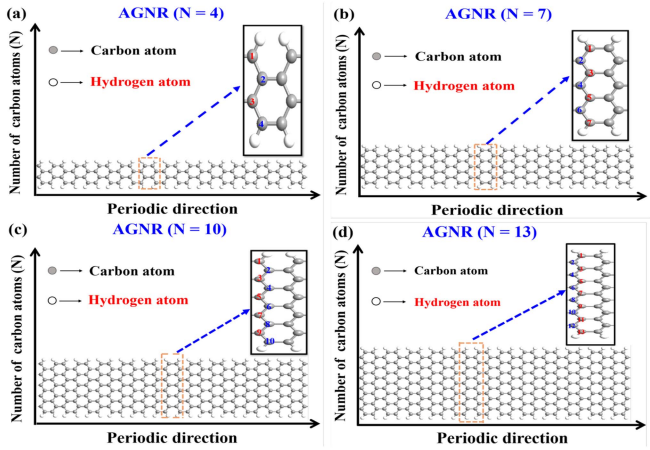


Fig. 2. (a)–(d) Optimized atomic structure of AGNR having different number of atoms used as channel material.

affecting the device's effective oxide thickness. The high-κ dielectric in GS will permit the increase in physical oxide thickness to prevent gate tunneling [15] and thus improves the carrier efficiency, reduce off current, and enhance the device performance [16]. Taking everything into consideration, we propose a Gate Stack Armchair Graphene Nanoribbon Field Effect Transistor (GS-AGNR (N) FET), where N is the number of carbon atoms along AGNR width used in the channel material of AGNR. In this research article, the authors primarily analyzed the electronic properties of AGNR having N = 4, 7, 10, and 13 using the DFT approach. These values of N belong to the 3p+1 group of AGNR because it has a better bandgap as compared to other groups. Further, the performance analysis of all device configurations is studied in terms of transmission coefficient, off current, switching ratio, and projected local density of states. All these parameters are calculated using the SE calculator with the Slater-Koster model having Density Functional Tight Binding (DFTB) parameterization.

II. Device Design and Description

The simulated architecture of the proposed device GS-AGNR FET is shown in Fig. 1. An AGNR layer of 50 Å is used as channel material considering the fact of scaling down [17]. The value of N in the AGNR varies for different types of device configurations. The optimized atomic structure for AGNR having values of N is shown in Fig. 2. From the above Fig. 2, it can be observed that the periodic direction of all widths is kept the same. The highlighted box in Fig. 2(a)–(d) shows one atom view, including the number of atoms along AGNR width. Further, the AGNR is rotated 30° to make a perfect interface with the ZGNR-type contacts, as shown in Fig. 1(b). The length of the drain and source is also taken 50 Å. A gate stacking architecture is used after a metallic gate. Step-increasing doping profile of n-type is used in all regions to make a junctionless configuration having a concentration of 3×10^{16} in the channel region and 5×10^{18} in the source and drain region. Gate potential plays an important role in calculating the electronic properties of a device, and hence, a metallic gate is employed over the dielectric layers [18]. Additional benefits of metal gates are their great thermal stability, high level of purity, and compatibility with Complementary Metal-Oxide-Semiconductor (CMOS) processing. Here, the thickness of the oxide layer (t_{ox}) is taken as 1 nm (0.50 nm $LaAlO_3$ layer + 0.50 nm SiO_2 layer). The t_{ox} is inversely related to the gate oxide capacitance (C_{ox}), i.e., reducing t_{ox} will increase the C_{ox}, which provides better control of the gate over the channel as the potential drop on gate oxide (Φ_{ox}) decreases. Further, the parameters can be co-related by a mathematical expression as follows [19]:

$$t_{ox} = \frac{\varepsilon_0 \varepsilon_r}{C_{ox}} \tag{1}$$

$$\Phi_{ox} = \frac{Q_0}{C_{ox}} \tag{2}$$

Now, putting the value of C_{ox} from (1) in (2)

$$\Phi_{ox} = \frac{Q_0 t_{ox}}{\varepsilon_0 \varepsilon_r} \tag{3}$$

where,

Φ_{ox} – gate oxide potential
Q_0 – surface charge density
t_{ox} – thickness of the oxide layer
C_{ox} – gate oxide capacitance
ε_0 – permeability coefficient
ε_r – the dielectric constant of the gate insulation

III. Simulation Methodology and Fabrication Techniques

The Atomistic Tool Kit (ATK) is used to carry out all the simulations. The semi-empirical (SE) calculator uses the Slater-Koster model with the DFTB parameter group in ATK to perform all the analyses. The Slater-Koster model is frequently used to create the tight-binding Hamiltonian matrix for a given system [20], [21]. This includes determining the matrix elements of the Hamiltonian between orbitals of atoms. The equation for the matrix elements of the tight-binding Hamiltonian in the Slater-Koster model is given as [21]:

$$H_{ij} = \sum_{\alpha\beta} t_{\alpha\beta}\left(R_{ij}\right) c_{i\alpha}^{\dagger} c_{j\beta} \tag{4}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:10:44 UTC from IEEE Xplore. Restrictions apply.

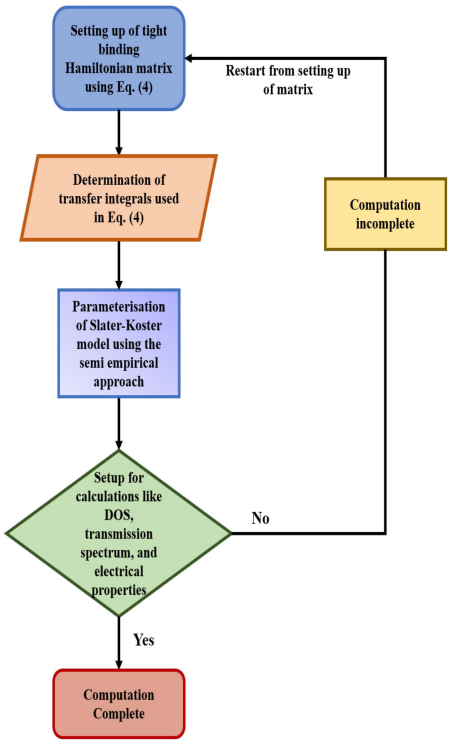


Fig. 3. Process flow chart for simulation approach.

where,

H_{ij} – matrix element of the Hamiltonian between atoms i and j

$t_{\alpha\beta}(R_{ij})$ – Slater-Koster transfer integral

$c^{\dagger}_{i\alpha}$ – creation operators for atomic orbitals

The non-equilibrium Green's function (NEGF) method without a self-consistent tight-binding function is adopted for the present study. The non-self-consistent portion of the tight-binding Hamiltonian is parametrized in SE using a two-center approximation, meaning that the matrix components are independent of the positions of the other atoms and only depend on the distance between two atoms. The Hartree potential is kept constant at 40 Ha, and the K-point sampling is taken $1\times1\times100$. The electrode temperature is kept at 300 K. Bulk configured AGNR electronic properties are calculated using the Fast Fourier Transform (FFT) Poisson solver, whereas device properties are calculated using the multi-grid type Poisson solver. Firstly, all structures are optimized using the Brenner_CH_2002 with a force tolerance of 0.05 eV/Å. Further, the electrical properties of a fully optimized structure are computed using DFT and NEGF formalisms. The process flow chart for the simulation approach is shown in Fig. 3. The step-by-step fabrication techniques for GS-GNRFET [22] is illustrated in Fig. 4.

IV. Model Calibration

The authors validated the simulation model by previously reported research findings in the literature [23]. To calibrate the simulation models, the MOSFET has been designed using

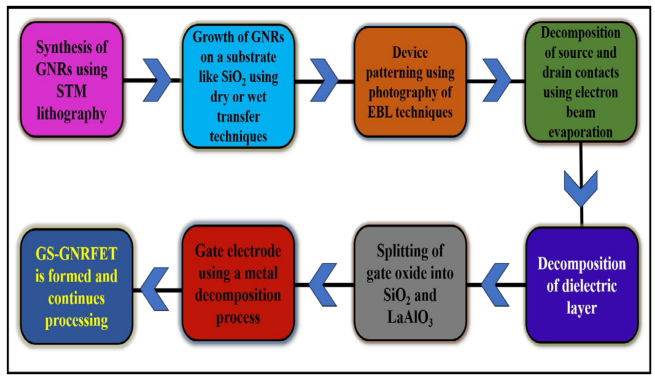


Fig. 4. Proposed fabrication flow for GS-GNRFET.

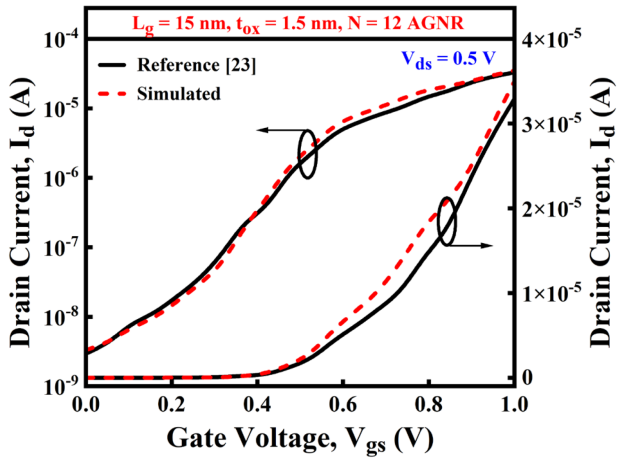


Fig. 5. Transfer characteristics calibration of simulated and reference data [23].

AGNR (N = 12) material in the entire channel region and fixed device dimensions (L_g = 15 nm, t_{ox} = 1.5 nm), the source and the drain are doped extensions of GNRs, and SiO_2 (κ = 3.9) as gate oxide as shown in reference [23]. The NEGF approach with a self-consistent tight-binding function is used to solve the Schrödinger equation to obtain device characteristics for MOSFET. Fig. 5 depicts the transfer characteristics at drain voltage, V_{ds} = 0.5 V, of the extracted data from the reference [23] and results obtained from the simulations. The left side of Fig. 5 shows the change in drain current (I_d) on the logarithmic scale with gate voltage (V_{gs}), while the right side of Fig. 5 depicts the linear change of I_d with V_{gs}. The logarithmic scale of I_d with V_{gs} depicts leakage current (I_{off}) in the device, aiding in the prediction of subthreshold slope, while the linear change of I_d with V_{gs} predicts the value of threshold voltage (V_{th}).

It can be observed from the left and right sides of Fig. 5 that the results of the simulations are consistent with previous findings. The near concordance between the simulated and reference data sets validates the choice of simulation models. Once the credibility of the simulation with the approach is established, a GS architecture is employed in the GNR-FET. The detailed simulation process of the present study for GS-AGNR (N) FET is discussed in Fig. 3.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:10:44 UTC from IEEE Xplore. Restrictions apply.

V. Result and Discussion

In this section, the DFT analysis of AGNR with varied N belonging to the (3p+1) family is studied in terms of band structure, density of states (DOS), and transmission coefficient. Further, the AGNR is used as a channel material in GS- AGNR FET with the varied number of carbon atoms along AGNR width, and then the performance analysis of the proposed devices has been carried out using the DFT and NEGF approach.

A. DFT Investigation of Electronic Properties of AGNR With Different Values of N

The band structure and DOS for AGNR (N = 4, 7, 10, 13) is depicted in Fig. 6. The band structure is used to compute the band gap, which shows the number of energy levels that an electron can have as illustrated in Fig. 6(a), (c), (e), and (g). High symmetric k-points, or Dirac points, are plotted on the x-axis, and energy (eV) is plotted on the y-axis to represent band structures. The values of bandgap are 1.98, 1.21, 0.87, and 0.68 eV for AGNR (N = 4), AGNR (N = 7), AGNR (N = 10), and AGNR (N = 13), respectively. From the bandgap analysis, it is observed that the 3p+1 family of AGNR, i.e., N = 4, 7, 10, and 13 having values of p = 1, 2, 3, and 4, are semiconductor in nature and can be used as a channel material in FETs. It is also observed that the value of the bandgap decreased as the widths of AGNR increased. The width of AGNR is given by [24]:

$$W_{AGNR} = \frac{\sqrt{3}a_{cc}\ (N+1)}{2} \tag{5}$$

where,

W_{AGNR} – width of AGNR
acc – bond length of carbon-carbon atom in AGNR
N – number of carbon atoms along AGNR width

The width of AGNR is 1.8, 2.88, 3.96, and 4.2 nm for N = 4, 7, 10, 13 respectively. From the bandgap analysis, we observed that W_{AGNR} and bandgap have an adverse relationship. The number of unoccupied sites that are accessible for electrons to fill at a given energy level determines the DOS, as illustrated in Fig. 6(b), (d), (f), and (h). From these figures, the authors observed that there is no peak at the Fermi level in DOS. Also, in accordance with the energy states that are available, DOS exhibits similarities to band structure. The semiconducting nature of the AGNR is verified by the fact that there is no peak at the Fermi level in DOS. Further, the number of DOS peaks increases as the value of width (W_{AGNR}) increases. In AGNR (N = 4), the DOS peaks in the conduction band (CB) are observed at 1.46 eV and in the valance band (VB) at −0.68, −1.68 eV. In AGNR (N = 7), the peaks in CB are at 0.80 and 1.30 eV, while in VB, peaks are observed at −0.84 and −1.22 eV. In the case of AGNR (N = 10) and AGNR (N = 13), DOS peaks in CB are observed at 1.94 and 1.68 eV, and in VB peaks are observed at −1.74 and −1.48 eV. From the above analysis, the authors observed that as the value of N in AGNR increased, the large number of peaks shifted in CB. Further, an increase in electron energy opens more states to occupy an electron in the CB. The

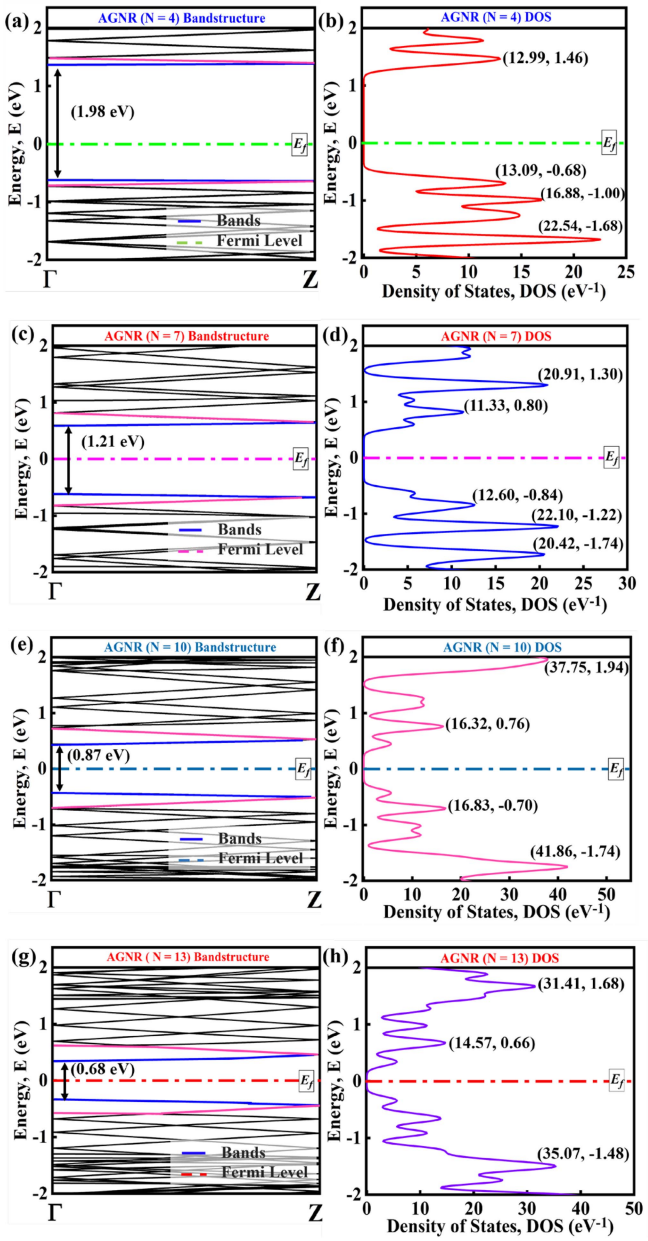


Fig. 6. Quantum transport properties for bulk configuration of AGNR with different N that are used as a channel material (a) AGNR (N = 4) band structure, (b) AGNR (N = 4) DOS, (c) AGNR (N = 7) band structure, (d) AGNR (N = 7) DOS, (e) AGNR (N = 10) band structure, (f) AGNR (N = 10) DOS, (g) AGNR (N = 13) band structure and, (h) AGNR (N = 13) DOS.

relation between DOS and energy is given by [25]:

$$D\ (E) = \frac{N\ (E)}{dE} \tag{6}$$

where,

D (E) - The density of states for the available energy range

The analysis of the transmission spectrum for diff6erent N in AGNR is illustrated in Fig. 7. The transmission spectrum of a nanoribbon (GNR) describes how the transmission of electrons changes as their energy varies while passing through the nanoribbon. This transmission is affected by the GNR structure,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:10:44 UTC from IEEE Xplore. Restrictions apply.

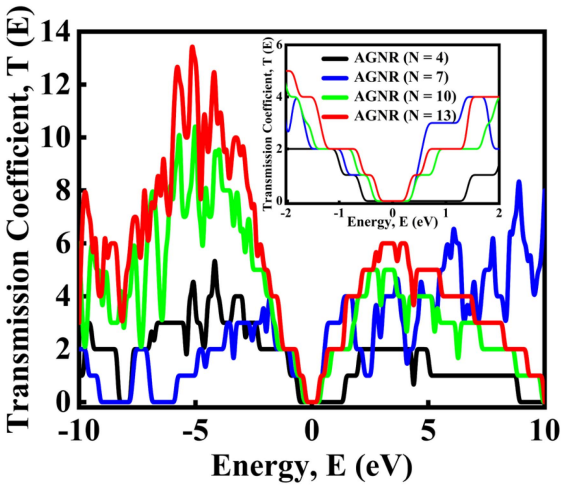


Fig. 7. Transmission spectrum for varied N in the AGNR.

which in turn depends on its width and other structural features. The energy range for the transmission coefficient is taken from -10 to $+10$ eV. It is observed from Fig. 7 that the peaks of the transmission spectrum increased as the value of N in AGNR increased. Also, no peak is observed at Fermi energy (i.e., for E = 0 eV), which points out that there is a possibility of conduction between the CB and VB. From Fig. 7, it is observed that the AGNR (N = 4) has no peaks in the energy range of - 0.5 to + 1.5 eV, which describes that an electron can easily tunnel through VB to CB. From the above discussions, the authors lead to an observation that the AGNR (N = 4) has improved transmission spectrum peaks around Fermi energy and has a better value of bandgap.

B. Performance Analysis of the Proposed Devices Using the DFT Approach

From the DFT analysis of AGNR, as discussed in Figs. 6 and 7, it is observed that the AGNR (N = 4) shows improved results in terms of band structure and transmission spectrum. Further, the transport properties for different device configurations with varying values of N in AGNR for the channel material are calculated. The authors have analyzed the impact of drain voltage (V_{ds}) on the transmission spectrum for each device configuration as depicted in Fig. 8. The transmission coefficient for a device configuration is obtained by the without self-consistent solution of the Schrodinger equation and 3-D Poisson's equations using the Slater-Koster model with NEGF method and it can be mathematically expressed as [26]:

$$T\ (E) = Trace\left[G\,(E)\,\Gamma_d\,(E)\,G^{\dagger}\,(E)\,\Gamma_s\,(E)\right] \qquad (7)$$

where,

$T(E)$ – transmission coefficient
$G(E)$ – retarded Green's function
$G^{\dagger}(E)$ – advanced Green's function
$\Gamma_L(E)$ – broadening function of the left electrode
$\Gamma_R(E)$ – broadening function of the right electrode

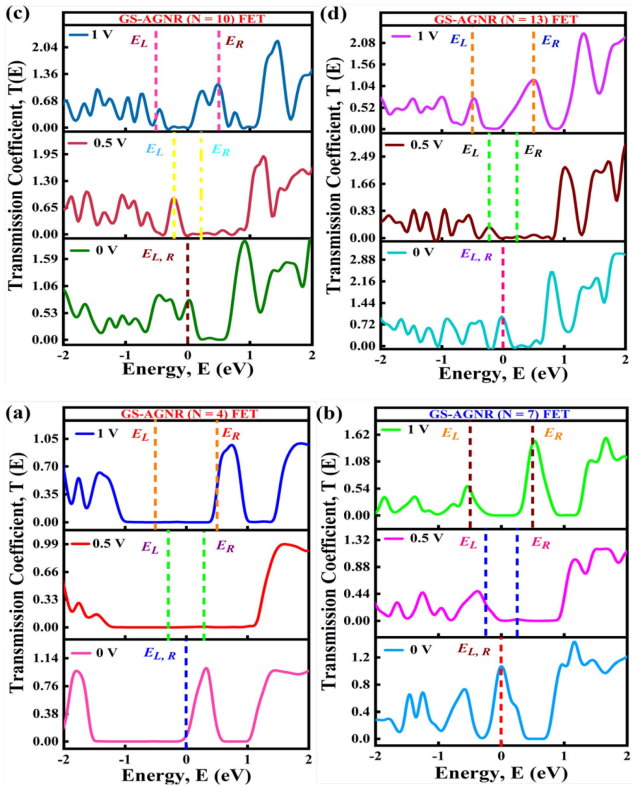


Fig. 8. Transmission spectrum at input voltages 0 V, 0.5 V, and 1 V for (a) GS-AGNR (N = 4) FET, (b) GS-AGNR (N = 7) FET, (c) GS-AGNR (N = 10) FET, (d) GS-AGNR (N = 13) FET.

The value of V_{ds} is changed from 0 to 1 V in 0.5-step increments. A new without self-consistent calculation needs to determine the associated transmission spectrum for every drain voltage. Once the transmission spectrum is calculated, it is observed that the Fermi level shift at the VB and CB edges is the primary distinction between the various spectra calculated from the parameters $\Gamma_L(E)$ and $\Gamma_R(E)$. The terms E_L and E_R in Fig. 8 denote the shift in the Fermi level corresponding to the left and right electrodes, respectively. From Fig. 8(a), it is observed that the GS-AGNR (N = 4) FET at 0 V shows a minimum resonance peak. The width and electrical characteristics of the GNR would determine the number of resonance peaks and their specific positions.

It is evident from Fig. 8(a) that the possibility of electrons tunneling for 0.5 V in GS-AGNR (N = 4) FET is greater than the other device configurations at this voltage (0.5 V). The reason can be understood by the Fermi levels shift in the left and right electrodes. The Fermi level shown in the presented work governs the range of electron transport inside a GS-AGNR (N) FET. The electron transmission inside the GNR channel increases with the value of the Fermi level range in GNR. It can be clearly observed from Fig. 8(a)–(d) that for 0.5 V, the range of Fermi level shift for both electrodes are higher in GS-AGNR (N = 4) FET as compared to the other three device configurations. Also, in Fig. 8(a) at 0.5 V, there is no peak present in the barrier of E_L and E_R, which points out the higher possibility of conductance as compared to the transmission spectrum, as shown in Fig. 8(b)–(d).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:10:44 UTC from IEEE Xplore. Restrictions apply.

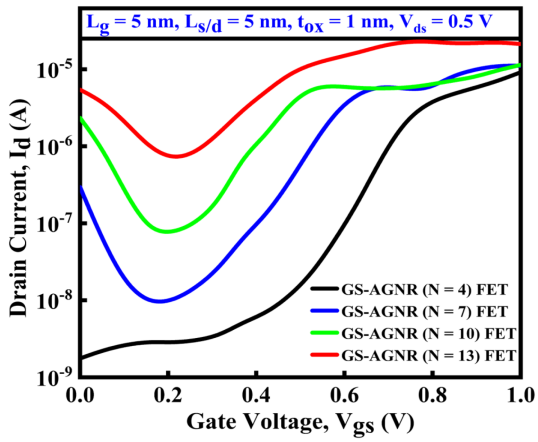


Fig. 9. Transfer characteristics of GS-AGNR (N) FETs with different values of N in the channel material.

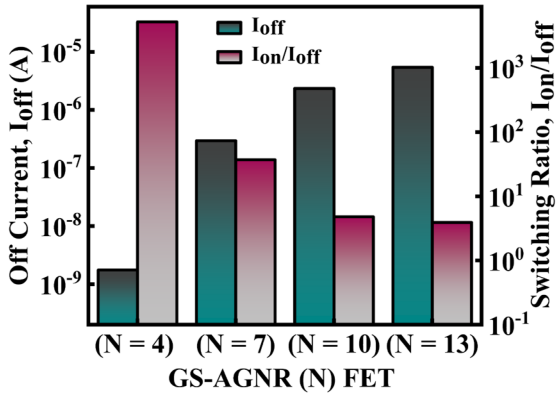


Fig. 10. Variation in off current and switching ratio for GS-AGNR FET.

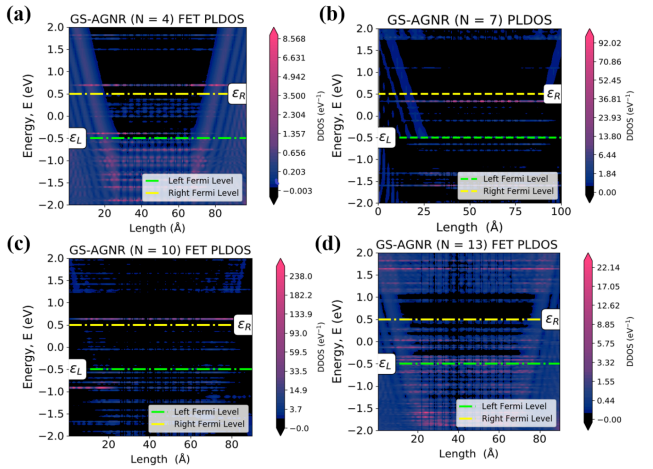


Fig. 11. Projected local density of states (PLDOS) for (a) GS-AGNR (N = 4) FET, (b) GS-AGNR (N = 7) FET, (c) GS-AGNR (N = 10) FET and, (d) GS-AGNR (N = 13) FET.

The values of T(E) at 0 V and 0 eV are 1.09 in GS-AGNR (N = 7) FET, which is higher than in other device configurations. When the drain voltages change from 0 to 1 V, it is observed that the value of T(E) decreases, and the area under the curve also decreases. In GS-AGNR (N = 10) FET and GS-AGNR (N = 13) FET, as depicted in Fig. 8(c) and (d), the values of T(E) at 0 V and 0 eV are 0.76 and 0.97. Also, as the V_{ds} change from 0 to 0.5 V, the resonance peak value decreases. Further, the authors observed that the variation of the transmission spectrum for GS-AGNR (N) FET having different values of N in channel material shows the transport of electrons and its dependence on external voltage. Moreover, the DFT analysis of GS-AGNR FET in terms of T(E) gives a unique approach to explain the device performance, which can be very useful in many applications.

The transfer characteristics of GS-AGNR (N) FETs with different values of N are shown in Fig. 9. The drain current is calculated after using the without self-consistent solution of tight-binding Hamiltonian matrix after the calculation T(E) and mathematically expressed by Landauer-Büttiker Formalism [27]:

$$I_d = \frac{2e}{h} \int_{-\infty}^{\infty} T(E) \left[f_{FD}(E - \mu_d) - f_{FD}(E - \mu_s) \right] dE \tag{8}$$

where,

I_d – drain current
f_{FD} – Fermi-Dirac distribution
μ_d – chemical potential of the drain electrode
μ_s – chemical potential of the source electrode

The (8) depicts the dependence of I_d on T (E). The GS-AGNR (N = 4) FET shows improved value of I_{off} as compared to other device configurations. The value of I_{off} changed from 1.75 nA to 5.42 μA for GS-AGNR (N = 4) FET and GS-AGNR (N = 13) FET. The narrower width of AGNR shows improved results in terms of I_{off}, switching ratio (I_{on}/I_{off}), and other parameters like transmission probability and conductance, which can be easily understood by analyzing the transmission spectrum for each device configuration. The transmission spectrum and transfer characteristics of a GNRFET are related through the electronic properties of the GNR and the operation of the FET. The discussion about the variation of transmission spectrum in terms of T(E) for different device configurations led to an observation that the transfer characteristics, as shown in Fig. 9, have been improved significantly for GS-AGNR (N = 4) FET. From the transfer characteristics, the switching ratio (I_{on}/I_{off}) and the off current (I_{off}) are calculated.

The variation in I_{off} and I_{on}/I_{off} ratio for all purposed device configurations is illustrated in Fig. 10 . The switching ratio is an important factor that determines the performance of a device [28], [29]. A high I_{on}/I_{off} ratio indicates that the device can reliably and rapidly switch between conducting (on) and non-conducting (off) states, enabling it to amplify and process electrical signals with minimal power consumption. In this present study, the GS-AGNR (N = 4) FET shows a higher switching ratio value than other device configurations. Based on the I_{on}/I_{off} parameter, the authors have observed that the GS-AGNR (N = 4) FET can be used in high-speed electronics and low-power integrated IC because of reduced off-current and better switching ratio.

Fig. 11 portrays the contour plot of the projected local density of states (PLDOS) for all device configurations. These contour

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:10:44 UTC from IEEE Xplore. Restrictions apply.

TABLE I
COMPARISON AMONG PROPOSED GS-AGNR(N)FETS WITH PREVIOUSLY REPORTED GNRFETS

Device	V_{ds} (V)	Channel AGNR (N)	I_{on} (μA)	I_{off} (μA)	I_{on}/I_{off}	Bandgap (eV)
GS-AGNR (N = 4) FET*	0.5	N = 4	9.14	1.75×10^{-3}	5.22×10^{3}	1.98
GS-AGNR (N = 7) FET*	0.5	N = 7	0.11	0.295	37	1.21
GS-AGNR (N = 10) FET*	0.5	N = 10	0.12	2.34	4.8	0.87
GS-AGNR (N = 13) FET*	0.5	N = 13	0.21	5.42	3.9	0.68
SB-GNRFET [4]	0.5	N = 12	-	-	33.7	-
AG-GNRFET [30]	0.5	N = 12	-	-	304	-
DMG-GNRFET [31]	0.6	N = 13	-	-	5823	-
SLD-GNRFET [32]	0.4	N = 12	3.66	15.7×10^{-4}	2.3×10^{3}	-

*Present Work

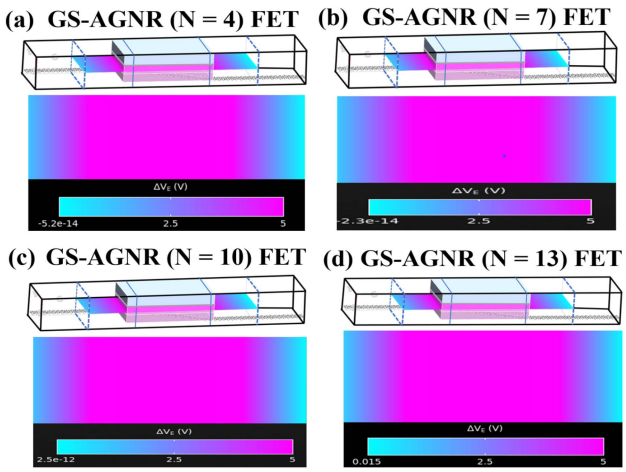


Fig. 12. Electrostatic difference potential profile for (a) GS-AGNR (N = 4) FET, (b) GS-AGNR (N = 7) FET, (c) GS-AGNR (N = 10) FET, and (d) GS-AGNR (N = 13) FET.

plots of PLDOS in GS-AGNR (N) FETs provide crucial insights into the electronic structure and behavior of these atomic-scale devices. In Fig. 11. the PLDOS is typically represented as a function of energy and position within the GNR. Each contour line in the graph corresponds to a specific energy level, and the intensity or color bar of the contour lines reflects the density of electronic states at that energy level. The magnitude of these contour lines associated with device density of states (DDOS) varies from -0.03 to 238 eV^{-1}. The contour plots of electrostatic difference potential in Fig. 12 show crucial insights into the distribution of electric charges within different device configurations. The contour plots clearly demonstrate where electric charges are concentrated and how they affect the device's overall performance. The blue color contour depicts the electronic distribution near the electrodes, whose magnitude ranges from 0.015 to -5.2×10^{-14} V, while the pink color illustrates the distribution of electric charges inside the channel. These contour plots can be used by scientists and engineers to improve the design of GNR FETs. They can fine-tune the devices having a narrower number of carbon atoms along AGNR width (like N = 4, as presented in this work) for certain uses, like high-speed computing or low-power electronics, by adjusting the charge distribution and electric field with exact atomic-level modifications.

Table I summarizes the comparison among proposed GS-AGNR(N) FETs with previously reported GNRFETs. The variation of N from 4 to 13 in GS-AGNR (N) FETs brings down the magnitude of the on current (9.14 μA to 0.21 μA) and increases the off current (1.75×10^{-3} to 5.4 μA). These variations in the on and off currents due to a varied N value in GS-AGNR(N) FETs imparts a switching ratio of 3.9 to 5.22×10^{3} corresponding to bandgap in the range of 0.68 to 1.98 eV.

VI. CONCLUSION

The Nobel prize-winning discovery of a 2D-monolayer structure in 2004 by Geim and Novoselov at Manchester University paved the path for this research. In this study, the armchair configuration of graphene with a varied number of carbon atoms along AGNR width is adopted to design GS-AGNR(N)FET. The density functional theory (DFT) assisted with a solution of 3-D Poisson's equations and a non-self-consistent solution of the Schrodinger equation using the Slater-Koster model with the NEGF method adopted for evaluating all the parameters. These parameters include band structure, the density of states, and the transmission spectrum of AGNR, which can potentially be used as channel materials. Further, the impact of V_{ds} on the transmission spectrum is analyzed for proposed device configurations. The DFT analysis of GS-AGNR FET in terms of T(E) gives a unique approach to explain the device performance, which can be very useful in many applications. The band structure and DOS for AGNR (N = 4, 7, 10, 13) concluded that the 3p+1 family of AGNR with a value of N ranging from 1 to 4 is semiconducting in nature and confirmed the potential alternative as a channel material in FETs. Moreover, the semiconducting nature of the AGNR is verified by the fact that there is no peak at the Fermi level in DOS. The value of I_{off} changed from 1.75 nA to 5.42 μA for GS-AGNR (N = 4) FET and GS-AGNR (N = 13) FET, respectively. The narrower width of AGNR shows improved results in terms of I_{off}, switching ratio, transmission probability, conductance, etc. The results of the transmission spectrum in terms of T(E) for varied input device configurations led to the conclusion that the transfer characteristics have been improved significantly for GS-AGNR (N = 4) FET. The contour plots of electrostatic difference potential show the distribution of electric charges in device configuration. In the context of GS-AGNR (N = 4) field-effect transistors, the off current (I_{off}) experiences a substantial reduction, while the switching ratio (I_{on}/I_{off}) exhibits significant improvement when

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:10:44 UTC from IEEE Xplore. Restrictions apply.

compared with other device configurations. The electronic and quantum transport properties studied in this research provide valuable insights into performance evaluation in terms of I_{on}, I_{off}, switching ratio, and bandgap, along with resonance peaks of transmission coefficient and electronic structure. Thus, from this study, the authors conclude that the GS-AGNR (N = 4) FET has better-switching performance and can be a potential alternative to conventional FETs in low-power applications such as low-power sensors, wireless communication, and medical devices.

ACKNOWLEDGMENT

The authors extend a warm thanks to the MRL facility in the Applied Physics Department at Delhi Technological University for their assistance with this research.

REFERENCES

[1] T. Radsar, H. Khalesi, and V. Ghods, "Graphene nanoribbon field effect transistors analysis and applications," *Superlattices Microstructures*, vol. 153, 2021, Art. no. 106869.

[2] B. Hoefflinger, "IRDS—International roadmap for devices and systems, rebooting computing, s3s," in *NANO-CHIPS 2030: -Chip AI for an Efficient Data-Driven World*. Berlin, German, Springer, 2020, pp. 9–17.

[3] B. Kumar and R. Chaujar, "Numerical simulation of analog metrics and parasitic capacitances of GaAs GS-GAA FinFET for ULSI switching applications," *Eur. Phys. J. Plus*, vol. 137, no. 1, 2022, Art. no. 110.

[4] Y. Y. Chen et al., "A SPICE-compatible model of MOS-type graphene nano-ribbon field-effect transistors enabling gate-and circuit-level delay and power analysis under process variation," *IEEE Trans. Nanotechnol.*, vol. 14, no. 6, pp. 1068–1082, 2015.

[5] A. S. Silva, S. M. Correa, C. A. B. Silva, and J. D. Nero, "Electronic transport and its inelastic effects for a doped phagraphene device," *J. Appl. Phys.*, vol. 128, no. 5, 2020, Art. no. 055104.

[6] K. S. Novoselov et al., "Electric field effect in atomically thin carbon films," *Science*, vol. 306, no. 5696, pp. 666–669, 2004.

[7] H. Shen, Y. Shi, and X. Wang, "Synthesis, charge transport and device applications of graphene nanoribbons," *Synthetic Met.*, vol. 210, pp. 109–122, 2015.

[8] A. Hazra and S. Basu, "Graphene nanoribbon as potential on-chip interconnect material—A review," *J. Carbon Reasearch*, vol. 4, no. 3, 2018, Art. no. 49.

[9] F. Schwierz, "Graphene transistors," *Nature Nanotechnol.*, vol. 5, no. 7, pp. 487–496, 2010.

[10] J. M. M. Tejada and J. V. Medina, "Review on graphene nanoribbon devices for logic applications," *Microelectronics J.*, vol. 48, pp. 18–38, 2016.

[11] B. Kumar, M. Sharma, and R. Chaujar, "Junctionless-accumulation-mode stacked gate GAA FinFET with dual-k spacer for reliable RFIC design," *Microelectronics J.*, vol. 139, 2023, Art. no. 105910.

[12] T. Radsar, H. Khalesi, and V. Ghods, "Improving the performance of graphene nanoribbon field-effect transistors by using lanthanum aluminate as the gate dielectric," *J. Comput. Electron.*, vol. 19, no. 4, pp. 1507–1515, 2020.

[13] N. Gupta and R. Chaujar, "Optimization of high-k and gate metal workfunction for improved analog and intermodulation performance of Gate Stack (GS)-GEWE-SiNW MOSFET," *Superlattices Microstructures*, vol. 97, pp. 630–641, 2016.

[14] B. Kumar, M. Sharma, and R. Chaujar, "Gate electrode work function engineered JAM-GS-GAA FinFET for analog/RF applications: Performance estimation and optimization," *Microelectronics J.*, vol. 135, 2023, Art. no. 105766.

[15] B. H. Lee, J. Oh, H. H. Tseng, R. Jammy, and H. Huff, "Gate stack technology for nanoscale devices," *Mater. Today*, vol. 9, no. 6, pp. 32–40, 2006.

[16] Anshul and R. Chaujar, "Numerical evaluation of the gate stacked GNR FET for improved device switching," in *Proc. Int. Conf. Recent Trends Microelectronics, Automat., Comput. Commun. Syst.*, 2022, pp. 88–91.

[17] K. L. Wong et al., "Carrier transport of rough-edged doped GNRFETs with metal contacts at various channel widths," *Superlattices Microstructures*, vol. 143, 2020, Art. no. 106548.

[18] B. Kumar and R. Chaujar, "Analog and RF performance evaluation of junctionless accumulation mode (JAM) gate stack gate all around (GS-GAA) FinFET," *Silicon*, vol. 13, pp. 919–927, 2021.

[19] J. Robertson, "High dielectric constant oxides," *Eur. Phys. J.-Appl. Phys.*, vol. 28, no. 3, pp. 265–291, 2004.

[20] K. Stokbro, D. E. Petersen, S. Smidstrup, A. Blom, M. Ipsen, and K. Kaasbjerg, "Semiempirical model for nanoscale device simulations," *Phys. Rev. B*, vol. 82, no. 7, 2010, Art. no. 075420.

[21] D. A. Papaconstantopoulos and M. J. Mehl, "The Slater–Koster tight-binding method: A computationally efficient and accurate approach," *J. Phys.: Condens. Matter*, vol. 15, no. 10, 2003, Art. no. R413.

[22] L. Tapasztó, G. Dobrik, P. Lambin, and L. P. Biro, "Tailoring the atomic structure of graphene nanoribbons by scanning tunnelling microscope lithography," *Nature Nanotechnol.*, vol. 3, no. 7, pp. 397–401, 2008.

[23] Y. Yoon, G. Fiori, S. Hong, G. Iannaccone, and J. Guo, "Performance comparison of graphene nanoribbon FETs with Schottky contacts and doped reservoirs," *IEEE Trans. Electron Devices*, vol. 55, no. 9, pp. 2314–2323, Sep. 2008.

[24] D. Gunlycke and C. T. White, "Tight-binding energy dispersions of armchair-edge graphene nanostrips," *Phys. Rev. B*, vol. 77, no. 11, 2008, Art. no. 115116.

[25] R. Jyoti, M. Chauhan, B. C. Choudhary, and R. K. Sharma, "Density functional theory study of manganese doped armchair graphene nanoribbon for effective carbon dioxide gas sensing," *ES Energy Environ.*, vol. 18, pp. 47–55, 2022.

[26] A. S. Silva, D. F. Ferreira, C. A. B. Silva Jr, and J. D. Nero, "Hydrogenation, width and strain effect in me-graphene devices," *Comput. Mater. Sci.*, vol. 210, 2022, Art. no. 111456.

[27] S. Smidstrup et al., "QuantumATK: An integrated platform of electronic and atomic-scale modelling tools," *J. Phys.: Condens. Matter*, vol. 32, no. 1, 2019, Art. no. 015901.

[28] Y. Pathak, B. D. Malhotra, and R. Chaujar, "Analog/RF performance and effect of temperature on ferroelectric layer improved FET device with spacer," *Silicon*, vol. 14, no. 18, pp. 12269–12280, 2022.

[29] R. Mann and R. Chaujar, "TCAD investigation of ferroelectric based substrate MOSFET for digital application," *Silicon*, vol. 14, no. 9, pp. 5075–5084, 2022.

[30] M. Gholipour, N. Masoumi, Y. Y. C. Chen, D. Chen, and M. Pourfath, "Asymmetric gate Schottky-barrier graphene nanoribbon FETs for low-power design," *IEEE Trans. Electron Devices*, vol. 61, no. 12, pp. 4000–4006, Dec. 2014.

[31] A. Naderi, "Theoretical analysis of a novel dual gate metal–graphene nanoribbon field effect transistor," *Mater. Sci. Semicond. Process.*, vol. 31, pp. 223–228, 2015.

[32] H. Sarvari and R. Ghayour, "Design of GNRFET using different doping profiles near the source and drain contacts," *Int. J. Electron.*, vol. 99, no. 5, pp. 673–682, 2012.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:10:44 UTC from IEEE Xplore. Restrictions apply.

2024 3rd International Conference for Innovation in Technology (INOCON)
Karnataka, India. Mar 1-3, 2024

Skin Cancer Detection using Knowledge Distillation

Rashmi Yadav[1]
Department of Computer Science and Engineering
Delhi Technological University, India
rashmiyadav.1512@gmail.com

Dr. Aruna Bhat[2]
Department of Computer Science and Engineering
Delhi Technological University, India
aruna.bhat@dtu.ac.in

2024 3rd International Conference for Innovation in Technology (INOCON) | 979-8-3503-8193-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/INOCON60754.2024.10512006

***Abstract* - Skin conditions are increasingly prevalent worldwide nowadays. For this, early detection and diagnosis are effective for the treatment and prevention of skin disorders. In this research, we proposed the distilling knowledge to identify skin diseases. Our work is in two parts. Firstly, we train the deep neural network (teacher model) on large dataset which is publicly available on the Kaggle platform. While the teacher model has high accuracy, due to its computational cost it can't be used on low-power devices. So, to resolve this problem, we transfer the knowledge of the higher model i.e. teacher model to the lower model i.e. student model by the knowledge distillation (KD) technique. As the student model is simple and requires few calculations, we can deploy the model on low-power devices. We evaluate the proposed framework on the publicly available dataset representing binary classification. Our results show that the teacher model has 0.9531 training accuracy and 0.8197 validation accuracy with 0.8758 test accuracy. Adding to this, the distilled model has 0.7970 accuracy on test set.**

Keywords: Skin lesion, detection, Knowledge distillation, DenseNet

I. INTRODUCTION

A considerable section of the world's population is afflicted by skin illnesses, which have recently raised serious health issues. Skin disorders may have a significant pessimistic impact on a person's psychology and result in long-term health problems if they are not addressed. To be properly treated and kept from spreading, skin problems must be diagnosed and treated as soon as feasible [2]. However, current techniques for skin disease detection frequently call for expensive tools and knowledge, making them unavailable in environments with limited resources.

The discipline of computer vision has made considerable strides in recent years in creating precise and effective models for picture categorization applications. Deep neural networks (DNNs) have become an effective method for classifying images and have been used in several fields, including medical image analysis [4]. In any case, because of their high computational requests, sending DNNs on low-power gadgets like cell phones is much of the time troublesome.

One system that can be utilized to manage the above issue involves the information refining idea for sending the model to this present reality. In information refining, an easier understudy model — a more modest, less convoluted brain organization — is prepared to imitate the way of behaving of a considerably more mind-boggling educator model — a greater, more complex brain organization [5]. The understudy model is made to have more clear engineering and needs fewer estimations, making it suitable for organizations on low-power devices. Information refining has been exhibited to be a helpful system for creating exact and compelling models for picture order undertakings.

Detailed insights about how Knowledge Distillation:

- Using a process called knowledge distillation, information is moved from a huge, complicated neural network called the instructor model to a more compact, straightforward neural network called the student model. Since Hinton et al. [1] initially suggested the method, but it has been widely applied across several industries.
- The fundamental idea of KD's is to use the output of teacher model as a soft target for the training of student model. In other words, instead of focusing just on training the student model to predict class labels accurately, we trained it to duplicate the teacher's model output probabilities. The soft targets compared to hard target of class labels are more informative and beneficial for the student model as shown in Figure 1.
- Mainly for the training of the knowledge distillation, there are two steps. At the initial step, the teacher model is trained on the dataset using some loss function like cross-entropy (we are using). In the following step, the student model is trained on teacher's model soft target. This step creates the loss function by combining the standard loss function such as cross entropy, with Kullback-Leibler(KL) divergence joining teacher and student model output. The KL divergence enables the student model to learn from the output of the teacher model by measuring the difference between the two output distributions.
- The advantages of Knowledge distillation are: first it can be used to improve the accuracy of student model, especially when there is a shortage of training data. Secondly, it can reduce the computational costs of the student model by making low-power compatible devices. Thirdly, it can transfer teacher-model information to student model, allowing student model to benefit fromteacher model experience.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:59:37 UTC from IEEE Xplore. Restrictions apply.

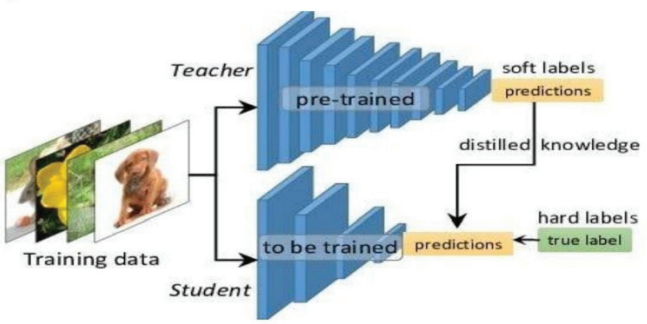


Fig. 1. Basic architecture of Knowledge Distillation [1]

In this study, we suggest a knowledge distillation-based research on skin lesion detection. The experiment consists of two steps: firstly, the training of DNN as a teacher model on Kaggle dataset images consists of benign and malignant moles. Then the knowledge of the teacher model is transferred to the student model using the above concept (KD). As the size of student model is small and requires less computational cost so, this model can be implemented on low-power devices.

II. METHODOLOGY

The proposed framework for skin disease detection using knowledge distillation consists of several steps, as described below:

A. Data Preparation and Preprocessing

For this, we utilized the Skin Cancer: Malignant vs Benign dataset [3]. Some of sample images are shown in Figure 2 The dataset is balanced mixture of malignant and benign skin moles, consisting of 3297 skin lesions images. The dataset splits into 80:20 ratio for the training and testing process. Training consists of 2637 images and the test has 660 images. All the images are labelled with binary classification labels indicating whether the image is malignant or benign. The images range in color and size some contain noises like hairs, ink marks, or rulers which are removed in preprocessing step, the preprocessed images are shown in Figure 3.

For our research, we first downloaded the dataset and organised it as per the given folders of training and testing. As mentioned above the splitting of the dataset is in 80:20 ratio.

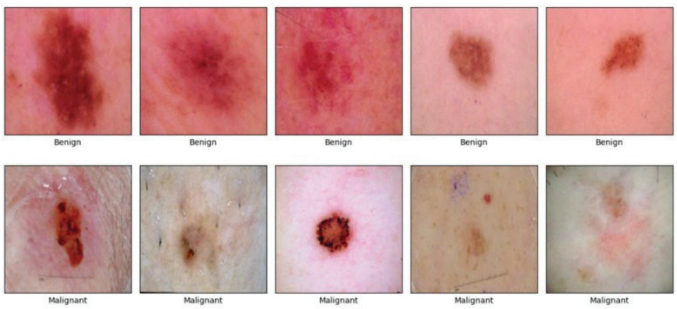


Fig. 2. Samples images from the dataset

Subsequently, the training dataset go through to data augmentation process where zoom, flips and random rotations were applied. Also, it helps to increase the training dataset and avert the model from overfitting. Data augmentation can better the generalization capability of the model.

Following data augmentation, all images are reshaped into 224*224 pixels. For training of model or neural network, it was ensured that all images were of the same dimensions and color channels. In the end, we normalize all the pixels values of each image in the dataset. With this step, the model was able to converge more quickly during training and prevent the gradient from becoming too large or too small.



Fig. 3. Preprocessed images from the dataset

B. Teacher Model Training

In the proposed framework for skin disease detection using knowledge distillation, to serve as the reference model for simple student model, we first trained a teacher model. The teacher model was designed using transfer learning with the DenseNet121 architecture as shown in Figure 4, which is already trained with the help of the ImageNet dataset. To this, we added a global average pooling layer and a dense layer with 512 units having ReLU activation function. Dropout at the rate of 0.5 was used to prevent any over-fitting in the model.

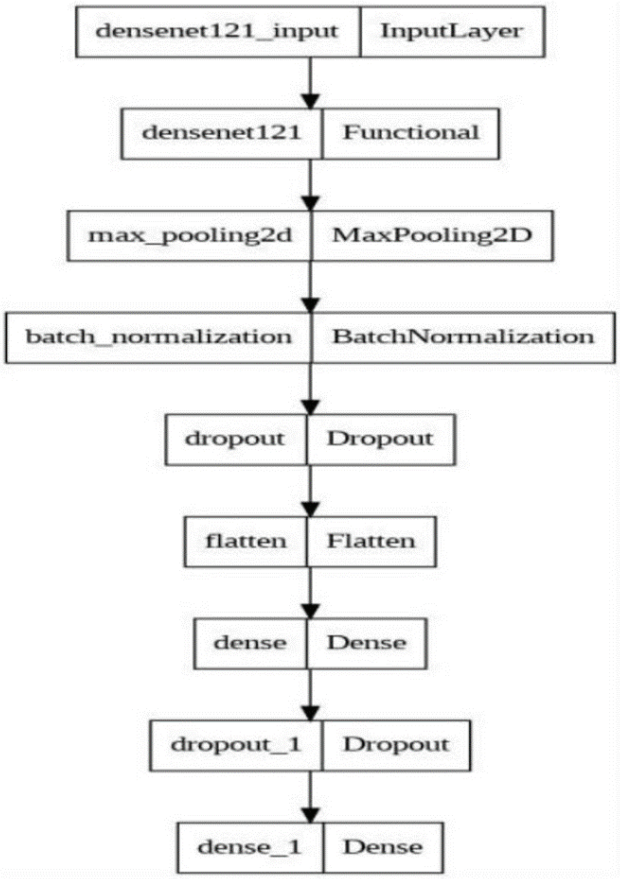


Fig. 4. DenseNet Model Architecture

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:59:37 UTC from IEEE Xplore. Restrictions apply.

The final output layer consisted of two units with softmax activation for binary classification of malignant and benign lesions. The model was learned using the training dataset followed by validation on the validation dataset.

Categorical cross-entropy was used for calculating loss and Adam was used as optimizer with 0.001 learning rate. When the validation loss didn't reduce for one epoch, the learning rate was limited, we employed early stopping with a tolerance of three epochs and a factor of 0.5.

The teacher model was run for 20 epochs on training data and it was evaluated against the vali- dation set. When the validation loss did not decrease for one epoch, the best model was saved using the ModelCheckpoint callback, and the learning rate was decreased using the Reduce LR On Plateau callback. To compare the effectiveness of the student models' performance, the training history of the instructor model was documented and utilised as a reference.

After the teacher model was trained, it was saved to a file using the Keras save function. The model was trained for 10 epochs, and during the final epoch, the model achieved a loss of 0.1117 and an accuracy of 0.9531 on the training data. On the validation data, the model achieved a loss of 0.5543 and an accuracy of 0.8197. The graphical representation of model accuracy and loss is shown in Figure 5

With the help of Matplotib's plot, the model's training and validation accuracy and loss were plotted. The accuracy and loss were plotted over 10 epochs. The model's training accuracy and loss decreased and converged, while the validation accuracy and loss plateaued.

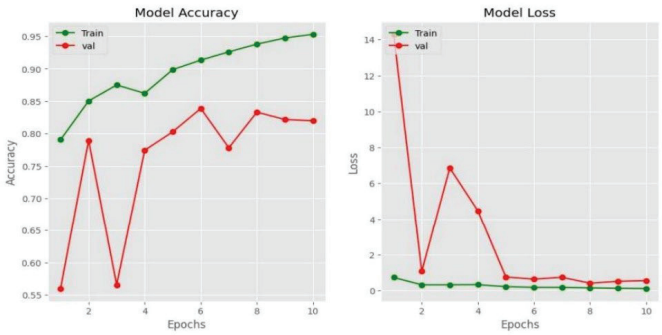


Fig. 5. Teacher model accuracy and loss

C. Student Model Training

In the student model training, a custom convolutional neural network is used. The model consists of convolution layer and max pooling layers with increasing filter size.

The input layer is defined using the Input function from Keras. The first convolutional block consists of 2 convolutional layers with 8 each layer having 8 filters and max-pooling layer following it. The second and third convolutional blocks have 16 and 32 filters respectively, with max-pooling layer after each block. The fourth convolutional block has 64 filters with a max-pooling layer following it. The fifth convolutional block has two convolutional layers with 64 filters each, along with max-pooling layers. Two dense layers with 256 and 2 units each follow the convolution blocks. To reduce over-fitting, a dropout with a rate of 0.5 is used for the second dense layer.

D. Model Distillation

The distillation process involves training the student model to predict not only the correct output labels but also the probabilities of the teacher model's predictions. This is done by minimizing the differences between the two models, as measured by the distillation loss function.

The Distiller class is initialized with the student and teacher models and allows the user to configure various parameters for the distillation process, such as the loss functions, weight coefficients, and temperature for softening probability distributions. In this particular implementation, the categorical crossentropy loss function is used for the student model's output and the KL Divergence loss function is used for the distillation loss. The temperature parameter is set to 6 and the weight coefficient alpha is set to 0.5. During the distillation process, we use the Adam optimizer with a learning rate of 1e-4 followed by beta parameters of 0.9 and 0.999 and epsilon as 1e-8. The training process consisted of 20 epochs with a batch size of 32.

During the process of training, ModelCheckpoint callback is used to save the best model data and Reduce LR On Plateau is used to lower the learning rate as and when the validation student loss plateaus.

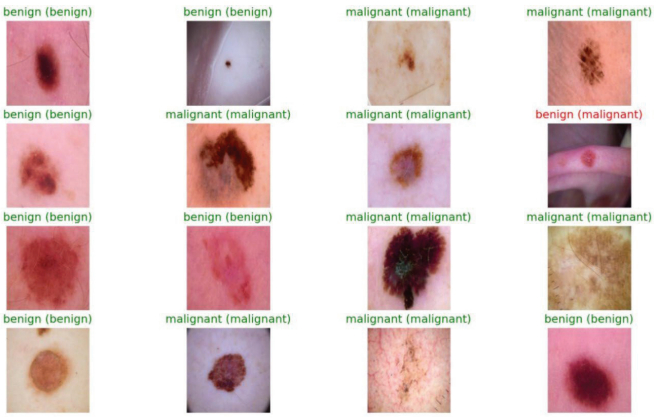


Fig. 6. Samples of predicted images and actual images

During the final epoch of the model distillation process, the training set accuracy was 0.8024, with a student loss of 0.4165 and a distillation loss of 0.0075. The validation set accuracy was 0.7837, and the validation student loss was 0.5160. These results suggest that the student model has learned to approximate the behaviour of the teacher model, but still, there is scope for improvement in validation performance. Overall, the process of knowledge distillation has proven to be effective in improving the performance of smaller student models using a larger and more complex teacher model.

III. Result and Analysis

The evaluation of the performance of the teacher model on the test set showed a promising accuracy of 0.8758 and a loss of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:59:37 UTC from IEEE Xplore. Restrictions apply.

0.3394, which served as the baseline for evaluating the performance of the student models trained using the knowledge distillation framework. Using the evaluate function of Keras, the student models' performance was assessed following training. The evaluation's findings revealed that the distilled student model had a 0.7970 accuracy on the test set and a 0.3858 loss. This illustrates how the knowledge distillation technique may be used to enhance student model performance.

The test set was loaded using the scikit-learn LoadFiles method, and the file names were preprocessed into arrays of photos. The trained model's performance was then assessed using the loaded test set and Keras predict function. Using matplotlib, a random selection of the test pictures, the predicted labels for those images, and the ground truth was plotted to show the predictions. For successful fore- casts, the model's predictions were highlighted in green; for wrong predictions, they were highlighted in red as shown in Figure 6.

IV. Conclusion and Future Work

Finally, it can be said that knowledge distillation has been successful in enhancing the performance of a smaller student model trained on a sizable dataset using a larger teacher model. The findings demonstrate that the student model may operate as accurately and effectively as the bigger instructor model when taught through knowledge distillation. As a result, less computationally intensive, smaller, and more effective models may be deployed.

There are various potential directions for improvement in future work. To evaluate if more performance enhancements can be made, one strategy that may be used is to experiment with various hyper-parameters and architectures. Exploring the usage of ensemble models, in which numerous smaller models are integrated to produce greater performance than a single large model, might be an- other field of research. Additionally, it might be advantageous to look into the use of various distillationmethods, like attention-based distillation, to enhance the functionality of the student models.

Overall, using knowledge distillation to deploy large models in real-world scenarios on devices with resource limitations is a promising strategy. We can make deep learning advantages available to a larger variety of applications and consumers by enabling the deployment of smaller and more effective models.

References

[1] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. "Distilling the knowledge in a neural network". In: arXiv preprint arXiv:1503.02531 (2015).

[2] Md Shakib Khan et al. "Knowledge distillation approach towards melanoma detection". In: Com- puters in Biology and Medicine 146 (2022), p. 105581.

[3] Skin Cancer: Malignant vs Benign dataset. https://www.kaggle.com/datasets/fanconic/ skin-cancer-malignant-vs-benign. Accessed November 30, 2023. 2018.

[4] Lin Wang and Kuk-Jin Yoon. "Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks". In: IEEE transactions on pattern analysis and machine intelligence 44.6 (2021), pp. 3048–3068.

[5] Yongwei Wang et al. "Ssd-kd: A self-supervised diverse knowledge distillation method for lightweight skin lesion classification using dermoscopic images". In: Medical Image Analysis 84 (2023), p. 102693.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 10,2024 at 04:59:37 UTC from IEEE Xplore. Restrictions apply.

Journal of Physics D: Applied Physics

ACCEPTED MANUSCRIPT

Spectral studies of thermally stable Dy3+/Sm3+ co-doped Li2Ba5W3O15 phosphors for warm white LEDs

To cite this article before publication: Anu . *et al* 2024 *J. Phys. D: Appl. Phys.* in press https://doi.org/10.1088/1361-6463/ad4566

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**.

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 10/05/2024 at 06:14

Spectral studies of thermally stable Dy^{3+}/Sm^{3+} co-doped $Li_2Ba_5W_3O_{15}$ phosphors for warm white LEDs

Anu[1], Sheetal Kumari[1], Nisha Deopa[2], A. S. Rao[1]*

[1]Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi 110 042, India

[2]Department of Physics, Chaudhary Ranbir Singh University, Rohtak Bypass Road, Jind, Haryana 126102, India

Abstract

In the current study, a series of $Li_2Ba_5W_3O_{15}$: RE^{3+} (RE=Dy, Sm) [LBW:Dy^{3+}/Sm^{3+}] phosphors were prepared using a high-temperature solid-state method. X-ray diffraction (XRD), Scanning electron microscopy (SEM) with Energy-dispersive X-ray analysis (EDX) scans showed that the crystal form was consistent with the standard LBW and comprised small irregularly shaped particles. Diffuse reflectance spectral (DRS) data was utilized to calculate the band gaps. Fluorescence study shows that LBW material doped with Dy^{3+} and Sm^{3+} yield distinct colors at 496 nm (blue) for Dy^{3+} and 582 nm (green-yellow), 612 nm (yellow), and 669 nm (red) for Sm^{3+} when excited by near-ultraviolet (n-UV) (336 nm) light. The observation of energy transfer (ET) between Dy^{3+} and Sm^{3+} ions play a role in modifying the luminescence of LBW:Dy^{3+}/Sm^{3+} co-doped phosphors. With a constant excitation wavelength (λ_{Ex}), different levels of activator doping lead to a change in the emission colors from their neutral white light to a deep orange-red region for LBW:Dy^{3+}/Sm^{3+} phosphors. The decay curves demonstrate a decrease in lifetime with an increase in the concentration of activator ions (Sm^{3+}). For D3S5 phosphor, the temperature-dependent photoluminescence (TDPL) characteristics were analysed under λ_{Ex} = 336 nm excitation. The results indicate excellent luminescence thermal stability with an activation energy of 0.16 eV at λ_{Ex} = 336 nm. With its low color-correlated temperature (CCT) and good thermal stability, the prepared phosphor sample shows potential as a solid-state emitting phosphor that can be used with UV chip stimulation for warm white LED applications.

***Corresponding author:** E-mail : drsrallam@gmail.com (Prof. A.S. Rao)

Tel+ :91 85860 39007, Fax+ :91 01127871023

Keywords: Dy^{3+}/Sm^{3+}, warm white, Reisfeld Approximation, CIE, TDPL.

Accepted Manuscript

1. Introduction:

Nowadays n-UV/blue light-emitting diodes (LEDs) have become increasingly popular because of their many benefits over other solid-state lighting sources, including their long lifespan, stability, high efficiency, and environmental friendliness [1–4]. Commercial LED technology was first introduced in 1960, and with the 1997 release of the InGaN blue LED, research interest in white LEDs strengthened. White light is produced by combining three primary Color combinations: the LED chips emit red, green, and blue (RGB); the $Y_3Al_5O_{12}$: Ce^{3+} (YAG:Ce^{3+}) phosphor combines blue and yellow color to provide a blue hue; and the UV or n-UV LED excitable phosphors emit a mixture of RGB hues [5]. RGB LED combinations for creating white light are less common due to their complex production and use, as well as their varied rates of deterioration and fluctuating driving current [6]. Therefore, the use of phosphor materials excitable by LED chips is encouraged to overcome these complexities. At present, YAG:Ce^{3+} phosphors, stimulated by blue LEDs, are extensively commercialised for white LEDs; however, because of insufficient red emission in the spectrum, they have a low Color rendering index (CRI) of less than 80 and a high correlated color temperature (~4500 K). Red-spectral area enhancement of YAG:Ce^{3+} can be achieved by co-doping with red-emitting ions such as Mn^{2+}, Eu^{3+}, and Sm^{3+} [7]. As an alternative, the use of RGB phosphors and UV or n-UV LED chips is becoming more common because of their high CRI values [8,9]. It is possible to produce white light with more uniformity and stable emission using this technology, since the output of the white LEDs produced depends on the quality of light emitted from phosphors [10]. Recent reports have indicated that a number of RGB phosphors, such as $KSrPO_4$:Eu^{2+} [11], $BaAl_2S_4$:Eu^{2+} [12], and $CaAlSiN_3$:Eu^{2+} [13], decrease blue emission efficiency by partially absorbing blue light. To increase the effectiveness of white LEDs, it is important to investigate novel UV or n-UV excitable phosphors [14]. Warm white lighting is great for areas like living rooms, bedrooms, and hallways as it creates a cozy and enjoyable lighting ambiance [8]. Studies have demonstrated that certain transitions of Dy^{3+}: $^4F_{9/2} \rightarrow {}^6H_{15/2}$, $^4F_{9/2} \rightarrow {}^6H_{13/2}$ [15], Sm^{3+}: $^4G_{5/2} \rightarrow {}^6H_{7/2}$, $^4G_{5/2} \rightarrow {}^6H_{9/2}$ [16], can cause Dy^{3+}/Sm^{3+} ions co-doped phosphors [17,18] or glasses [19,20] emission in the warm white light range. Now, selecting the appropriate host medium is also essential when preparing luminous materials for a variety of applications. Recent researches has focused on inorganic luminescent materials using substrates like molybdate [21,22], tungstate [23,24], phosphate [25], vanadate [26] and silicate [27] etc. Since tungstates have strong covalent interaction, high stability, and strong visible luminescence, more suitable as host materials for its potential applications in lasers, optical fibers, w-LEDs, plasma display panels, quantum electronics, phosphors, scintillators when

Accepted Manuscript

doped with suitable RE^{3+} ions [28]. Also, the alkali metal ions, for instance Li^+, Na^+, K^+, etc. improves the luminescence emission intensity and color purity of the host matrix. It is evident from literature that the presence of Li^+ ions in the host matrix improves the luminescence property as modifier in the host matrix [23]. Additionly, tungstate with alkaline-earth metal ions (i.e., Ba^{2+}, Sr^{2+}, Ca^{2+}, Mg^{2+} etc.) are a fascinating group of compounds that exhibit blue or green luminescence associated with the tetrahedral WO_4 groups [29]. These substances find extensive applications as X-ray phosphors, scintillators, laser crystals, and matrices for electron spin resonance [30]. All the aforementioned scientific patronages offered by the constituent chemical species prompted us to prepared the $Li_2Ba_5W_3O_{15}$ (LBW) host lattice co-doped with Dy^{3+} and Sm^{3+} ions and analysed for phase purity, microscopic morphology, photoluminescence (PL) properties, and temperature-dependent PL properties.

2. Synthesis and instrumentations:

2.1. Synthesis method:

Multiple techniques exist for synthesizing phosphor materials, such as high temperature solid-state method, combustion process, co-precipitation and sol-gel. However, in this particular investigation, the high-temperature solid-state method was used for synthesis purpose.

Multicomponent lithium barium tungstate phosphors, co-doped with Dy^{3+} and Sm^{3+} ($(Li_2CO_3+(5-x-y)BaCO_3+3WO_3+xDy_2O_3+ySm_2O_3$, where $x = 0.03$ and $0.01 \leq y \leq 0.07$, $\Delta y = 0.02$), were synthesized in this study. The starting materials, finely ground powders, were combined in an agate mortar, with 5 grams of each batch composition admixed till homogeneous. Subsequently, the thoroughly mixed sample was subjected to a heating process in an alumina crucible, maintaining a temperature of 1200 °C for 6 hours at a heating rate of 5°C/minute. After this, letting the mixture naturally cool to ambient temperature, the final products were reground to undergo further characterizations.

2.2. Instrumentations:

Various analytical methods were employed to examine the structural, optical, and temperature dependent characteristics of the synthesized phosphors. XRD was utilized to assess the sample structural features, employing a Bruker D-8 Advance diffractometer (accuracy ±0.01), 2θ angles ranging from 20˚ to 70˚, and a nickel (Ni) filter, with a radiation source of Cu-K_α. The surface morphology was analysed using FE-SEM with a JEOL JSM-6490 operating at a 10 kV accelerating voltage. DRS analysis was conducted using a JASCO V-770 spectrophotometer, which utilizes two different light sources, i.e., halogen and deuterium, for the UV and visible regions, with a resolution of ±0.1 nm. PL excitation (PLE) and PL emission spectra were

Accepted Manuscript

examined using a JASCO 8300 FP spectrofluorophotometer, equipped with a Xenon lamp (300 W) as the excitation source and a resolution of ±1.0 nm. Time-resolved PL (TRPL) investigations were carried out using a 150 MHz digital oscilloscope (HM 1507 Hameg Instruments). The above measurements were performed at ambient atmosphere. The thermal behavior of the optimized phosphor sample was investigated using temperature dependent PL (TDPL) study with an Ocean optics Flame-S-XR1-ES spectrometer.

3. Results and discussion:

3.1. Structural and morphological investigations

3.1.1. XRD study: structural investigations

X-ray diffraction analyses were conducted on all the powdered samples, including both un-doped LBW and co-doped LBW ($Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$; where x = 0.03 and $0.01 \leq y \leq 0.07$, $\Delta y = 0.02$). The measurements were performed in $20^{\circ} \leq 2\theta \leq 70^{\circ}$ and the results were illustrated in **Fig. 1 (a)**. The diffraction patterns align closely with the JCPDS database file no. 01-072-1717, corresponding to LBW sample with a monoclinic structure and $P6_3mc$ space group (186). The lattice parameters were determined as: a = b = 5.760 Å, c = 23.742 Å, α = β = 90°, γ = 120°, and the volume (V) was evaluated as 682.17 $Å^3$. The host's crystal structure and the emergence of an impurity phase do not seem to be affected by increasing doping concentrations, suggesting that the activator ions (Dy^{3+} and Sm^{3+}) have been effectively integrated into the host matrix to take the position of the Ba^{2+} in the lattice structure.

Williamson-Hall (W-H) plot was utilized for the un-doped LBW sample and $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and $0.01 \leq y \leq 0.07$, $\Delta y = 0.02$) phosphors to calculate the strain (∈) and the average crystallite size (D) [31]:

$$\beta \cos\theta = \frac{k\lambda}{D} + 4 \in \sin\theta \quad (1)$$

'K' is structure factor with a value of 0.94. '2θ' and 'β' refers to the diffraction angle full width at half maximum (FWHM) consistent to the XRD peaks, respectively and 'λ' stand for the wavelength of the Cu-K_α source. The graph (straight line) between $\beta \cos\theta$ and $4 \sin\theta$ is drawn to calculate the values of "D" and "∈", as shown by the W-H plot in **Fig. 1 (b)**.

Accepted Manuscript

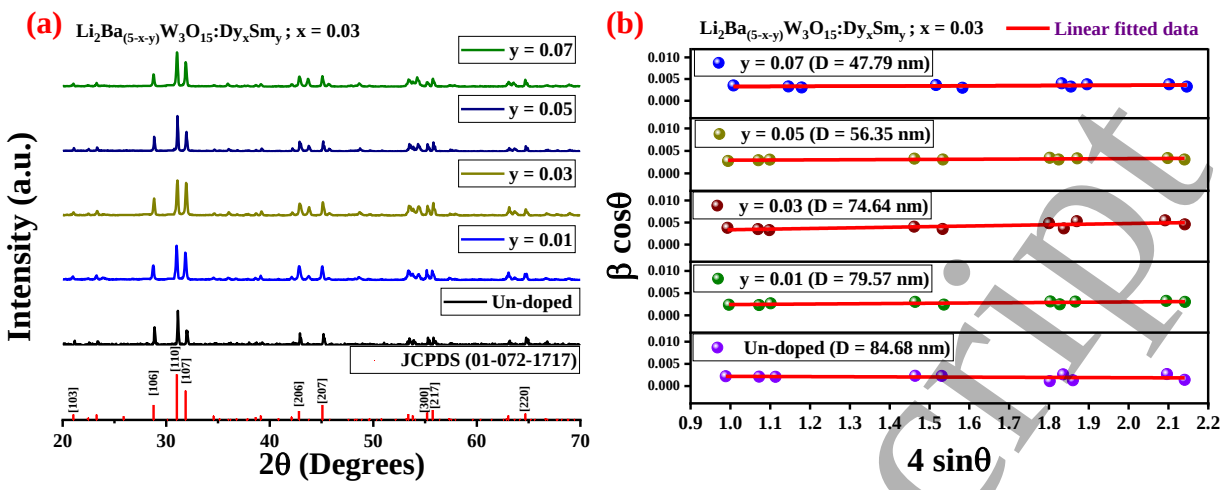


Fig. 1 (a) The XRD patterns and **(b)** A graph between $\beta_{hkl}cos\theta$ and $sin\theta$ for un-doped LBW and $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and $0.01 \leq y \leq 0.07$, $\Delta y = 0.02$) phosphors.

3.1.2 SEM and EDX studies: morphological investigations

The surface morphology of the prepared phosphor sample was analysed using SEM and EDX. The SEM image in **Fig. 2 (a)** depicts the surface of a $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and y = 0.05) phosphor. It is evident that there is inconsistency in both the shape and size of the particles. This variability is attributed to the agglomeration phenomenon observed in the prepared material, likely due to the high-temperature synthesis process and hand-grinding. The whole consists of irregular micron-sized particles, which play a pivotal role in enhancing the luminous intensity in solid-state lighting devices [32]. As shown in **Fig. 2 (b)**, the EDX analysis was utilized for the confirmation of the elements present in the prepared phosphor material. Clear peaks corresponding to Ba, O, W, Dy, and Sm are observed in the spectrum, validating the uniform distribution of these components in the prepared phosphors. This confirms the synthesis of LBW:Dy^{3+}/Sm^{3+} phosphors, as further illustrated in **Fig. 2 (c)-(h)**. Notably, elements with low energy, like Li, show no peaks in their EDX spectra, which is consistent with observations found in the literature [33].

Accepted Manuscript

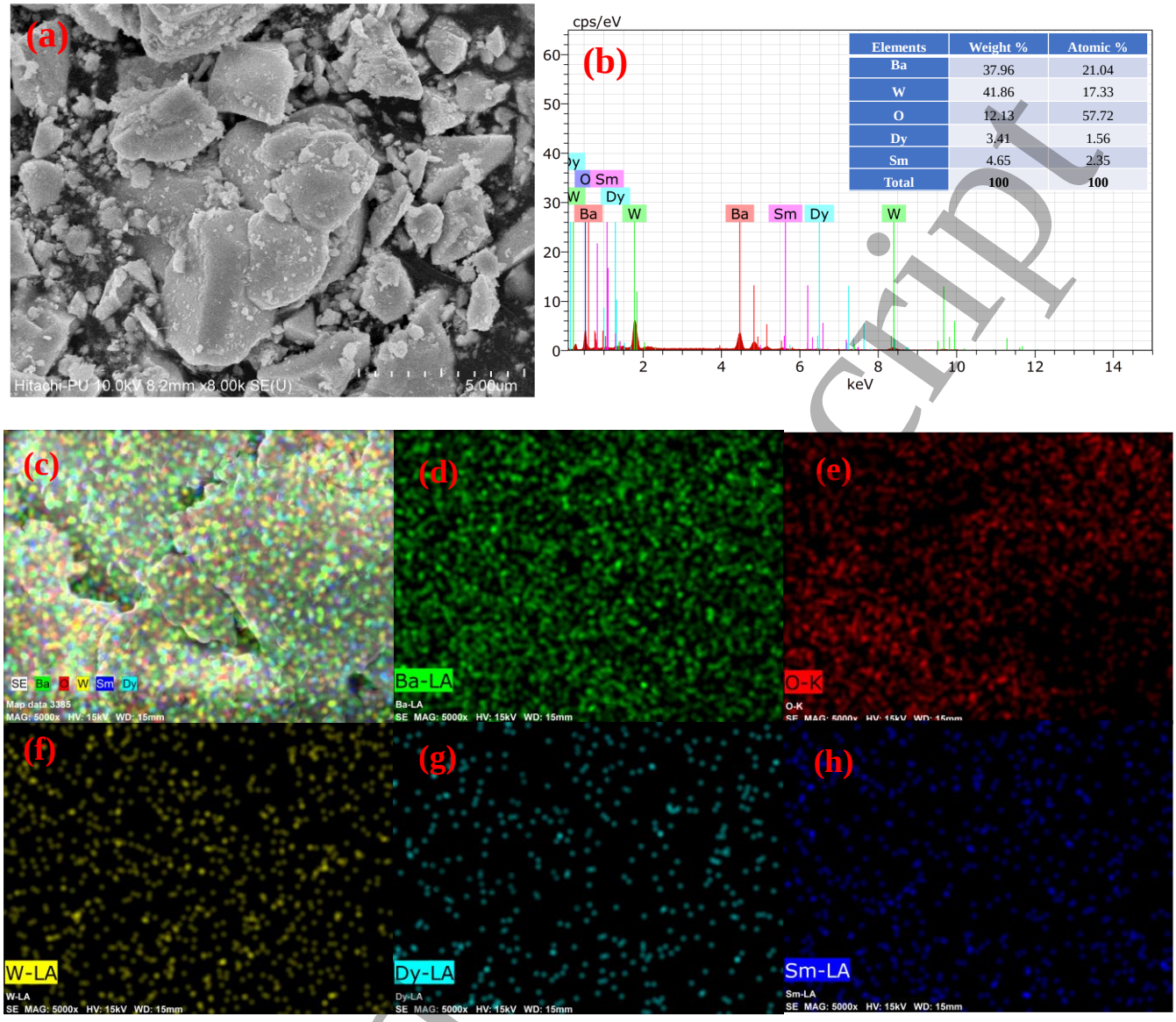


Fig. 2 (a) The SEM picture, **(b)** EDX pattern (Insert figure show the weight% and atomic% of individual atoms present in the phosphor) and **(c)-(h)** elemental distribution (Ba, W, O, Dy and Sm) of $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and y = 0.05) phosphor.

3.2 Optical studies:

3.2.1 Study of DRS and band gap (E_g) energy evaluation

Fig. 3 represents the DRS of un-doped LBW and Dy^{3+}/Sm^{3+} ions co-doped LBW phosphors in the 200-1600 nm range. The peaks corresponding to 472, 946, 1072, 1272, 1378, 1474 and 1530 nm wavelength shows the transition from the ground state, i.e., $^6H_{5/2}$ to various excited states $^4I_{11/2}$, $^6F_{11/2}$, $^6F_{9/2}$, $^6F_{7/2}$, $^6F_{5/2}$, $^6H_{13/2}$ and $^6H_{11/2}$, related to Sm^{3+} ions [34], however, the transitions corresponding to $^6F_{3/2}$, $^6F_{5/2}$ and $^6F_{7/2}$ excited states from the ground state $^6H_{15/2}$ are represented by the remaining peaks at 736, 792 and 888 nm, were related to Dy^{3+} ions [35,36].

Accepted Manuscript

The optical absorption near the band edge is represented by the Kubelka-Munk (K-M) function, which can be utilised to address the band gap energy (E_g) [37]:

$$\alpha h\nu = A\left(h\nu - E_g\right)^{n/2} \quad (2)$$

where, the absorption coefficient, photon energy, energy band gap, and constant are represented by the symbols α, $h\nu$, E_g and A. Equation (2) uses the values n/2 = ½ for direct and 2 for indirect transitions, respectively. The function F(R) can be expressed as follows [38]:

$$F(R) = \frac{(1-R^2)}{2R} = \frac{K}{S} \quad (3)$$

which relates the ratio of the absorption coefficient (K) and scattering coefficient (S) of the sample with diffuse reflectance (R). The inset in **Fig. 3** shows the band gap value, which was determined by extrapolating the linear section of the plot between $[F(R)\,h\nu]^2$ and $h\nu$ and found to be equal to 3.74, 3.64 and 3.61 eV for the un-doped LBW, $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and y = 0, y = 0.05) phosphors. The 4f energy level of Dy^{3+} and Sm^{3+} ions may be the cause of insignificant fall in the E_g values from 3.74 to 3.61 eV with the doping of RE^{3+} ions. Conversely, the slight decline also implies that the studied materials crystallized effectively [36].

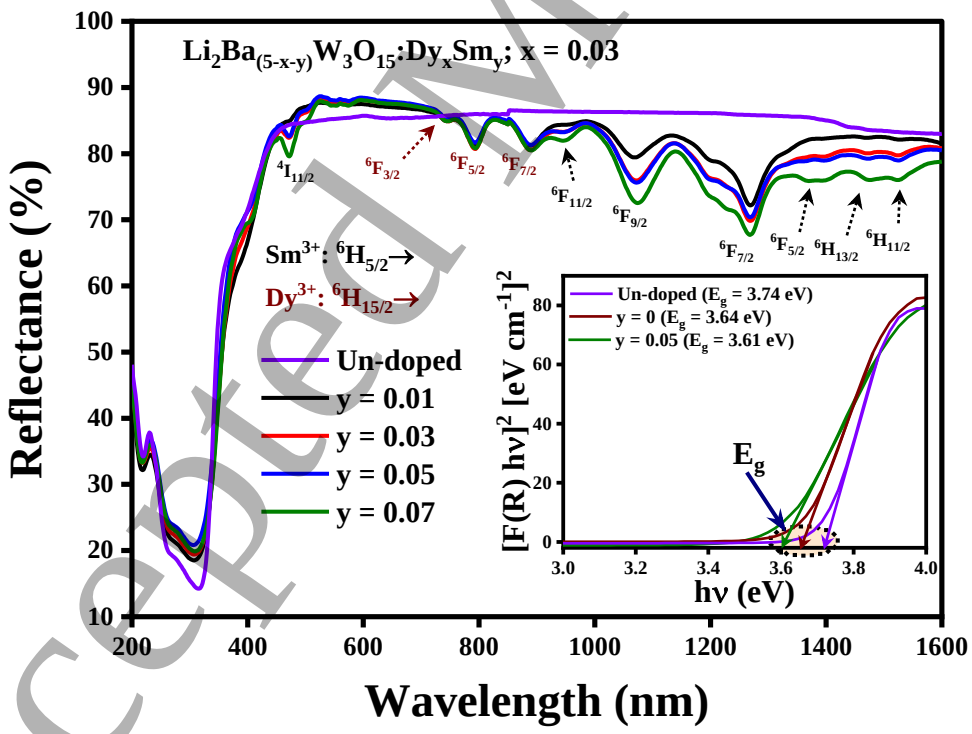


Fig. 3 DRS of the un-doped LBW, $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and $0.01 \leq y \leq 0.07$, $\Delta y = 0.02$). [Inset shows the band gap variations of un-doped LBW, $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and y = 0, y = 0.05) phosphors].

Accepted Manuscript

3.2.2 Study of PL excitation (PLE) and PL emission spectra

3.2.2.1 PLE and PL of LBW:Dy^{3+} phosphors

The PLE spectra of LBW phosphor doped with 3.0 mol% of Dy^{3+} ions (D3) at a PL emission wavelength (λ_{Em}) of 493 nm in the 300-460 nm spectral range is shown in **Fig. 4 (a)**. Nine PLE bands were detected at 338, 347, 357, 368, 378, 396, 425, 447 and 456 nm, corresponding to the transition from the ground state $^6H_{15/2}$ to different excited states $^6P_{3/2}$, $^6P_{7/2}$, $^4I_{13/2}$, $^6P_{5/2}$, $^4K_{17/2}$, $^4F_{7/2}$, $^4G_{11/2}$, $^4I_{15/2}$ and $4F_{9/2}$ [39,40]. $^6P_{5/2}$, the strongest excitation band, is situated at around 368 nm. Consequently, PL emission spectra were recorded, as depicted in **Fig. 4 (b)**, monitoring the excitation wavelength (λ_{Ex}) at 368 nm. The PL spectra showed three emission peaks corresponding to the transitions $^4F_{9/2}\rightarrow{}^6H_{13/2}$, $^6H_{11/2}$, and $^6H_{15/2}$. The emission intensities exhibited a gradual increase to their peak values when the concentration of Dy^{3+} ions reach to 3 mol% shown in the inset of **Fig. 4 (b)** illustrates the concentration quenching effect. The increase in the Dy^{3+} ion's concentration might lead to the non-radiative ET due to very less distance between the adjacent luminous centres [41]. Blasse suggested that the critical separation (R_c) and the ET process among adjacent dopant (Dy^{3+}) ions can be determined using the following expression [31]:

$$R_c \approx 2\left[\frac{3V}{4\Pi X_c N}\right]^{1/3} \quad (4)$$

In this case, "N", "X_c" and "V" signifies the number of atoms, optimum Dy^{3+} ions concentration and unit cell volume. The R_c was evaluated for X_c = 0.03, $N = 2$ and V = 681.182 Å^3 and found to be equal to 27.88 Å, after investigating the experimental statistics. Van Uitert claims that ET is connected to other phenomena, such as multipolar interactions, re-absorption of radiation, and energy exchange [42]. Since the current investigation's evaluated value of the essential distance between the adjacent Dy^{3+} ions was found to be greater than 5 Å [43]. Consequently, non-radiative energy transfer among Dy^{3+} ions typically happens due to electric multipole-multipole interactions, which is closely linked to the distance between them according to Dexter's theory [44,45]. The variation in photoluminescence intensity is commonly employed to explain the type of multipole interaction if the energy transfer occurs between activators of the same type. Again, the following expression can indicate a correlation among PL intensity (I) and the concentration of the activator ions (x in moles) [46];

$$\frac{I}{x} = k\,[1 + A(x)^{\frac{Q}{3}}]^{-1} \quad (5)$$

Accepted Manuscript

here, "A" is a constant, and "Q" denotes the type of interaction amid RE^{3+} ions. Q = 3 indicates the energy migration among nearest neighbour or next nearest neighbour activators. It can also take values 6, 8 and 10 for electric d-d (dipole-dipole), d-q (dipole-quadrupole), and q-q (quadrupole-quadrupole) interaction [45]. By simplifying Eq. 5, we can obtain that the relationship between $log\left(\frac{I}{x}\right)$ and $log(x)$ for LBW:Dy^{3+} phosphors under λ_{Ex} = 368 nm for λ_{Em} = 493 nm, with a slope of Q/3, well-illustrated in **Fig. 4 (c)**. Using λ_{Em} = 493 under λ_{Ex} = 368 nm, the values for the "Q" parameter are found at 5.4, which is close to 6. This indicates that the reason of ET between neighbouring Dy^{3+} ions is electric d-d interaction, which leads to concentration quenching. Additionally, the CIE chromaticity coordinates of D3 were evaluated using the PL data and lies in white region, shown in **Fig. 4 (d)**.

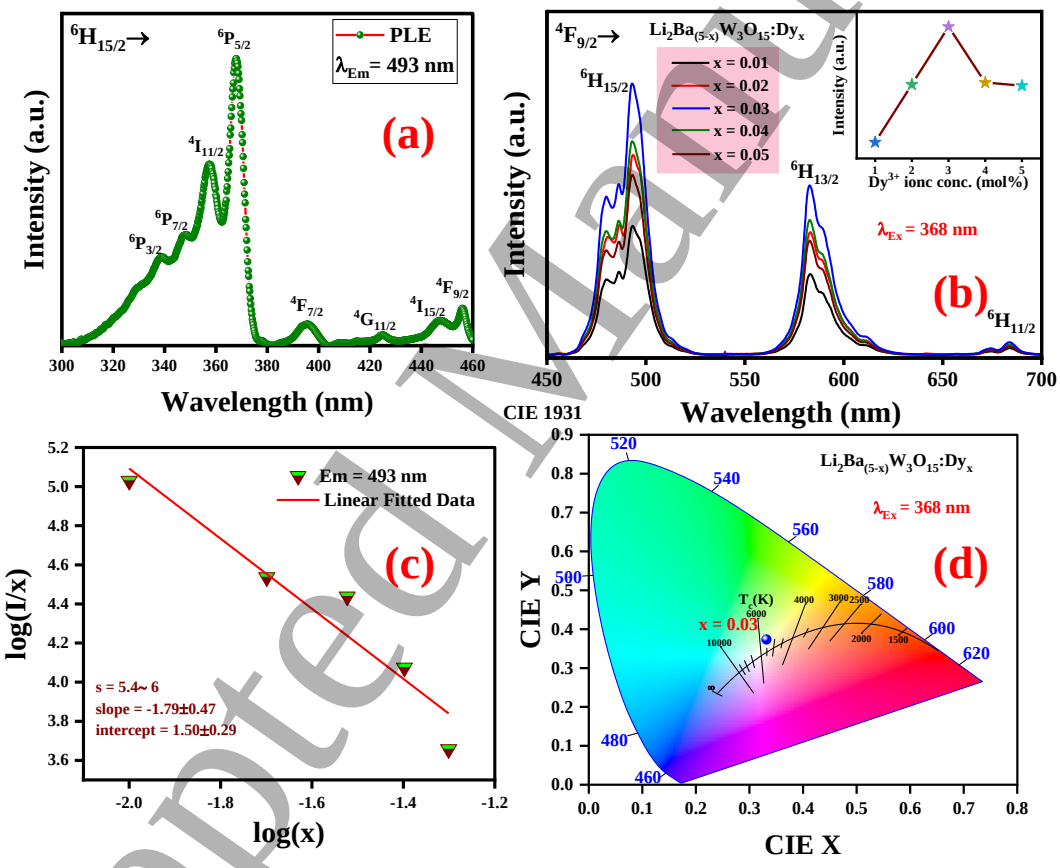


Fig. 4 (a) PLE for $Li_2Ba_{(5-x)}W_3O_{15}:Dy_x$ (x = 0.03), **(b)** PL emission spectra, **(c)** Dexter plot for LBW:xDy^{3+} ($0.01 \leq x \leq 0.05$, $\Delta x = 0.01$) phosphor and **(d)** CIE chromaticity coordinates for $Li_2Ba_{(5-x)}W_3O_{15}:Dy_x$ (x = 0.03) at λ_{Ex} =368 nm.

3.2.2.2 Effect of co-doping of Sm^{3+} ions on LBW:Dy^{3+} phosphor

Accepted Manuscript

In the pursuit of synthesising single-phase warm white phosphors, we utilized $Li_2Ba_{(5-x)}W_3O_{15}:Dy_x$ (x = 0.03) phosphors co-doped with varying concentrations of Sm^{3+} ions. Another objective of this modification was to explore the ET mechanism between the two RE^{3+} ions. In RE^{3+} ions, the down-conversion through ET involves the transfer of energy between a high-energy RE^{3+} ion and low-energy state RE^{3+} ion resulting in luminescence. It is much more probable that the ET process will go from Dy^{3+} to Sm^{3+} ions due to the higher energy state of the Dy^{3+} ions as compare to the Sm^{3+} ions. The spectrum overlap between the acceptor (Sm^{3+}) excitation and the donor (Dy^{3+}) emission is the primary requirement for the ET process analysis [47]. **Fig. 5** shows the alignment of the PL emission peak of Dy^{3+} ions at 495 nm (from $^4F_{9/2} \rightarrow {}^6H_{15/2}$) with the PLE peak of Sm^{3+} ions at 468 nm (from $^6H_{5/2} \rightarrow {}^4I_{11/2}$). These findings inspired us for further investigation into the ET process from Dy^{3+} to Sm^{3+} by combining the ideal concentration of Dy^{3+} ions with various concentration levels of Sm^{3+} ions (from 1 to 7 mol% in increments of 2 mol%).

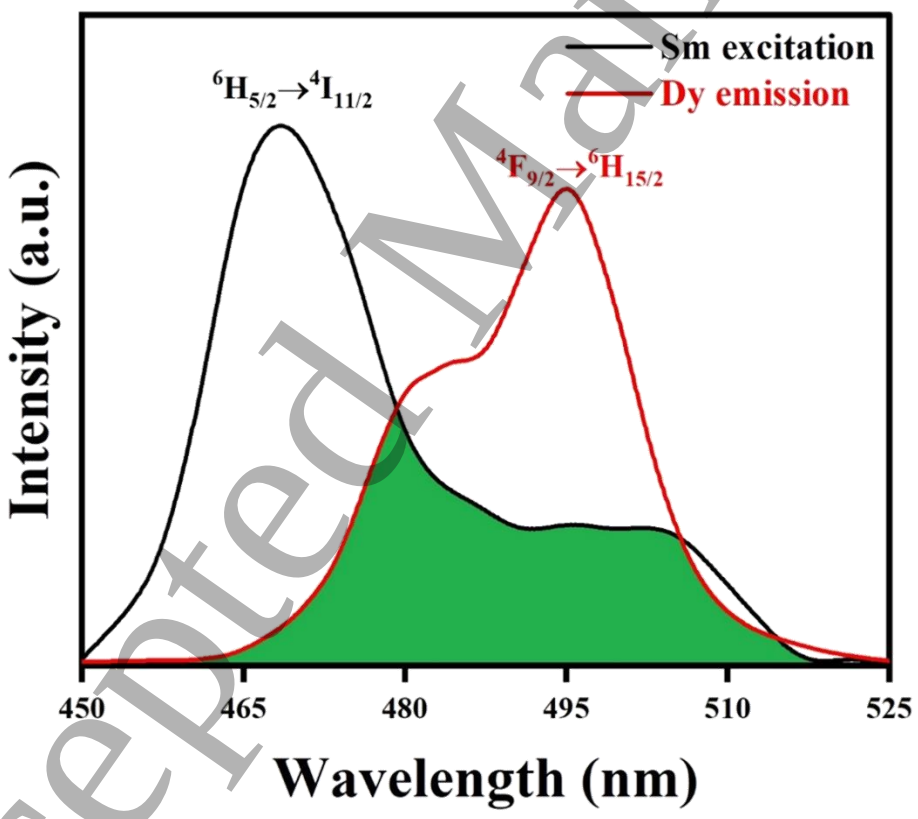


Fig. 5 The overlapped region between the Sm^{3+}excitation band and Dy^{3+} emission band.

Before commencing this investigation, the PLE spectrum of solely Dy^{3+} (3 mol% doping of Dy^{3+}-D3) and Sm^{3+} (1 mol% doping of Sm^{3+}-S1) ions doped phosphors can be observed at λ_{Ex} = 336 nm, as depicted in **Fig. 6 (a)**. **Fig. 6 (b)** displays the PL spectra of three samples

Accepted Manuscript

individually doped with 3 mol% Dy^{3+} (D3) and 1 mol% Sm^{3+} (S1) ions, as well as co-doped with 3 mol% Dy^{3+} and 1% mol Sm^{3+} (D3S1) ions. Under λ_{Ex} = 336 nm, the D3S1 samples exhibit characteristic emission peaks from both Dy^{3+} and Sm^{3+} ions. These include an emission peak at 495 nm for Dy^{3+} ions and emission peaks at 581 nm, 612 nm, and 667 nm corresponding to $^4G_{5/2} \rightarrow {}^6H_{5/2}$, $^6H_{7/2}$, and $^6H_{9/2}$ transitions for Sm^{3+} ions [32,48]. It is evident from the literature that the commercial LED has its excitation wavelength from near UV to UV region [49]. The 336 nm excitation wavelength used in the present work is near to UV region, the commercial LEDs can be used practically to excite the phosphors studied in the present work.

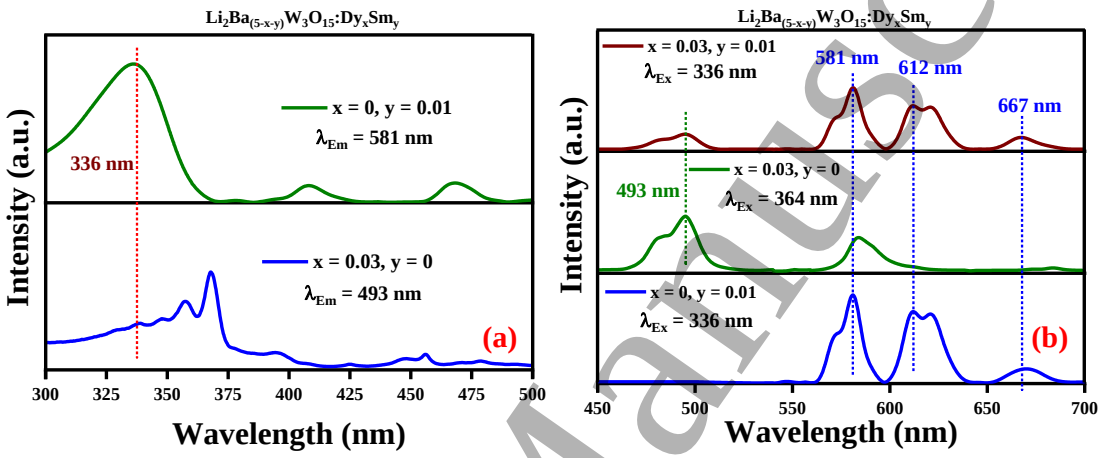


Fig. 6 (a) PLE spectrum of $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03, y = 0 & x = 0, y = 0.01) phosphors, **(b)** PL spectrum of $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0, y = 0.01, x = 0.03, y = 0 and x = 0.03, y = 0.01) phosphors.

In **Fig. 7 (a)**, the intensity of PL emission peaks of Sm^{3+} at 581, 612, and 669 nm significantly increases by about 3-4 times when doped with Dy^{3+}, as compared to single doped with Sm^{3+} ions. This suggests the effective ET from Dy^{3+} to Sm^{3+} under λ_{Ex} = 336 nm. **Fig. 7 (b)** displays the PL emission spectra of $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and $0.01 \leq y \leq 0.07$, Δy = 0.02) phosphors. Under λ_{Ex} = 336 nm excitation, both characteristic peaks of Sm^{3+} and Dy^{3+} are simultaneously visible. Specifically, the peaks at 493 nm (blue light) and 583 nm (yellow light) correspond to the characteristic transitions of $^4F_{9/2} \rightarrow {}^6H_{15/2}$ and $^6H_{13/2}$ for Dy^{3+}, respectively. Meanwhile, the peaks at 581 nm (yellow light), 612 nm (orange-red light), and 669 nm (red light) correspond to the characteristic transitions of $^4G_{5/2} \rightarrow {}^6H_{5/2}$, $^6H_{7/2}$, and $^6H_{9/2}$ for Sm^{3+} ions.The luminous intensity of Dy^{3+} ions gradually diminish with an increase in the concentration of Sm^{3+} ions taking Dy^{3+} ions concentration constant which approves the ET between Dy^{3+} and Sm^{3+} ions. After 5 mol% doping concentration of Sm^{3+} ions, the fluorescence intensity peaks subsequently decrease due to concentration quenching.

Accepted Manuscript

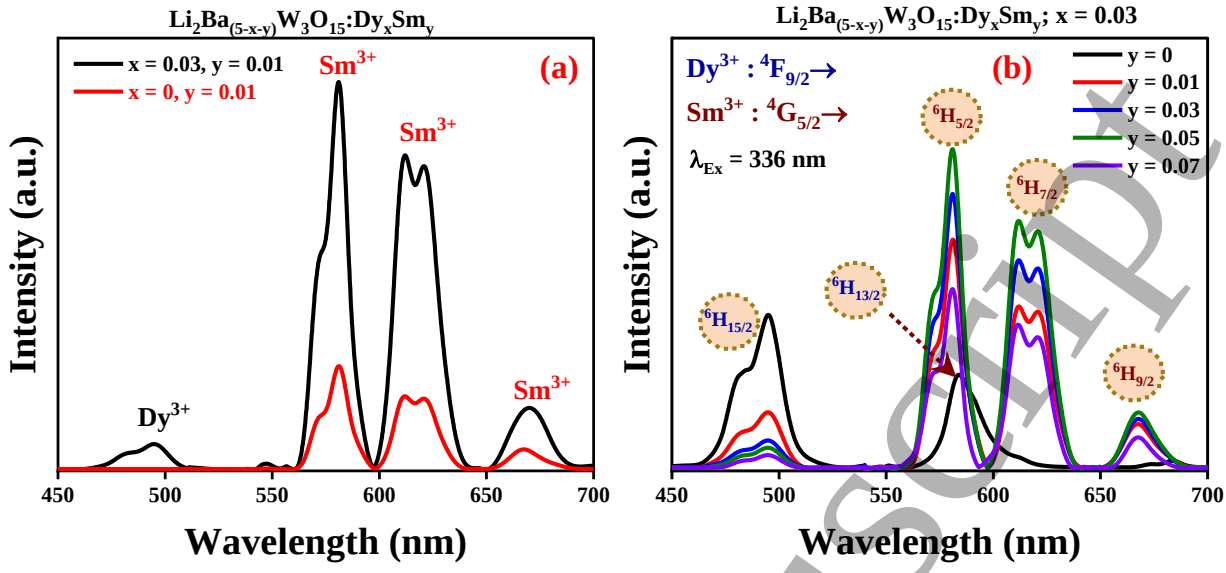


Fig. 7 (a) PL emission spectra comparison of $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (x = 0.03, y = 0.01 & x = 0, y = 0.01) under 336 nm excitation, **(b)** PL emission spectra for LBW, $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and $0.01 \leq y \leq 0.07$, $\Delta y = 0.02$) phosphors.

3.2.2.3 Energy transfer calculation

Fig. 8 (a) illustrates the schematic representation of the ET mechanism from Dy^{3+} to Sm^{3+} within the LBW host material. The excited state of Dy^{3+} ($^4F_{9/2}$) ions is slightly higher than those of Sm^{3+} ($^4I_{19/2}$) ions, suggested that the non-radiative mechanisms are primarily responsible for the ET from Dy^{3+} to Sm^{3+} ions. These non-radiative ET mechanisms involve radiation reabsorption, exchange interaction, and electric multi-pole interaction [17]. The multipole-multipole interaction has a higher chance of causing ET amid all the interactions, in the present study. Reisfeld's approximation was utilized to identify the precise interaction mechanism that exists in the prepared Dy^{3+}/Sm^{3+} co-doped phosphor samples [17]:

$$\frac{\eta_0}{\eta} = C^{a/3} \qquad (6)$$

Here, the symbols η and η_0 represent the radiant quantum efficiencies in the presence and in the absence of the sensitizer (Sm^{3+}) ion, respectively. The type of multipolar interaction is indicated by the variable "a". For d-d, d-q and q-q interaction, a = 6, 8 and 10, respectively. C represents the total concentration of the activator (Dy^{3+}) and the sensitizer (Sm^{3+}) ions (in moles), i.e., C = (3+0), (3+1), (3+3), (3+5) and (3+7) for $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (x = 3, y = 0, 1, 3, 5 & 7), respectively. The following is the relationship between the values of radiant quantum efficiencies and the luminous intensities [47]:

Accepted Manuscript

$$\frac{I_{S0}}{I_S} \propto \frac{\eta_0}{\eta} = C^{a/3} \quad (7)$$

I_S & I_{S0} represent the PL emission peak intensities of activator (Dy^{3+}) ions in the presence and absence of sensitizer ions (Sm^{3+}), respectively. **Fig. 8 (b)** illustrates the changes in the ratio $\frac{I_{S0}}{I_S}$ as a function of $C^{a/3}$. For, a = 6, the data showed the strongest linear fit, suggesting non-radiative ET through d-d interaction. The energy levels for the dopant ions were determined by referring to Carnall's paper [34].

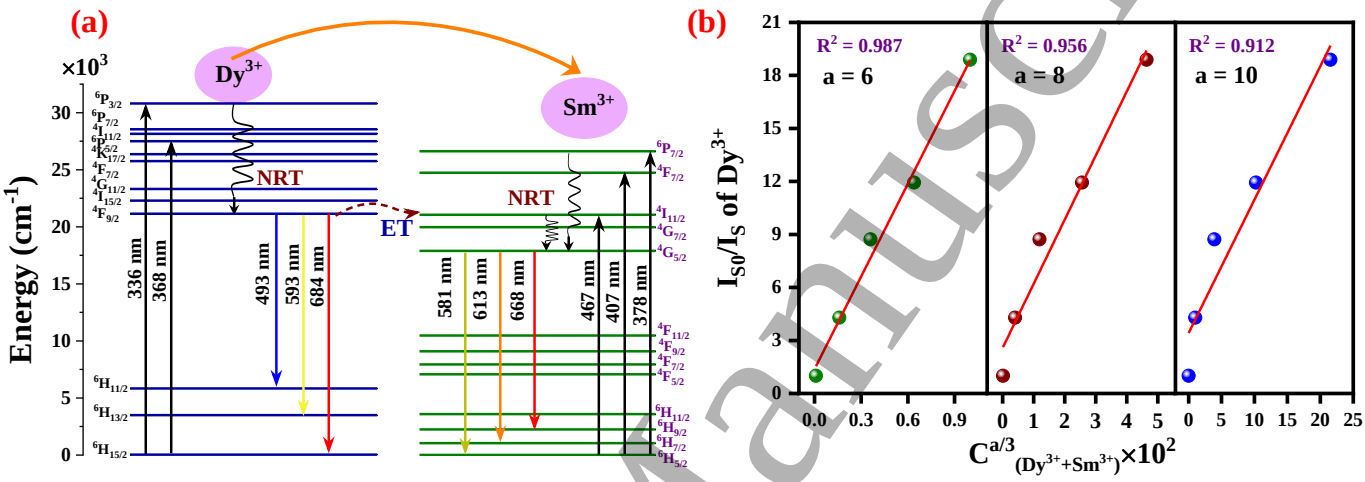


Fig. 8 (a) Schematic diagram of energy transfer from Dy^{3+} to Sm^{3+} in LBW host material, **(b)** The variation in the ratio I_{S0}/I_S as a function of $C^{a/3}$.

3.2.2.4 Study of time resolved photoluminescence (TRPL)

Fig. 9 (a) shows the decay curves of LBW, $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and 0.01 ≤ y ≤ 0.07, Δy = 0.02) co-doped LBW phosphors when the excitation and emission wavelengths are fixed at 336 nm and 581 nm, respectively. This allows for a deeper exploration of the ET process between Dy^{3+} and Sm^{3+} ions. Single exponential fitting is not appropriate for the measured data; however, double exponential fitting provides a useful assessment method [50]*:*

$$I_t = I_0 + A_1 \exp\left(-\frac{t}{\tau_1}\right) + A_2 \exp\left(-\frac{t}{\tau_2}\right) \quad (8)$$

where, I_0 and I_t denote the intensity at initial time and at time t, respectively. The exponential component's rapid and slow lifetimes are denoted by τ_1 and τ_2, respectively. The correlative parameters that resulted after fitting are A_1 and A_2. Therefore, the following equation can be

Accepted Manuscript

used to confirm the average lifetime of the Dy^{3+}/Sm^{3+} co-doped ions in the LBW phosphor samples [41]:

$$\tau = \frac{A_1\tau_1^2 + A_2\tau_2^2}{A_1\tau_1 + A_2\tau_2} \quad (9)$$

Thus, for the as-prepared $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and $0.01 \leq y \leq 0.07$, $\Delta y = 0.02$) co-doped phosphor samples, the computed average lifetime is 0.646 ms, 0.641 ms, 0.637 ms, and 0.624 ms, respectively. The lifetime (τ), energy transfer probability (P_{ET}), and energy transfer efficiency (ETE; η_{ET}) are plotted against the Sm^{3+} ions concentration in **Fig. 9 (b)**. As seen in **Fig. 9 (b)**, adding more Sm^{3+} ions, shortens the luminescence lifetime values, indicating the possibility of ET from $Dy^{3+} \rightarrow Sm^{3+}$. The following expression can be used to estimate the ETE based on the measured lifetime values [51]:

$$\eta_{ET} = 1 - \frac{\tau}{\tau_0} \quad (10)$$

here, τ_0 represents the lifetime for Dy^{3+} single doped sample while for Dy^{3+}/Sm^{3+} co-doped samples, the lifetime is represented by τ. When y = 0.07, the maximum η_{ET} increases to 18%. The P_{ET}, which measures the chance of ET from the donor ions to the acceptor ions, is computed using the formula given below [52]:

$$P_{ET} = \frac{1}{\tau} - \frac{1}{\tau_0} \quad (11)$$

Fig. 9 (b) shows the variation of τ, η_{ET} and P_{ET} with the concentration of Sm^{3+} ions and **Table 1** provides an overview of the above evaluated parameters.

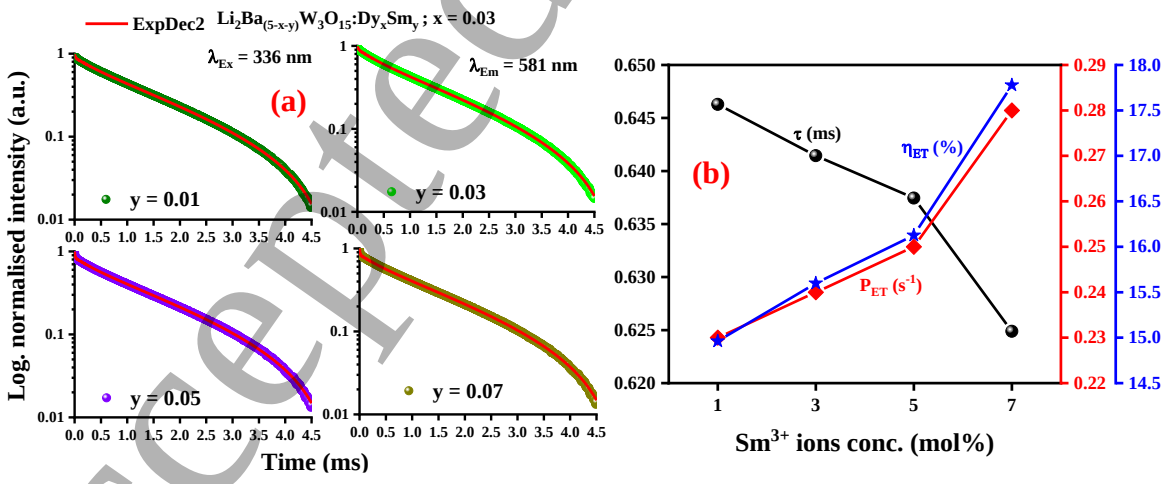


Fig. 9 (a) The decay curves of $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and $0.01 \leq y \leq 0.07$, $\Delta y = 0.02$) phosphors at λ_{Em} = 581 nm & λ_{Ex} = 336 nm and **(b)** variation of τ_{av}, η_T and P_T with the Sm^{3+} ions concentration.

Accepted Manuscript

Table 1

The average value of decay lifetime (τ, ms), energy transfer efficiency (η_{ET}) and energy transfer probability (P_{ET}) of $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and 0.01 ≤ y ≤ 0.07, Δy = 0.02) co-doped phosphors.

System Name	τ (ms)	η_{ET} (%)	P_{ET} (s^{-1})
D3S1	0.646	14.96	0.23
D3S3	0.641	15.60	0.24
D3S5	0.637	16.12	0.25
D3S7	0.624	17.78	0.28

The energy transfer efficiency can be evaluated using the emission spectra, represented in Fig. 8 (b). The following expression can be used to evaluate the $\eta_{ET(emssion)}$ using emission spectra of the as-prepared $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and 0.01 ≤ y ≤ 0.07, Δy = 0.02) co-doped phosphors [53]:

$$\eta_{ET(emssion)} = \frac{I^{Dy^{3+}} + I^{Dy^{3+}+Sm^{3+}}}{I^{Dy^{3+}}} \quad (12)$$

here, $I^{Dy^{3+}}$ and $I^{Dy^{3+}+Sm^{3+}}$ are the luminescence intensities of $^4F_{9/2} \rightarrow {}^6H_{15/2}$ for 0.03 moles of Dy^{3+} single ions doped and Dy_xSm_y (where x = 0.03 and 0.01 ≤ y ≤ 0.07, Δy = 0.02) co-doped LBW samples. The ETE were evaluated by taking the integrated luminescence intensity data in Fig. 8 (b) into Eq. (12). The energy transfer efficiency increases with increasing Sm^{3+} ions concentration. A comparison of ETE values calculated using the emission spectra and fluorescence decay analysis has been represented in Table 2.

Table 2

A comparison of ETE values calculated using the emission spectra and fluorescence decay analysis for the as-prepared $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and 0.01 ≤ y ≤ 0.07, Δy = 0.02) co-doped phosphors.

System Name	η_{ET} (%) (Using emission spectra)	η_{ET} (%) (Using Decay analysis)
D3S1	0.32	0.23
D3S3	0.41	0.24
D3S5	0.43	0.25
D3S7	0.46	0.28

Accepted Manuscript

3.2.2.5 Evaluation of color coordinates

The chromaticity coordinate diagram for Dy^{3+} single ions doped and Dy^{3+}/Sm^{3+} ions co-doped LBW phosphors, designed by the Commission International de I 'eclairage (CIE) in 1931, evaluated using the formula given in literature [47,52] and shown in **Fig. 10 (a)**. The CIE color coordinates of white luminescent phosphor are (0.333, 0.333), which are in the white light area, based on the emission spectrum of the substance [47]. CIE of D3 sample lie in yellow-white area. When Sm^{3+} ions are co-doped with D3 phosphors, the emission gradually moves to the warm white light region. **Table 3** represents the CIE chromaticity coordinates, Correlated Color Temperature (CCT), and emission light color information for the prepared samples. The CCT values are determined using the Mc-Camy equation [54]:

$$CCT = a\eta^3 + b\eta^2 + c\eta + d \quad (13)$$

here, $a = -449$, $b = 3525, c = -6823.2$ and $d = 5520.3$ are the constants; and $\eta = \frac{x-x_e}{y-y_e}$; $x_e = 0.3320, y_e = 0.1858$, (x, y) are the CIE color coordinates of each sample.

Low CCT values imply warmer light, while high CCT values seem to be colder [41]. **Tables 3** and **Fig. 10 (a)** illustrate that the introduction of a specific quantity of Sm^{3+} doping into the D3 host leads to a notable decrease in color temperature, shifting the emission from the yellow-white sunlight region to the warmer white light region. While the D3 phosphor itself, emitting cold white light, achieves white light emission through blue and yellow light emissions, however, the absence of red light can cause prolonged discomfort such as eye strain, fatigue, and piercing sensations. Consequently, a phosphor incorporating a three-color mixture is developed to emit warm white light, thereby reducing the detrimental effects of blue light on human eyes and significantly cutting costs. Hence, the synthesized $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and $0.01 \leq y \leq 0.07$, $\Delta y = 0.02$) co-doped phosphors, owing to its low CCT value and emission of warm white light, holds promise as potential candidate for indoor and ambient lighting applications, alleviating eye fatigue. **Table 4** shows the comparison of CCT of the optimised phosphor $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03, y = 0.05) sample with previously reported literature [17,32,55,56].

Color purity (CP) is an additional significant factor in colorimetric calculations that analyse the color saturation. To assess the CP, following equation was used [48]:

$$\text{Color purity (CP)} = \frac{\sqrt{(x-x_{ee})^2+(y-y_{ee})^2}}{\sqrt{(x_d-x_{ee})^2+(y_d-y_{ee})^2}} \quad (14)$$

where, (x_{ee}, y_{ee}) is equal energy point, (x, y) is color coordinates of the prepared Dy^{3+}/Sm^{3+} co-doped LBW phosphors, and (x_d, y_d) is the dominant wavelength point. The CP values, shown in inset of **Fig. 10 (b)**, increases from 15% to nearly 80%, which are relatively high, indicate

Accepted Manuscript

the use of the prepared phosphor material as a reddish-orange emitting element in w-LEDs when excited by n-UV or blue light, presented in **Table 3**.

Table 3

The CIE chromaticity coordinates, CCT (K), emission color and CP (%) of the $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and 0.01 ≤ y ≤ 0.07, Δy = 0.02) phosphor samples.

System Name	CIE		CCT (K)	Emission color	CP (%)
	X	Y			
D3	0.3327	0.3739	5495	Neutral white	14.89
D3S1	0.5397	0.4047	1882	Warm white	68.73
D3S3	0.5655	0.4101	1791	Warm white	76.64
D3S5	0.5746	0.4120	1768	Warm white	79.31
D3S7	0.5726	0.4143	1782	Warm white	79.57

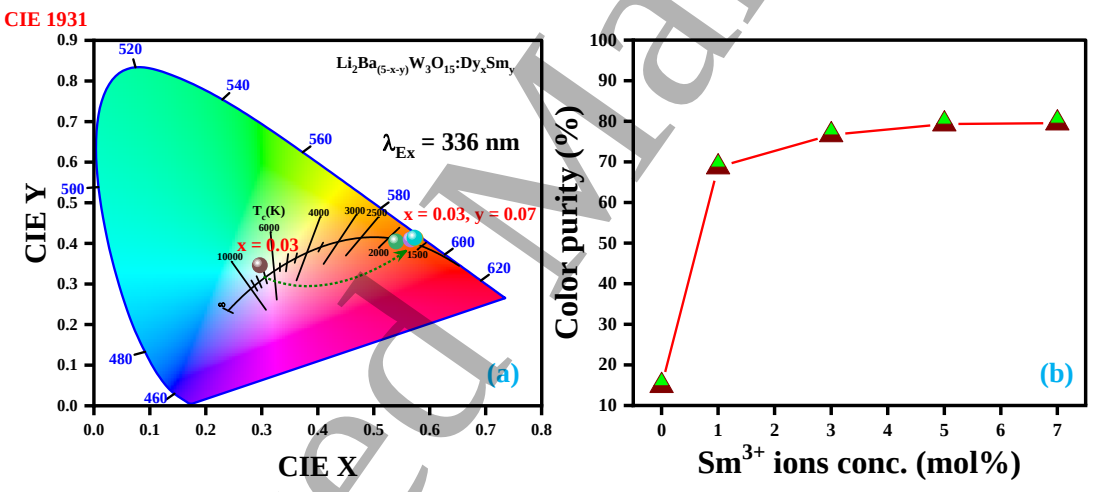


Fig. 10 (a) CIE chromaticity coordinates, **(b)** Color purity values for $Li_2Ba_{(5-x)}W_3O_{15}:Dy_x$ (where x = 0.03) and $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03 and 0.01 ≤ y ≤ 0.07, Δy = 0.02) phosphors.

Accepted Manuscript

Table 4

Comparison of CCT value of the optimised phosphor $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03, y = 0.05 (D3S5)) sample with previously reported literature.

Material	CCT (K)	References
$NaY(MoO_4)_2$:Dy10Sm15	3705	[17]
CYN:Dy5Sm7	3241	[32]
NYF:Dy1Sm1.6	4174	[55]
$CaLaTiO_3$:Dy4Sm1	2591	[56]
D3S5	**1768**	**This work**

3.2.2.6 Study of temperature dependent PL

The thermal stability of the phosphor samples prepared can be well understood using thermogravimetric analysis (TGA) and the temperature-dependent photoluminescence (TD-PL) studies. We have studied the TGA of the un-doped LBW phosphor and reported it in our previous paper [50]. To complement the encouraging TGA result, the TD-PL studies have been recorded in the present work for the co-doped LBW phosphors. It is well known that, temperature has a significant effect on w-LED performance; UV/blue LED chips lose optical performance as the temperatures rises. Thus, for the optimized phosphor sample $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03, y = 0.05) at λ_{Ex} = 336 nm, TDPL experiments were performed, as shown in **Fig. 11**, to investigate its thermal behavior by raising the temperature from ambient temperature to 443 K. As seen in **Fig. 11 (a)**, the PL emission peak intensity seems to be slightly decreasing as the temperature increases from 298 K to 443 K. However, the peak position remains unchanged. **Fig. 11 (b)** displays the shift of CIE chromaticity coordinates with temperature ranges from 293 K to 443 K. The CIE chromaticity coordinates are almost unchanged for $\Delta x < 0.001$ and $\Delta y < 0.006$, although the temperature rises from 298 K to 443 K.

The activation energy must be investigated in order to better understand the thermal quenching effect, and it can be estimated using the Arrhenius equation [57]:

$$I_T = \frac{I_0}{[1+c\ exp\ (-\Delta E/KT)} \quad (15)$$

where, I_0 is the emission intensity at room temperature and I_T is the emission intensity for each changing temperature [58], A is a constant and the Boltzmann constant, k, is equal to 8.62×10^{-5} eV/K. The activation energy (ΔE), can be obtained using the relationship between

Accepted Manuscript

$ln\,[(I_0/I_T)-1]$ and $1/KT$, as seen in **Fig. 11 (c)**. Consequently, following the linear fitting for these sites, ΔE was calculated to be 0.16 eV, exceeding earlier researches [55,59–62], presented in **Table 5**.

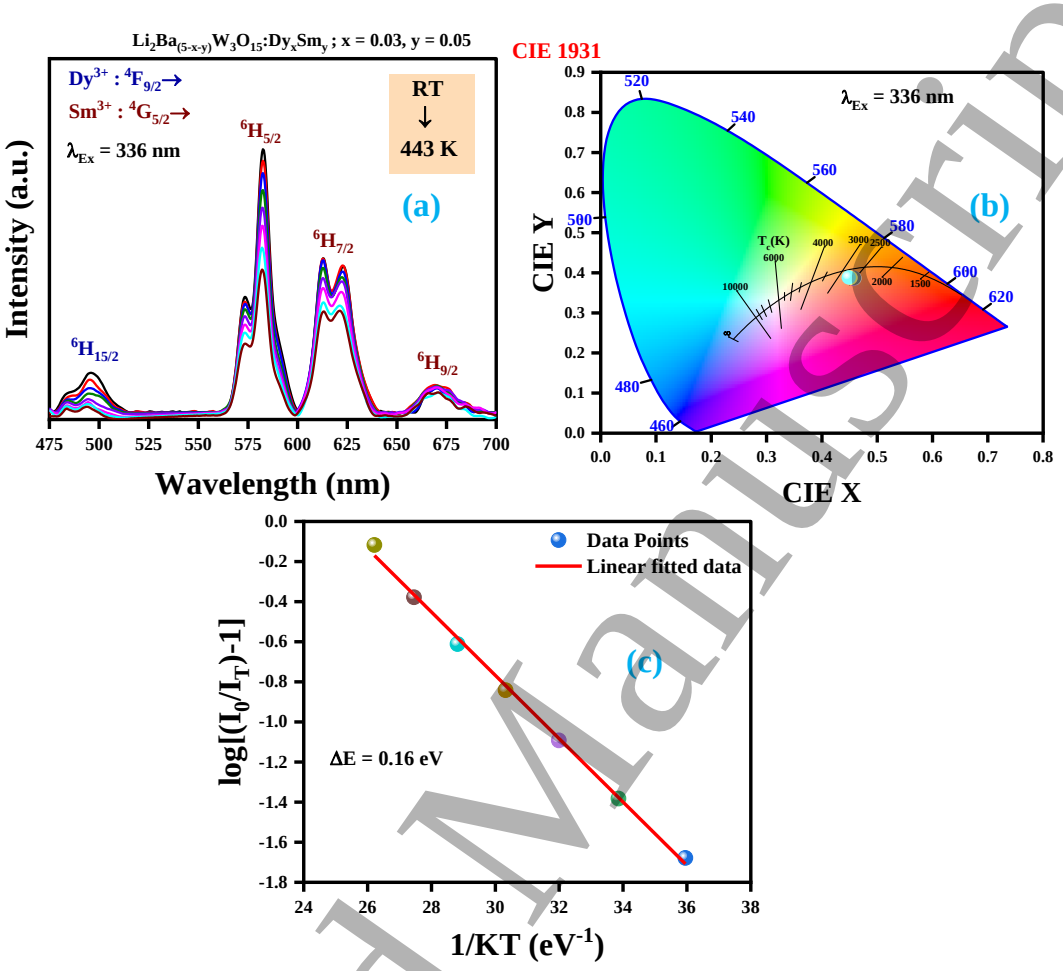


Fig. 11 (a) TDPL, **(b)** CIE color chromaticity coordinates and **(c)** The linear fitting curve of $ln\,[(I_0/I_T)-1]$ against $1/KT$ for $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03, y = 0.05) phosphor sample.

Accepted Manuscript

Table 5

Comparison of activation energy of the prepared $Li_2Ba_{(5-x-y)}W_3O_{15}:Dy_xSm_y$ (where x = 0.03, y = 0.05 (D3S5)) sample with some previously prepared phosphor samples.

Sensing material	Activation energy (ΔE, eV)	References
Na5Y9F32:Dy1Sm0.8	0.14	[55]
CLGP:Dy0.20	0.09	[59]
Li0.97NbO3: Dy2Sm1	0.15	[60]
Na5Y9F32:Dy1Sm2.4	0.15	[61]
Lu1.93Ge2O5:Bi6Sm1	0.11	[62]
D3S5	**0.16**	**This work**

4. Conclusions

In conclusion, the high-temperature solid-state reaction technique was used to synthesize a series of Dy^{3+}/Sm^{3+} co-doped warm white LBW phosphors. The addition of dopant (Dy^{3+}/Sm^{3+}) ions in the LBW host lattice without inducing structural changes was shown by the slight shift in bandgap with the change in its concentration. Warm white light emission with a low CCT (1768 K) value can be achieved by simply changing the concentration of Sm^{3+} which makes the light emission appropriate for indoor lighting applications. Co-doping of LBW:Dy^{3+} phosphors with Sm^{3+} ions allow for the modification of luminescent colors from a neutral white to orange-red region. The concentration quenching of singly Dy^{3+} ions doped and Dy^{3+}/Sm^{3+} co-doped LBW phosphors is caused by d-d interaction, suggested by Dexter theory and Reisfeld approximation. In this study, the maximum η_{ET} was determined to be 17.78%. The optimised phosphor sample's activation energy of 0.16 eV was discovered to be higher than that of previous reported phosphors, indicating the better thermal stability of the prepared phosphor. According to our findings, the Dy^{3+}/Sm^{3+} co-doped LBW phosphor has a wide range of potential applications in warm light w-LED technological advances because of its superior solubility for RE^{3+} ions, high thermal stability, and high luminous efficiency.

Accepted Manuscript

Declaration of Competing Interest

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

Acknowledgement

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.

REFERENCES:

[1] Wang J, Guo J, Lv Q, Ma Z, Feng X, Lu Y, Gao J, Chen W, Deng B and Yu R 2022 Spectroscopic investigation of the novel orange-red phosphor $Ca_3La_2W_2O_{12}:Sm^{3+}$ with the high color purity for w-LED applications *J. Lumin.* **241** 118459

[2] Wu Q, Li Y, Wang Y, Liu H, Ye S, Zhao L, Ding J and Zhou J 2020 A novel narrow-band blue-emitting phosphor of Bi^{3+}-activated $Sr_3Lu_2Ge_3O_{12}$ based on a highly symmetrical crystal structure used for WLEDs and FEDs *Chem. Eng. J.* **401** 126130

[3] Gao H, Wang M, Mao Y, Zhao Q, Zhang C, Jin L, Xia Y, Li W, Zhang K and Zeng H 2023 Intramolecular-Catalyzed Vitrimers with White-Light Emission for Light-Emitting Diode Encapsulation *Adv. Mater.* **35**

[4] Zhang X, Cui R, Zhang J, Yuan G, Qi X and Deng C 2021 Structural and photoluminescence properties of a novel Ba_2GdNbO_6: Sm^{3+} phosphor *Optik (Stuttg).* **245** 167646

[5] Shao B, Huo J and You H 2019 Prevailing Strategies to Tune Emission Color of Lanthanide-Activated Phosphors for WLED Applications *Adv. Opt. Mater.* **7**

[6] Chen D, Xiang W, Liang X, Zhong J, Yu H, Ding M, Lu H and Ji Z 2015 Advances in transparent glass–ceramic phosphors for white light-emitting diodes-A review *J. Eur. Ceram. Soc.* **35** 859–69

[7] Zhang J, Gu G, Di X, Xiang W and Liang X 2019 Optical characteristics of Ce,Eu:YAG single crystal grown by Czochralski method *J. Rare Earths* **37** 145–50

[8] Sun L, Devakumar B, Liang J, Wang S, Sun Q and Huang X 2020 A broadband cyan-emitting $Ca_2LuZr_2(AlO_4)_3$:Ce^{3+} garnet phosphor for near-ultraviolet-pumped warm-white light-emitting diodes with an improved color rendering index *J. Mater. Chem. C* **8** 1095–103

Accepted Manuscript

[9] Ta C, Wen H, Mo Z, Huo Y, Mu Z, Dawi E A, Luo L and Hakeem D A 2023 Preparation and luminescence investigations of garnet-based $Y_2Ca_2Ga_3VO_{12}$:RE (RE = Eu^{3+}, Sm^{3+}, and Dy^{3+}) phosphors *Mater. Res. Bull.* **164** 112254

[10] Meyer J and Tappe F 2015 Photoluminescent Materials for Solid-State Lighting: State of the Art and Future Challenges *Adv. Opt. Mater.* **3** 424–30

[11] Su Y-K, Peng Y-M, Yang R-Y and Chen J-L 2012 Effects of NaCl flux on microstructure and luminescent characteristics of $KSrPO_4$:Eu^{2+} phosphors *Opt. Mater. (Amst).* **34** 1598–602

[12] Yu R, Mi Noh H, Moon B K, Choi B C, Jeong J H, Jang K, Lee H S and Yi S S 2013 Enhanced photoluminescence of $BaAl_2S_4$:Eu^{2+} phosphor by Mg^{2+} ions doping *Ceram. Int.* **39** 9709–13

[13] Wang H, Liu Y, Zhu X, Wei L, Jiang X, Chen Y and Li L 2020 Preparation of $CaAlSiN_3$:Eu^{2+} red-emitting phosphor by a two-step method for solid-state lighting applications *Ceram. Int.* **46** 23035–40

[14] Jo D S, Luo Y, Senthil K, Toda K, Kim B S, Masaki T and Yoon D H 2012 Synthesis and photoluminescence properties of new $NaAlSiO_4$:Eu^{2+} phosphors for near-UV white LED applications *Opt. Mater. (Amst).* **34** 696–9

[15] Long J, Chu F, Wang Y, Zhao C, Dong W, Yuan X, Ma C, Wen Z, Ma R, Du M and Cao Y 2017 $M_8MgSc(PO_4)_7$: x Dy^{3+} (M = Ca/Sr) Single-Phase Full-Color Phosphor with High Thermal Emission Stability *Inorg. Chem.* **56** 10381–6

[16] Sun X, Huang Z, Fu X, Xu L, Liu K and Yuan H 2020 Generation of warm white light by doping Sm^{3+} in Ca_3TeO_6:Dy^{3+} fluorescent powders *Ceram. Int.* **46** 14252–6

[17] Mu B, Huang J, Song R, Dong Y and Qi B 2023 Optical analysis of Dy^{3+}/Sm^{3+} co-doped $NaY(MoO_4)_2$ phosphors for warm white LED *J. Alloys Compd.* **945** 168916

[18] Ouertani G, Maciejewska K, Piotrowski W, Horchani-Naifer K, Marciniak L and Ferhi M 2024 High thermal stability of warm white emitting single phase $GdPO_4$: Dy^{3+}/Sm^{3+} phosphor for UV excited wLEDs *J. Lumin.* **265** 120228

[19] Vijayakumar M and Marimuthu K 2023 Effective role of Co-dopants (Dy^{3+}, Sm^{3+}) on Tunable multicolour emission of Tb^{3+} doped telluroborate glasses for Natural White light applications *J. Non. Cryst. Solids* **620** 122591

[20] Malchukova E 2022 RE-codoping effect on emission ability of Sm,Gd- and Sm,Eu-codoped complex borosilicate glasses *Mater. Res. Bull.* **152** 111847

[21] Wang Y, Song M, Xiao L and Li Q 2021 Upconversion luminescence of Eu^{3+} and Sm^{3+} single-doped $NaYF_4$ and $NaY(MoO_4)_2$ *J. Lumin.* **238** 118203

Accepted Manuscript

[22] Wang X, Zhao Z, Deng M, Yang Z, Wu Q and Zhang W 2022 Color-tunable luminescence performance of single-component $Na_5Y(MoO_4)_4:Dy^{3+}$, Tm^{3+} white-emitting phosphor for white light-emitting diodes *Ceram. Int.* **48** 22869–76

[23] Zhu H, Liang C and Huang W 2020 Crystal structure and luminescent properties of Li^+ ions doped double tungstate $KEu(WO_4)_2$ red emitting phosphors *Phys. B Condens. Matter* **582** 411999

[24] Huang D, Zhou Y, Xu W, Yang Z, Hong M and Yu J 2012 Synthesis and luminescent properties of $Na_5La_{1-x}Eu_x(WO_4)_4$ red phosphors *J. Lumin.* **132** 2788–93

[25] Bi Y, Yang H, Yun X, Zhang H, Ma X, Li G, Liang Y and Chen Y 2020 Tailoring narrow-band blue emitting $Ca_7(Mg, Zn)_2P_6O_{24}$: Eu^{2+} phosphate phosphor via crystal lattice distortion for WLEDs *Opt. Mater. (Amst).* **107** 110094

[26] Zhou J, Yao Y, Chen Y, Wang B, Huang X, Lai Y, Zhang Y and Wu Q 2020 Synthesis, energy transfer mechanism, and tunable emissions of novel $Na_3La(VO_4)_2:Re^{3+}$ (Re^{3+} = Dy^{3+}, Eu^{3+}, and Sm^{3+}) vanadate phosphors for near-UV-excited white LEDs *Ceram. Int.* **46** 6276–83

[27] Khaidukov N M, Kirm M, Feldbach E, Mägi H, Nagirnyi V, Tõldsepp E, Vielhauer S, Jüstel T, Jansen T and Makhov V N 2017 Luminescence properties of silicate apatite phosphors $M_2La_8Si_6O_{26}$:Eu (M = Mg, Ca, Sr) *J. Lumin.* **191** 51–5

[28] Yan S A, Chang Y S, Wang J W, Hwang W S and Chang Y H 2011 Synthesis and luminescence properties of color-tunable $Ba_{1-y}Sr_yLa_{4-x}Tb_x(WO_4)_7$ (x = 0.02-1.2, y = 0-0.4) phosphors *Mater. Res. Bull.* **46** 1231–6

[29] Barros B S, de Lima A C, da Silva Z R, Melo D M A and Alves-Jr S 2012 Synthesis and photoluminescent behavior of Eu^{3+}-doped alkaline-earth tungstates *J. Phys. Chem. Solids* **73** 635–40

[30] Thorp J S and Ammar E A E 1976 Spin-lattice relaxation in gadolinium-doped calcium tungstate *J. Mater. Sci.* **11** 1215–9

[31] Jana S, Mondal A, Manam J and Das S 2020 Pr^{3+} doped $BaNb_2O_6$ reddish orange emitting phosphor for solid state lighting and optical thermometry applications *J. Alloys Compd.* **821** 153342

[32] Shi Y, Cui R, Liu X, Yang L, Li Y, Zhang J and Deng C 2023 Ca_2YNbO_6: Dy^{3+}, Sm^{3+}/Eu^{3+} warm white phosphors with tunable light color and FIR optical temperature measurement characteristics *Opt. Mater. (Amst).* **136**

[33] Kumari C, Kumar A, Sharma S K and Manam J 2023 Sr_3LiSbO_6: Er^{3+} phosphors for green LEDs and solar cell applications *Vacuum* **207** 111599

Accepted Manuscript

[34] Carnall W T, Fields P R and Rajnak K 1968 Spectral Intensities of the Trivalent Lanthanides and Actinides in Solution. II. Pm^{3+}, Sm^{3+}, Eu^{3+}, Gd^{3+}, Tb^{3+}, Dy^{3+}, and Ho^{3+} *J. Chem. Phys.* **49** 4412–23

[35] Ning Liu, Lefu Mei, Jianxiong Bin, Ze Zhang Z P 2021 Effect of anionic group $[SiO_4]^4/[PO_4]^{3-}$ on the luminescence properties of Dy^{3+}-doped tungstate structural compounds 843–51

[36] Xiang Y, Zheng Y, Yang L, Li M, Mao Y and Zhu J 2022 Exploring a new Dy^{3+}-activated borotellurate phosphor with thermally stable photoluminescence *J. Alloys Compd.* **919** 165837

[37] Huang H, He Y, Lin Z, Kang L and Zhang Y 2013 Two novel Bi-based borate photocatalysts: Crystal structure, electronic structure, photoelectrochemical properties, and photocatalytic activity under simulated solar light irradiation *J. Phys. Chem. C* **117** 22986–94

[38] Rohilla P and Rao A S 2022 Synthesis optimisation and efficiency enhancement in Eu^{3+} doped barium molybdenum titanate phosphors for w-LED applications *Mater. Res. Bull.* **150** 111753

[39] Monika D L, Nagabhushana H, Krishna R H, Nagabhushana B M, Sharma S C and Thomas T 2014 Synthesis and photoluminescence properties of a novel Sr_2CeO_4:Dy^{3+} nanophosphor with enhanced brightness by Li+ co-doping *RSC Adv.* **4** 38655–62

[40] Carnall W T, Fields P R and Rajnak K 1968 Electronic Energy Levels in the Trivalent Lanthanide Aquo Ions. I. Pr^{3+}, Nd^{3+}, Pm^{3+}, Sm^{3+}, Dy^{3+}, Ho^{3+}, Er^{3+}, and Tm^{3+} **4424**

[41] Anu, Deopa N and Rao A S 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

[42] Uitert L G Van, Soc J E and Uitert I G Van 1967 Characterization of Energy Transfer Interactions between Rare Earth Ions Characterization of Energy Transfer Interactions between Rare Earth Ions *J. Electrochem. Soc.* **114** 1048–53

[43] Anu and Rao A S 2023 Luminescence and optical thermometry strategy based on emission spectra of $Li_2Ba_5W_3O_{15}$: Pr^{3+} phosphors *Opt. Mater. (Amst).* **145** 114476

[44] Dexter D L and Schulman J H 1954 Theory of Concentration Quenching in Inorganic Phosphors *J. Chem. Phys.* **22** 1063–70

[45] Li H, Zhao R, Jia Y, Sun W, Fu J, Jiang L, Zhang S, Pang R and Li C 2014 $Sr_{1.7}Zn_{0.3}CeO_4$: Eu^{3+} Novel Red-Emitting Phosphors: Synthesis and Photoluminescence Properties *ACS Appl. Mater. Interfaces* **6** 3163–9

Accepted Manuscript

[46] Deopa N, Saini S, Kaur S, Prasad A and Rao A S 2019 Spectroscopic investigations on Dy^{3+} ions doped zinc lead alumino borate glasses for photonic device applications *J. Rare Earths* **37** 52–9

[47] Anu, Kumar A, Ravina, Deopa N and Rao A S 2023 Efficient tunable photoluminescence of Dy^{3+}/Eu^{3+} co-doped OFSZBS glasses for warm white LEDs *J. Non. Cryst. Solids* **616** 122421

[48] Anu and Rao A S 2023 Synthesis , Structural and Fluorescence Investigations of Novel - $Li_2Ba_5W_3O_{15}$: Sm^{3+} Phosphors for Photonic Device Applications *J. Fluoresc.*

[49] Bachmann V, Ronda C and Meijerink A 2009 Temperature Quenching of Yellow Ce^{3+} Luminescence in YAG:Ce *Chem. Mater.* **21** 2077–84

[50] Anu and Rao A S 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**

[51] Rao A S 2022 Effective sensitization of Eu^{3+} visible red emission by Sm^{3+} in thermally stable potassium zinc alumino borosilicate glasses for photonic device applications *J. Lumin.* **244** 118689

[52] Kaur S, Rao A S and 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

[53] Gao D, Chen B, Sha X, Zhang Y, Chen X, Wang L, Zhang X, Zhang J, Cao Y, Wang Y, Li L, Li X, Xu S, Yu H and Cheng L 2024 Near infrared emissions from both high efficient quantum cutting (173%) and nearly-pure-color upconversion in $NaY(WO_4)_2$:Er^{3+}/Yb^{3+} with thermal management capability for silicon-based solar cells *Light Sci. Appl.* **13** 17

[54] Rohilla P and Rao A S 2022 Linear and non-linear photoluminescence studies of Ho^{3+}/Yb^{3+} co-doped titanate phosphors for photonic applications *J. Alloys Compd.* **928** 167156

[55] Xu F, Zhou X, Xia H, Song H and Chen B 2021 Highly thermally stable Dy^{3+}/Sm^{3+} co-doped $Na_5Y_9F_{32}$ single crystals for warm white LED *J. Phys. Chem. Solids* **158** 110240

[56] Fu S and Tian L 2020 Tunable cool to warm white emission in $Ca_{1-x}La_{2x/3}TiO_3$ coped with Dy^{3+} and Sm^{3+} for white LED application *Optik (Stuttg).* **219** 165305

[57] Ding Y, Guo N, Zhu M, Lv W, Ouyang R, Miao Y and Shao B 2020 Luminescence and temperature sensing abilities of zincate phosphors co-doped bismuth Bi^{3+} and lanthanide Eu^{3+}/Sm^{3+} *Mater. Res. Bull.* **129** 110869

Accepted Manuscript

[58] Lee G, Han J Y, Im W Bin, Cheong S H and Jeon D Y 2012 Novel Blue-Emitting Na x $Ca_{1-x}Al_{2-x}Si_{2+x}O_8$:Eu^{2+} (x = 0.34) Phosphor with High Luminescent Efficiency for UV-Pumped Light-Emitting Diodes *Inorg. Chem.* **51** 10688–94

[59] Liu S, Liu S, Wang J, Sun P, Zhong Y, Jeong J H, Deng B and Yu R 2018 Preparation and investigation of Dy^{3+}-doped $Ca_9LiGd_{2/3}(PO_4)_7$ single-phase full-color phosphor *Mater. Res. Bull.* **108** 275–80

[60] Min Z, Zeng Q, Chen S, Qin Y and Yao C 2022 Tunable photoluminescence of $LiNbO_3$: RE^{3+} (RE^{3+} = Dy^{3+}, Sm^{3+}, Dy^{3+}/Sm^{3+}) single-phase phosphors for warm white LEDs *J. Alloys Compd.* **924** 166497

[61] Chen C, Yu C, Xu F, Li Q and Zhang Y 2021 High energy transfer in Dy^{3+}/Sm^{3+}co-doped transparent borosilicate glass-ceramics containing novel $Na_5Y_9F_{32}$ nanocrystals for w-LEDs applications *Ceram. Int.* **47** 1–9

[62] Li Q, Zhang S, Lin W, Li W, Li Y, Mu Z and Wu F 2020 A warm white emission of Bi^{3+}-Eu^{3+} and Bi^{3+}-Sm^{3+} codoping $Lu_2Ge_2O_7$ phosphors by energy transfer of Bi^{3+}-sensitized Eu^{3+}/Sm^{3+} *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **228** 117755

Accepted Manuscript

Signal, Image and Video Processing
https://doi.org/10.1007/s11760-024-03269-z

ORIGINAL PAPER

SSR-GAN: super resolution-based generative adversarial networks model for flood image enhancement

Vinay Dubey[1] · **Rahul Katarya**[1]

Received: 18 September 2023 / Revised: 27 April 2024 / Accepted: 5 May 2024
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

Abstract
Floods, a common natural disaster, it affects more than half of all natural disasters, primarily due to high floods, high tides, heavy rainfall, and human activity. Distinguishing between the flooded and non-flood areas during the disaster is dependent on the quality of images collected from different sources. Since synthetic aperture radar (SAR) images can see through cloud cover and capture photos of the Earth's surface during bad weather, they are excellent for detecting floods. Image Enhancement is the approach that helps to increase the resolution of the SAR images, which helps to more accurately categorize the flooded and non-flooded areas. The quality of satellite and aerial imagery can be enhanced to recognize flooded and non-flooded areas. Our proposed approach uses a super-resolution technique to enhance the resolution of SAR images. We develop a super resolution-based generative adversarial network, or SSR-GAN (Satellite Super Resolution-based Generative adversarial network). In our approach to estimating generative models through an adversarial process, we concurrently train two different models: a generative model G and a discriminative model D. An image with low resolution is transformed into high pixel density. The peak signal-to-noise ratio (PSNR), the structural similarity index(SSIM), Multiscale Structural Similarity (MSSIM) and mean squared error (MSE) are the performance metrics used to compare the interpolation methods for enhancing resolution. In comparison to the existing approaches with our proposed model SSR-GAN, the values of PSNR, SSIM, and MSSIM are greater, and the MSE error value is lower.

Keywords Flood disaster · Generative adversarial network · Image enhancement · Super-resolution

1 Introduction

Flood catastrophes pose a significant threat to urban areas, posing a greater threat to diversity and community than other natural or man-made disasters. Vulnerable cities worldwide face extreme tides, hurricanes, river overflow, and inadequate drainage systems. Numerous cities across the globe are vulnerable to this threat, which is mostly caused by isolated areas of intense precipitation. Because they are unable to handle the floods caused by heavy precipitation and inadequate drainage systems frequently cause floods.

1.1 Problem statement

A flood occurs when the water level of the river rises above the river bank due to the flow of the river. High river levels are typically caused by river precipitation and the backwater resulting from high tides near the river's mouth [1]. Floods cause significant financial losses and require efficient detection and prediction systems. Accurate mapping using SAR photos and satellite imagery is crucial for detecting and predicting flooded areas. Detecting urban flooded areas helps cities evolve flood maps, reducing risk. Mapping flooded areas is complex due to discontinuous, shallow, and ephemeral flooding [2]. For better disaster response and conditional awareness on the ground is very essential. Nowadays Satellite imagery is crucial for disaster response, with Synthetic Aperture Radar (SAR) being a promising technique for detecting Earth's surface, even in cloudy environments, and studies have shown its effectiveness in detecting floods [3].

✉ Rahul Katarya
rahuldtu@gmail.com

Vinay Dubey
vinaydubeyamit@gmail.com

[1] Delhi Technological University, New Delhi, India

Published online: 21 May 2024

SAR's all-weather and night-time accessibility make it crucial for efficient disaster prevention and rescue operations, providing clear, detailed data under harsh conditions.

1.2 Major contribution

However, some progress has been made in using SAR for urban flood scenarios when utilized with high-quality data as well as creative image processing techniques [4]. There is some issue with satellite images because some satellite images have low contrast images. Blurry, noise and low contrast can cause the image to deteriorate, which Reduces the target detection ability of SAR images [5]. Research has indicated that social sensing could be a useful data source for guiding flood relief and emergency response decisions, as well as advances in science and technology that make use of innovative methods for data collection and processing [6]. To improve image quality, we propose our SSR-GAN model which utilizes a single generator, discriminator, and VGG net, in our proposed GAN model to calculate perceptual loss and structural similarity. The experimental results show that our model outperforms current methodologies in numerous image quality enhancement matrices.

- We present a new, competent super resolution-based approach for improving the quality of satellite-based flood images.
- We use generative adversarial networks GAN for super-resolution. Our model can restore fine details and enhance the visual quality of the photographs in comparison to conventional methods, which fail to provide visually realistic results.
- VGG NET is used to calculate perceptual difference, utilizing intermediate feature maps to measure perceptual loss between generated and target images, guiding image generation process optimization.
- We worked on a large-scale multi-temporal dataset for the past ten years, incorporating images before and after the event under different weather conditions.
- The comprehensive comparison performed between the proposed SSR-GAN and the state-of-art approaches shows that our model outperforms existing techniques.

The rest of the article is structured as follows. Section 2 discusses the related work. Section 3 proposed methodology is explained. The experimental findings, analyses, and dataset summary are included in Sect. 4. Finally, Sect. 5 presents the work's and conclusion.

2 Related work

The literature review examines various methodologies to enhance Synthetic Aperture Radar (SAR) images, focusing mainly on flood control and surveillance applications [7]. Speckle noise reduction is crucial for improving SAR image interpretation, target detection, and classification. Techniques such as fuzzy ARTMAP and deep neural networks enhance image quality and facilitate flood prediction [8–10]. Edge detection filters and region classification aid in identifying potential groundwater locations, while mathematical models analyze the effects of geological features on runoff and soil erosion [11, 12]. Moreover, models like MA, NSCT, and MP-CMA are prominent for SAR image enhancement, integrating speckle reduction with feature techniques such as ℓp-LSCE and GAN-FEM address noise enhancement [13]. Automatic target recognition (ATR) reduction and improved target.

feature extraction [14, 15]. Integrating artificial intelligence into change detection approaches has gained traction, particularly for identifying flood-affected regions using supervised classification methods [16, 17]. Post-event flood mapping relies on hydraulic model simulations, SAR, and optical satellite imagery [18, 19]. Innovative approaches like generative adversarial networks (GANs) with perceptual and structural loss functions are proposed for SAR image quality improvement [20]. Automated methods and machine learning models accelerate flood map creation during emergencies, leveraging tools like Google Earth Engine for satellite imagery acquisition and processing [21–23]. Unscented Kalman Filter (UKF), and Super Resolution (SR) are used to reduce noise from images [24]. Additionally, techniques such as Histogram Equalization (HE), Adaptive Histogram Equalization (AHE), and super-resolution reconstruction contribute to image enhancement [25–27]. Furthermore, for non-uniformly illuminated fundus images, GANs with illumination-guided attention are suggested, along with machine learning methods like Principal Component Analysis (PCA) and Random Forest (RF) [28, 29]. Dual-branch neural networks and compressed sensing methods aid in single-image rain removal and image compression, respectively [30, 31]. Intelligent prediction models like Deep-SIQRV for disease spread and flood index insurance algorithms are also proposed [32, 33]. Multisource remote sensing approaches and meta-heuristic feature selection techniques enhance flood forecasting accuracy [34]. Lastly, image enhancement methodologies for underwater river crab target recognition, COVID-19 patient CT scans, low-light picture enhancement, and satellite-based inference models are explored [35–39]. An overview of relevant works that are comparative is given in Table 1

Some significant research articles are given in the table above. SAR pictures are the primary source of study. The

Table 1 A summary of existing methods for SAR image quality enhancement

Ref. No	Dataset	Methodology	Description	Merit	Limitations
[8]	Land sat 8 ETM satellite images	Fuzzy ARTMAP, artificial neural network method	ARTMAP network is made of two modules (ARTb and ARTa) that provide stable identification classes	The ARTMAP approach can produce maps with the desired level of accuracy	For geological regions with a lot of variety, like mountains, the geological maps are not very accurate
[9]	Sentinel- 1A images of ganga basin	Cloud-based google earth engine (GEE) framework	The enhanced EVI and NDVI generated from satellites are used to evaluate the direct effects of floods on plant cover	Quickly determine the particular flood-affected areas and provide a quicker response	As a ratio-based index, (NDVI) and (EVI) both capture intrinsic on-linearity
[10]	Sentinel-1 (SAR satellite) and Sentinel-2 (optical satellite)	Deep neural networks, Generative adversarial networks, edge enhancement	Corresponding positions between SAR and optical pictures were determined using an image translator based on (cGANs)	The proposed methodology could effectively estimate the intersections of optical and SAR pictures	Some special land coverings and textural information are poorly transferred from SAR photos to optical images
[13]	Synthetic aperture radar images	Nonsubsampled contourlet transform, memetic algorithm	An objective enhancement criterion then tries to identify the (nearly) ideal image that uses that criterion	As the speckle noise is reduced, it can enhance the edge characteristics and contrast of SAR images	The algorithm's complexity is increasing, and operating efficiency must be improved
[14]	Synthetic aperture radar images	ℓp-regularization, low-rank, space-angle continuity,	Introduce the ℓp-LSCE, a feature-oriented algorithm to improve the target in a noisy environment	It can use the relationships between pictures to decrease speckles	The suggested 3-D image-generating model will require more processing power
[15]	Monochromatic synthetic aperture images	Generative adversarial network (GAN)	The GAN-driven focusing-enhancement approach (GAN-FEM) is used to fit an unknown out-of-focus kernel	This approach used the 2-D monochromatic imaging system's capacity for focusing enhancement for 3-D targets	The GAN-FEM also introduces some minor issues, such as the addition of energy noise toward the target scatter
[17]	Synthetic aperture radar images (sentinel-1, Sentinel-2 data)	Change detection, Supervised classification	A novel probabilistic change detection method for exposing various types of disruptions caused by excessive precipitation	The advantages of adding the SWIR channel to the visual and near-infrared indexes were only a few percentage points greater in size	Results will be improved by providing refined identification capacity at different types of deteriorations
[21]	RADARSAT-2 Synthetic Aperture Radar (SAR) images	Threshold and polarization	Flood water can be distinguished more effectively using HH polarization	This can facilitate the flood mapping process by ensuring that flood maps are distributed to the relevantly	Combining the information about the backscattering fluctuation might improve the accuracy

Table 1 (continued)

Ref. No	Dataset	Methodology	Description	Merit	Limitations
[24]	Synthetic aperture radar (SAR)	Unscented kalman filter (UKF), super resolution (SR)	A novel approach to super-resolving Satellite image utilizing the ISUKF, which can withstand multiplicative noise, is proposed	An approach to both improve resolution and reduce inherent noise, since traditional super-resolution techniques are not effective	If the transformation is linear then the unscented kalman filter will be slower than the conventional Kalman filtering

survey has highlighted the importance of SAR-based flood image improvement and has noted numerous significant findings.

Speckle noise: Speckle noise, which emerges as granular interference and can reduce image quality, frequently affects SAR photos.

Incorporating multi-temporal data: During a flood event, SAR images taken at various times might reveal important details about the flood dynamics. It would be helpful to develop algorithms that can integrate and analyze time-series SAR data.

Improved contrast enhancement methods: Poor contrast is another issue with SAR images. Novel contrast enhancement methods are required to improve the visibility of flood-related characteristics and the discrimination between flooded and non-flooded areas.

Limited image quality improvement: Traditional image enhancement approaches may struggle to adequately enhance the quality of SAR flood photos, especially in terms of noise reduction, and maintaining essential features.

3 Proposed methodology

This section outlines the fundamental ideas that are required for an exhaustive comprehension and efficient implementation of our work. The following essay gives a general introduction to generative adversarial networks, super-resolution, and image enhancement.

3.1 Super-resolution

Super-resolution (SR) is a term used to describe the process of improving the resolution and amount of detail in low-resolution pictures or videos. By using its lower-resolution counterpart, it is possible to rebuild a higher-resolution version of an image. When using satellite imagery for disaster relief, a high-resolution image is helpful for improved region categorization in multi-spectral remote sensing images [40]. It can be categorized into single-image and multi-image super-resolution techniques.

SISR is a crucial image processing technique that optimizes features and textures to improve visual perception and reconstruct high-resolution images from low-resolution ones [41]. Multi-image SR (MISR) combines numerous LR images, each of which has a unique subset of HR images. As a result, MISR can improve reconstruction accuracy compared to single-image SR (SISR) [42].

3.2 Generative adversarial networks

Generative modeling is an unsupervised machine learning task that automatically learns patterns in incoming data, enabling the model to generate new instances from the original dataset. To create new examples, we train a generator model, and we train a discriminator model to determine whether an example is genuine (coming from the domain) or fake (being generated). Generative Adversarial Networks (GANs) belong to generative models [23]. Currently, the most used data generation has been trained on the dataset. GANs consist of two networks Generator (G) and Discriminator (D). The discriminator is provided with real and generated data points that it must learn to distinguish between. The Generator has to make data points such that it becomes difficult for the discriminator to distinguish. The discriminator learns to separate the underlying distribution between real and generated points, and the Generator needs to make data points with a similar distribution to the real data points. Both the networks are trained simultaneously, thus at the end, having a Generator that captures the underlying real data distributions.

$$\min_{G}\max_{D} V(D, G) \quad \mathrm{E}_{x\sim \mathrm{Pdata}(x)}\left[\log D(x)\right] + \mathrm{E}_{z\sim \mathrm{Pz}(z)}\left[\log(1 - D(G(z)))\right] \tag{1}$$

In the above 1, Eq. 1, P_data is the real data distribution and P_z is the distribution followed by the latent variable

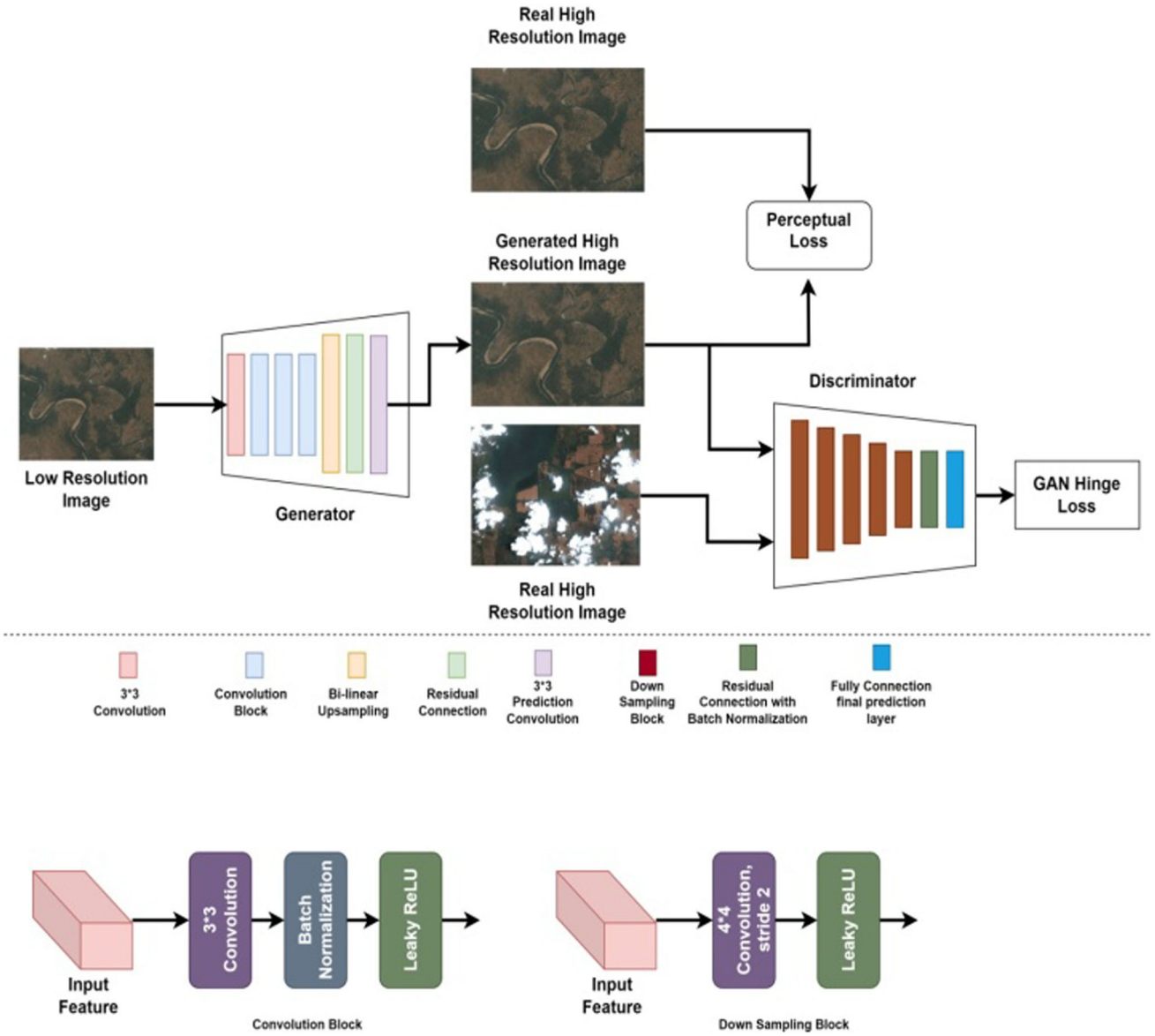


Fig. 1 The architecture of our proposed (SSR-GAN) model for SAR image enhancement

which the generator maps to generated data points mimicking the real data distribution.

3.2.1 Need of VGG net for perception loss

Perceptual loss, a loss term in picture synthesis applications, uses a convolutional neural network to preprocess input mages and calculate the similarity between input and output images. [43]. The usage of VGG16 for feature extraction for calculating perceptual differences is what connects perceptual loss and VGG16. The plan is to use the pre-trained VGG16 network and its intermediate feature maps to calculate the perceptual loss among generated and target images. The perceptual loss can be estimated and used as a guide for optimizing the picture production process by targeting pictures at different levels of VGG16. To compute the comparing the representations of features of the generated perceptual loss and get an astonishing result, multiple convolution layers from VGG-Net, a pre-trained neural.

network on a natural picture dataset was employed in the maximal effort. While we focus on additional image synthesis challenges, such as super-resolution, colorization, and other picture-building tasks [44].

3.3 Model architecture

Our model is inspired by [45] for super-resolution generations to generate high-resolution satellite data images. The

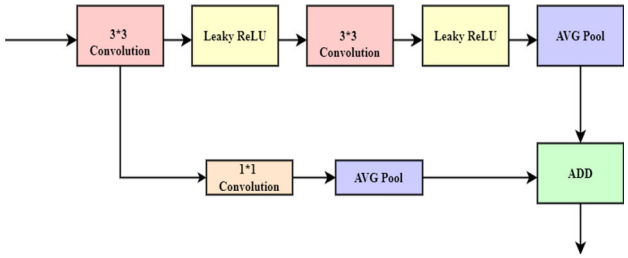


Fig. 2 Block diagram of down-sample block

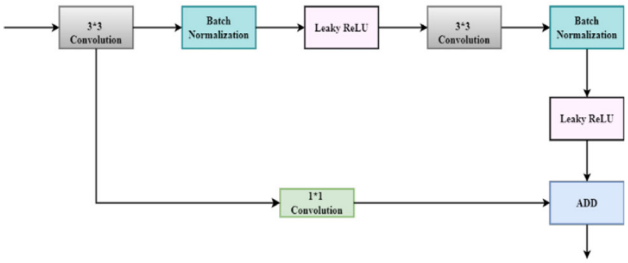


Fig. 3 Block diagram of residual block with batch normalization

block diagram of our suggested SSR-GAN model is shown in Fig. 1. Our model (SSR-GAN) consists of a Generator, Discriminator, and VGG net for preserving structure. At the time of training, we pass a low-resolution image (128*128) which is provided as input to the generator, which consists of a series of convolution and residual blocks followed by a single bilinear upsampling convolution and predict the following output using a residual block and final prediction layer. Our generator is a full convolution network that can upsample images of any resolution. The discriminator is provided with a real high resolution (256*256) and Generated high-resolution image(256*256), which the generator has to distinguish between by predicting real or fake. We also use VGG-NET for the perceptual loss using features extracted from generated high-resolution images and real-pair high-resolution images to preserve the structure of the image. Our model learns distribution to upsample locally, preserving global structure at the time of testing, using 256*256 resolution image input and 512*512 resolution output, enhancing local fine-tuning over predefined steps.

3.3.1 Loss for discriminator

$$\mathcal{L}_{Adv}^{D} = E_{x\sim p_{data}}\left[\max\left(0,\ 1 - D\left(x\right)\right)\right] + E_{\widehat{x}\sim p_G}\left[\max\left(0,\ 1D\left(x\right)\right)\right] \quad (2)$$

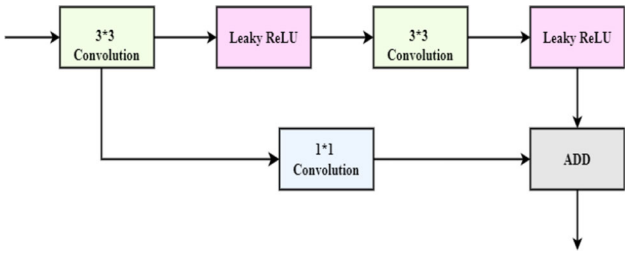


Fig. 4 Block diagram of residual block without batch normalization

In Eq. 2, p_{data} is the data distribution of high-resolution images and p_G is the Generator learned distribution to generate high-resolution images. We use Hinge Adversarial loss introduced in our model [46].

3.3.2 Loss for generator

$$\mathcal{L}_{Adv}^{G} = E_{\widehat{x}\sim p_G}[-D(\widehat{x})] \quad (3)$$

In Eq. 3, which is used to preserve the global structure of the images generated we use perceptual loss [23]. We use features extracted from the 9th layer of the Vgg-16 Network [47].

3.3.3 Perceptual loss

$$\ell_{feat}^{\varnothing,j}(\widehat{y},\ y) = \frac{1}{C_J H_J W_J}||\varnothing_j(\widehat{y}) - \varnothing_j(y)||_2^2 \quad (4)$$

In above Eq. 4,y and $\widehat{y}$ are the targeted images and synthesis images. $\varnothing_j$ represents the perceptual function that outputs the activation of the jth layer in the perceptual loss network. C_J,H_J, W_J are the dimensions of the tensor feature map [48].

The architecture of the downsample block is depicted in above Fig. 2. The discriminator's downsample block is sab-block, followed by the residual block with batch normalization.

In the above Fig. 3, the inner layered architecture of the residual block with batch normalization (BN) layer is explained. Residual block with BN is part of the generator of our proposed model. In below Fig. 4, the inner layered architecture of the residual block is explained. Residual block without BN is also part of the discriminator block of our proposed model.

4 Result and analysis

4.1 Dataset

Images with a size of 256 × 256 are used as inputs in our experiment. The experiment uses 3000 Sentinel-2 images from India over the past 20 years to construct a dataset of landscapes, including lakes, terraces, villages, and snow-capped mountains. The data is primarily prone to flooding, ensuring accurate resolution and geographic makeup.

We will be using data from the previous 20 years for various geographical locations of India because the satellite has a 12-day return cycle. The majority of the area we consider in our study is quite prone to flooding.

Algorithm 1 A detailed description of our proposed SSR-GAN model is provided below Algorithm 1.

Algorithm: Procedure for the SAR image enhancement using the proposed (SSR-GAN) model
Input: Raw (low resolution) SAR images
Output: High-resolution SAR images

Adam Optimizer: Learning Rate:= 0.0001, beta_1 = 0.5,beta_2 = 0.999
Spectral Normalization is used for all Convolution and Fully Connected layers in Discriminator and Generator

---

For the number of training iterations do:
Begin
{
For t = 0,....,n do: \\ Training the discriminator
Sample $\{X_{high(i)}\} \sim P_{ra}$ batch from real data \\ images at high resolution
Sample $\{X_{low(i)}\} \sim P_{ra}$ batch from a real date
\\ images at low resolution
$L_d^{adv} = E_{x \sim P_high}[\max(0,1\text{-}D(x))] + E_{x \sim P_low}[\max(0,1+D(G(x`)))]$
Update the weights of the discriminator using the Adam Optimizer
For t = 0,....,n do: \\ Training the generator

Sample $\{X_{low}\} \sim P_r$ a batch from a real date
\\ images at low resolution
$L_g^{adv} = E_{x' \sim P_high}[-D(G(x`))]$
Update the weights of the generator using Adam Optimizer
}
At the time of inference
For number of samples do:
{
Sample $\{X_{high(i)}\} \sim P_{ra}$ batch from real data
\\ images at high resolution
\\ Generate super Resolution images by providing high-resolution images as inputs

$X_{super(i)} = G(X_{high(i)})$
}
End
Return High-Resolution Image

4.2 Evaluation metrics

Pixel statistics are suitable for evaluating image quality using traditional measures like PSNR, SSIM, MULTI-SCALE SSIM, and MSE.

4.2.1 Peak signal-to-noise ratio

$$\mathrm{PSNR} = 20\log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right) \tag{5}$$

4.2.2 Structural similarity index measure

$$\mathrm{SSIM}(x,\ y) = [(x,\ y)]^{\alpha}\left[c(x,\ y)\right]^{\beta}\left[s(x,\ y)\right]^{\gamma} \tag{6}$$

4.2.3 Multi-scale SSIM

$$\mathrm{MSSIM}(x,\ y) = [l_m\,(x,\ y)]^{\alpha_M}\prod_{j=1}^{M}\left[C_j\,(x,\ y)\right]^{\beta_j}.\left[S_j\,(x,\ y)\right]^{\gamma_j} \tag{7}$$

4.2.4 Mean squared error

$$\mathrm{MSE} = \frac{1}{MN}\sum_{n-1}^{M}\sum_{m-1}^{N}\left[\hat{y}(n,\ m) - y(n,\ m)\right]^2 \tag{8}$$

4.3 Analysis of results

Figure 5 provides the visual results of our proposed model and the existing approaches.

Table 2, shows the statistical measure of the existing model with our proposed model. The proposed model outperforms existing models in terms of PSNR, SSIM, MSSIM, and MSE parameters, with a low MSE value, indicating superior performance.

4.4 Analysis of result

In this section, The proposed model and traditional techniques were compared in terms of PSNR, SSIM, MSSIM, and MSE, with the original image size being (256*256) resolution.

Figure 6 displays the PSNR calculation and comparison of existing and proposed approaches, with epochs displayed on the X-axis and ratio values listed on the Y-axis.

Figure 7 compares existing approaches with the proposed

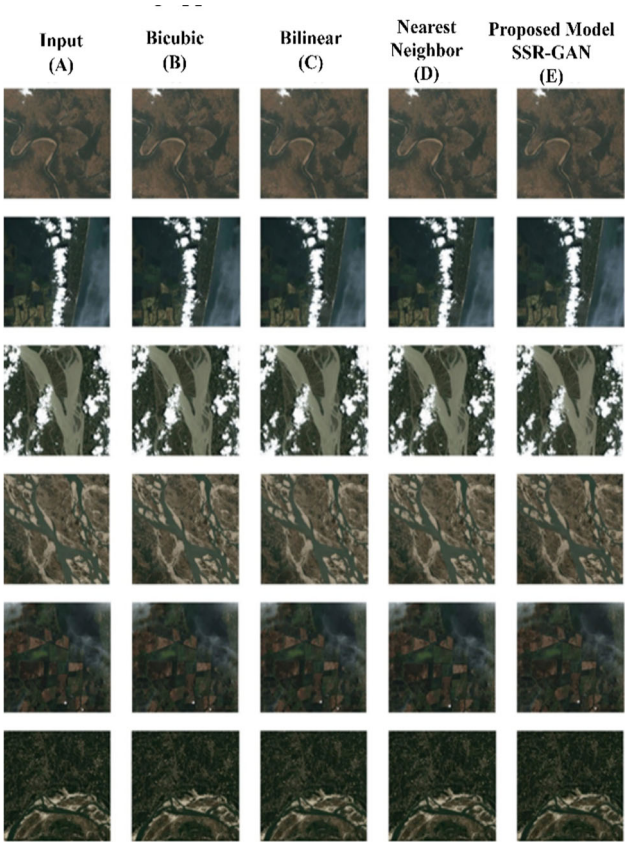


Fig. 5 Comparative image analysis between existing approaches and our proposed model **A** input raw images **B** Bicubic model output **C** Bilinear model output **D** Nearest Neighbour model output **E** Our proposed (SSR-GAN) model output

Table 2 Performance comparison table of the existing model with our proposed model

Method	PSNR↑	SSIM ↑	M-SSIM ↑	MSE ↓
Bilinear interpolation	26.89	0.72	0.89	0.030
Bicubic interpolation	26.34	0.67	0.89	0.031
Nearest neighbor	27.21	0.61	0.84	0.053
Proposed model (SSR-GAN)	**28.11**	**0.81**	**0.97**	**0.012**

Bold values indicate our proposed model results

one, displaying the number of epochs on the X-axis and the percentage of the Structural Similarity Index on the Y-axis.

Figure 8 compares existing approaches with the proposed approach. The X-axis contains the number the epochs and the Y-axis contains the percentage of the Structural Similarity Index, more accurately.

Figure 9 compares existing approaches with the proposed

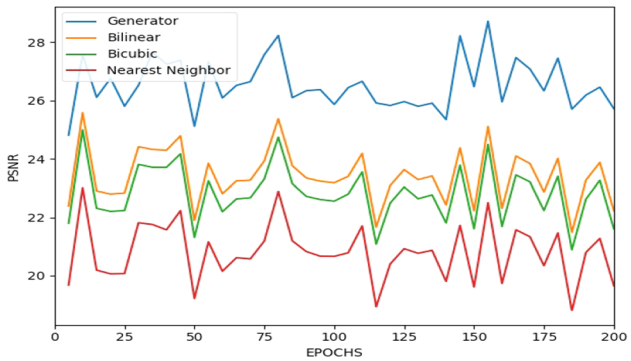


Fig. 6 The graph representation between PSNR and epochs

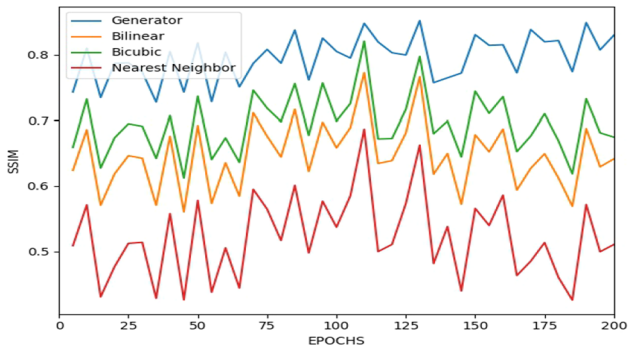


Fig. 7 The graph representation between SSIM and epochs

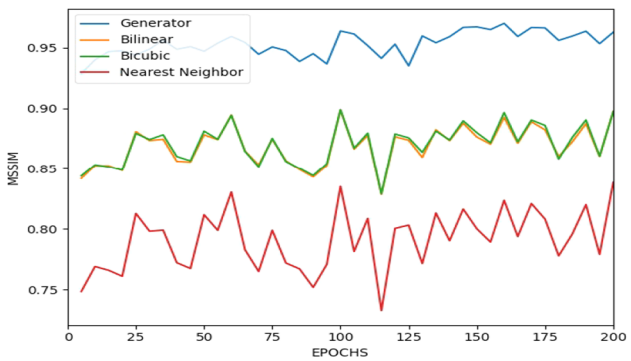


Fig. 8 The graph representation between MSSIM and epochs

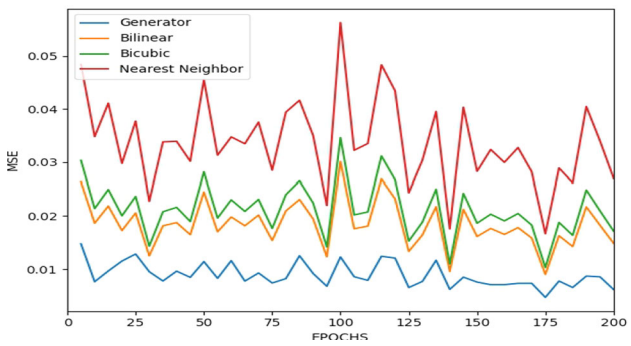


Fig. 9 The graph representation between MSE and epochs

method, with the X-axis representing epochs and the Y-axis representing percentage inaccuracy (MSE).

5 Conclusion and future scope

This paper presents a GAN-based super-resolution approach for flood-based SAR image enhancement, enhancing the quality and spatial resolution of SAR images, particularly for flood management and monitoring applications. The proposed SSR-GAN model included both image quality improvement and a fake-image detection step, signifying significant development in image processing and verification. The loss is computed via VGG-NET. In comparison to other traditional approaches already in use, the proposed SSR-GAN model produces improved PSNR, SSIM, and M-SSIM scores. In comparison to previous methods, our model has a lower MSE value. We supply a (256*256) resolution image as the input, and our model predicts the outcome with an enhanced (512*512) resolution image since it has been taught to learn distribution to upsample locally while maintaining the overall structure. GAN-based SAR image enhancement has shown satisfactory outcomes and has the potential to improve the quality and interpretability of SAR images significantly. Further research in SAR analysis could lead to advancements, but challenges include large datasets and careful hyperparameter adjustment to ensure model resilience and adaptability. Furthermore, a balance between image enhancement and the preservation of important information required careful hyperparameter adjustment.

Author contributions Vinay Dubey Worked on the Conceptualization of the Idea to reach the research goals and aims, conducted a research and investigation process, and Conceived and designed the GAN-based model's methodology and creation. Rahul Katarya helps in Validation, Formal analysis of the research idea, analysis of the collected Dataset, and Preparation for the manuscript's presentation.

Funding No funding was obtained for this study.

Data availability Data is available on request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Tingsanchali, T.: Urban flood disaster management. Procedia Eng. **32**, 25–37 (2012). https://doi.org/10.1016/j.proeng.2012.01.1233
2. Dumitru, C.O., Cui, S., Faur, D., Datcu, M.: Data Analytics for rapid mapping: Case study of a flooding event in Germany and the tsunami in Japan using very high resolution SAR images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.. **8**, 114–129 (2015). https://doi.org/10.1109/jstars.2014.2320777
3. Tong, X., Luo, X., Liu, S., Xie, H., Chao, W., Liu, S., Liu, S., Makhinov, A.N., Makhinova, A.F., Jiang, Y.: An approach for flood monitoring by the combined use of Landsat 8 optical imagery and Cosmo-skymed radar imagery. ISPRS J. Photogramm. Remote. Sens. **136**, 144–153 (2018). https://doi.org/10.1016/j.isprsjprs.2017.11.006
4. Mason, D.C., Speck, R., Devereux, B., Schumann, G.J.-P., Neal, J.C., Bates, P.D.: Flood detection in urban areas using terrasar-X. IEEE Trans. Geosci. Remote Sens. **48**, 882–894 (2010). https://doi.org/10.1109/tgrs.2009.2029236
5. Ponmani, E., Saravanan, P.: Image denoising and despeckling methods for SAR images to improve image enhancement performance: a survey. Multimed. Tool. Appl. **80**, 26547–26569 (2021). https://doi.org/10.1007/s11042-021-10871-7
6. Young, J.C., Arthur, R., Spruce, M., Williams, H.T.P.: Social sensing of flood impacts in India: a case study of Kerala 2018. Int. J. Disaster Risk Reduct. **74**, 102908 (2022). https://doi.org/10.1016/j.ijdrr.2022.102908
7. Dellepiane, S.G., Angiati, E.: Quality Assessment of despeckled SAR images. IEEE J Sel Top. Appl. Earth Obs. Remote Sens. **7**, 691–707 (2014). https://doi.org/10.1109/jstars.2013.2279501
8. Arabi Aliabad, F., Shojaei, S., Zare, M., Ekhtesasi, M.R.: Assessment of the fuzzy ARTMAP neural network method performance in geological mapping using satellite images and boolean logic. Int. J. Environ. Sci. Technol. **16**, 3829–3838 (2018). https://doi.org/10.1007/s13762-018-1795-7
9. Ghosh, S., Kumar, D., Kumari, R.: Evaluating the impact of flood inundation with the cloud computing platform over vegetation cover of Ganga Basin during COVID-19. Spat. Inf. Res. **30**, 291–308 (2022). https://doi.org/10.1007/s41324-022-00430-z
10. Toriya, H., Dewan, A., Ikeda, H., Owada, N., Saadat, M., Inagaki, F., Kawamura, Y., Kitahara, I.: Use of a DNN-based image translator with edge enhancement technique to estimate correspondence between SAR and optical images. Appl. Sci. **12**, 4159 (2022). https://doi.org/10.3390/app12094159
11. Ardakani, A.H., Shojaei, S., Siasar, H., Ekhtesasi, M.R.: Heuristic evaluation of groundwater in arid zones using remote sensing and Geographic Information System. Int. J. Environ. Sci. Technol. **17**, 633–644 (2018). https://doi.org/10.1007/s13762-018-2104-1
12. Shojaei, S., Kalantari, Z., Rodrigo-Comino, J.: Prediction of factors affecting activation of soil erosion by mathematical modeling at Pedon scale under laboratory conditions. Sci. Report. (2020). https://doi.org/10.1038/s41598-020-76926-1
13. Li, Y., Hu, J., Jia, Y.: Automatic sar image enhancement based on nonsubsampled contourlet transform and memetic algorithm. Neurocomputing **134**, 70–78 (2014). https://doi.org/10.1016/j.neucom.2013.03.068
14. Chen, L., Jiang, X., Li, Z., Liu, X., Zhou, Z.: Feature-enhanced speckle reduction via low-rank and space-angle continuity for circular SAR target recognition. IEEE Trans. Geosci. Remote Sens. **58**, 7734–7752 (2020). https://doi.org/10.1109/tgrs.2020.2983420
15. Ye, G., Zhang, Z., Ding, L., Li, Y., Zhu, Y.: Gan-based focusing-enhancement method for monochromatic synthetic aperture imaging. IEEE Sens. J. **20**, 11484–11489 (2020). https://doi.org/10.1109/jsen.2020.2996656

16. Chandran, D.V., Anitha, J.: Change detection & flood water mapping from remotely sensed images- a survey. 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS). (2022). doi: https://doi.org/10.1109/icscds53736.2022.9761015
17. Cerbelaud, A., Roupioz, L., Blanchet, G., Breil, P., Briottet, X.: A repeatable change detection approach to map extreme storm-related damages caused by intense surface runoff based on optical and SAR Remote Sensing: Evidence from three case studies in the south of France. ISPRS J. Photogramm. Remote. Sens. **182**, 153–175 (2021). https://doi.org/10.1016/j.isprsjprs.2021.10.013
18. Scotti, V., Giannini, M., Cioffi, F.: Enhanced flood mapping using synthetic aperture radar (SAR) images, hydraulic modelling, and social media: A case study of hurricane harvey. J. Flood Risk Manag. (2020). https://doi.org/10.1111/jfr3.12647
19. Di, Z., Chen, X., Wu, Q., Shi, J., Feng, Q., Fan, Y.: Learned compression framework with pyramidal features and quality enhancement for SAR Images. IEEE Geosci. Remote Sens. Lett. **19**, 1–5 (2022). https://doi.org/10.1109/lgrs.2022.3155651
20. Luo, Y., Pi, D.: Sar-to-optical image translation for Quality Enhancement. J. Ambient. Intell. Humaniz. Comput. **14**, 9985–10000 (2022). https://doi.org/10.1007/s12652-021-03665-0
21. Manjusree, P., Prasanna Kumar, L., Bhatt, C.M., Rao, G.S., Bhanumurthy, V.: Optimization of threshold ranges for rapid flood inundation mapping by evaluating backscatter profiles of high incidence angle SAR images. Int. J. Disaster Risk Sci. **3**, 113–122 (2012). https://doi.org/10.1007/s13753-012-0011-5
22. Shen, X., Wang, D., Mao, K., Anagnostou, E., Hong, Y.: Inundation extent mapping by synthetic aperture radar: a review. Remote Sensing. **11**, 879 (2019). https://doi.org/10.3390/rs11070879
23. Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature Pyramid Networks for Object Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017). doi: https://doi.org/10.1109/cvpr.2017.106
24. Kanakaraj, S., Nair, M.S., Kalady, S.: SAR image super resolution using importance sampling unscented Kalman filter. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sen. **11**, 562–571 (2018). https://doi.org/10.1109/jstars.2017.2779795
25. Dubey, V., Katarya, R.: Adaptive histogram equalization based approach for SAR Image Enhancement: A Comparative Analysis. 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). (2021). doi: https://doi.org/10.1109/iciccs51141.2021.9432287
26. Zhan, Q., Chen, Y., Chen, Y., Lu, Y., Xu, C.: SAR image super-resolution reconstruction based on an optimize iterative method for regularization. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. (2021). doi: https://doi.org/10.1109/igarss47720.2021.9554072
27. Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. IEEE Trans. Image Proc. **19**, 2861–2873 (2010). https://doi.org/10.1109/tip.2010.2050625
28. Zhang, W., Gao, Y., Cao, L., Zhang, Y., Huang, Z., Wang, B.: A fundus image enhancer based on illumination-guided attention and optic disc perception gan. Optik **279**, 170729 (2023). https://doi.org/10.1016/j.ijleo.2023.170729
29. Shah, N.H., Priamvada, A., Shukla, B.P.: Random Forest-based nowcast model for rainfall. Earth Sci. Inf. **16**, 2391–2403 (2023). https://doi.org/10.1007/s12145-023-01037-0
30. Zheng, Z., Chen, Z., Wang, W., Huang, M., Wang, H.: Dual parallel multi-scale residual overlay network for single-image rain removal. SIViP **18**, 2413–2428 (2023). https://doi.org/10.1007/s11760-023-02917-0
31. El-Ashkar, A.M., Taha, T.E., El-Fishawy, A.S., Abd-Elnaby, M., Abd El-Samie, F.E., El-Shafai, W.: Simultaneous compressed sensing and single-image super resolution for SAR image reconstruction. Optical and Quantum Electronics (2023). https://doi.org/10.1007/s11082-022-04407-y
32. Gupta, A., Katarya, R.: A deep-SIQRV epidemic model for COVID-19 to access the impact of prevention and control measures. Comput. Biol. Chem. **107**, 107941 (2023). https://doi.org/10.1016/j.compbiolchem.2023.107941
33. Thomas, M., Tellman, E., Osgood, D.E., DeVries, B., Islam, A.S., Steckler, M.S., Goodman, M., Billah, M.: A framework to assess remote sensing algorithms for satellite-based Flood Index Insurance. IEEE J Sel. Top. Appl. Earth Obs. Remote Sens. **16**, 2589–2604 (2023). https://doi.org/10.1109/jstars.2023.3244098
34. Farooq, M.S., Tehseen, R., Qureshi, J.N., Omer, U., Yaqoob, R., Tanweer, H.A., Atal, Z.: FFM: Flood forecasting model using Federated Learning. IEEE Access. **11**, 24472–24483 (2023). https://doi.org/10.1109/access.2023.3252896
35. Gupta, A., Singh, A.: edl-nsga-ii: ensemble deep learning framework with nsga-ii feature selection for heart disease prediction. Exp. Syst. (2023). https://doi.org/10.1111/exsy.13254
36. Mirza, M.W., Siddiq, A., Khan, I.R.: A comparative study of medical image enhancement algorithms and quality assessment metrics on COVID-19 CT Images. SIViP **17**, 915–924 (2022). https://doi.org/10.1007/s11760-022-02214-2
37. Liu, X., Ma, W., Ma, X., Wang, J.: Lae-Net: a locally-adaptive embedding network for low-light image enhancement. Pattern Recogn. **133**, 109039 (2023). https://doi.org/10.1016/j.patcog.2022.109039
38. Zhang, J., Hao, S., Rao, Y.: Pre-trained low-light image enhancement transformer. IET Image Proc. (2024). https://doi.org/10.1049/ipr2.13076
39. Selvam, C., Jebadass, R.J., Sundaram, D., Shanmugam, L.: A novel intuitionistic fuzzy generator for low-contrast color image enhancement technique. Information Fusion. **108**, 102365 (2024). https://doi.org/10.1016/j.inffus.2024.102365
40. Anwar, S., Khan, S., Barnes, N.: A deep journey into super-resolution. ACM Comput. Surv. **53**, 1–34 (2020). https://doi.org/10.1145/3390462
41. Ye, S., Zhao, S., Hu, Y., Xie, C.: Single-Image Super-resolution challenges: A brief review. Electronics **12**, 2975 (2023). https://doi.org/10.3390/electronics12132975
42. Kowaleczko, P., Tarasiewicz, T., Ziaja, M., Kostrzewa, D., Nalepa, J., Rokita, P., Kawulok, M.: A real-world benchmark for sentinel-2 multi-image Super-Resolution. Sci. Data. (2023). https://doi.org/10.1038/s41597-023-02538-9
43. Li, M., Hsu, W., Xie, X., Cong, J., Gao, W.: SACNN: Self-attention convolutional neural network for low-dose CT denoising with self-supervised Perceptual Loss Network. IEEE Trans. Med. Imaging **39**, 2289–2301 (2020). https://doi.org/10.1109/tmi.2020.2968472
44. Santos, M.S., Ren, T.I., Kalantari, N.K.: Single image HDR reconstruction using a CNN with masked features and Perceptual Loss. ACM Transactions on Graphics (2020). https://doi.org/10.1145/3386569.3392403
45. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Commun. ACM **63**, 139–144 (2020). https://doi.org/10.1145/3422622
46. Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017). doi: https://doi.org/10.1109/cvpr.2017.19
47. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and Super-Resolution. Computer Vision—ECCV

2016, 694–711 (2016). https://doi.org/10.1007/978-3-319-46475-6_43
48. Liu, Y., Chen, H., Chen, Y., Yin, W., Shen, C.: Generic perceptual loss for modeling structured output dependencies. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (2021). doi: https://doi.org/10.1109/cvpr46437.2021.00538

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.

Croatian Operational Research Review
CRORR **15**(2024), 1–11

Stock assessment using cumulative prospect theory in DEA cross-efficiency model: A case study of the Indian stock market

Reenu Kumari[1,2,*], **Anjana Gupta**[2] and **Abha Aggarwal**[3]

[1] *Department of Applied Sciences, Maharaja Surajmal Institute of Technology, Delhi, India*
E-mail: ⟨reenu_kumari@msit.in⟩

[2] *Department of Applied Mathematics, Delhi Technological University, Delhi, India*

[3] *University School of Basic and Applied Sciences, Guru Gobind Singh Indraprastha University, Delhi, India*
E-mail: ⟨{Gupta, Aggarwal}anjanagupta@dce.ac.in⟩

Abstract. Market volatility is becoming increasingly common as numerous factors are implemented in the financial system. As a result, portfolio managers and individual investors require reliable methods to assess stock performance. This study examines stock assessments using cross-efficiency evaluations in cases where negative data is present. An alternative approach to achieve this goal is to use an RDM DDF-based cross-efficiency model which oversees the negative data. We expand the RDM-based cross-efficiency analysis, which uses row and column average values to select portfolios and identify different groups for stock management. To explore the psychological factors that influence the choices made by stock market investors, we incorporate the cumulative prospect theory value for each stock as an output and the variance as an input to evaluate the overall efficiency of the assets. For the empirical analysis, our study focuses on a sample of 30 stocks listed on the Nifty-50 on India's National Stock Exchange. The results of our empirical study verify that the proposed method can serve as an effective tool for stock selection. This demonstrates how the chosen portfolio gives companies a more diversified and well-balanced approach for selecting stocks, thus improving logical decision-making.

Keywords: cross-efficiency, cumulative prospect theory, data envelopment analysis, diversification, stock selection

Received: November 3, 2023; accepted: March 28, 2024; available online: May 27, 2024

DOI: 10.17535/crorr.2024.0001

1. Introduction

After Markowitz [14] introduced the mean-variance optimization (MVO) model in the 1950s, the financial market grew exponentially and consistently. The Mean-Variance Optimization (MVO) model represents a crucial element in modern finance, offering a balance between portfolio returns and risks, thereby quantifying portfolio performance. With rapid advancements in artificial intelligence, the MVO selection process has now been integrated into diverse methodological frameworks within contemporary investment science. In the global market, asset assessment and ranking are vital for investors. In this study, we employ data envelopment analysis (DEA) with cross-efficiency analysis in the presence of negative input-output to improve portfolio selection by enhancing the stock evaluation and management process.

DEA stands out as one of the distinguished methodologies for examining the performance of decision-making units (DMUs) that manage a variety of inputs and outputs. After the

*Corresponding author.

This is an open access article under the CC BY-NC 4.0 license
http://www.hdoi.hr/crorr-journal ©2024 Croatian Operational Research Society, 1–11

groundbreaking research of Charnes et al. [3], who introduced non-parametric programming to approximate the production frontier, there has been a notable increase in publications concerning DEA principles and their practical implications in recent years. Pioneering studies, in this field, such as by Murthi et al. [15], were instrumental in applying DEA and crafting the DEA portfolio efficiency index (DPEI). Similarly, Chen et al. [4] introduced three distinct DEA models that incorporate various risk indicators for assessing portfolio efficiency in a possibilistic environment. Gupta et al. [7] developed an integrated approach for efficient portfolio evaluation, featuring a dynamic re-balancing strategy to empower investors in constructing optimal portfolios within a credibilistic framework. Kumar et al. [10] presented an innovative approach to multiperiod efficient portfolio selection (MPEPS), which involves a two-stage process. Additionally, Zhou et al. [26] proposed a portfolio re-balancing strategy that capitalizes on DEA frontier improvement within the mean-variance (MV) framework.

In DEA, a Decision Making Unit (DMU) evaluates itself by optimizing input and output weights to attain maximum relative efficiency. However, DEA has been criticized for its flexible weight assignment and reliance on self-assessment. To overcome this challenge, Sexton et al. [20] introduced a cross-efficiency method that integrates peer evaluation, shifting focus from self-assessment. Oukil and Amin [16] employed the concept of maximum cross-efficiency to evaluate the performance of baseball team players, presenting an alternative approach to team management. Wu [24] applied cross-efficiency to stock organization and portfolio optimization, and Lim et al. [11] introduced a novel methodology for portfolio selection on the Korean stock exchange using DEA cross-efficiency. Amin and Hajjami [1] demonstrated the effect of alternative optimal solutions in the construction of a cross-efficiency matrix (CEM) and validated that the performance of the MV portfolio selection method is enhanced. Deng and Fang [6] incorporated a DEA prospect cross-efficiency evaluation into an innovative MVM (mean-variance-maverick) framework for fuzzy portfolio selection. Amin and Oukil [2] and Sanei and Banihashemi[19] effectively applied the cross-efficiency evaluation for portfolio section.

Instead of utilizing DEA methods, researchers have also employed multi-criteria decision-making (MCDM) methods that focus on analyzing and ranking decision alternatives based on a set of criteria. Kovač and Podrug [9] provide practical perspectives on utilizing MCDM and the MPT theory for share selection and portfolio optimization. In addition, Poklepovi and Babi [17] proposed a hybrid MCDM technique to evaluate the shares of the Zagreb stock exchange. DEA offers an advantage over MCDM because MCDM relies on a uniform set of weights, whereas DEA utilizes a linear programming model to determine individual weights for each DMU, along with an enhancement in efficiency.

Cross-efficiency analysis is frequently employed in scenarios where the production frontier exhibits constant returns to scale (CRS), as occurrences of negative efficiencies are uncommon across different situations. Lim et al. [12] and Wu et al. [25] introduced two notable methods for evaluations of the cross-efficiencies in cases involving variable returns to scale (VRS) production frontier. Their primary aims are either to reduce negative efficiencies or transform them into positive values. Recognizing the limitations in prior research, Kao and Liu [8] proposed a slack-based measure model for assessing the cross-efficiency value of DMUs without eliciting negative efficiencies. Their concept provides reasonable efficiencies for only weakly efficient DMUs. Soltanifar and Sharafi [21] presented a non-radial cross-efficiency model integrated with a hybrid MCDM technique. In contrast, Lin [13] proposed an approach for cross-efficiency evaluation based on the Directional Distance Function (DDF), utilizing the Range Directional Measure (RDM) model introduced by Portela et al. [18] and duality theory within the framework of VRS technology. The model introduced by Lin [13] can handle situations with negative input-output variables and successfully addresses the issue of negative cross-efficiency. Shrivastava et al.[22] introduced a portfolio selection approach that leverages cross-efficiency, by employing the model from [13] and Cumulative Prospect Theory (CPT). They calculated the CPT value of the assets from their weekly returns using the Cumulative probability (uniform

probability) weighting function and the utility function in CPT. The CPT-based model then constructs effective portfolios by utilizing the CEM as the basis for return calculations. Chen et al. [5] explored a portfolio selection challenge that incorporated fuzzy DEA cross-efficiency evaluation, considering both undesirable fuzzy inputs and outputs.

Typically, the mean values of the CEM columns are the standard for evaluating the performance of a specific DMU. The simultaneous assessment of both the conventional cross-efficiency score (from the CEM column average) and the performance diversity score (from the CEM row average) offers a holistic method to identify stocks that outperform their peers and possess distinctive capabilities absent in others.

This study addresses the issue of robust stock selection using a cross-efficiency evaluation, that accommodates both positive and negative data. The CA and RA are used for robust portfolio selection and the identification of various types of assets. When implemented in a strategic stock selection process, this approach enables managers to pinpoint stocks/assets that not only deliver strong overall performance but also offer diverse capabilities to set them apart from other alternatives. The notable contributions of this research are as follows:

- Firstly, only a limited number of studies in the current body of literature make use of the DEA cross-efficiency model to handle negative input-output data in portfolio selection, as seen in the works of Chen et al. [5] Talluri et al.[23] and Shrivastava et al. [22], and among others.
- Additionally this study addresses asset cross-efficiency by incorporating the CPT value and return of each stock as outputs, and variance as an input. Our method combines cross-efficiency with CPT, exploring the psychological aspects that affect decision-makers when shaping portfolios. It is important to highlight that only a few studies (e.g., Shrivastava et al. [22] have employed CPT values as an output.
- Finally, in contrast to the methods employed by Talluri et al. [23], this study employs a DDF-based RDM cross-efficiency model to classify the assets. In contrast to, the classification employed by Shrivastava et al.[22], this study classifies stocks into distinct categories, to enable the inclusion of well-performing stocks with both similarity and variation in their performance. This approach offers financial experts a more comprehensive and detailed understanding of the subject.

The rest of the study comprises four sections, including the preliminary section which discusses the DEA cross-efficiency techniques to deal with negative values in input/output. The proposed work methodology is described in section 3. Next, the practical application of the proposed approach is demonstrated to highlight its validity and effectiveness in section 4. The paper concludes with Section 5, which examines the results and proposes potential avenues for future research.

2. Preliminaries

This section provides an overview of the key studies that motivated us to initiate this research. Let a group of n homogeneous DMUs be evaluated. Assume that each of these DMUs consumes m different inputs x_{ij} $(i = 1, 2, \ldots, m)$ to generate s different outputs y_{rj} $(r = 1, 2, \ldots, s)$. Specifically, j^{th} DMU consumes amount x_{ij} of input i and generates amount y_{rj} of output r. Cross-efficiency is a concept employed in ranking to tackle the issue of inconsistency among the efficiencies of a set of DMUs assessed using different weightings in DEA. The traditional cross-efficiency method relies on the CCR model and cannot handle the negative input-output data.

To address this challenge, Lin [13] introduces a VRS cross-efficiency evaluation approach based on DDF, allowing input-output variables to take on negative values in line with the approach proposed by Portela et al. [18], as outlined below:

$$\textbf{Model I} \qquad \begin{aligned} &\text{Min } \sum_{i=1}^{m} v_{ik}(x_{ik}+d_{ik}^{-}) - \sum_{r=1}^{s} \mu_{rk}(y_{rk}-d_{rk}^{+}) + \xi \\ &\text{s.t.} \\ &\sum_{i=1}^{m} v_{ik}x_{ij} - \sum_{r=1}^{s} \mu_{rk}y_{rj} + \xi \geq 0, \\ &\sum_{i=1}^{m} v_{ik}d_{ik}^{-} + \sum_{r=1}^{s} \mu_{ik}d_{rk}^{+} = 1, \quad j = 1,2,\ldots,n, \\ &v_{ik},\ \mu_{rk} \geq 0, \quad r = 1,2..s \ \ i = 1,2...m. \end{aligned}$$

where the directional vectors are computed as

$$\begin{aligned} d_{rk}^{+} &= \max_{j=1,2,\ldots,n} y_{rj} - y_{rk}, \qquad r = 1,2,\ldots,s, \\ d_{ik}^{-} &= x_{ik} - \min_{j=1,2,\ldots,n} x_{ij}, \qquad i = 1,2,\ldots,m. \end{aligned}$$

are the possible non-negative improvements for inputs and outputs. Using the weights of DMUk derived from M2, established the cross-efficiency of DMUj, in the following manner:

$$\beta_{kj} = \frac{\sum_{i=1}^{m} v_{ik}^{*}(x_{ij}+d_{ik}^{-}) - \sum_{r=1}^{s} \mu_{rk}^{*}(y_{rj}+d_{rk}^{+}) - \xi^{*}}{\sum_{i=1}^{m} v_{ik}^{*}d_{ik}^{-} + \sum_{r=1}^{s} \mu_{rk}^{*}d_{rk}^{+}}. \tag{1}$$

Thus, the cross-efficiency score of jth DMU is given by $\beta_j = \frac{1}{n}\sum_{k=1}^{n} \beta_{kj}$ which is the simple column average of the cross-efficiency matrix. Similarly, we obtain the row average cross-efficiency as $\eta_j = \frac{1}{n}\sum_{k=1}^{n} \beta_{jk}$. The self-efficiency of DMU_k will be obtained by putting $j = k$ in the equation (1) and denoted as β_{kk}^{*}. The row average cross-efficiency provides insights into the individual characteristics of each DMU. When the value of a DMU's row mean is relatively lower compared to other DMUs, it distinguishes that DMU from the rest, indicating a high level of performance diversity.

Calculation of the "Maverick index" (MI)

The cross-efficiency scores can be utilized even further by defining the extent to which there is variation between the peer appraisal and the self-appraisal efficiency of DMUs. In finance, the Maverick index serves as a valuable tool for assessing the false positives among financial assets. This measure of concordance between the two efficiencies is called the Maverick index [**?**] and it is defined as follows:

$$MI_k = \frac{\beta_{kk}^{*} - \beta_k}{\beta_k} \tag{2}$$

MI quantifies the extent of disparity between a self-evaluation score and a peer-evaluation score. A higher MI value suggests a higher likelihood of the DMU being mistakenly perceived as efficient. Conversely, a lower index value indicates a more favorable outcome.

Calculation of The "Performance Diversity Index" (PDI)
The calculation of the PDI follows a methodology similar to that of MI. To determine the performance of an evaluated DMU (stock), its basic efficiency is compared with the average cross-efficiency of all other DMUs, using the weights specific to the evaluated DMU. Consequently, all DMUs involved in the assessment procedure are subject to the same criteria. The PDI is computed as follows:

$$PDI_k = \frac{\beta^*_{kk} - \eta_k}{\eta_k}. \tag{3}$$

3. Proposed Methodology

DEA cross-efficiency is the most appropriate approach for portfolio selection because it rationalizes equities more thoroughly and analyses stocks based on efficiency rather than absolute output. As financial decision-making often involves risks and uncertainties, decision analysis should consider the behavioral aspects of decisions. From this perspective, this study introduces a resilient approach to stock management, which encompasses the following stages:
Choice of input and output variables: Input: The variance of each asset is computed to assess its associated risk, which will serve as an input in the proposed method.
Output: To account for the irrational behaviors of decision-makers (DMs) in stock selection, we incorporate the CPT value of stocks using the procedures outlined by Shrivastava et al. [22] as an output. The second output is defined as the mean return of the stocks and is computed using the following formula: $R_{jt} = \frac{C_{j(t+1)} - C_{jt}}{C_{jt}}$ where C_{jt} and $C_{j(t+1)}$are the closing price of the jth stock at time t and $t+1$ respectively.
Assessing the cross-efficiency of the stocks: Model (M2) is utilized to calculate the cross-efficiency scores for the chosen stocks. Subsequently, we determine the self-efficiency, the row average efficiency, and the column average efficiency from the obtained Cross-Efficiency Matrix (CEM).
Categorizing the Stocks: Based on the CA and RA scores obtained from M2 and equation, we categorize the stocks into the following groups: Group G1: High Column Average/High Row Average (HC/HR)
Group G2: High Column Average/Low Row Average (HC/LR)
Group G3: Low Column Average/High Row Average (LC/HR)
Group G4: Low Column Average/Low Row Average (LC/LR)
The RA value describes the unique characteristics of each stock. When a stock's row average value is comparatively lower than that of other stocks, it stands out from the rest, indicating a higher degree of performance diversity. When we depict the means of the rows and columns derived from the cross-efficiency analysis in a quadrant format, stocks in G2 exhibit superior efficiency than the other stocks, setting them apart. Stocks originally in G1 and G3 should be relocated to G2 to improve their performance. In contrast, stocks situated in G4 demonstrate subpar efficiency and fail to distinguish themselves from other stocks.

4. Empirical illustration

To demonstrate the effectiveness of the proposed DEA cross-efficiency evaluation approach addressing negative data, we employ it to analyze real data from the Indian stock market. We use data from Shrivastava et al. [22] focusing on the performance of the top 30 assets included in the Nifty 50 index. For each asset, one input variable variance, and two output variables, CPT value and long mean return are considered. For the long mean return, we consider the data from 05-03-2018 to 25-02-2019. To calculate the CPT value and variance, we select the weekly return of stocks for the period from 05-03-2018 to 02-08-2018. We select the outputs

for stock evaluation because of their direct relevance to portfolio assessment, and, importantly, their ability to account for the DM's behavioral characteristics.

To evaluate the performance of the listed stocks, the M1 model is used to evaluate the assessment of the stocks. Table 1 shows the CPT value, mean return, and variance with self-efficiency obtained by M1 of the thirty stocks.

S.No	Stocks	Variance	CPT value	Return	Self-efficiency
S_1	BAJAJ-AUTO	0.0013	-0.0273	0.0001	0.4933
S_2	BRITANNIA	0.0007	-0.0065	0.0043	0.8914
S_3	HEROMOTOCO	0.0008	-0.0248	-0.0049	0.4717
S_4	ICICIBANK	0.0024	-0.0221	0.006	0.4848
S_5	INDUSINDBK	0.0004	-0.0067	0.0013	0.8383
S_6	KOTAK BANK	0.0007	-0.0080	0.0043	0.8914
S_7	L & T	0.0007	-0.0122	0.0016	0.7138
S_8	NESTLEIND	0.0009	-0.0041	0.007	1.0000
S_9	NTPC	0.0005	-0.0159	0.0001	0.7125
S_{10}	ULTRACEMCO	0.0009	-0.0176	0.0011	0.6200
S_{11}	WIPRO	0.0011	-0.0135	0.0041	0.6902
S_{12}	TCS	0.0009	0.0008	0.006	0.9569
S_{13}	ITC	0.0008	0.0042	0.0027	0.8092
S_{14}	TITAN	0.001	-0.0175	0.0059	0.8333
S_{15}	BAJFINANCE	0.002	0.0227	0.0125	1.0000
S_{16}	BAJAJFINSV	0.0011	-0.0038	0.0078	0.9756
S_{17}	SHREECEM	0.0009	-0.0134	0.0037	0.7399
S_{18}	TATASTEEL	0.0017	-0.0291	-0.0011	0.4177
S_{19}	BHARTIARTL	0.0011	-0.0267	-0.0013	0.4959
S_{20}	GRASIM	0.0011	-0.0213	-0.0037	0.4594
S_{21}	TATACONSUM	0.002	-0.0581	-0.0033	0.3361
S_{22}	M&M	0.0006	0.0086	-0.0010	1.0000
S_{23}	ONGC	0.0009	-0.0226	-0.0016	0.5309
S_{24}	MARUTI	0.0004	-0.0115	-0.0028	0.6319
S_{25}	RELIANCE	0.0016	0.0039	0.008	0.7770
S_{26}	TECHM	0.0012	-0.0059	0.0046	0.7109
S_{27}	CIPLA	0.0012	-0.0100	-0.0007	0.5287
S_{28}	COALINDIA	0.0014	-0.0324	-0.0040	0.4031
S_{29}	HINDALCO	0.0034	-0.0344	0.0008	0.2996
S_{30}	HDFCLIFE	0.001	-0.0180	-0.0010	0.5384

Table 1: *Input Output variable with Self-efficiency after solving Model I.*

4.1. Stock management using the proposed method

The model M1 computes the cross-efficiency of the stocks. The cross-efficiency matrix (CEM) is generated using the input and output weights from M1 and Equation (1). The resulting CEM offers unique insights pertinent to this analysis. A widely recognized use of CEM involves averaging the efficiency scores down a column and designating that average as the peer appraisal score, commonly known as the CA efficiency.

Stocks	S_1	S_1	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
CA efficiency	0.4816	0.8650	0.4521	0.4550	0.8027	0.8588	0.6902	0.9871	0.6750	0.6022
RA efficiency	0.6614	0.6332	0.6614	0.6182	0.6332	0.6332	0.6614	0.6182	0.6332	0.6614
Stocks	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
CA efficiency	0.6759	0.9237	0.7644	0.7953	0.9591	0.9386	0.7219	0.4053	0.4816	0.4396
RA efficiency	0.6614	0.6614	0.6614	0.6182	0.6182	0.6182	0.6332	0.6614	0.6614	0.6614
Stocks	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
CA efficiency	0.3290	0.6737	0.5130	0.6007	0.7510	0.6877	0.5037	0.3894	0.2915	0.5179
RA efficiency	0.6182	0.4965	0.6614	0.6614	0.6614	0.6614	0.6614	0.6614	0.6182	0.6614

Table 2: *The column and row average efficiency obtained from CEM.*

A stock with a high CA performs well because it outperforms the other units. Also, the low Row average (RA) value, indicates that the specific stock possesses distinctive capabilities and expertise that other stocks lack. The CA efficiency and RA efficiency obtained from CEM are presented in Table 2. We categorize the stocks based on their CA and RA values. Figure 1 shows the grouping of stocks based on the CA and RA scores for all 30 stocks. The data is categorized as either high or low for both CA and RA. The mean of these scores is employed to split the categories into high and low (in this instance, the average resulted in a value of 0.641). Although the classification thresholds are subjective, they can also be determined by decision makers or managers engaged in stock evaluation..

Discussion on the different groups of stocks The suggested stock classification aids investors in formulating portfolio optimization approaches. Supreme or exceptional performers are prime candidates for long-term investments because of their outstanding performance in various aspects and unique skills that contribute to building a robust and diversified stock selection, capable of meeting the current and future demands of the financial market. However, all-round good performers and niche performers exhibit differences. While the former excel overall compared to their peers, they lack the specific distinctive skills that distinguish them. Detailed descriptions of these groups are as follows.

1. All-round good-performing stocks (HC/HR): Six stocks are assigned to the HC/HR quadrant (G1). Stocks in this group are referred to as all-round good performers because they possess strengths that align with the performance of other assets. A portfolio generated with these stocks may perform well, but it may not be stable. The RA and CA efficiencies along with MI and PDI are presented in Table 3. The average MI and PDI for this portfolio are 0.0364 and 0.1739, respectively.

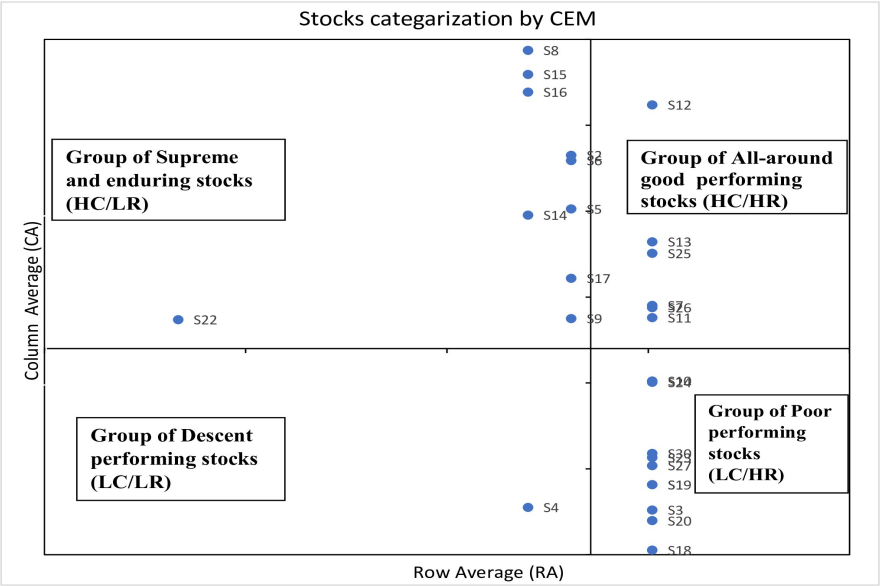


Figure 1: *Different group of stocks based on CA & RA.*

Stocks	CA	RA	MI	PDI
S_{12}	0.9237	0.6614	0.0359	0.4469
S_{13}	0.7644	0.6614	0.0587	0.2236
S_{25}	0.7510	0.6614	0.0346	0.1749
S_7	0.6902	0.6614	0.0341	0.0793
S_{26}	0.6877	0.6614	0.0337	0.0749
S_{11}	0.6759	0.6614	0.0211	0.0436
Average	0.7488	0.6614	0.0364	0.1739

Table 3: *Statistics of group HC/HR.*

2. Supreme and enduring stocks: Table 4, describes the list of stocks in the HC/LR (G2) group. The stocks within this category possess distinctive resources and strengths to set them apart or align them with the performance of other stocks. Consequently, assets from this group can be regarded as long-term stability portfolios. The portfolio obtained from this group is well-diversified with a PDI value of 0.4609.

3. Descent-performing stocks: Only stock 4 comes under the category G3. The stock in this group has unique resources that set them apart from other stocks but lack the strength to match the performance of their peers. But these are less diversified.

4. Poor performing stocks: Within the high LC/HR (G4) quadrant, there are thirteen assets, which are typically considered subpar performers. These assets lack distinctive resources or strengths that can outperform their market counterparts. Consequently, portfolios comprising assets from this group may not exhibit strong performance, and the overall diversification of such portfolios is notably low. The results are shown in Table 6.

After examining the tables, it is evident that G2 (HC/LR) assets have the best performance

Stocks	CA	RA	MI	PDI
S_8	0.98713	0.6182	0.01304	0.61758
S_{15}	0.95908	0.6182	0.04266	0.61758
S_{16}	0.93862	0.6182	0.039398	0.57811
S_2	0.86501	0.6332	0.03050	0.40787
S_6	0.85878	0.6332	0.03798	0.40787
S_5	0.80265	0.6332	0.04441	0.32400
S_{14}	0.79534	0.6182	0.04772	0.34793
S_{17}	0.72189	0.6332	0.02495	0.16859
S_9	0.67497	0.6332	0.05560	0.12532
S_{22}	0.6737	0.4965	0.48431	1.0141
Average	0.8277	0.6135	0.0820	0.4609

Table 4: *Statistics of group HC/LR.*

Stocks	CA	RA	MI	PDI
S_4	0.4550	0.618	0.06541	-0.2158

Table 5: *Statistics of group LC/LR.*

Stocks	CA	RA	MI	PDI
S_{10}	0.6022	0.6614	0.0296	-0.06253
S_{24}	0.6007	0.6614	0.05200	-0.04454
S_{30}	0.51788	0.6614	0.03962	-0.1859
S_{23}	0.5130	0.6614	0.034913	-0.19725
S_{27}	0.5037	0.6614	0.049605	-0.20058
S_{19}	0.4816	0.6614	0.02968	-0.25017
S_1	0.4816	0.6614	0.02435	-0.25411
S_3	0.45208	0.6614	0.04339	-0.28677
S_{20}	0.43963	0.6614	0.04498	-0.30536
S_{18}	0.40534	0.6614	0.03049	-0.368416
S_{28}	0.38944	0.6614	0.035076	-0.39049
S_{21}	0.32902	0.61821	0.02151	-0.45633
S_{29}	0.29147	0.61821	0.027893	-0.51537
Average	0.46212	0.65472	0.035624	-0.27060

Table 6: *Statistics of group LC/HR.*

with an average MI score of 0.0820 and a higher average PDI score of 0.4609. G3 and G4 have significantly lower PDI scores when assessing overall diversification with less MI. It should be

noted that the portfolio can also be chosen from a combination of the above groups of stocks or from the entire group. Similar to each category the stocks with low MI or high PDI can be selected.

To the best of our knowledge and based on our extensive literature review, there is currently no available approach to directly compare the outcomes of the method proposed in this research. Existing studies that utilize DEA cross-evaluations for portfolio selection, stocks are typically ranked in descending sequence of their cross-efficiency values, and the top n stocks are selected, where n corresponds to the number of stocks desired in portfolio. In contrast, our method involves the categorization of stocks into distinct groups based on the average values of the rows and columns in the Cross-Efficiency Matrix (CEM). Therefore, a direct comparison between these two methods is not meaningful.

5. Conclusion

We used the RDM DDF-based cross-efficiency analysis to categorize stocks with negative data. Although various approaches are available for categorizing and grouping stocks, the proposed classification methodology stands out as distinctive and holds significance in the context of portfolio construction. CA values in CEM are commonly employed in the standard practice of evaluating a unit's overall comparative performance. We established an alternative approach that involves the RA values of CEM to determine the breadth of performance diversity. These two ideas are merged to categorize financial assets into different groups, which further enables a more effective stock rationalization process that considers comparative performance and diversity of performance concurrently.

The limitations and potential future directions of this study include the Indian stock market with a restricted dataset. To enhance the significance of efficiency comparisons and benchmark selection, it would be valuable to encompass major markets in other countries. In future research, broadening the modeling scope to include ordinal and interval data values would enable the consideration of the impact of fuzzy input-output data.

References

[1] Amin, G. R. and Hajjami, M. (2021). Improving DEA cross-efficiency optimization in portfolio. Expert systems with applications, 168, 114280. doi: 10.1016/j.eswa.2020.114280

[2] Amin, G. R. and Oukil, A. (2019). Gangless cross-evaluation in DEA: an application to stock selection. RAIRO-Operations Research,53, 645–655. doi: 10.1051/ro/2018060

[3] Charnes, A., Cooper, W. W. and Rhodes, E. (1978). Measuring the efficiency of decision-making units. European journal of operational research, 429–444. doi: 10.1016/0377-221790138-8

[4] Chen, W., Gai, Y. and Gupta, P. (2018). Efficiency evaluation of fuzzy portfolio in different risk measures via DEA. Annals of Operations Research, 269, 103–127. doi: 10.1007/s10479-017-2411-9

[5] Chen, W., Li, S. S., Mehlawat, M. K., Jia, L. and Kumar, A. (2021). Portfolio selection using data envelopment analysis cross-efficiency evaluation with undesirable fuzzy inputs and outputs. International Journal of Fuzzy Systems, 23, 1478–1509. doi: 10.1007/s40815-020-01045-y

[6] Deng, X. and Fang, W. (2019). A novel mean-variance-maverick DEA prospect cross-efficiency approach for fuzzy portfolio selection. Journal of Intelligent & Fuzzy Systems, 37, 8113–8130. doi: 10.3233/JIFS-190568

[7] Gupta, P., Mehlawat, M. K., Kumar, A., Yadav, S. and Aggarwal, A. (2020). A credibilistic fuzzy DEA approach for portfolio efficiency evaluation and rebalancing toward benchmark portfolios using positive and negative returns. International Journal of Fuzzy Systems, 22, 824–843. doi: 10.1007/s40815-020-00801-4

[8] Kao, C. and Liu, S. T. (2020). A slacks-based measure model for calculating cross efficiency in data envelopment analysis. Omega, 95, 102192. doi: 10.1016/j.omega.2020.102192

[9] Kovač, D. and Podrug, D. (2023). Improving portfolio liquidity: MCDM approach to share selection on the Zagreb Stock Exchange. Croatian Operational Research Review 14, 29–39. doi: 10.17535/crorr.2023.0003

[10] Kumar, A., Yadav, S., Gupta, P. and Mehlawat, M. K. (2021). A credibilistic multiobjective multiperiod efficient portfolio selection approach using data envelopment analysis. IEEE Transactions on Engineering Management. doi: 10.1109/TEM.2021.3072543

[11] Lim, S., Oh, K. W. and Zhu, J. (2014). Use of DEA cross-efficiency evaluation in portfolio selection: An application to Korean stock market. European journal of operational research, 236, 361–368. doi: 10.1016/j.ejor.2013.12.002

[12] Lim, S. and Zhu, J. (2015). DEA cross-efficiency evaluation under variable returns to scale. Journal of the Operational Research Society, 66, 476–487. doi: 10.1057/jors.2014.13

[13] Lin, R. (2020). Cross-efficiency evaluation capable of dealing with negative data: A directional distance function based approach. Journal of the Operational Research Society, 71, 505–516. doi: 10.1080/01605682.2019.1567652

[14] Markowitz, H. (1952). The utility of wealth. Journal of Political Economy, 60, 151–158. doi: 10.1086/257177

[15] Murthi, B. P. S., Choi, Y. K. and Desai, P. (1997). Efficiency of mutual funds and portfolio performance measurement: A non-parametric approach. European Journal of Operational Research, 98, 408–418. doi: 10.1016/S0377-2217 00356-6

[16] Oukil, A. and Amin, G. R. (2015). Maximum appreciative cross-efficiency in DEA: A new ranking method. Computers & Industrial Engineering, 81, 14–21. doi: 10.1016/j.cie.2014.12.020

[17] Poklepović, T. and Babić, Z. (2014). Stock selection using a hybrid MCDM approach. Croatian Operational Research Review, 273–290. Retrieved from: hdoi.hr/crorr-journal

[18] Portela, M. S., Thanassoulis, E. and Simpson, G. (2004). Negative data in DEA: A directional distance approach applied to bank branches. Journal of the Operational Research Society, 55, 1111–1121. doi: 10.1057/palgrave.jors.2601768

[19] Sanei, M. and Banihashemi, S. (2014). Selecting the best of portfolio using OWA operator weights in cross-efficiency-evaluation. International Scholarly Research Notices, 2014 doi: 10.1155/2014/978314

[20] Sexton, T. R., Silkman, R. H. and Hogan, A. J. (1986). Data envelopment analysis: Critique and extensions. New directions for program evaluation, 1986, 73–105. doi: 10.1002/ev.1441

[21] Soltanifar, M. and Sharafi, H. (2022). A modified DEA cross-efficiency method with negative data and its application in supplier selection. Journal of Combinatorial Optimization, 43, 265–296.doi: 10.1007/s10878-021-00765-7

[22] Srivastava, S., Aggarwal, A. and Bansal, P. (2022). Efficiency Evaluation of Assets and Optimal Portfolio Generation by Cross Efficiency and Cumulative Prospect Theory. Computational Economics, 1–30.; doi: 10.1007/s10614-022-10334-7

[23] Talluri, S., DeCampos, H. A. and Hult, G. T. M. (2013). Supplier rationalization: A sourcing decision model. Decision Sciences, 44, 57–86. doi: 10.1111/j.1540-5915.2012.00390.x

[24] Wu, D. (2016). Robust decision support system for asset assessment and management. IEEE Systems Journal, 11, 1486–1491. doi: 10.1111/j.1540-5915.2012.00390.x

[25] Wu, J., Chu, J., Sun, J., Zhu, Q. and Liang, L. (2016). Extended secondary goal models for weight selection in DEA cross-efficiency evaluation. computers & industrial engineering, 93, 143–151. doi: 10.1016/j.cie.2015.12.019

[26] Zhou, Z., Xiao, H., Jin, Q. and Liu, W. (2018). DEA frontier improvement and portfolio rebalancing: An application of China mutual funds on considering sustainability information disclosure. European Journal of Operational Research, 269, 111–13. doi: 10.1016/j.ejor.2017.07.010

Check for updates

Original Research Article

Composites and Advanced Materials

Composites and Advanced Materials
Volume 33: 1–17
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/26349833241255957
journals.sagepub.com/home/acm
Sage

Sustainable building materials: A comprehensive study on eco-friendly alternatives for construction

Yonatan Ayele Abera[1,2]

Abstract

An in-depth review of a varied range of eco-friendly options for construction is done in this thorough research on sustainable building materials. With increased worries about environmental deterioration, the building sector is increasingly focusing on long-term solutions. This research looks into a variety of materials, including bamboo, engineered wood products, recycled composites, and optimal concrete mixtures. The study underlines the importance of compressive, tensile, and flexural strengths, which are a basic feature for structural integrity. The study also gives useful insights into their practical applicability in real-world building projects by assessing the flexural strength of various materials. Furthermore, the study examines the environmental effect of these materials, taking into account characteristics such as renewability, recyclability, and energy efficiency. Laboratory tests were conducted to determine the fundamental properties of selected materials as part of the investigations. The study emphasizes the ecological advantages of adopting these sustainable alternatives through life cycle assessments and comparative studies. The research output can be served as a thorough reference for architects, engineers, and policymakers, providing a complete knowledge of sustainable building materials and their critical role in developing a greener, more resilient built environment.

Keywords

sustainable building materials, eco-friendly alternatives, construction, environmental impact, life cycle assessment, green building practices

Introduction

The building sector is important in constructing the contemporary world, but it also poses substantial environmental issues. The vast demand for building materials, as well as the resource-intensive processes involved, lead to significant carbon emissions, resource depletion, and environmental deterioration.[1] In response to these environmental issues, there is an urgent need to adopt sustainable building techniques that reduce the sector's environmental effect. One of the most important ways for accomplishing this aim is the widespread use of environmentally friendly and sustainable construction materials.[2]

This article gives a detailed analysis on sustainable building materials, with an emphasis on eco-friendly construction choices. The major goal is to offer a complete grasp of various sustainable materials, their features, environmental benefits, and prospective building applications.[3] We hope to shed light on the potential solutions available to transform the way we build and construct our built environment by a comprehensive examination of

[1]Department of Construction Technology and Management, Dilla University, Dilla, Ethiopia
[2]Department of Civil Engineering, Delhi Technological University, Delhi, India

Date received: 22 August 2023; accepted: 29 April 2024

Corresponding author:
Yonatan Ayele Abera, Department of Construction Technology and Management, Dilla University, 419, Dilla, Ethiopia.
Email: aberas2010@gmail.com

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

scientific literature, experimental assessments, and life cycle evaluations.[4]

The first section of this study delves into an extensive literature review, capturing the latest advancements and developments in the realm of sustainable building materials.[5] We explore a diverse range of eco-friendly alternatives, including recycled materials, bio-based composites, and low-carbon options.[6] By delving into the production processes, material properties, and applications of each alternative, we seek to uncover the unique environmental advantages they offer and the challenges they might present in real-world applications.[7]

The second stage of the investigation comprises rigorous experimental analysis to supplement the conclusions from the literature review. We evaluate the mechanical, thermal, and durability qualities of selected sustainable materials using extensive laboratory experiments. These assessments give vital insights into the material's performance characteristics, assisting in determining its appropriateness for various building applications and informing design decisions.[8]

The third component of our research is a life cycle assessment (LCA) of sustainable construction materials.[9] The LCA investigates each material's complete life cycle, from raw material extraction to manufacture, transportation, consumption, and final disposal. We may conduct educated comparisons between eco-friendly materials and conventional alternatives by measuring environmental consequences such as carbon footprint, energy usage, and waste generation, supporting evidence-based decision-making in sustainable construction practices.[10]

The four eco-friendly materials created from diverse building wastes are rooted as sustainable construction materials, as shown in the figure below, and these materials are mostly produced from construction and demolition wastes. Tiles of various sorts and sizes can be mined from numerous sources of building byproducts (see Figure 1). Bricks and blocks are also important construction materials that may be sourced as recycled solutions from local garbage. These materials are considered sustainable construction materials because they add value to the use of recycled resources.[11]

We realize the revolutionary potential of sustainable construction materials in producing a more eco-conscious constructed environment as we go forward with this thorough research.[12] Our goal is to give architects, engineers, politicians, and industry stakeholders with a thorough grasp of the various environmentally friendly choices accessible. We hope to contribute to a greener, more robust, and peaceful future for our planet and its people by encouraging the use of sustainable materials in building.[13]

Sustainable building materials

Because of their environmental benefits and contribution to the creation of eco-friendly structures, sustainable building materials have grown in popularity in recent years. Reclaimed wood is one significant type of these materials. Reclaimed wood, which is salvaged from ancient houses, barns, or other structures, not only lessens the demand for virgin lumber but also gives character to new creations. Because of its strength and durability, it is a fantastic choice for flooring, furniture, and ornamental pieces. The building sector supports recycling and minimizes the depletion of natural forests by reusing wood that would otherwise wind up in landfills.[14]

Another important sustainable building resource is recycled metal. Steel and aluminum, for example, may be recycled endlessly without losing their quality. In green buildings, recycled metal materials are often utilized for



Figure 1. Sustainable building materials.

structural framing, roofing, and ornamental features.[15] Metal recycling eliminates the need for mining, conserves energy, and lessens the environmental effect of metal manufacturing. Furthermore, employing recycled metal reduces the carbon impact, making it a crucial choice for sustainable building methods (see Figure 2(a) and (b)).

Bamboo, a fast renewable material, has emerged as a critical component of sustainable building. Its rapid development rate, which allows it to attain maturity in a few of years, making it a very sustainable material for a variety of uses. Bamboo is an extremely flexible material that may be used for flooring, roofing, wall cladding, and even structural elements.[16] Because of its tensile strength and flexibility, it is an excellent substitute for conventional hardwoods. Furthermore, bamboo production aids in the fight against deforestation, lowers greenhouse gas emissions, and provides employment possibilities for local populations.[17]

Recycled glass is a long-lasting material that is often used in green construction projects. Crushed glass is an aggregate that may be used in concrete, glass worktops, tiles, and insulation. Incorporating recycled glass into building components not only saves natural resources but also decreases the energy necessary to manufacture new glass. The building sector helps to trash reduction initiatives and fosters a circular economy by diverting glass debris from landfills.[18]

Rammed earth construction is a centuries-old building style that has seen a resurgence in sustainable architecture. To build walls, layers of soil, chalk, lime, or gravel are compacted inside a frame. Rammed earth constructions have high thermal mass, which automatically regulates inside temperatures. This approach makes use of locally available resources, which reduces transportation emissions. Furthermore, with proper care, rammed earth

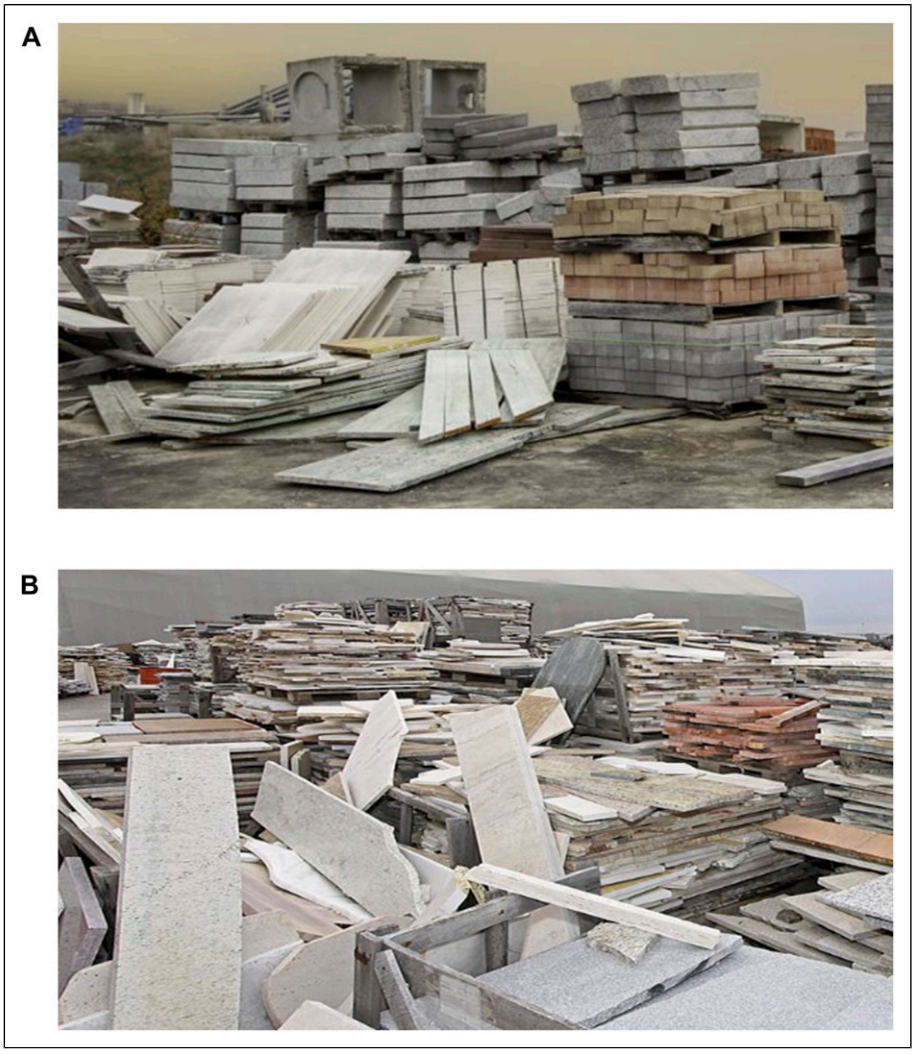


Figure 2. Sustainable building materials.

buildings have a low-carbon footprint and may survive for millennia, making them an environmentally benign alternative for sustainable construction.[4]

Hempcrete, a combination of hemp fibers, lime, and water, has gained popularity in recent years as a sustainable alternative to standard concrete. Hemp growing absorbs CO_2, making it a carbon-negative crop. Hempcrete is a lightweight, insulating, and non-toxic material that creates a healthy interior atmosphere. It also has great moisture-regulating qualities, which reduces the likelihood of mold and mildew growth.[19] The construction sector promotes sustainable agriculture and decreases dependency on energy-intensive building materials by using hempcrete into construction projects.

To summarize, the use of sustainable construction materials not only addresses environmental problems, but also helps to the production of healthier, more energy-efficient, and longer-lasting structures. The building sector encourages sustainable methods, conserves natural resources, and reduces the environmental effect of construction operations by adopting recovered wood, recycled metal, bamboo, recycled glass, rammed earth, and hempcrete.[20] As the demand for sustainable buildings grows, the development and use of these materials will be critical in crafting a greener and more sustainable future for the construction industry.

Literature review

The first section of this research focuses on completing a thorough literature assessment on sustainable construction materials. We go through recent research articles, studies, and industry publications to find a wide choice of environmentally friendly products.[21] The evaluation of literature covers recycled materials such as recycled aggregates and recovered wood, bio-based materials such as bamboo and hempcrete, and low-carbon solutions such as geopolymers and fly ash-based goods.[22] We investigate each material's manufacturing methods, features, and uses in order to comprehend its environmental benefits and limits.[23]

The pursuit of sustainable building materials has emerged as a critical component of the construction industry's attempts to solve environmental concerns and adopt more environmentally friendly methods.[5] A thorough investigation of eco-friendly alternatives for construction materials has become critical in the industry's desire to lower its ecological imprint and migrate to a greener and more sustainable built environment.[24]

Several studies have emphasized the negative environmental implications of traditional building materials, such as high carbon emissions, excessive resource usage, and trash creation.[25] As a result, academics and industry practitioners have shifted their focus to sustainable building materials, which provide potential ways to alleviate these environmental burdens.[26]

Recycled materials have received a lot of interest in recent study. Recycling aggregates from building and demolition debris has showed promise as a viable alternative to conventional aggregates, decreasing the need for massive mining and landfilling (see Figure 3). Additionally, recovered wood has developed as a sustainable solution, offering an environmentally benign alternative to virgin lumber and helping to forest conservation.[27]

Bio-based products have gained popularity due to their recyclability and minimal environmental effect. Bamboo has been a popular alternative for structural applications because of its quick growth rate and remarkable strength-to-weight ratio.[28] Moreover, hempcrete, a hemp fiber and lime combination, has shown remarkable thermal and acoustic qualities, making it an appealing choice for sustainable building insulation.[29]

Low-carbon alternatives are being intensively researched in order to reduce the carbon footprint of construction materials. Geopolymers, which are generated by the interaction of industrial byproducts such as fly ash and slag with alkaline activators, provide a cementitious alternative with much lower greenhouse gas emissions than typical Portland cement. Moreover, fly ash-based goods, which use discarded fly ash as a key constituent, have the potential to reduce environmental impacts while also contributing to the effective use of industrial byproducts.[2]

The incorporation of environmentally friendly building materials into the construction sector necessitates a thorough grasp of their qualities and prospective uses. Yet, there are still issues to address in terms of material performance, cost-effectiveness, regulatory compliance, and public acceptance. Addressing these impediments and providing evidence-based data to promote the informed selection and integration of eco-friendly options in construction methods is critical.[30]

There is a rising interest in sustainable building materials in the literature, with considerable research concentrating on material characterization, performance evaluation, and environmental analyses.[31] Many studies have been undertaken to estimate the total environmental effect of these materials, allowing for direct comparisons with traditional counterparts.

Finally, the literature illustrates the importance of promoting sustainable building materials in the construction business. This extensive research seeks to add to the current body of knowledge by conducting a detailed evaluation of eco-friendly alternatives and their potential to revolutionize the way we design our built environment.[32] This study intends to contribute to the paradigm shift toward greener and more sustainable construction methods by giving information on the environmental benefits, performance features, and prospective uses of these materials.[33]

Experimental methodology

By combining diverse research methodologies, the methodology used in this study intended to give a thorough



Figure 3. Partial views of sustainable building materials from recycled wastes.

knowledge of sustainable construction materials. To provide a multi-dimensional perspective on the viability and effect of eco-friendly alternatives in construction, the study approach included a combination of Interview for focused group, literature review, material assessment, life cycle analysis (LCA), and economic analysis.

The technique began with an exhaustive literature research to generate a solid understanding of sustainable construction principles. This entailed studying academic publications, industry reports, and case studies to identify important materials, manufacturing techniques, and applications that are environmentally benign. This literature assessment aided in the selection of resources for deeper investigation in following phases of the project.

Material evaluation followed, with an emphasis on selected sustainable construction materials. At this phase, material qualities such as strength, durability, thermal performance, and fire resistance were thoroughly examined. A comparison with traditional materials revealed technical capability and potential benefits of eco-friendly alternatives. This stage was critical in determining how these materials aligned with building needs and where they might be used most effectively.

Life cycle analysis (LCA), a thorough technique to analyze the environmental effect of materials throughout their full life cycle, was at the heart of the methodology. This included the extraction of raw materials, manufacture, transportation, usage, and end-of-life stages. Quantitative data on greenhouse gas emissions, energy consumption, and trash creation allowed for a comprehensive assessment of the environmental footprint of sustainable materials vs traditional alternatives. LCA enabled the identification of possible environmental "hotspots" and directed the evaluation of products' overall environmental advantages. The research also included economic analysis, which assessed the cost implications of using sustainable construction materials. This investigation provides insights into the economic viability of sustainable materials and their potential long-term advantages by assessing economic repercussions.

Finally, the research used a multi-dimensional approach to thoroughly investigate eco-friendly options for construction materials. This study intended to provide a comprehensive knowledge of the technical, environmental, and economic aspects of sustainable building materials by integrating a literature review, material assessment, life cycle analysis, and economic analysis. The holistic character of the technique guarantees that stakeholders may make educated decisions about the incorporation of these materials, furthering sustainable building practices in the construction sector.

Results and discussions

Experimental analysis

The second section of this study is an experimental investigation of the mechanical, thermal, and durability properties of selected sustainable construction materials. We do laboratory studies on recycled concrete, bio-based composites, and low-carbon cementitious materials samples. Compressive strength, flexural strength, and tensile strength are all mechanical tests. Thermal conductivity and heat storage capacity measures are used to assess thermal

characteristics. Weathering, chemical assault, and fire resistance are all tested for in durability testing.

The current work integrates rigorous experimental analysis to describe the mechanical, thermal, and durability features of selected eco-friendly alternatives in the quest of understanding and evaluating the feasibility of sustainable building materials. This investigation seeks to give critical data and insights to support the acceptance and integration of these materials in real-world building applications through laboratory experiments and simulations.

Data analysis and interpretation. The results of the experimental investigation are submitted to extensive statistical analysis and interpretation. This includes evaluating the material's mechanical properties, thermal properties, and durability performance using graphical representations, regression analysis, and other statistical techniques. The data are then critically reviewed to discover the sustainable building materials under consideration's strengths, limitations, and potential for development.

In the context of "Sustainable Building Materials: A Comprehensive Research on Eco-friendly Alternatives for Construction," data analysis and interpretation entail reviewing the findings of numerous tests and experiments done on various eco-friendly building materials. The goal is to use the data to develop relevant inferences and insights about the performance and appropriateness of these materials for building applications. Data analysis and interpretation are important parts of the research process because they guide the transformation of raw data into relevant insights and practical findings. This critical stage entails a methodical approach to organizing, analyzing, and identifying patterns from acquired data. To condense complicated information into understandable narratives, several methods such as statistical techniques, visualization tools, and qualitative analysis are used.

The process starts with data cleaning and preparation, which involves identifying and correcting inconsistencies, mistakes, and anomalies. Following that, statistical analysis techniques are used to discover linkages, trends, and correlations in the data. Inferential statistics allow researchers to derive inferences about larger populations based on sampling data, whereas descriptive statistics provide a picture of the data's fundamental patterns and dispersion.

In contrast, qualitative data analysis entails evaluating non-numerical data such as text, photographs, or interviews. This method uses coding and thematic analysis to uncover repeating patterns, themes, and insights that provide light on the research issue. Data visualization tools, such as graphs, charts, and maps, help both expert and non-technical audiences communicate complicated results. Graphic representations improve comprehension and emphasize crucial themes, assisting in the dissemination of study findings.

The interpretation step entails developing relevant conclusions from the data that has been studied. Researchers contextualize their findings within the framework of current literature, theory, and study aims, providing insights, explanations, and implications. This synthesis bridges the gap between empirical findings and theoretical frameworks, so contributing to field knowledge progress.

Finally, data analysis and interpretation are critical steps in the research process because they structure acquired information and translate it into insights that influence decision-making and motivate additional investigation. The rigor and precision of these processes support the legitimacy and validity of research findings, allowing researchers to make significant contributions to their respective fields.

Mechanical testing. The mechanical characteristics of sustainable building materials play an important role in determining their structural performance and appropriateness for various construction applications. The primary mechanical characteristics tested in this experimental investigation are compressive strength, tensile strength, and flexural strength. Compressive strength tests entail submitting samples to compressive stresses until they break, revealing important information about the material's load-bearing capabilities and structural integrity. Tensile strength tests assess the material's resistance to tension pressures, whereas flexural strength tests assess the material's capacity to handle bending loads. To guarantee accuracy and consistency, these tests are carried out in accordance with industry standards.

Compressive strength of sustainable building material. Compressive strength is an important feature to consider when evaluating sustainable construction materials since it determines a material's capacity to endure axial loads or pressure. Sustainable construction materials have a low environmental effect throughout their entire life cycle, from extraction and manufacture to use and disposal. These materials are intended to be resource- and energy-efficient, as well as ecologically benign. One of the primary benefits of sustainable construction materials is their ability to retain high compressive strength while utilizing environmentally responsible manufacturing procedures. This feature protects the structural integrity of buildings, reducing the need for repairs and replacements, and thereby lowering the total environmental effect. Tables 1–10

The use of high compressive strength sustainable building materials adds greatly to the construction industry's efforts to cut carbon emissions and combat climate change. Construction experts may make high-strength concrete mixes that meet or surpass typical concrete criteria by incorporating resources such as recycled aggregates, fly ash, or slag, which are byproducts of other industrial processes. These materials not only save natural

Table 1. Compressive strength of sustainable and conventional building materials.

Sample no.	Sustainable building material	Compressive strength (MPa)
1	Recycled concrete	35.2
2	Bamboo composite	48.6
3	Hempcrete	22.8
4	Geopolymer	42.1
5	Fly ash-based concrete	39.4
6	Conventional concrete	42.9

Table 2. Tensile strength of sustainable and conventional building materials.

Sample ID	Material type	Tensile strength (MPa)
1	Recycled concrete	15.2
2	Bamboo	45.6
3	Hempcrete	8.9
4	Geopolymer	25.1
5	Conventional concrete	20.3
6	Steel	550.2

Table 3. Flexural strength of sustainable building materials.

Sample ID	Material	Flexural strength (MPa)
1	Bamboo	100-150
2	Engineered wood	40-100
3	Recycled concrete	3-10
4	Hempcrete	0.2-1.0
5	Straw bale	1-2
6	Rammed earth	0.1-0.5

resources but also reduce the need for energy-intensive manufacturing procedures. As a consequence, building projects' overall carbon impact is lowered, harmonizing with global sustainability goals.

Furthermore, sustainable building materials with high compressive strength provide architects and engineers with additional design and construction freedom. This adaptability enables the creation of new and efficient structural systems, encouraging the use of smaller parts and optimal designs. These materials result in lighter and more resource-efficient structures by lowering the volume of material required without sacrificing strength. Furthermore, sustainable materials frequently have higher durability and tolerance to harsh climatic conditions, enhancing the lifespan of structures while reducing maintenance requirements. This long-lasting performance not only benefits the environment by minimizing trash, but it also benefits building owners financially by lowering life cycle expenses.

In conclusion, the compressive strength of sustainable building materials is critical to the advancement of ecologically aware construction techniques. By using high compressive strength materials into building projects, the construction industry may produce long-lasting, robust, and energy-efficient buildings, greatly contributing to the worldwide trend toward sustainable development and a greener future.

Tensile strength of sustainable building materials. Tensile strength, which determines a material's capacity to bear pulling or stretching forces without breaking or deforming, is an important attribute in sustainable construction materials. Bamboo, engineered wood products, and certain natural textiles have exceptional tensile strength in the domain of sustainable building. Bamboo, for example, is a quickly renewable material with high tensile strength, making it an excellent choice for a variety of structural applications. Because of their layered and bonded structure, engineered wood products such as laminated veneer lumber (LVL) and cross-laminated timber (CLT) have great tensile strength, offering architects and builders with sustainable alternatives to traditional lumber and steel.

Furthermore, natural fibers such as hemp and jute have high tensile strength, making them desirable components in sustainable composite materials. When these natural fibers are mixed with biopolymers or resins, they form strong and environmentally friendly composites that may be used in building components such as panels, tiles, and even structural elements. The use of these sustainable materials not only decreases reliance on traditional, energy-intensive resources, but it also dramatically reduces the carbon footprint of construction projects.

Furthermore, investigating the tensile strength of recycled materials is a critical component of sustainable construction strategies. For example, recycled steel keeps its tensile strength even after reprocessing, making it a great choice for reinforcing constructions. Similarly, recycled plastic composites have high tensile strength, making them a potential alternative to traditional building materials. By utilizing recycled materials into construction projects, the construction industry may actively contribute to waste reduction while retaining high structural integrity, so supporting the built environment's sustainability.

Table 4. Thermal conductivity of building materials.

Material	Sample size (mm)	Test method	Thermal conductivity (W/mK)
Recycled concrete	150 × 50 × 10	ASTM C518	0.075
Bamboo	100 × 40 × 8	ISO 8302	0.060
Hempcrete	120 × 45 × 9	EN 12664	0.095
Geopolymer	130 × 35 × 12	ASTM C177	0.080
Conventional concrete	180 × 60 × 15	ISO 8301	0.065

Table 5. Specific heat capacity of building materials.

Material	Sample size (mm)	Test method	Specific heat capacity (J/kg-K)
Recycled concrete	150 × 50 × 10	ASTM E1952	1000
Bamboo	100 × 40 × 8	ISO 11357	900
Hempcrete	120 × 45 × 9	EN 12524	1100
Geopolymer	130 × 35 × 12	ASTM E1269	950
Conventional concrete	180 × 60 × 15	ISO 11357	1050

Table 6. Thermal diffusivity of building materials.

Material	Sample size (mm)	Test method	Thermal diffusivity (mm^2/s)
Recycled concrete	150 × 50 × 10	Laser flash analysis	1.20 × 10^-6
Bamboo	100 × 40 × 8	Transient plane source	1.05 × 10^-6
Hempcrete	120 × 45 × 9	Laser flash analysis	0.98 × 10^-6
Geopolymer	130 × 35 × 12	Transient plane source	1.15 × 10^-6
Conventional concrete	180 × 60 × 15	Laser flash analysis	1.30 × 10^-6

Table 7. Accelerated aging test for weathering resistance.

Material	Sample size (mm)	Test method	Exposure time (weeks)	Color change (ΔE)	Surface degradation
Recycled concrete	150 × 50 × 10	ASTM G155	12	3.2	Minimal
Bamboo	100 × 40 × 8	ISO 4892-3	8	2.5	Slight
Hempcrete	120 × 45 × 9	EN ISO 4892-2	16	4.0	Moderate
Geopolymer	130 × 35 × 12	ASTM G154	10	3.8	Moderate
Conventional concrete	180 × 60 × 15	ISO 4892-2	14	2.9	Slight

Table 8. Accelerated aging test for chemical resistance.

Material	Sample size (mm)	Test method	Exposure time (days)	Weight change (%)	Surface integrity
Recycled concrete	150 × 50 × 10	ASTM D1308	30	0.5	Intact
Bamboo	100 × 40 × 8	ISO 175	20	0.3	Intact
Hempcrete	120 × 45 × 9	EN ISO 2812-2	40	1.2	Slight damage
Geopolymer	130 × 35 × 12	ASTM D543	25	0.8	Minimal damage
Conventional concrete	180 × 60 × 15	ISO 175	35	0.4	Intact

Table 9. Comparative analysis of sustainable building materials.

Material type	Environmental impact	Mechanical properties	Cost-effectiveness
Recycled aggregates	Low	Varies	Cost-saving
Bamboo	Low	High tensile strength	Varies
Rammed earth	Low	Good compressive strength	Cost-competitive
Straw bales	Low	Good insulation properties	Cost-competitive
Engineered wood	Low	Varies	Varies

Table 10. Comparative analysis of eco-friendly building materials.

Material	Environmental impact (kg CO2 eq/m3)	Compressive strength (MPa)	Tensile strength (MPa)	Production cost ($/m3)
Recycled aggregates	150	30	3	80
Bamboo	50	40	6	120
Rammed earth	80	25	2.5	90
Straw bales	100	15	1.5	70
Engineered wood	120	35	4	110

In conclusion, knowing and harnessing the tensile strength of sustainable building materials is critical for the construction industry's transition to environmentally friendly methods. Builders and architects can design robust and ecologically responsible structures by embracing elements such as bamboo, engineered wood products, natural fibers, and recycled materials, paving the way for a greener future in construction.

Flexural strength of sustainable building materials. Flexural strength is an important feature of sustainable construction materials, since it determines structural integrity and endurance. Unlike typical building materials, sustainable alternatives are designed to endure bending stresses while being environmentally benign. One of the key advantages of these materials is their capacity to properly distribute loads, reducing the danger of fractures and failures under stress. Wood, for example, is a renewable resource with high flexural strength, making it a popular choice for green building. Furthermore, engineered wood products such as cross-laminated timber (CLT) improve flexural strength while effectively utilizing wood, encouraging sustainable forestry methods.

Furthermore, novel sustainable materials like bamboo have a high flexural strength, making them perfect for construction applications. Bamboo's quick growth rate and renewability make it an environmentally aware choice, providing a sturdy and versatile alternative for a variety of construction components. Concrete, a frequently used building material, may also be enhanced for flexural strength in environmentally friendly methods. The environmental effect of concrete manufacturing can be greatly decreased while improving flexural performance by using recycled aggregates or supplemental cementitious materials such as fly ash and slag.

Advances in composite materials have resulted in the creation of high-performance alternatives with excellent flexural strength in the field of sustainable building. Natural fibers like jute or hemp are frequently combined with bio-based resins to create materials that are both robust and ecologically beneficial. Furthermore, recycled plastic composites made from post-consumer waste have high flexural strength, making them a viable alternative for a variety of building applications while addressing the issue of plastic pollution. The construction industry may further improve the flexural strength of sustainable materials by constantly studying and adopting novel approaches, opening the way for eco-conscious and durable structures that match the expectations of the future.

Thermal characterization. The thermal characteristics of sustainable building materials are critical in evaluating their energy efficiency and thermal insulation potential. Thermal conductivity tests reveal the material's ability to conduct heat, whereas specific heat capacity measures determine the material's ability to store thermal energy. These tests give vital information on the thermal performance of the material, assisting architects and engineers in creating energy-efficient and climate-responsive structures.

Thermal characterization is essential for comprehending the thermal characteristics and behavior of sustainable building materials. This evaluation is critical for improving

building design, assuring energy efficiency, and providing a comfortable indoor atmosphere. Thermal conductivity, specific heat capacity, and thermal diffusivity of sustainable building materials are frequently measured during thermal characterization.

Thermal conductivity. Thermal conductivity is an important attribute that governs how well a substance transmits heat. It's critical to understand how heat moves through building components like walls, roofs, and floors. Low thermal conductivity sustainable building materials may efficiently insulate buildings, lowering heat loss during colder months and minimizing heat gain during warmer months. Materials' thermal conductivity is measured in watts per meter-kelvin (W/mK) and assessed using standard test techniques such as ASTM C518 or ISO 8302.

Specific heat capacity. The quantity of heat energy required to increase the temperature of a material by one degree Celsius is measured as specific heat capacity, also known as heat capacity. Sustainable building materials with a greater specific heat capacity may store more heat energy, resulting in increased thermal inertia. Its feature aids in the regulation of temperature changes within buildings, resulting in a more stable and comfortable interior environment. Specific heat capacity is tested using methods such as ASTM E1952 or ISO 11357 and is expressed in joules per kilogram-kelvin (J/kg-K).

Thermal diffusivity. Thermal diffusivity is a metric that describes how quickly heat is transported through a material in comparison to its ability to store heat. It is a mix of thermal conductivity and specific heat capacity that determines how rapidly a material responds to temperature changes. High thermal diffusivity materials transfer and disperse heat quickly, making them ideal for applications requiring rapid thermal reaction. Typically, thermal diffusivity is determined using techniques such as laser flash analysis (ASTM E1952) or the transient plane source method (ISO 22007-2).

Thermal characterization of sustainable building materials provides designers and engineers with vital information about their thermal performance. This information aids in the selection of materials that maximize energy efficiency, improve occupant comfort, and contribute to the construction of sustainable and resilient structures. Additionally, knowing the thermal behavior of these materials is critical for the design of passive heating and cooling techniques that result in lower energy consumption and a lower carbon footprint. Incorporating thermal characterization into material selection and building design procedures is an important step toward a more sustainable and energy-efficient built environment.

Durability assessment. Durability is an important component of sustainable building materials since it directly affects service life and long-term environmental effect. To imitate real-world environmental challenges, the experimental study includes accelerated aging testing, exposure to harsh weather conditions, and chemical resistance tests. We examine the materials' robustness, resistance to degradation, and propensity for weathering and deterioration by submitting them to extreme environments. This data is critical for determining the acceptability of materials for various building applications and assuring the long-term durability of created structures.

Durability testing is an important part in determining the performance and lifetime of sustainable construction materials. It entails putting the materials to a variety of environmental pressures, such as severe environments and aging variables, in order to assess their capacity to endure these difficulties over time. The purpose of durability testing is to guarantee that the eco-friendly alternatives chosen have resilience, resistance to deterioration, and longevity, making them appropriate for use in real-world building circumstances.

The assessment of durability is a critical component in determining the long-term performance and resilience of materials, components, and structures in the context of varied environmental and operational situations. It entails systematically testing a material's capacity to endure wear, degradation, and probable deterioration over time. This evaluation is critical for assuring the lifespan of structures, lowering maintenance costs, and decreasing the environmental effect of frequent replacements.

A variety of elements are considered in the durability evaluation, including exposure to external stressors such as temperature variations, moisture, chemicals, and mechanical pressures. Accelerated aging experiments, field exposure studies, and simulation approaches are frequently used to simulate real-world situations and forecast how materials would perform in various scenarios.

Durability evaluations help in the identification of possible weaknesses, allowing for the adoption of preventative actions and informed material selection. Assessing the corrosion resistance of steel reinforcements in concrete structures, for example, helps avoid structural damage due to rust, hence increasing the construction's lifespan.

Material science, non-destructive testing, and predictive modeling advancements have broadened the scope of durability evaluation. Life cycle evaluations are also used to examine the total environmental effect of a material's performance across its lifetime.

Finally, durability testing ensures that materials and structures not only satisfy functional requirements but also provide long-term value by withstanding the rigors of their operational environment. This technique adds to the

building industry's commitment to sustainability and appropriate resource management by generating long-lasting structures that contribute to a resilient built environment.

Designers, engineers, and builders may make educated judgments about the selection and integration of sustainable building materials by completing a full durability evaluation. A thorough assessment of the material's performance under diverse situations permits the construction of long-lasting and ecologically friendly structures, therefore contributing to a more sustainable built environment.

Comparative analysis. The experimental research comprises a comparison of sustainable construction materials and conventional alternatives to give significant findings. We can make direct comparisons and evaluate the environmental benefits and performance advantages of eco-friendly alternatives by conducting parallel testing on traditional materials typically used in building.

The "Sustainable Building Materials: A Comprehensive Study on Eco-friendly Alternatives for Construction" comparative study focuses on analyzing the performance and applicability of various eco-friendly building materials as alternatives to conventional construction materials. The study's goal is to determine each material's benefits and disadvantages in terms of environmental impact, mechanical qualities, and cost-effectiveness. Recycled aggregates, bamboo, rammed earth, straw bales, and engineered wood are among the materials evaluated for the investigation (see Figure 4).

Environmental impact. In the construction industry's desire to reduce its environmental imprint, sustainable building materials have emerged as critical components. These materials have the potential to drastically reduce resource use, waste creation, and stress on natural ecosystems. The environmental effect of sustainable building materials is diverse, spanning numerous phases of their lifespan, including extraction and production, shipping, installation, usage, and final disposal.

One of the key benefits of sustainable construction materials is that they have a lower embodied energy and carbon footprint. These materials are frequently acquired locally, reducing transportation energy requirements and helping area economies. Furthermore, many sustainable building materials are sourced from renewable resources,

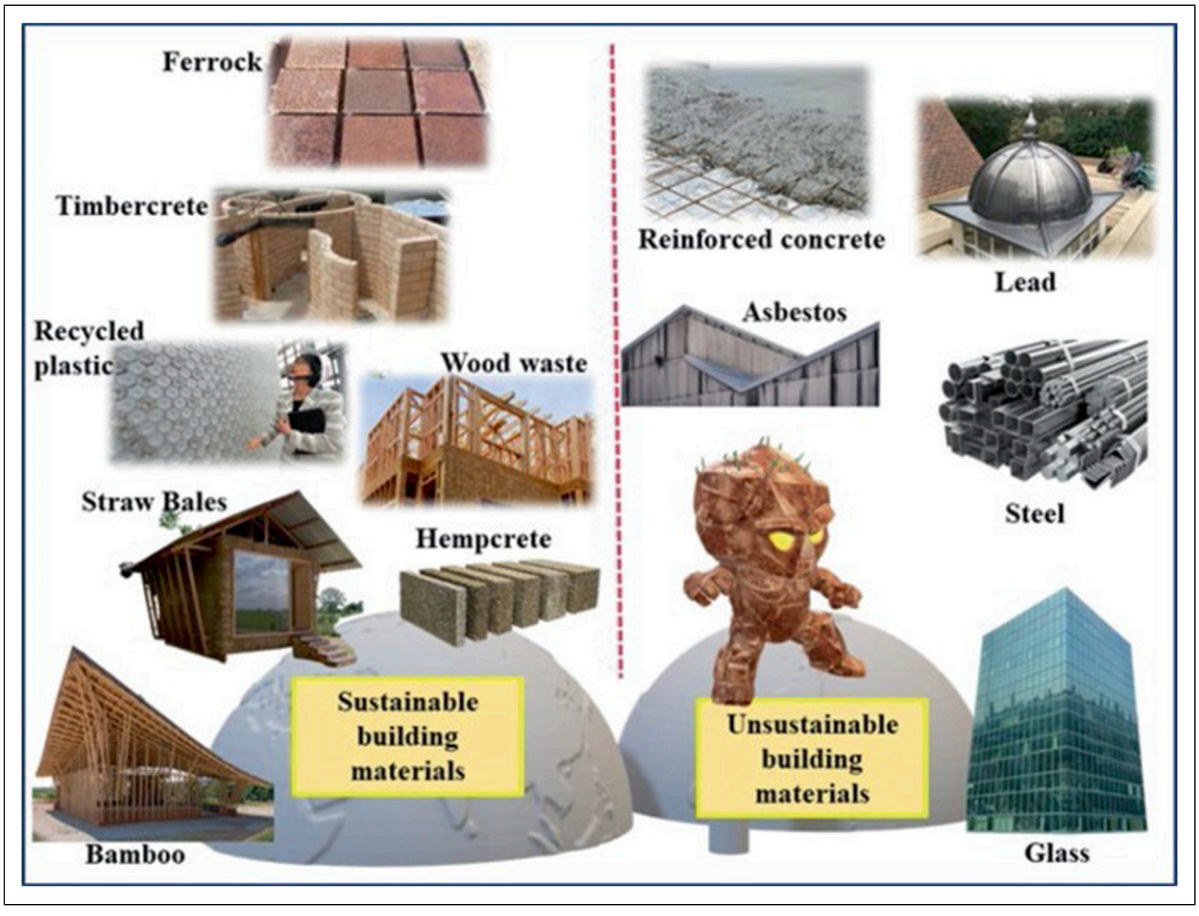


Figure 4. Comparative view of sustainable building materials and unsustainable building materials.

such as bamboo or cork, which replenish more quickly than standard building materials. The building sector may reduce its reliance on finite resources and help conserve natural ecosystems by prioritizing the use of renewable resources.

Furthermore, sustainable construction materials are built to last, decreasing the need for frequent replacements and preserving resources throughout the structure's existence. This longevity, along with recyclability, increases the usable life of the material and reduces the volume of building debris sent to landfills. For example, using recycled concrete aggregates not only diverts trash from landfills but also minimizes the environmental effect of mining virgin materials.

In terms of indoor environmental quality, sustainable materials frequently emit low levels of volatile organic compound (VOC), resulting in better indoor settings. This is especially important since poor indoor air quality can harm occupiers' health and well-being. Low-VOC paints, formaldehyde-free insulation, and non-toxic finishes assist to preserve indoor air quality and the well-being of building inhabitants.

However, it is critical to recognize that the environmental effect of sustainable construction materials varies by option. Life cycle evaluations are crucial tools for assessing a material's overall effect, taking into consideration elements such as resource extraction, production, transportation, and disposal. Furthermore, the availability and scalability of sustainable materials may differ by region, necessitating careful consideration of local context and resources.

Finally, sustainable building materials are critical in lowering the environmental effect of the construction sector. These materials contribute to a constructed environment that is not only ecologically friendly but also favorable to occupant well-being by prioritizing resource efficiency, carbon reduction, durability, and indoor air quality. The use of sustainable building materials demonstrates a conscientious commitment to a more sustainable future, helping to shape a construction sector that balances its role in urban expansion with a duty to protect the planet's natural resources.

Mechanical qualities. Sustainable building materials provide a harmonic balance of environmental responsibility and structural integrity, with mechanical qualities playing an important part in molding their feasibility for construction applications. These characteristics include a material's capacity to tolerate external pressures, stresses, and deformations while retaining overall structural performance over time. Understanding and utilizing the mechanical characteristics of sustainable building materials is critical for designing robust and long-lasting buildings that meet sustainability objectives.

Many sustainable construction materials have mechanical qualities that rival, if not outperform, their traditional equivalents. Cross-laminated timber (CLT), for example, has excellent strength-to-weight ratios, making it appropriate for load-bearing applications while lowering the requirement for energy-intensive steel and concrete. Furthermore, natural fibers like bamboo and hemp have high tensile strength, making them suitable reinforcing materials in a variety of building aspects.

Durability is an important aspect of sustainable materials, which is frequently ascribed to their capacity to withstand degradation caused by environmental variables such as moisture, temperature changes, and chemical exposure. This resilience contributes to longer service lifespans and lower maintenance requirements, which aligns with resource efficiency and waste reduction concepts.

Innovative sustainable materials, such as self-healing concrete, go beyond mechanical qualities by having the potential to mend cracks and extend their longevity on their own. This method represents the marriage of material science and structural engineering, offering increased longevity and sustainability for concrete structures.

The mechanical qualities of sustainable building materials, on the other hand, might vary greatly depending on factors such as material composition, production procedures, and quality control. Thorough testing and validation are required to verify that these materials fulfill the performance specifications of specific applications. Sustainable material standards and norms must adapt to accommodate their particular properties and promote their incorporation into mainstream construction practices.

Finally, the mechanical features of sustainable building materials support their importance in altering the construction sector. Sustainable materials provide equivalent or greater strength, durability, and resilience to traditional materials, paving the path for structurally sound and ecologically responsible construction solutions. Because of the synergy between mechanical performance and sustainable principles, these materials are positioned as significant contributors to the building industry's progress toward a more robust, efficient, and sustainable future.

Cost-effectiveness. As a more thorough understanding of sustainable construction materials' long-term cost-effectiveness emerges, the idea that they are more expensive has steadily altered. While the initial investment in sustainable materials may appear to be costlier than that of traditional counterparts, a careful examination reveals that the lifespan cost of these materials frequently produces considerable savings.

Sustainable construction materials help to save costs in a variety of ways. For starters, their durability and resistance lower maintenance and replacement costs during the life of the structure. Materials such as recycled steel, which has equivalent strength to virgin steel, can increase construction lifetime and reduce maintenance costs without sacrificing structural integrity.

Furthermore, sustainable materials usually improve energy efficiency, resulting in lower operational costs. Thermally insulated materials, such as insulated concrete forms (ICFs), for example, minimize heating and cooling needs, resulting in cheaper utility expenditures over time. Similarly, energy-efficient windows and smart building systems lead to significant energy savings.

The natural waste and resource consumption reduction inherent in sustainable materials coincides with trash disposal rules, potentially resulting in cheaper disposal costs and reducing the environmental effect associated with waste management. Furthermore, the incorporation of renewable energy systems, which is frequently enabled by sustainable construction materials, can result in energy savings that more than equal the original investment costs. As green construction approaches grow more common, economies of scale drive down the cost of green materials. Manufacturing process advancements, more competition, and rising demand have all led to a reduction of the cost gap between sustainable and conventional choices.

When analyzing the cost-effectiveness of sustainable construction materials, it is critical to include not just the initial expenses but also the overall financial impact. The financial argument for sustainable materials becomes persuasive when lifespan costs, decreased operational expenditures, higher property value, and environmental benefits are included in. The cost-effectiveness of sustainable building materials emerges as a strategic investment in both the physical environment and the future in a world more alert to the need for resource efficiency and environmental responsibility.

The comparison research reveals that each eco-friendly construction material has its own set of advantages and disadvantages. The material used is determined by project-specific considerations such as location, climate, resource availability, and desired level of sustainability. Using various sustainable materials and technologies together can provide synergistic benefits, resulting in more environmentally conscious and resilient construction methods. Further research and development in these materials will help to promote future sustainable construction solutions.

Reliability and reproducibility. The experimental analysis is carried out with a strict emphasis on reliability and repeatability. All tests are repeated numerous times to assure accuracy, and statistical averages are employed to reduce potential mistakes. The methodologies and protocols used adhere to industry norms and guidelines, guaranteeing that the experimental findings can be repeated and validated by other researchers.

Finally, the experimental analysis is an important part of this comprehensive study on sustainable construction materials. This research provides critical data to aid building industry decision-making by assessing the mechanical, thermal, and durability qualities of eco-friendly alternatives. This study intends to contribute to the adoption and implementation of sustainable building materials, promoting a more ecologically conscious and resilient built environment, through rigorous and reliable experimental methodologies.

Life cycle analysis

A quick overview. Life Cycle Assessment (LCA) is a thorough technique for assessing the environmental effect of goods, processes, or systems over the course of their entire lifespan, from raw material extraction and manufacture to use, maintenance, and final disposal. LCA considers indirect and cumulative consequences in addition to immediate ones, providing for a comprehensive knowledge of a product's environmental footprint.

LCA is a systematic way to quantifying environmental issues such as energy use, greenhouse gas emissions, water consumption, air pollution, and resource depletion. The evaluation attempts to identify possible hotspots and areas for improvement across the lifespan stages, assisting in the direction of sustainable design, decision-making, and policy formation.

Goal and scope definition, life cycle inventory analysis, life cycle impact assessment, and interpretation are the four essential steps of the process. Data on material inputs, energy consumption, and emissions are gathered and categorized throughout the life cycle inventory study. The impact assessment assesses the possible environmental impacts of these inputs and outputs and converts them into environmental indicators.

LCA provides useful information for informed decision-making in a variety of industries, including building, manufacturing, and consumer goods (see Figure 5). LCA in construction aids in the identification of materials and designs that reduce environmental effect while enhancing performance. Comparing the life cycle implications of different insulating materials, for example, might help guide decisions that lead to energy efficiency and lower emissions.

LCA has gained significance as a technique for attaining environmentally responsible practices as sustainability becomes increasingly important in numerous businesses. It allows stakeholders to examine the larger implications of their decisions, promoting a more balanced approach that adheres to the concepts of resource efficiency, circular economy, and decreased ecological impact.

To conclude, Life Cycle Assessment is a strong tool that aids informed decision-making by offering a comprehensive perspective of the environmental implications of a product or process. LCA promotes a more sustainable approach to design, manufacturing, and consumption by analyzing the full lifespan, so contributing to a more responsible and resilient global economy.

The comparative analysis emphasizes the advantages and disadvantages of various eco-friendly construction materials. Each material has distinct features that make it

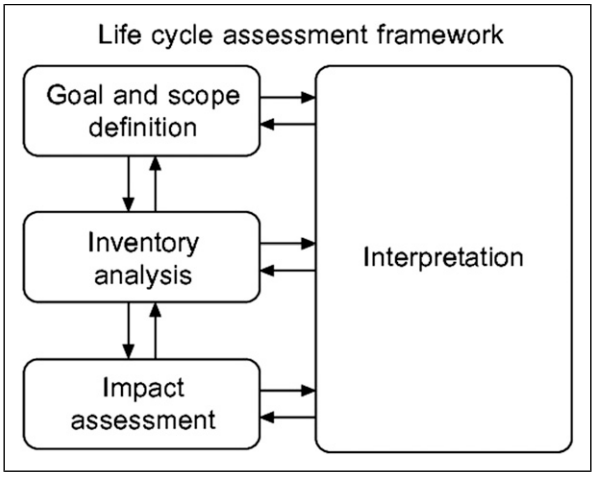


Figure 5. LCA: Sustainable building materials.

appropriate for various construction applications. Recycled aggregates are notable for their environmental advantages, while bamboo and engineered wood excel in mechanical characteristics. Rammed earth and straw bales perform admirably in terms of environmental and thermal performance. To obtain the best environmental, structural, and economic outcomes in construction projects, sustainable building materials should be chosen with specific project needs, local availability, and project goals in mind.

Bamboo has the lowest environmental effect and the highest tensile strength in this example, making it an appealing eco-friendly solution for particular applications. Recycled aggregates, on the other hand, have a reduced cost and a competitive compressive strength, making them acceptable for cost-conscious projects. Rammed earth provides an excellent combination of environmental impact and cost, whereas engineered wood has strong mechanical characteristics and a low environmental impact. Straw bales, while less expensive, have poorer strength and a bigger environmental effect than alternative materials.

The general equation for emission estimation used in this investigation is as follows

$$E = A \times EF \times (1 - ER/100) \quad (1)$$

where: E = emissions, A = activity rate, EF = emission factor, and ER = overall emission reduction efficiency, %.

This simplified table illustrates the many qualities of each eco-friendly construction material, allowing stakeholders to make educated selections based on their individual project needs, financial limits, and sustainability objectives. Additional criteria and data points may be incorporated in the comparison analysis for a more thorough assessment in real-world applications.

Policy implications and suggestions. The comprehensive study on eco-friendly options for construction materials emphasizes the critical need for policymakers, industry stakeholders, and researchers to work together to promote sustainable practices in the construction sector. This study has important policy implications that might accelerate the shift to eco-friendly construction materials and methods.

To begin, politicians should emphasize the creation and implementation of strong environmental standards and regulations for the building sector. These policies should stimulate the use of sustainable building materials by providing incentives, tax breaks, and subsidies to construction firms that utilize environmentally friendly alternatives. Governments may encourage market demand for sustainable materials by fostering a favorable policy climate, resulting in more research and development in this sector.

Furthermore, investment in research and development is critical. Governments and the commercial sector should sponsor research and development activities aimed at discovering new eco-friendly materials and upgrading current ones. Collaboration between academics, research institutes, and industry actors can lead to the development of novel materials that are both environmentally friendly and economically feasible for large-scale construction projects. Furthermore, research should focus on enhancing the durability, strength, and insulating features of sustainable materials in order to make them more appealing to builders and developers.

Furthermore, education and awareness programs should be undertaken to educate architects, engineers, contractors, and customers about the advantages of adopting environmentally friendly construction materials. Training programs and workshops may assist construction workers grasp the technical features and uses of sustainable materials. By raising knowledge, not just legislation but also educated consumer decisions may increase demand for sustainable materials. Furthermore, educational institutions should add modules on sustainable construction techniques into their curricula to ensure that the future generation of architects and engineers is well-versed in environmentally friendly solutions.

Green building certifications, such as LEED (Leadership in Energy and Environmental Design) and BREEAM (Building Research Establishment Environmental Assessment Method), can also encourage the use of environmentally friendly products. For buildings that acquire certain certifications, governments may grant tax breaks or quicker permitting processes. Policymakers may provide a competitive edge for eco-friendly construction by recognizing and rewarding sustainable building techniques, driving more stakeholders to invest in such materials and processes.

Furthermore, governments should consider encouraging cross-national research cooperation and knowledge exchange. International collaboration can help to encourage the exchange of best practices, technical developments, and effective policies relating to sustainable building materials. Countries may expedite their efforts to create more sustainable infrastructure and reduce the total environmental effect of the construction sector by learning from one another's experiences.

Finally, the study on eco-friendly construction materials gives useful insights into the policy implications required to promote sustainable practices in the construction sector. Policymakers can pave the way for a greener and more sustainable future in the construction sector by developing and enforcing stringent regulations, investing in research and innovation, launching educational initiatives, incentivizing green certifications, and fostering international collaborations. When implemented properly, these policy recommendations can stimulate broad use of eco-friendly construction materials, resulting in considerable reductions in environmental degradation and a more sustainable built environment for future generations.

Discussions

The significant research on environmentally friendly construction materials has yielded helpful insights that require serious debate and review. This part delves into important findings, potential ramifications, and opportunities for future research, highlighting the significance of sustainable building processes and the role of eco-friendly materials in shaping the construction industry's future.

One important discovery is the clear environmental benefit of sustainable building materials. The LCA analysis indicated the potential for lower greenhouse gas emissions, energy consumption, and resource depletion associated with the use of environmentally friendly alternatives. This aligns with global efforts to combat climate change and underscores the critical role that the construction industry can play in advancing environmental goals. The use of sustainable materials not only gives immediate environmental benefits, but also sets the standard for a more responsible approach to construction.

The economic research yielded mixed results, emphasizing the significance of a thorough understanding of the economic implications of sustainable materials. While some materials showed cost savings over time, others needed upfront investments, which may prevent early adoption. This stresses the importance of looking forward and accounting for concerns other than immediate building costs. Policymakers and industry stakeholders should consider innovative financing mechanisms and incentive structures to balance these constraints and promote the use of sustainable alternatives.

The debate also emphasizes the potential of technology and innovation to drive the use of sustainable construction materials. The study's focus on material characterization underlines the importance of technical feasibility in material selection. Emerging technology, such as enhanced manufacturing methods and digital tools like Building Information Modeling (BIM), offer opportunities to enhance the design, manufacture, and application of sustainable materials. Continued investment in research and development can lead to materials with higher performance qualities, allowing them to be employed in a wider range of construction projects.

However, the use of sustainable construction materials has concerns that must be carefully addressed. Barriers identified in the study included industry inertia, a lack of customer information, and potential quality disparities. To solve these difficulties, a multifaceted approach involving education, capacity building, and collaboration among many stakeholders is required. Policy and regulation are vital, as well-designed incentives and enabling structures can promote dramatic transformation in the industry.

Finally, the argument emphasizes the transformative power of environmentally friendly construction materials in creating a sustainable future. The study's findings show the interconnection of environmental, economic, and technological factors in driving the use of these materials. As the construction industry works for greater sustainability, it must adopt a comprehensive approach that includes innovation, regulatory support, and collaborative efforts to incorporate sustainable materials and practices, paving the way for a more resilient and ecologically conscious built environment.

Conclusion

In conclusion, this extensive study on eco-friendly options for construction materials provided numerous critical conclusions that highlight the necessity and promise of sustainable building methods. The examination of diverse materials' environmental, technological, and economic characteristics has offered significant insights into their feasibility and influence on the building industry.

- One of the most important conclusions is the significant environmental benefit provided by sustainable construction materials. Life cycle assessment (LCA) studies repeatedly demonstrated that these materials had lower carbon emissions, energy consumption, and resource depletion. This emphasizes their potential to greatly contribute to global sustainability goals and address the building sector's ecological footprint.
- The economic research found a complicated interaction between the initial expenses and the long-term

advantages. While some sustainable materials demonstrated significant cost advantages over their lives, others required substantial inputs. This highlights the significance of a holistic strategy that takes into account both short-term expenditures and long-term returns, emphasizing the importance of novel funding mechanisms and regulatory incentives.

- The study also highlighted the importance of technology and innovation in promoting the use of sustainable materials. Material characterization showed these materials' technical feasibility and performance characteristics, highlighting prospects for further improvement and refinement. The outcomes of the study imply that developing technologies can improve the design, manufacture, and application of sustainable materials.
- However, the study revealed certain barriers to wider adoption. Inertia in the industry, a lack of market knowledge, and possible quality variances surfaced as impediments that need focused actions. Effective legislation, incentives, and stakeholder collaboration are required to overcome these obstacles and speed the incorporation of sustainable materials into mainstream building methods.
- Finally, the study's findings highlight the critical significance of eco-friendly options in forging a more sustainable future for the construction sector. Stakeholders can work together to drive the adoption of sustainable building materials by addressing environmental concerns, economic considerations, and technological innovations, fostering a resilient and environmentally conscious built environment that meets the needs of current and future generations.

Acknowledgements

The author would like to thank the Department of Civil Engineering, DTU, for their valuable initiatives during practical laboratory works that were part of the current research work. And I also would like to thank all who facilitated site visits at Burari C&D waste recycling plants, Shastri Park C&D waste recycling plant, and Mundika C&D waste recycling plant, New Delhi. I also would like to thank Prof. Raju Sarkar and Prof. Amit Kumar Srivastava for their valuable guidance during the preparation of the work.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Yonatan Ayele Abera https://orcid.org/0000-0003-4366-7561

Data availability statement

Supporting data for this research is included in the manuscript.

References

1. Chakraborty D and Elzarka H. Advanced machine learning techniques for building performance simulation: a comparative analysis. *Journal of Building Performance Simulation* 2019; 12(2): 193–207. DOI: 10.1080/19401493.2018.1498538.
2. Xie X and Gou Z. Building performance simulation as an early intervention or late verification in architectural design: same performance outcome but different design solutions. *Journal of Green Building* 2017; 12(1): 45–61. DOI: 10.3992/1552-6100.12.1.45.
3. Endo N, Shimoda E, Goshome K, et al. Simulation of design and operation of hydrogen energy utilization system for a zero emission building. *Int J Hydrogen Energy* 2019; 44(14): 7118–7124. DOI: 10.1016/j.ijhydene.2019.01.232.
4. Charai M, Mghazli MO, Channouf S, et al. Lightweight waste-based gypsum composites for building temperature and moisture control using coal fly ash and plant fibers. *Construct Build Mater* 2023; 393. DOI: 10.1016/j.conbuildmat.132092.
5. Pearlmutter D, Theochari D, Nehls T, et al. Enhancing the circular economy with nature-based solutions in the built urban environment: green building materials, systems and sites. *Blue-Green Systems* 2020; 2(1): 46–72. DOI: 10.2166/BGS.2019.928.
6. Farrukh B, Younis I and Longsheng C. The impact of natural resource management, innovation, and tourism development on environmental sustainability in low-income countries. *Resour Pol* 2023; 86. DOI: 10.1016/j.resourpol.2023.104088.
7. Ruggerio CA. Sustainability and sustainable development: a review of principles and definitions. *Sci Total Environ* 2021; 786. DOI: 10.1016/j.scitotenv.2021.147481.
8. Alqahtani FK, Ghataora G, Khan MI, et al. Lightweight concrete containing recycled plastic aggregates. 2015 international conference on electromechanical control technology and transportation, Montréal, QC, Canada, 12–15 May 2015. DOI: 10.1002/pola.
9. Emad MZ, Mitri HS and Henning JG. Effect of blast vibrations on the stability of cemented rockfill. *Int J Min Reclamat Environ* 2012; 26: 233–243. DOI: 10.1080/17480930.2012.707527.
10. Jayasinghe G, Tokashiki Y and Kitou M. Development of synthetic lightweight soil aggregates utilizing coal fly ash and mine clay as waste materials. *Tropical Agricultural Research Extension* 2005; 8. https://www.agri.ruh.ac.lk/tare/pdf/V_8/AG.8.1.pdf

11. Walker R and Pavía S. Moisture transfer and thermal properties of hemp-lime concretes. *Construct Build Mater* 2014; 64: 270–276. DOI: 10.1016/j.conbuildmat.2014.04.081.
12. Kymäläinen HR and Sjöberg AM. Flax and hemp fibres as raw materials for thermal insulations. *Build Environ* 2008; 43(7): 1261–1269. DOI: 10.1016/j.buildenv.2007.03.006.
13. Feng SJ, Li J, Zheng QT, et al. Utilization potential of waste residue and dust powder from C&D waste. *Case Stud Constr Mater* 2023; 19: e02513. DOI: 10.1016/j.cscm.2023.e02513.
14. Vázquez-López A, Ao X, del Río Saez JS, et al. Triboelectric nanogenerator (TENG) enhanced air filtering and face masks: recent advances. *Nano Energy* 2023: 108635. DOI: 10.1016/j.nanoen.2023.108635.
1.5. Chakraborty D and Elzarka H. Advanced machine learning techniques for building performance simulation: a comparative analysis. *Journal of Building Performance Simulation* 2019; 12(2): 193–207. DOI: 10.1080/19401493.2018.1498538.
16. Shaheen N, Khushnood RA, Memon SA, et al. Feasibility assessment of newly isolated calcifying bacterial strains in self-healing concrete. *Construct Build Mater* 2023; 362. DOI: 10.1016/J.CONBUILDMAT.2022.129662.
17. Ye X and Rasoulinezhad E. Assessment of impacts of green bonds on renewable energy utilization efficiency. *Renew Energy* 2023; 202: 626–633. DOI: 10.1016/j.renene.2022.11.124.
18. Corinaldesi V and Moriconi G. Use of synthetic fibers in self-compacting lightweight aggregate concretes. *J Build Eng* 2015; 4: 247–254. DOI: 10.1016/j.jobe.2015.10.006.
19. Teotónio I, Cabral M, Cruz CO, et al. Decision support system for green roofs investments in residential buildings. *J Clean Prod* 2020; 249. DOI: 10.1016/J.JCLEPRO.2019.119365.
20. de Brito J and Robles R. Recycled aggregate concrete (RAC) methodology for estimating its long-term properties. *Indian J Eng Mater Sci* 2010; 17(6): 449–462. DOI: 10.1063/1.4953383.
21. Zhou X, Li H and Lu C. An environment-friendly thermal insulation material from cotton stalk fibers. *Energy Build* 2010; 42(7): 1070–1074. DOI: 10.1016/j.enbuild.2010.01.020
22. Brzyski P and Widomski M. The influence of partial replacement of hemp shives by expanded perlite on physical properties of hemp-lime composite. *AIP Conf Proc* 2017; 1866(1). DOI: 10.1063/1.4994486.
23. Sai Trivedi S, Snehal K, Das BB, et al. A comprehensive review towards sustainable approaches on the processing and treatment of construction and demolition waste. *Construct Build Mater* 2023; 393. DOI: 10.1016/j.conbuildmat.2023.132125.
24. Bheel N, Ali MOA, Khahro SH, et al. Experimental study on fresh, mechanical properties and embodied carbon of concrete blended with sugarcane bagasse ash, metakaolin, and millet husk ash as ternary cementitious material. *Environ Sci Pollut Control Ser* 2022; 29(4): 5224–5239. DOI: 10.1007/S11356-021-15954-4
25. Rukhaiyar S and Samadhiya NK. Strength behaviour of sandstone subjected to polyaxial state of stress. *Int J Min Sci Technol* 2017; 27(6): 889–897. DOI: 10.1016/j.ijmst.2017.06.022.
26. Vennes I and Mitri H. Geomechanical effects of stress shadow created by large-scale destress blasting. *J Rock Mech Geotech Eng* 2017; 9(6): 1085. DOI: 10.1016/j.jrmge.2017.09.004.
27. Pisello AL, Castaldo VL, Pignatta G, et al. Experimental in-lab and in-field analysis of waterproof membranes for cool roof application and urban heat island mitigation. *Energy Build* 2016; 114: 180–190. DOI: 10.1016/j.enbuild.2015.05.026.
28. Abadie A, Diamond A and Hainmueller AJ. Synthetic control methods for comparative case studies: estimating the effect of California's Tobacco control program. *J Am Stat Assoc* 2010; 105(490). DOI: 10.1198/jasa.2009.ap08746.
29. Ma W, Liu T, Hao JL, et al. Towards a circular economy for construction and demolition waste management in China: critical success factors. *Sustainable Chemistry and Pharmacy* 2023; 35. DOI: 10.1016/j.scp.2023.101226.
30. Maderuelo-Sanz R, García-Cobos FJ, Sánchez-Delgado FJ, et al. Thermal and acoustical evaluation of biocomposites made of agricultural waste for ceiling tiles. *Appl Acoust* 2022; 191. DOI: 10.1016/j.apacoust.2022.108689
31. Farrant WE, Babafemi AJ, Kolawole JT, et al. Influence of sugarcane bagasse ash and silica fume on the mechanical and durability properties of concrete. *Materials* 2022; 15(9). DOI: 10.3390/MA15093018.
32. Imam S, Coley DA and Walker I. The building performance gap: are modellers literate? *Build Serv Eng Res Tecnol* 2017; 38(3): 351–375. DOI: 10.1177/0143624416684641.
33. Minapu LK, Ratnam MV, Rangaraju U, et al. Experimental study on light weight aggregate concrete with pumice stone, silica fume and fly ash as a partial replacement of coarse aggregate. *International Journal of Innovative Research in Science, Engineering and Technology (An ISO Certified Organization)* 2007; 3297(12): 18130–18138. DOI: 10.15680/IJIRSET.2014.0312051.

J Mater Sci: Mater Electron **(2024) 35:867**

Synthesis, analysis and characterizations of Dy^{3+} ions-doped $CaBi_2Nb_2O_9$ phosphors for optoelectronic device applications

Sheetal Kumari[1], Pooja Rohilla[1], Samarthya Diwakar[2], Rupesh A. Talewar[3], Ankur Shandilya[4], Yasha Tayal[5], K. Swapna[6], Aman Prasad[2,*], and A. S. Rao[1]

[1] *Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi 110042, India*
[2] *Department of Physics and Computer Science, Dayalbagh Educational Institute (DEI), Deemed University, Dayalbagh, Agra 282005, India*
[3] *Department of Physics, Shri Ramdeobaba College of Engineering and Management, Katol Road, Nagpur 440013, India*
[4] *Department of Physics, Hindu College, University of Delhi, Delhi 110007, India*
[5] *ABES Engineering College, Ghaziabad 201009, India*
[6] *Department of Physics, Koneru Lakshmaiah Education Foundation, Vaddeswaran, Guntur 522502, India*

Received: 9 November 2023
Accepted: 21 April 2024
Published online: 28 April 2024

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

ABSTRACT

The current study focuses on the synthesis and other detailed studies on $CaBi_2Nb_2O_9$ (abbreviated CBNO) phosphor doped with Dy^{3+} ions. The aforementioned phosphors were prepared via a solid-state reaction procedure and a comprehensive analysis of the crystal structure, particle morphology, vibrational bands, luminous characteristics, decay and thermal stability of the phosphors was done. Phase identification of the synthesized material was carried out using X-ray diffraction (XRD) technique and the optical bandgap was determined using diffuse reflectance spectra (DRS). The morphological studies and particle size estimation of the phosphor material were conducted using a scanning electron microscope (SEM). Under 452 nm excitation, the photoluminescence (PL) spectra of Dy^{3+}-doped CBNO phosphor exhibit three peaks at 482, 576, and 666 nm, which correspond to the transitions $^4F_{9/2} \rightarrow {}^6H_{15/2}$, $^6H_{13/2}$, and $^6H_{11/2}$. The electric dipole–dipole interaction was the main mechanism for concentration quenching for Dy^{3+} ions. The estimated CIE chromaticity coordinates for Dy^{3+} ion-doped CBNO phosphors fall in the white region. The temperature-dependent PL (TD-PL) findings reveals that CBNO phosphors have a considerably superior thermal stability. The results of these investigations suggest that the CBNO phosphors doped with Dy^{3+} ions exhibit promising properties that could be used for prospective w-LEDs and other optoelectronic device applications.

Address correspondence to E-mail: amanprasad2891@gmail.com

https://doi.org/10.1007/s10854-024-12640-2

1 Introduction

The development of a variety of inorganic phosphors appropriate for white light-emitting diodes (w-LEDs) has progressively drawn a great deal of interest to promote the area of solid-state lighting and other display related applications by virtue of their excellent luminescence, low operating voltage, and chemical and thermal durability. [1–7]. It is well known that the w-LEDs are regarded as the fourth generation of SSL and are a better alternative to conventional incandescent, and fluorescent lamps. w-LEDs have received plenty of attention and have been the subject of extensive research because of their unique benefits, which include environmental protection, short response times, high luminous efficiencies, long lifespans, reliability, and lack of mercury. Usually $Y_3Al_5O_{12}$: Ce^{3+} (YAG: Ce) phosphors, which emit yellow light, are combined with blue light from n-UV LED chips to produce commercially available w-LEDs [1, 6, 8–11]. The lack of red colour in this setup results in a low colour rendering index (CRI) and a high colour temperature (CCT). This is inefficient for general room lighting as well. Researchers are concentrating on pc-w-LEDs through tricolour (blue, green, and red) phosphors coated on n-UV chips to get greater CRI and lower CCT in order to solve this drawback [12–14]. For significant red nitride phosphors used in w-LEDs ($SrLiAl_3N_4$: Eu^{2+}, $CaAlSiN_3$: Eu^{2+}, and $Ca_2Si_5N_8$: Eu^{2+}), the synthesis requires a demanding environment, such as high pressure and high temperature. Commercial red phosphor (Y_2O_2S: Eu^{3+}) used in it, however, has several intrinsic drawbacks due to colour re-absorption and distinct degradation such as low luminous efficiency and unstable CCT respectively [15–19]. The mild synthesis environment and generally stable chemical properties of the oxide materials make them useful for a variety of applications. A single-phase white-light-emitting phosphor stimulated effectively with n-UV LED chip is widely preferred to solve the challenges mentioned above, and in the last few decades, this desire has attracted the attention of numerous research groups.

Dysprosium oxide (Dy_2O_3), one of the known rare earth ions, is crucial for emission in white region since it may be easily activated by UV or blue LED chips. Dy^{3+} exhibits blue ($^4F_{9/2} \rightarrow {}^6H_{15/2}$), yellow ($^4F_{9/2} \rightarrow {}^6H_{13/2}$) and red ($^4F_{9/2} \rightarrow {}^6H_{11/2}$) emission bands at wavelengths of 482 nm, 575 nm, and 666 nm, respectively [9, 20–24]. Also, $^4F_{9/2} \rightarrow {}^6H_{13/2}$ has a higher intensity but, it is less sensitive to the environment of the ligand. $^4F_{9/2} \rightarrow {}^6H_{15/2}$ is hypersensitive but, it is greatly impacted by the environment of the host, and $^4F_{9/2} \rightarrow {}^6H_{11/2}$ is least affected by the environment of the host. Many phosphor materials are being studied for their yellow to blue (Y/B) intensity ratio since it must be equal to one to satisfy the need for pure white light emission. Various Dy^{3+} ions-doped phosphors, including $SrAl_2O_4$ [16], $KBaY(BO_3)_2$ [15], and $SrLaAlO_4$ [17], that possess a high absorption in the n-UV range and emit white light, have already been reported.

Nowadays, a wide variety of host materials, including phosphates, silicates, borates, vanadates, fluorides, ceramics, niobates, and tungstates can be doped with RE ions. Due to its great chemical stability, good thermal stability, and eco-friendliness, niobate is regarded as a dependable host material. In general, niobium oxides have a low phonon frequency [25, 26]. The likelihood of the non-radiative transition can be efficiently lowered, that lends the phosphors their strong fluorescence [27]. The Aurivillius-type bismuth layered structure ferroelectrics (BLSFs), whose chemical formula is $(Bi_2O_2)^{2+}$ $(A_{m-1}B_mO_{3m+1})^{2-}$, have drawn a lot of interest because of their comparatively high Curie point (T_c), which translates to high thermal stability [28, 29]. CBNO is a member of the BLSF family with $m = 2$, where Ca^{2+} is present at site A and Nb^{5+} is present at site B of the pseudo-perovskite $(CaNb_2O_7)^{2-}$ blocks. CBNO has a stacking $(Bi_2O_2)^{2+}$ layer and perovskite $(CaNb_2O_7)^{2-}$ units with double NbO_6 octahedral layers [30]. It was proposed that the perovskite blocks could accommodate a wide range of cations, but the bismuth oxide layers were difficult to alter [28]. However, no research has been done on the CBNO: Dy^{3+} phosphor for fabricating white LEDs.

The main aim of this research work is to generate white light with improved luminosity. In this work, the high temperature solid-state method was used to synthesis a series of Dy^{3+} ions-doped CBNO phosphors. The structural and spectroscopic characteristics of phosphors were investigated using a variety of characterisation techniques, including XRD, UV–Vis, PL emission, decay profiles and TD-PL. FT-IR studies were employed to understand various vibrational groups present in the CBNO phosphor. From the PL emission data, the CIE chromaticity coordinates were carefully examined. As far as we are aware, there hasn't been any information published about the luminescence characteristics of Dy^{3+}-doped CBNO phosphors. The aforementioned characteristics suggest that

white light-emitting CBNO phosphor may be utilised in the fabrication of w-LEDs and other optoelectronic device applications.

2 Material preparation and instrumentation

The CBNO: x Dy^{3+} ($x = 0.01, 0.02, 0.03, 0.04,$ & 0.05 mol) ions-doped phosphors were prepared by solid-state reaction process. Starting raw materials $CaCO_3$ (Thermo fisher 98%), Bi_2CO_3 (Loba Chemie 99%), Nb_2O_5 (Loba Chemie 99.9%), Dy_2O_3 (Central Drug House 99.9%), are used. These materials were weighed with their stoichiometric ratio and ground for an hour in an agate mortar. Acetone is employed as a wetting medium. The ground powder was put in an alumina crucible, heated in a high temperature programmable muffle furnace at 1250 °C for 5.5 h, and then allowed to cool at ambient temperature. After that process, the final collected powder was reground again for 10 min to obtain the homogeneous mixture for further characterization purposes.

The chemical reaction of synthesis process can be described by the following chemical formula:

$$CaCO_3 + Nb_2O_5 + Bi_2CO_3 = CaBi_2Nb_2O_9 + \text{gases}$$

The XRD patterns of the as-titled samples were recorded using the Bruker Model D8 Advance instrument with Nickel filtered Cu-Kα radiation (1.54 Å). Using the JEOL 7610 F Plus machine, SEM records for surface morphology were taken. A Perkin-Elmer Frontier FT-IR spectrometer was employed for analysing the FT-IR spectra. A JASCO model V-923 UV–Vis-NIR spectrophotometer (± 2.0 nm resolution) was utilized to record the DRS for the as-titled phosphors. A Jasco FP-8300 spectrofluorometer with a Xenon excitation source having 1.0 nm resolution was used to study the PL excitation and emission spectra. The PL decay profiles were analysed using an Edinburgh FL920 fluorescence spectrometer. The Flame-S-XR1-ES spectrometer from Ocean Optics was utilised to study TD-PL spectra. All measurements except TD-PL were made at the room temperature. The flowchart of the steps involved in the preparation of the phosphor by the solid-state reaction method is shown in Fig. 1.

3 Results and discussion

3.1 XRD analysis

The initial assessment of the purity of the as-prepared sample was conducted using XRD. Figure 2a illustrates the XRD patterns of the undoped CBNO host material. All the diffraction peaks in the sample can be attributed to the standard data for CBNO (ICDD no. 49-0608). No signs of any other phases were detected, indicating that the obtained samples consist of a single phase. $CaBi_2Nb_2O_9$ exhibits an

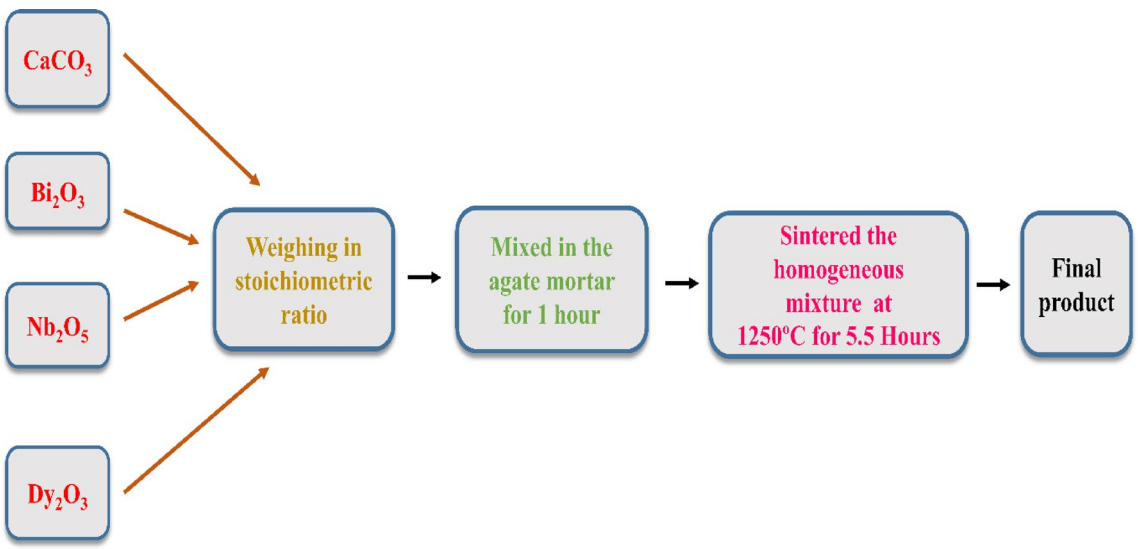


Fig. 1 Flow chart of the steps involved in the synthesis of CBNO phosphor

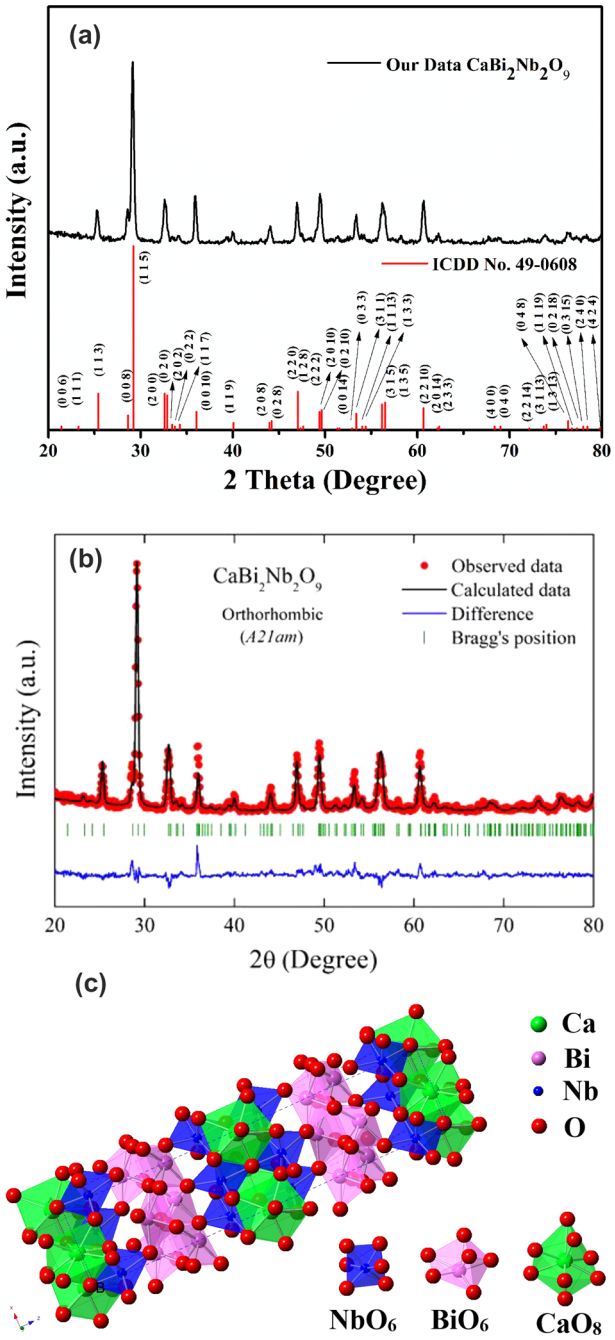


Fig. 2 **a** XRD pattern of the $CaBi_2Nb_2O_9$ and the reference data of ICDD no. 49–0608 **b** Reitveld refinement pattern of an un-doped CBNO phosphor **c** Crystal structure of $CaBi_2Nb_2O_9$, emphasizing the coordination environment of Ca, Bi and Nb atoms

orthorhombic crystal structure with a space group of $A2_1am$, featuring lattice parameters $a = 5.4827$ Å, $b = 5.4415$ Å, $c = 24.8974$ Å, and $\alpha = \beta = \gamma = 90°$.

Table 1 Calculated structural lattice and rietveld refinement parameters of CBNO phosphor

Parameters	Standard data JCPDS	CBNO (un-doped)	Reference
Formula	$CaBi_2Nb_2O_9$	$CaBi_2Nb_2O_9$	[18, 34]
Crystal system	$A2_1am$ (36)—Orthorhombic	$A2_1am$ (36)—Orthorhombic	
Crystal density (gm cm^{-1})	7.499	7.499	
Units, Z	4	4	
a (Å)	5.4827	5.4406(3)	
b (Å)	5.4415	5.4722(3)	
c (Å)	24.8974	24.8527(8)	
$\alpha = \beta = \gamma$	90°	90°	
Cell volume, V ($Å^3$)	742.79	739.9161(3)	
GOF		2.54	

To further assess the phase purity of the synthesized samples, Rietveld refinement was employed, using CBNO as a representative. Rietveld refinement is a method that involves analyzing profile measurements of powder intensity to approximate the phosphor's structural model. The initial model was constructed based on previously reported crystallographic data for CBNO [31]. Figure 2b depicts Rietveld refinement plots for $CaBi_2Nb_2O_9$ powders, which were prepared using the solid-state reaction method and heat-treated at 1250 °C for 5.5 h. The observed XRD patterns closely match the theoretical fit results, indicating the success of the Rietveld refinement method. The refined crystallographic parameters align well with the reflection conditions, and the fit is highly satisfactory with $\chi^2 = 2.54$, confirming the phase purity of the as-prepared sample and $\chi 2$ value is less than the other reported papers [32, 33]. The smaller value of χ^2 indicates that a single phase has been formed. Table 1 provides the structural parameters of the Rietveld refinement, including crystal structure, space group, lattice parameters and goodness of fit (χ^2) [18, 34]. The term χ^2 serves as a measure of the goodness of fit (GOF), indicating how well the experimental XRD patterns align with the calculated ones. Table 2 presents the atomic coordinates and the occupancy of cations and anions in the CBNO compound [31]. Following the refinement process, crystal structures of the CBNO compound were generated using the output.cif file with the Vesta software, as depicted in Fig. 2c.

Table 2 Atomic coordinates, cation and anion occupancies of a CBNO compound

Atom	x	y	z	Site occupancy factor (SOF)	References
Ca	− 0.0763	0.2526	0.0000	0.5	[31]
Bi	0.4543	0.7742	0.2000	1	
Nb	0.4910	0.7434	0.4152	1	
O1	0.5056	0.3009	0.0000	0.5	
O2	0.5854	0.6483	0.3311	1	
O3	0.7535	0.9968	0.2528	1	
O4	0.7345	0.9832	0.9303	1	
O5	0.7857	0.9676	0.5880	1	

Figure 2c illustrates the crystal structure of CBNO, highlighting the arrangement of Ca^{2+}, Bi^{3+}, and Nb^{5+} atoms in their respective coordination environments. CBNO belongs to the Aurivillius-type materials, characterized by a pseudo-perovskite structure. This category was initially identified by Aurivillius in 1949. As explained in the introduction, the crystal structure of CBNO has Ca^{2+} ions occupying the A site and Nb^{5+} ions occupying the B site. The crystal structure of CBNO is composed of chains of CaO_8 octahedra that share edges and run parallel to the a-axis. According to structural data, calcium ions have a coordination number of 8, while bismuth and niobium ions have a coordination number of 6. The ionic radii for 8-coordinated Ca^{2+} and 6-coordinated Bi^{3+} are 1.12 Å and 1.03 Å, respectively. Consequently, due to their similar ionic radii and valence, it is anticipated that Dy^{3+} ion dopants, with a 6-coordinated structure and an ionic radius of 0.912 Å, will substitute for Bi^{3+} ions within the host structure.

3.2 Morphological studies

The microscopic shape and particle size of the as-prepared phosphors have also been further examined using SEM. Figure 3a, b shows the SEM pictures of the

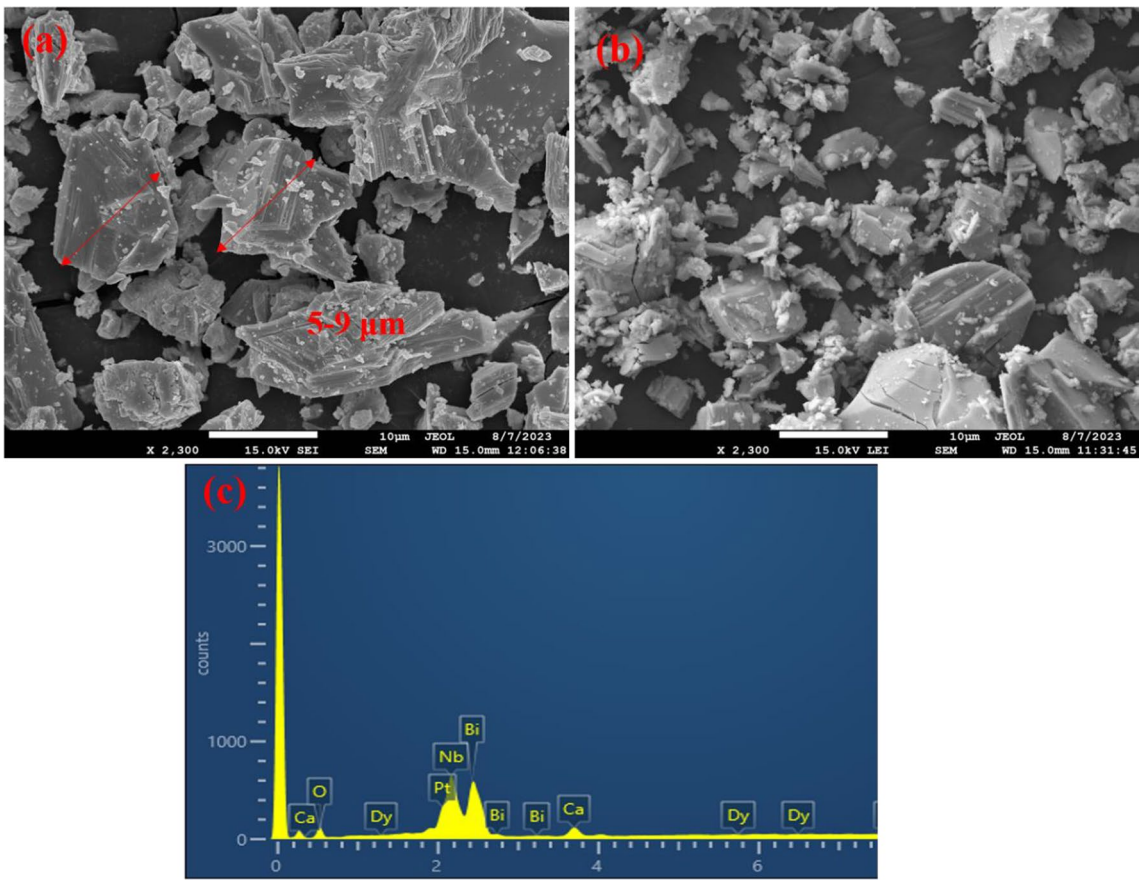


Fig. 3 **a**–**b** SEM images recorded for 0.04 mol of Dy^{3+} ions-doped CBNO phosphor, **c** corresponding EDS image

CBNO: 0.04 mol Dy^{3+}-doped phosphor. The irregular shape of the phosphor is evident in Fig. 3. Due of the high-temperature reaction and grinding procedure, the synthesised materials show agglomeration. Today's commercially available phosphor particles are typically between two 2–10 μm (μm) in size. The as-prepared phosphor samples have an average particle size in the 5–9 μm range as has been highlighted in the image. Also, the phosphor particles with size ranges in μm have been considered more convenient for use in pc w-LEDs. Energy dispersive spectrum (EDS) used to monitor the elemental composition of the as-titled phosphor is depicted in Fig. 3c. The EDS image shows that every precursor element is present in the material. The different colours, i.e. green, purple, yellow, and blue corresponding to Dy, Bi, Nb, and Ca, respectively are depicted in Fig. 4. These results demonstrate the equally distribution of the components in the as-prepared phosphors.

3.3 FT-IR analysis

Figure 5 shows the FT-IR spectrum for the CBNO: 0.04 mol Dy^{3+}-doped phosphor. FT-IR spectrum was captured to reveal information about the functional groups found in the CBNO phosphor. It contains peaks at 545, 610, 797, 1071 and 2974 cm^{-1}. The band centred at 610 cm^{-1} can be assigned to Nb–O of the NbO_6 octahedral stretching vibration [34]. The octahedral NbO_6 peak indicates single-phase formation.

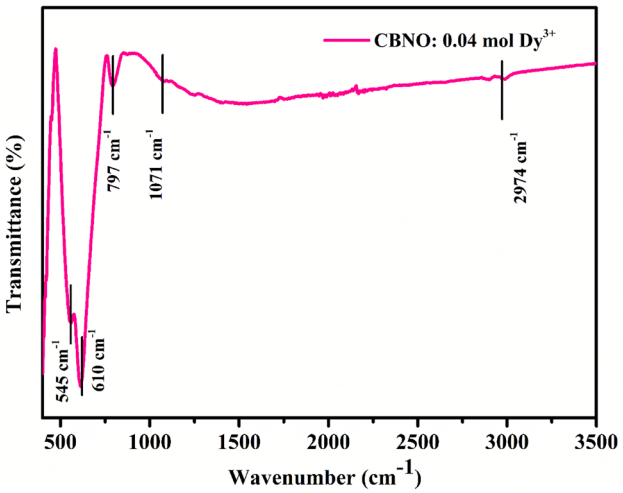


Fig. 5 FT-IR spectrum of CBNO: 0.04 mol Dy^{3+} ion-doped phosphor

The stretching vibration mode of Bi-O_3 is attributed to the shoulder peak at 545 cm^{-1} [35]. The stretching vibration mode of Nb–O_5 and the bending vibration mode of O_4–Nb–O_5 correspond to the shoulder peaks at 797 and 1071 cm^{-1}, respectively [18, 35]. The hydrogen bonding is attributed to the band was seen in the region 2700–3000 cm^{-1}.

3.4 DRS analysis

Phosphors used in w-LEDs depend significantly on the optical bandgap (Eg). Figure 6 displays the DRS

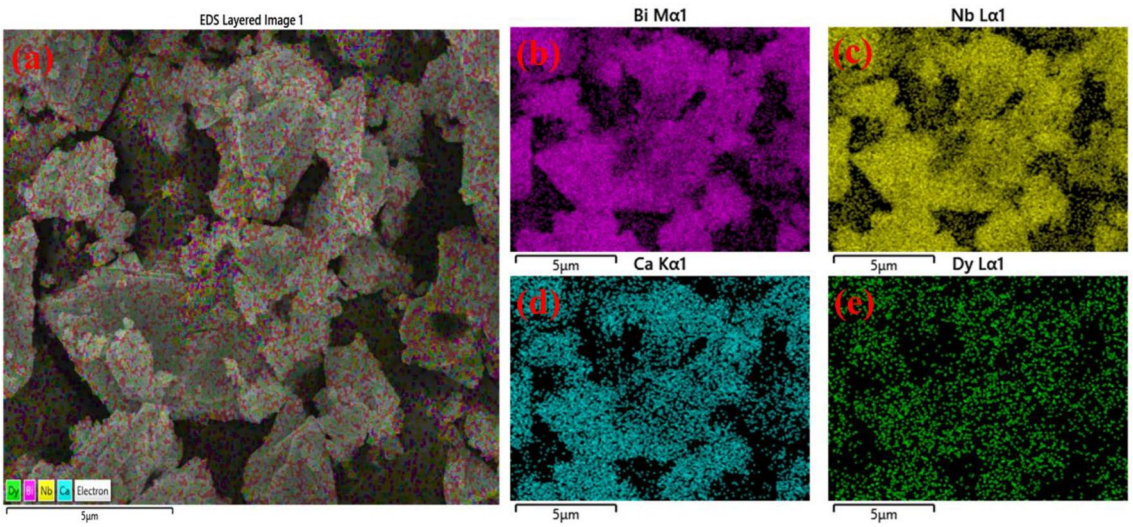


Fig. 4 **a** The total element distribution map, **b–e** shows the relative element points of various colours against a dark background; purple, yellow, blue and green, correspond to Bi, Nb, Ca, and Dy, respectively (Color figure online)

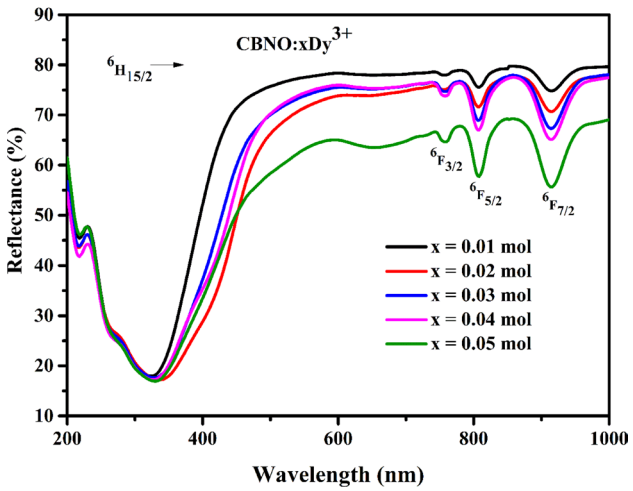


Fig. 6 DRS of CBNO: xDy^{3+} (x = 0.01, 0.02, 0.03, 0.04, & 0.05 mol) phosphors

for CBNO: x Dy^{3+} (x = 0.01, 0.02, 0.03, 0.04, & 0.05 mol) phosphors samples in the range 200–1000 nm. The charge transfer (CT) from the oxygen (2p) to the empty 4f orbital of the Dy^{3+} ion in the host can be seen by two absorption lines observed at wavelengths of 219 nm and 264 nm. The peaks observed in the spectra at 756 nm, 808 nm & 915 nm correspond to the transitions from $^6H_{15/2}$ to $^6F_{3/2}$, $^6F_{5/2}$, & $^6F_{7/2}$ respectively. Carnall et al. have validated each of the aforementioned band designations. [36].

Bandgap determination actually helps us to understand parameters like stability, lifetimes and degradation of the material. This knowledge is essential for deploying materials in various photonic, SSL and solar cell applications. Using the equation below, the E_g was estimated by transforming the DRS data into the Kubelka–Munk function [37].

$$F(R) = \frac{(1-R^2)}{2R} = \frac{\alpha}{S} \quad (1)$$

Here α stands for the absorption coefficient, R for the samples' reflectance, and S for the scattering coefficient. The Tauc equation, which determines the energy bandgap (E_g) and absorption coefficient (α), is shown below [8].

$$\alpha h\upsilon = C\left(h\upsilon - E_g\right)^n$$

Here C stands for the energy-independent constant (= 1), and hν represents the photon energy. n can have value of 2 for indirect or 1/2 for direct transition.

Since F(R) and α are directly related, the equation above can be rewritten as follows [8]:

$$F(R)h\upsilon = C\left(h\upsilon - E_g\right)^n \quad (2)$$

The optical bandgap for the Dy^{3+} ions-doped CBNO phosphors using a Tauc plot which was drawn between the $[F(R)\ h\upsilon]^2$ and hν. The E_g can be estimated by extrapolating the linear region of the plot. The direct bandgap values evaluated for CBNO: x Dy^{3+} (x = 0.01, 0.02, 0.03, 0.04, & 0.05 mol) phosphors are 4.59, 4.57, 4.56, 4.54, and 4.51 eV, respectively and shown in Fig. 7. Thus, it can be concluded that CBNO phosphor is viable host lattice because of its high bandgap and may find utility in w-LEDs applications. The E_g value slightly decreases with an increase in Dy^{3+} ion concentration, which may be due to the 4f energy level of Dy^{3+} ions [38].

3.5 PL studies

It is vital to know the proper excitation wavelength in order to analyse the PL spectra of Dy^{3+}-doped CBNO phosphor. Excitation spectrum of the as-prepared CBNO: 0.04 mol Dy^{3+}-doped phosphor measured at an emission wavelength of 576 nm in the 300–500 nm range is shown in Fig. 8. In that, PL excitation peaks were observed from the ground state $^6H_{15/2}$ to various excited states $^4M_{17/2}$, $^6P_{7/2}$, $^6P_{5/2}$, $^4F_{7/2}$, $^4G_{11/2}$, $^4I_{15/2}$ and $^4F_{9/2}$ at 311, 354, 367, 389, 428, 452 and 478 nm respectively [32, 39–41]. The strongest excitation peak was seen at a wavelength of 452 nm. Therefore, emission spectra for all of the Dy^{3+}-doped CBNO phosphors were analysed at this wavelength.

PL emission spectra for CBNO: x Dy^{3+} (x = 0.01, 0.02, 0.03, 0.04, & 0.05 mol) phosphors were recorded at 452 nm excitation in the range 470–700 nm as shown in Fig. 9. In the PL spectra, three emission peaks were seen at wavelengths of 482 nm, 576 nm, and 666 nm, corresponding to the $^4F_{9/2} \rightarrow {}^6H_{15/2}$, $^6H_{13/2}$, and $^6H_{11/2}$ transitions, respectively [9, 42]. The position of the peaks do not significantly alter with increasing Dy^{3+} ion concentration because the 5 s and 5p outer electrons shield the 4f electrons. Among three distinctive peaks, it can be seen that the emission at 576 nm dominates and is the source of yellow colour emission. As Dy^{3+} concentration rises up to 0.04 mol, the emission intensity also rises. Beyond 0.04 mol of Dy^{3+} ions, concentration quenching due to non-radiative energy transfer between

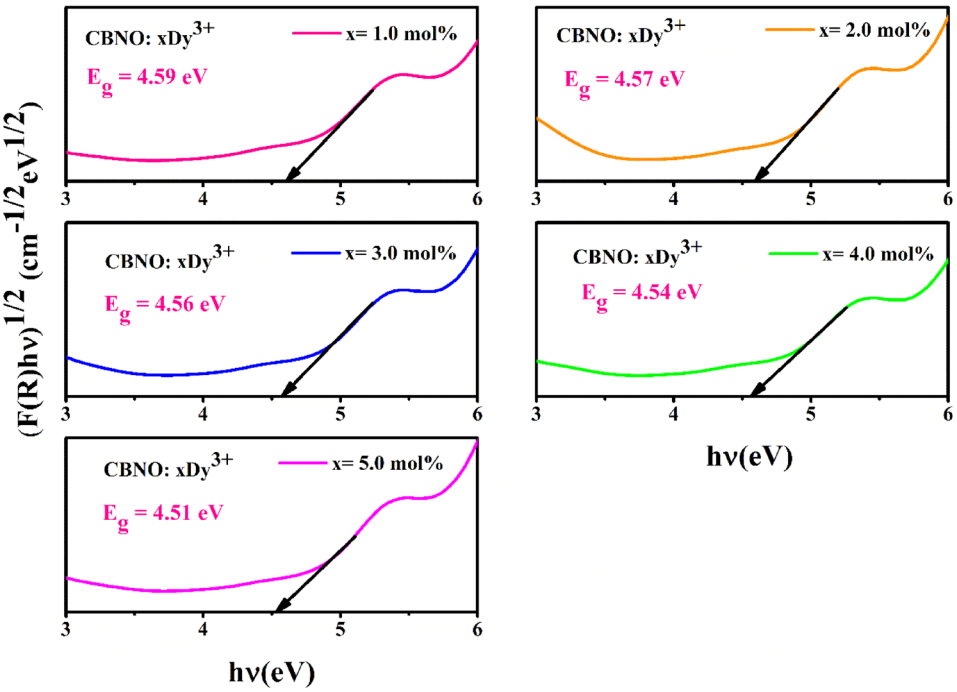


Fig. 7 Tauc plot for direct optical bandgap of CBNO phosphors at different concentrations of Dy^{3+} ions

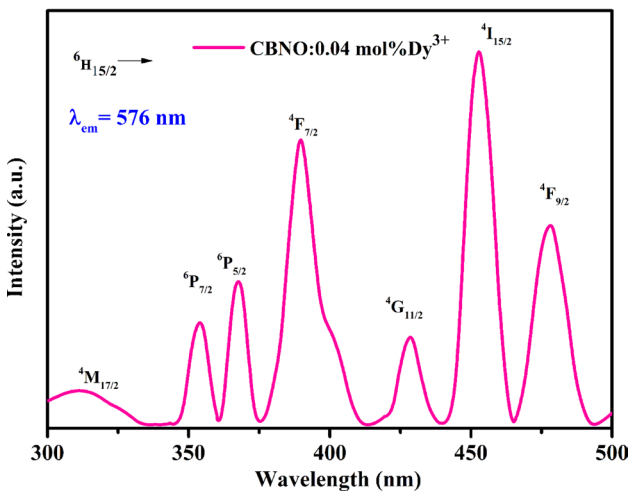


Fig. 8 Excitation spectrum of CBNO: 0.04 mol Dy3+-doped phosphor under 576 nm emission wavelength

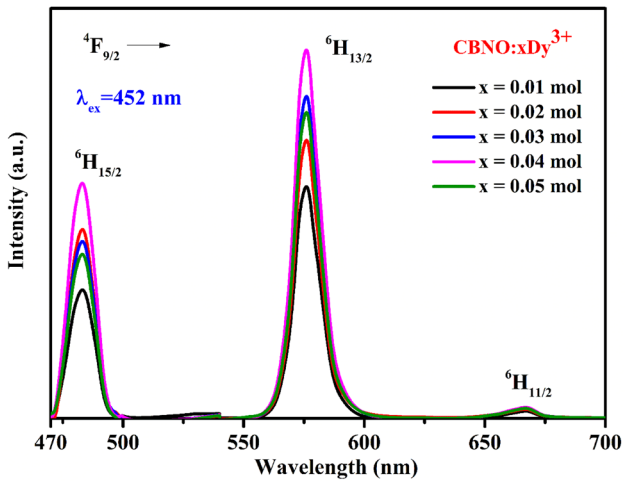


Fig. 9 PL emission spectra of CBNO: xDy^{3+} (x=0.01, 0.02, 0.03, 0.04, & 0.05 mol) at λ_{ex}=452 nm

Dy^{3+}–Dy^{3+} ions led to a drop in PL intensity. Radiation re-absorption, exchange contact, and multipole-multipole interaction between the activator ions are the reasons that cause concentration quenching.

The critical distance (R_c) is assessed to comprehend the kind of interaction that occurs between Dy^{3+} ions. Using Blasse's equation, the Rc is computed [43].

$$R_c \approx 2\left[\frac{3V}{4\pi x_c N}\right]^{1/3} \quad (3)$$

Here V stands for the unit cell volume, N is the number of cations per unit cell and x_c is the optimum concentration of the RE ions. After putting all the values, i.e. V = 748.68 Å^3, x_c = 0.04 mol, & N = 4, R_c was found to be 20.71 Å. A multipolar electric contact

between the Dy^{3+} ions arises during the energy transfer if the value of R_c is more than 5Å [44].

The type of multipolar interaction has been estimated from Dexter hypothesis, which suggests a relationship between variation in emission intensity and concentration of RE ions [45].

$$\log\left(\frac{I}{x}\right) = \log f - \frac{Q\log(x)}{3} \quad (4)$$

Here, Q stands for the type of interaction that takes place between the dopant (6, 8, or 10 denoting electric dipole–dipole (D–D), dipole-quadrupole (D–Q), and quadrupole–quadrupole (Q–Q), respectively) and A is a constant, too. The (Q/3) value is calculated using the slope of the linear fitting curve log(I/x) vs log(x), as demonstrated in Fig. 10. The computed Q value is 4.82 which is very near to 6, suggests that electric D–D interaction is the reason for concentration quenching and consequent energy transfer between Dy^{3+} ions.

Figure 11 displays the energy level diagram of the CBNO phosphor, which also covers the mechanisms for emission, excitation, energy transfer, and cross relaxation channels (CRC). The Dy^{3+} ions are first excited to higher excited levels as part of the emission mechanism, which is followed by a rapid relaxation to the metastable state $^4F_{9/2}$ level via non-radiative transfers. From the $^4F_{9/2}$ excited state, radiative transfers take place to various low energy states, such as $^6H_{15/2}$, $^6H_{13/2}$, and $^6H_{11/2}$. CRC and resonance energy transfer (RET) in the CBNO phosphor may result in concentration quenching and non-radiative energy transfer.

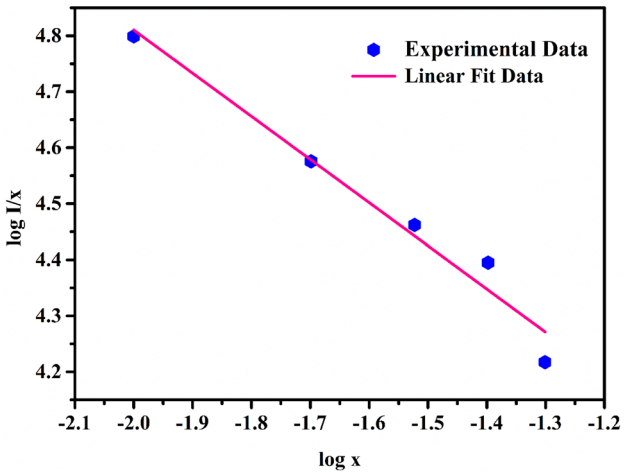


Fig. 10 Plot of log (I/x) with log(x)

3.6 CIE chromaticity coordinates

The colour of a phosphor is determined using Commission Internationale de l'Eclairclage (CIE) coordinates. The variations in the concentration of Dy^{3+} have no impact on the position of the emission peaks, but influence the intensity of the PL emission. Figure 12 displays the CIE coordinates of every Dy^{3+} ion-doped CBNO phosphor sample at λ_{ex} = 452 nm and λ_{em} = 576 nm, which are in the white area at ambient temperature. They have also been tabulated in Table 3 [17, 46, 47]. From the CIE chromaticity analysis, it can be concluded that these as prepared phosphors are acceptable for the manufacture of white LEDs.

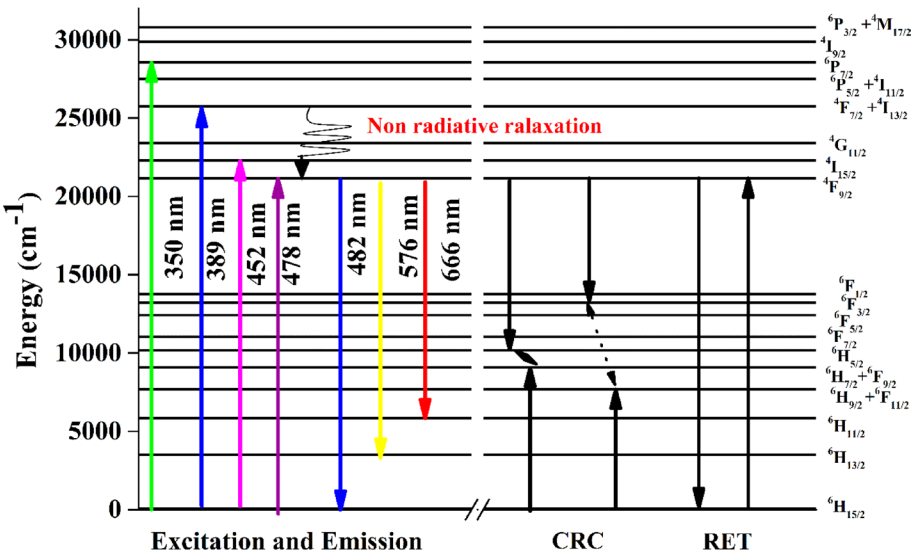


Fig. 11 Partial energy level diagram and possible cross-relaxation channels of Dy^{3+}-doped CBNO phosphor

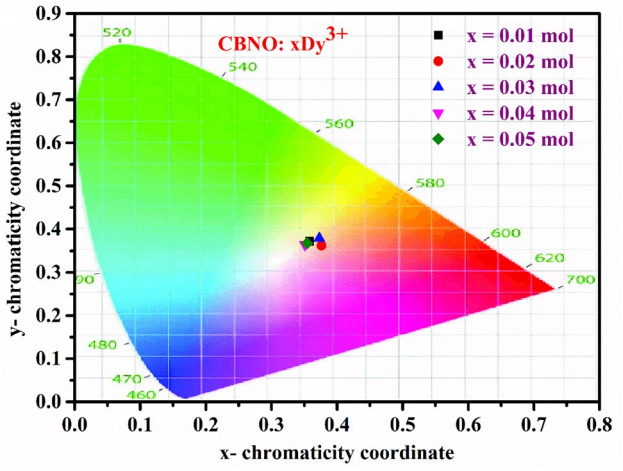


Fig. 12 CIE chromaticity coordinates for CBNO: xDy^{3+} (x = 0.01, 0.02, 0.03, 0.04, & 0.05 mol) phosphors under λex = 452 nm

Correlated colour temperature (CCT) is another significant indicator for emphasising the usage and significance of RE^{3+}-doped phosphor. CCT is the colour temperature of white light emitted by the phosphor. A low or high CCT indicates that the light being emitted is either warm or cold, respectively.

McCamy provided a formula for calculating CCT given below [9]:

$$CCT = -449n^3 + 3525n^2 - 6823.2n + 5520.3 \quad (5)$$

Here, $n = \frac{x-xe}{y-ye}$ and (x_e, y_e) refers to the colour coordinate of white illumination (x_e = 0.332, y_e = 0.186) and (x, y) stands for colour coordinates of the as-titled phosphor. Table 3 shows the CCT values of Dy^{3+}-doped CBNO phosphor at λ_{ex} = 452 nm. The calculated CCT values of the CBNO:xDy^{3+} phosphors is significantly less than 5000 K, indicating that warm white light can be created from these phosphors.

Colour purity (CP) is a significant factor in colourimetric calculations. The saturation and level of vibrant colour are described by CP. Following equation is used to evaluate the CP [48]:

$$CP = \frac{\sqrt{(x - x_{ee})^2 + (y - y_{ee})^2}}{\sqrt{(x_d - x_{ee})^2 + (y_d - y_{ee})^2}} \quad (6)$$

where (x_{ee}, y_{ee}) stand for equal energy point, (x, y) are colour coordinates of the optimized phosphor, and (x_d, y_d) denotes points of dominant wavelength. The calculated CP values for all the phosphors lie between 14 to 22%. These values are small and have been tabulated in Table 3. Owing to the low values of CP, it can be used in photonic devices as an n-UV pumped white emitting component [49].

3.7 Decay profiles

PL decay curves for CBNO: x Dy^{3+} (x = 0.01, 0.02, 0.03, 0.04, & 0.05 mol) phosphors under λ_{ex} = 452 nm and λ_{em} = 576 nm are shown in Fig. 13. The best fit to the measured decay curves is tri-exponential equation which is given below [50]:

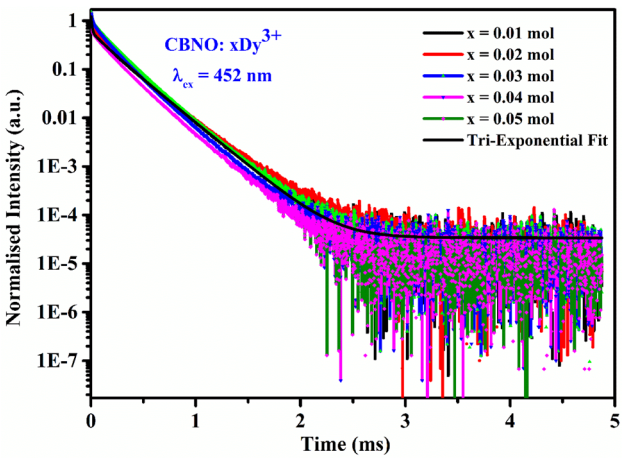


Fig. 13 PL decay profiles for CBNO: xDy^{3+} (x = 0.01, 0.02, 0.03, 0.04, & 0.05 mol) phosphors under λex = 452 nm

Table 3 CIE coordinates, CCT values & colour purity of CBNO: xDy^{3+} (x = 0.01, 0.02, 0.03, 0.04, & 0.05 mol) under 452 nm excitation wavelength

Sample ID CBNO: xDy^{3+}	CIE coordinates (x, y)	CCT values (K)	Colour purity (%)	τ_{avg} (μs)	References
x = 0.01 mol	(0.3579,0.3719)	4275.087	18.2	220.85	[17, 46, 47, 51]
x = 0.02 mol	(0.3761,0.3622)	4218.556	21.1	215.81	
x = 0.03 mol	(0.3729,0.3789)	3992.671	17.6	198.79	
x = 0.04 mol	(0.3508,0.3639)	4163.508	14.3	194.41	
x = 0.05 mol	(0.3543,0.3675)	3898.664	16.7	185.63	

$$I = I_0 + A_1 * \exp\left(-\frac{t}{\tau_1}\right) + A_2 * \exp\left(-\frac{t}{\tau_2}\right) + A_3 * \exp\left(-\frac{t}{\tau_3}\right) \quad (7)$$

where I_0 and I are the PL intensity at time $t = 0$ and time t, respectively. τ_1, τ_2 and τ_3 denote the decay lifetimes and A_1, A_2 & A_3 denote the constants for fitting. The average decay time (τ_{avg}) has been calculated with the formula given below [50, 51]

$$\tau_{avg} = \frac{A_1\tau_1^2 + A_2\tau_2^2 + A_3\tau_3^2}{A_1\tau_1 + A_2\tau_2 + A_3\tau_3} \quad (8)$$

The evaluated values of τ_{avg} have been presented in Table 3.The values of τ_{avg} decrease with the increase in doping ions demonstrating the possibility of energy transfer between Dy^{3+} ions [52]. As the doping concentration rises, the spacing between Dy^{3+} ions gets smaller. The luminescence lifetime decreases due to an increase in the nonradiative interaction between activator ions.

3.8 TD-PL studies

The TD-PL spectra of CBNO: 0.04 mol Dy^{3+}-doped phosphor were recorded under 452 nm excitation wavelength between 25 and 175ºC to study the thermal quenching behaviour and have been displayed in Fig. 14 along with the varying temperature. The PL emission intensity was seen to be decreasing with increasing temperature. At 175 ºC, the phosphor retained 48.2% of its maximal light intensity, demonstrating remarkable thermal stability.

To further assess the thermal characteristics of the sample, the activation energy (ΔE) was also measured. The value of ΔE was calculated by using the Arrhenius equation [53]

$$I_T = \frac{I_0}{1 + C\exp\left(-\frac{\Delta E}{K_B T}\right)} \quad (9)$$

Here, I_0 and I_T represent the intensity at temperature 25 ºC and the intensity at a specific temperature T and K_B denotes Boltzmann constant. The value of the activation energy (ΔE) is obtained from the slope of the of graph shown for $\ln((I_0/I_T)-1)$ versus $1/K_BT$, as depicted in Fig. 15. The calculated activation energy for CBNO: 0.04 mol Dy^{3+}-doped phosphor was 0.287 eV which is greater than other values in literature [12, 54, 55]. The high value of activation energy indicates that the as-prepared phosphor is more stable thermally. As a result, TD-PL spectra for the CBNO: 0.04 mol Dy^{3+}-doped phosphor demonstrated the excellent temperature stability [42].

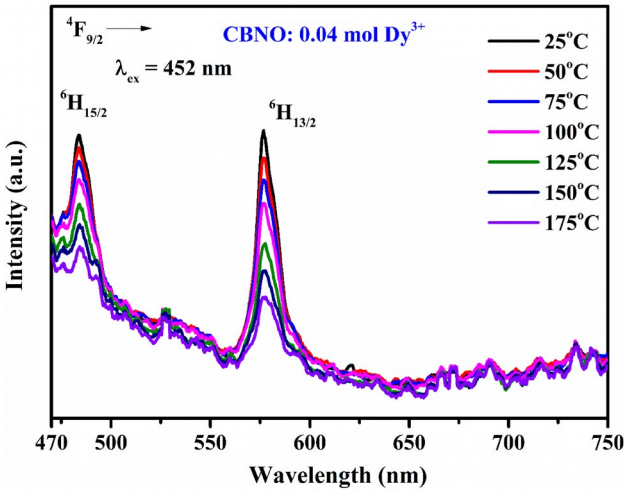


Fig. 14 TD-PL spectra for 0.04 mol Dy^{3+}-doped CBNO phosphor at 452 nm excitation wavelength from RT to 175 ºC

4 Conclusion

In conclusion, Dy^{3+}-doped CBNO phosphors were effectively synthesised using the solid-state reaction technique. The XRD analysis verifies the single phase formation of Dy^{3+}-doped CBNO phosphors. Morphological study was carried out using SEM, and the particle size was determined to be in the micrometre range. EDS indicates that all the elements were evenly distributed over the crystal's surface. FT-IR spectrum for Dy^{3+}-doped phosphor demonstrate the presence of vibrational bands. The optical bandgap of the as-titled phosphors was seen to decrease with an increase in the quantity of Dy^{3+} ions, owing to their 4f energy level. The Dy^{3+}-doped CBNO phosphors show emission peaks at wavelengths of 482, 576, and 661 nm when excited at a wavelength of 452 nm, which are attributable to distinct transitions of the Dy^{3+} ions. For CBNO phosphor, the optimal Dy^{3+} concentration was found to be 0.04 mol beyond which quenching takes place. The calculated CIE coordinates for all of the Dy^{3+} ion-doped CBNO phosphor samples at $\lambda_{ex} = 452$ nm and $\lambda_{em} = 576$ nm lie in the white region. A tri-exponential behaviour is observed in the decay curves of the as-prepared phosphors under stimulation at 452 nm.

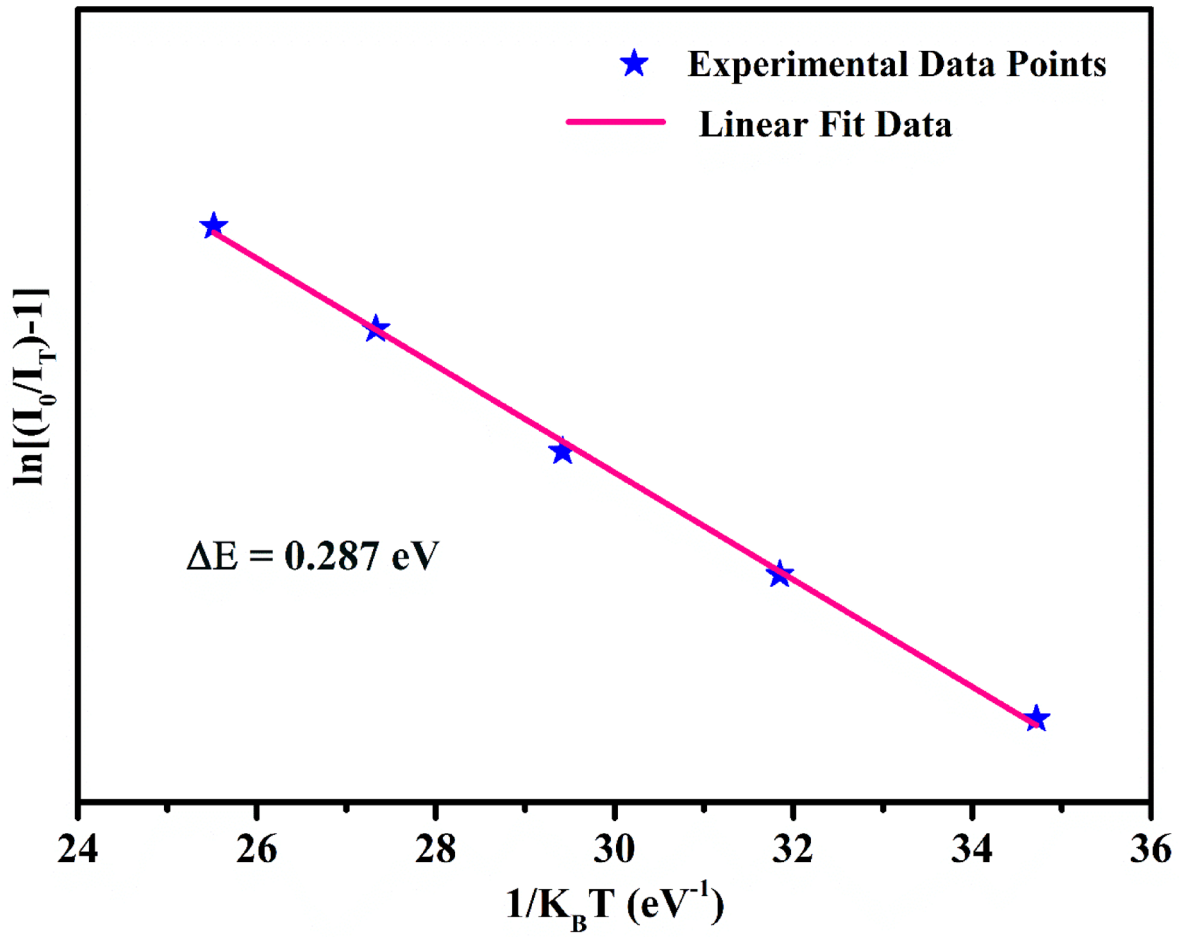


Fig. 15 Linear relationship of ln $[(I_0/I_T)-1]$ versus $1/K_BT$

The thermal stability of the phosphors was evaluated by TD-PL measurements. It was observed that a high activation energy was needed for thermal quenching in the phosphors. The experimental findings show that the CBNO phosphor doped with Dy^{3+} ions has at potential for usage in the fabrication of w-LEDs and other optoelectronic devices.

Acknowledgements

The author is very thankful to CSIR-UGC for giving her a Junior Research Fellowship (UGC Ref. 1431/CSIR-UGC NET JUNE 2019). Dr. Aman Prasad is grateful to Dayalbagh Educational Institute (DEI), Deemed University, Agra for providing a Research Seed Money Grant (DEI/GBMF(3052022)/518). We are also grateful to Dr. Shayama Prasad and Dr. Manju Srivastava in the Department of Chemistry at Dayalbagh Educational Institute (DEI), Deemed University, Agra for making the FESEM facility available to us.

Author contributions

Sheetal Kumari: Experimentation, data analysis, manuscript writing. Pooja Rohilla: Experimentation, data analysis. Samarthya Diwakar: Synthesis, experimentation. Rupesh Talewar: Data Analysis,
Ankur Shandilya: Data Analysis. Yasha Tayal: Experimentation, data analysis. K. Swapna: Data analysis. Aman Prasad: Work plan, supervision, review, data analysis. A.S. Rao: Data analysis, review.

Data availability

The data that support the findings of this study are available on request from the corresponding author Dr. Aman Prasad. The data are not publicly available due to restrictions, e.g. their containing information that could compromise the privacy of research participants.

Declarations

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

1. J. Wu, M. Li, H. Jia, M. Wang, Z. Liu, J. Lumin. **214**, 116607 (2019)
2. N. Deopa, M.K. Sahu, P.R. Rani, R. Punia, A.S. Rao, J. Lumin. **222**, 117166 (2020)
3. G. Wu, J. Xue, X. Li, Q. Bi, M. Sheng, Z. Leng, Ceram. Int. **49**, 10615 (2023)
4. J. Guo, J. Guo, J. Gao, J. Chen, Y. Yang, Y. Yang, L. Zheng, Y. Li, L. Zhao, B. Deng, R. Yu, Ceram. Int. **48**, 26992 (2022)
5. V.R. Prasad, S. Damodaraiah, S. Babu, Y.C. Ratnakaram, J. Lumin. **187**, 360 (2017)
6. S. Sarathkumar, S. Munimasthani, U.S.U. Thampy, R.V.S.S.N. Ravikumar, J. Mater. Sci. Mater. Electron. **31**, 11589 (2020)
7. X. Huang, J. Liang, S. Rtimi, B. Devakumar, Z. Zhang, Chem. Eng. J. **405**, 126950 (2021)
8. P. Rohilla, A.S. Rao, Mater. Res. Bull. **150**, 111753 (2022)
9. L. Zhou, L. Liu, R. Yang, J. Kong, Y. Li, Y. Zhao, Y. Du, C. Li, Z. Zou, B. Deng, R. Yu, Inorg. Chem. Commun. **148**, 110316 (2023)
10. R. Yu, Y. Guo, L. Wang, H.M. Noh, B.K. Moon, B.C. Choi, J.H. Jeong, J. Lumin. **155**, 317 (2014)
11. S. Dai, D. Zhao, R. Zhang, L. Jia, Q. Yao, J. Alloys Compd. **891**, 161973 (2021)
12. L. Wu, Y. Bai, L. Wu, H. Yi, Y. Kong, Y. Zhang, J. Xu, RSC Adv. **7**, 1146 (2017)
13. P. Rohilla, A.S. Rao, J. Mater. Sci. Mater. Electron. **34**, 1 (2023)
14. S. Adachi, J. Lumin. **202**, 263 (2018)
15. B. Han, Y. Dai, J. Zhang, B. Liu, H. Shi, Ceram. Int. **44**, 14803 (2018)
16. K.G. Kumar, P.B. Bhargav, K. Aravinth, R. Arumugam, P. Ramasamy, J. Lumin. **223**, 117126 (2020)
17. P. Sehrawat, A. Khatkar, P. Boora, M. Kumar, R.K. Malik, S.P. Khatkar, V.B. Taxak, Ceram. Int. **46**, 16274 (2020)
18. S.R. Alharbi, M. Alhassan, O. Jalled, S. Wageh, A. Saeed, J. Mater. Sci. **53**, 11584 (2018)
19. T. Sakthivel, G. Annadurai, R. Vijayakumar, X. Huang, J. Lumin. **205**, 129 (2019)
20. H. George, N. Deopa, S. Kaur, A. Prasad, M. Sreenivasulu, M. Jayasimhadri, A.S. Rao, J. Lumin. **215**, 116693 (2019)
21. M. Monisha, N. Mazumder, G. Lakshminarayana, S. Mandal, S.D. Kamath, Ceram. Int. **47**, 598 (2021)
22. W.J. Huang, J.H. Lei, Y.H. Chen, J.Y. Chen, Y. Li, L.P. Chen, H. Guo, Ceram. Int. **49**, 9799 (2023)
23. S. Damodaraiah, R.P. VL, Y.C. Ratnakaram, Optik (Stuttg) **192**, 162980 (2019)
24. A. Ichoja, S. Hashim, S.K. Ghoshal, Optik (Stuttg) **218**, 165001 (2020)
25. Z. Shen, C. Qin, W. Luo, F. Song, Z. Wang, Y. Li, S. Zhang, J. Mater. **6**, 459 (2020)
26. T.S. Sreena, P.P. Rao, A.K.V. Raj, T.R.A. Thara, J. Alloys Compd. **751**, 148 (2018)
27. H. Zou, Y. Hu, X. Zhu, D. Peng, X. Chai, J. Mater. Sci. Mater. Electron. **28**, 11921 (2017)
28. C. Qin, Z. Yang, S. Wen, Q. Luo, F. Sheng, S. Yan, H. Zhu, M. Wang, Y. Ming, J. Mater. Sci. Mater. Electron. **29**, 7801 (2018)
29. M. Ferroelectrics, D. Peng, H. Zou, C. Xu, J. Li, X. Wang, X. Yao, J. Appl. Phys. (2013). https://doi.org/10.1063/1.4818793
30. R. Tang, J. Kong, L. Zheng, Z. Wang, J. Guo, X. Jin, X. Hu, N. Gong, B. Deng, R. Yu, Mater. Res. Bull. **154**, 111925 (2022)
31. K. Miura, Appl. Phys. Lett. **80**, 2967 (2002)
32. P. Kumar, S. Singh, I. Gupta, K. Nehra, V. Kumar, D. Singh, Mater. Chem. Phys. **295**, 127035 (2023)
33. J. Song, T. Yang, H. Zhang, Y. Xiang, R. Song, B. Wang, J. Zhu, J. Lumin. **257**, 119664 (2023)
34. H.C. Gupta, V. Luthra, Vib. Spectrosc. **56**, 235 (2011)
35. O. Jalled, Z. Alharbi, S.R. Alharbi, A. Saeed, M. Alhassan, S. Al-Heniti, H.Y. Mohammed, Y. Al-Hadeethi, F. Al-Marzouki, A. Al-Mujtaba, J. Nanosci. Nanotechnol. **17**, 594 (2017)
36. W.T. Carnall, P.R. Fields, K. Rajnak, J. Chem. Phys. **49**, 4424 (1968)
37. S. Kumari, A.S. Rao, R.K. Sinha, Mater. Res. Bull. **167**, 112419 (2023)
38. N. Deopa, A.S. Rao, J. Lumin. **194**, 56 (2018)

39. E.A. Rathnakumari, S.M.M. Kennedy, J. Lumin. **235**, 118018 (2021)
40. S. Kaur, A.S. Rao, M. Jayasimhadri, Mater. Res. Bull. **116**, 79 (2019)
41. S. Kaur, D. Arora, S. Kumar, G. Singh, S. Mohan, P. Kaur, P.K. Kriti, D.P. Singh, J. Lumin. **202**, 168 (2018)
42. S. Kumari, A. Anu, P.R. Prasad, A.S. Rao, J. Mater. Sci. Mater. Electron. **34**, 907 (2023)
43. X. Meng, S. Huang, M. Shang, Inorg. Chem. Commun. **113**, 107768 (2020)
44. J. Zhao, S.X. Huang, D. Zhao, C.H. Huang, M.J. Ma, R.J. Zhang, Y.C. Fan, B.Z. Liu, J.W. Gao, Q. Zong, Y.P. Yu, Optik (Stuttg) **159**, 115 (2018)
45. S. Kaur, N. Deopa, A. Prasad, R. Bajaj, A.S. Rao, Opt. Mater. (Amst) **84**, 318 (2018)
46. X. Qin, X. Liu, W. Huang, M. Bettinelli, X. Liu, Chem. Rev. **117**, 4488–527 (2017)
47. E.A. Rathnakumari, S.M.M. Kennedy, Optik (Stuttg) **275**, 170601 (2023)
48. S. Kumari, A.S. Rao, R.K. Sinha, J. Mol. Struct. **1295**, 136507 (2024)
49. K. Poria, R. Lohan, S. Bhatia, A. Kumar, R. Singh, N. Deopa, R. Punia, J.S. Shahi, A.S. Rao, RSC Adv. **13**, 11557 (2023)
50. G. Jyothi, K.G. Gopchandran, Ceram. Int. **43**(15), 12044–12056 (2017)
51. X. Li, C. Liang, S. Guo, Y. Xiao, C. Chang, J. Lumin. **188**, 199 (2017)
52. X. Zhang, R. Cui, J. Zhang, G. Yuan, X. Qi, C. Deng, Optik (Stuttg). **245**, 167646 (2021)
53. K. Maheshwari, A.S. Rao, Opt. Mater. (Amst) **137**, 113533 (2023)
54. G. Zhu, Z. Li, F. Zhou, C. Wang, S. Xin, J. Lumin. **196**, 32 (2018)
55. B. Ma, X. Ma, T. Xu, K. Su, Q. Zhang, RSC Adv. **8**, 14164 (2018)

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

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

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

Techno-Economic, Environmental, and Policy Perspectives of Carbon Capture to Fuel Technologies

Aditya Singh[1], Vishesh Saini[2], Sambhav Jain[3], Anunay Gour[4]
[1,2,3]UG - Environmental Engineering, Delhi Technological University, Rohini, Delhi, India.
[4]Assistant Professor, Environmental Engineering, Delhi Technological University, Rohini, Delhi, India.
***Emails**: adityasingh_en20a18_21@dtu.ac.in[1], visheshsaini_en20a18_07@dtu.ac.in[2], sambhavjain_en20a17_58@dtu.ac.in[3], anunaygour@dtu.ac.in[4]*

Abstract
This research paper provides a comprehensive exploration of carbon capture to fuel technologies, covering various capture methodologies and conversion processes. The analysis begins by dissecting post-combustion, pre-combustion, and direct air capture technologies, elucidating their principles, advantages, and limitations. A focus on the conversion of captured carbon into usable fuels delves into synthetic fuels and hydrogen production methods, detailing chemical processes, catalysts, and energy requirements. Moving beyond technical aspects, the paper critically analyzes the efficiency and viability of carbon capture to fuel processes, employing case studies and real-world examples to illustrate the practical application of techno-economic assessments and life cycle analyses. Economic considerations further assess implementation costs, operational expenses, and potential revenue streams, drawing insights from existing economic models and case studies. Environmental impact and benefits take center stage, evaluating potential reductions in greenhouse gas emissions, resource efficiency, and ecological considerations associated with converting captured carbon to fuel. A comparative analysis with other carbon capture applications offers a holistic perspective on the environmental footprint. The regulatory landscape is thoroughly examined, encompassing existing policies, government incentives, and international agreements influencing the development and deployment of carbon capture to fuel technologies. The research concludes with reflections on the current status, challenges, and a roadmap for future advancements, serving as a comprehensive guide for researchers, policymakers, and industry stakeholders in the pursuit of sustainable energy solutions.

Keywords: *Carbon Capture, Fuel Production, Techno-Economic Assessment, Environmental Impact, Policy and Regulation, Sustainability.*

1. Introduction
The escalating threat of climate change necessitates urgent and innovative strategies to mitigate carbon dioxide (CO2) emissions. As industries and societies grapple with the imperative to transition towards sustainable energy sources, carbon capture technologies have emerged as pivotal tools in the fight against climate change. This paper focuses on the transformative potential of carbon capture to fuel technologies, offering a nuanced exploration of their technological foundations, economic viability, environmental impact, and the intricate policy landscapes shaping their implementation.

The primary objective of this research is to comprehensively explore the intricacies of carbon capture to fuel technologies. This includes delving into various capture methodologies such as post-combustion, pre-combustion, and direct air capture, and examining their principles, advantages, and limitations. The subsequent exploration of the conversion of captured carbon into usable fuels involves an in-depth examination of synthetic fuels and hydrogen production methods, including the chemical processes,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

catalysts, and energy requirements involved. Beyond the technical aspects, the paper critically evaluates the efficiency and viability of carbon capture to fuel processes, using real-world case studies to illustrate the practical application of techno-economic assessments and life cycle analyses. Economic considerations further dissect the implementation costs, operational expenses, and potential revenue streams, drawing insights from existing economic models and case studies. Environmental impact and benefits take center stage in the analysis, assessing potential reductions in greenhouse gas emissions, resource efficiency, and ecological considerations associated with converting captured carbon to fuel. The paper also includes a comparative analysis with other carbon capture applications, offering a holistic perspective on the environmental footprint. Navigating the complex policy and regulatory landscape, the study examines existing frameworks, government incentives, and international agreements influencing the development and deployment of carbon capture to fuel technologies. The paper concludes by reflecting on the current status, challenges, and presenting a roadmap for future advancements, contributing to the collective knowledge base for researchers, policymakers, and industry stakeholders in the pursuit of a sustainable and low-carbon future.

2. Carbon Capture Technologies

In this segment, we are able to now not only discover the concepts, blessings, and limitations of every carbon capture technology however additionally shed light on latest trends and innovations which have fashioned the evolving landscape of carbon capture techniques.

2.1. Post-Combustion Capture

Post-combustion capture is an extensively employed technique that focuses on capturing carbon dioxide (CO2) emissions after the combustion of fossil fuels. Exhaust gases from commercial procedures or electricity plants are subjected to separation technologies, which include amine scrubbing or chemical adsorption, to isolate and seize CO2. This technique is effective for retrofitting current facilities, as it could be carried out to flue gases from numerous assets.

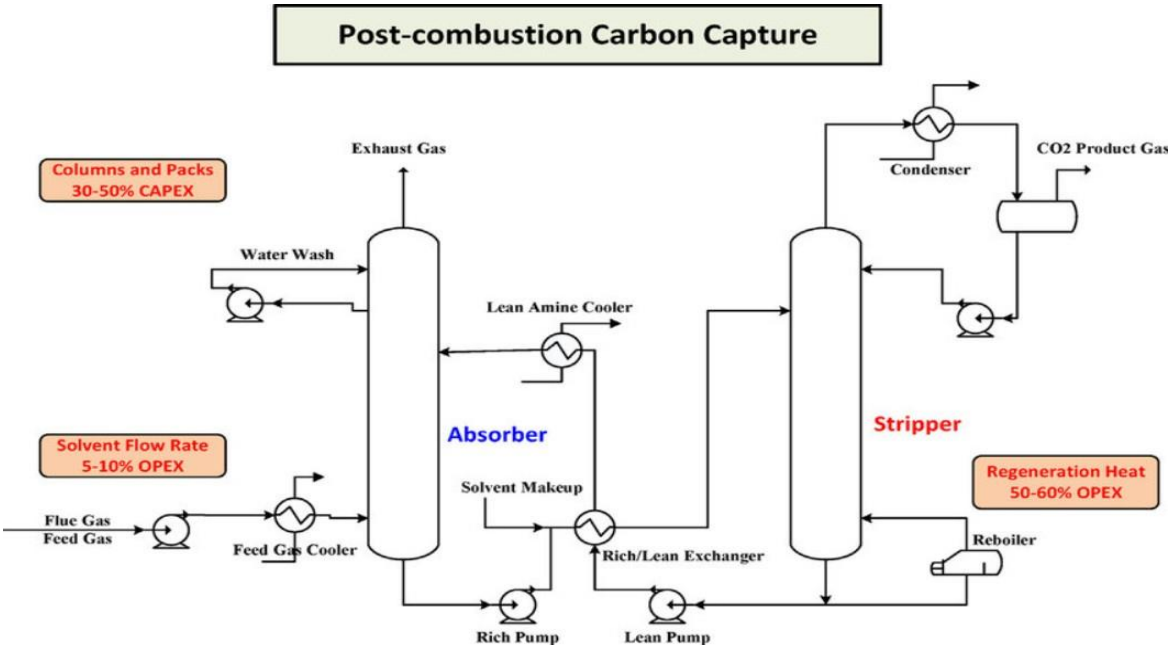


Figure 1 Post-Combustion Carbon Capture Process

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

Figure 1 A simplified representation of a post-combustion carbon capture process utilizing reactive solvents is depicted in the diagram [1].

2.1.1. Principles

Post-combustion capture by and large entails the separation of CO2 from flue gases emitted in the course of the combustion of fossil fuels. This is completed through absorption or adsorption tactics the usage of solvents or stable sorbents.

2.1.2. Advantages

Post-combustion seize is adaptable to existing business facilities, making it a viable option for retrofitting. It enables the reduction of emissions from strength flowers and business methods without large changes to existing infrastructure.

2.1.3. Limitations

The main demanding situations encompass the strength-extensive nature of the separation system and the want for large volumes of solvents or sorbents. This can result in accelerated operational prices and ability environmental effects.

2.1.4. Recent Developments

Recent advancements in post-combustion capture awareness on enhancing the performance of separation processes and developing novel solvents or sorbents with advanced overall performance and reduced environmental effect. Integration with renewable strength sources and optimization of process parameters are key areas of innovation [1].

2.2. Pre-Combustion Capture

Pre-combustion capture involves the extraction of carbon from fossil fuels before combustion occurs. Commonly applied in included gasification blended cycle (IGCC) electricity flora, this method converts fossil fuels right into a syngas, making an allowance for the separation and capture of carbon before electricity technology. The syngas, predominantly composed of hydrogen and carbon monoxide, can be applied for power production, at the same time as the captured carbon is stored or repurposed.

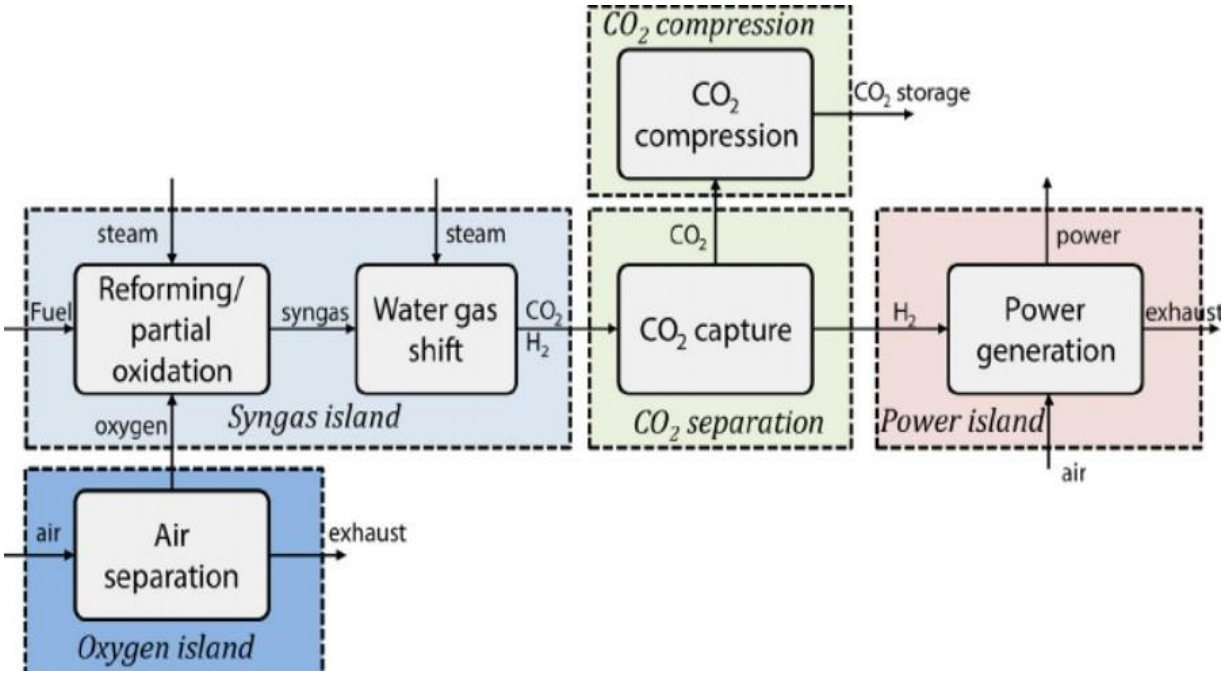


Figure 2 The Pre-Combustion Capture Process

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

Figure 2: A diagram illustrating the pre-combustion capture process designed for power generation.

2.2.1. Principles

Pre-combustion capture involves the conversion of fossil fuels into syngas earlier than combustion. Carbon is separated from the syngas, typically through water-gasoline shift reactions or bodily separation strategies.

2.2.2. Advantages

Pre-combustion seize offers higher performance in comparison to submit-combustion strategies and enables the manufacturing of hydrogen, a precious clean electricity carrier. It is mainly suitable for new power vegetation or industrial facilities.

2.2.3. Limitations

The complexity of the gasification manner and the need for specialized systems can bring about higher capital prices. Additionally, the downstream techniques for coping with and storing captured carbon can pose technical demanding situations.

2.2.4. Recent Developments

Ongoing studies in pre-combustion capture focuses on optimizing gasification techniques, exploring alternative feedstocks, and growing superior catalysts for syngas manufacturing. Integration with hydrogen production technologies is a key region of innovation [2].

2.3. Direct Air Capture

Direct air capture is a rising generation designed to seize CO2 at once from the ambient air. Utilizing specialized sorbents or solvents, direct air seizes structures selectively bind with CO2 molecules, allowing for their separation. This approach is in particular precious for addressing emissions from dispersed sources, together with transportation and agriculture. While nonetheless in the early stages of development, direct air seizes holds promise for attaining terrible emissions and fighting climate trade on a broader scale in Figure 3.

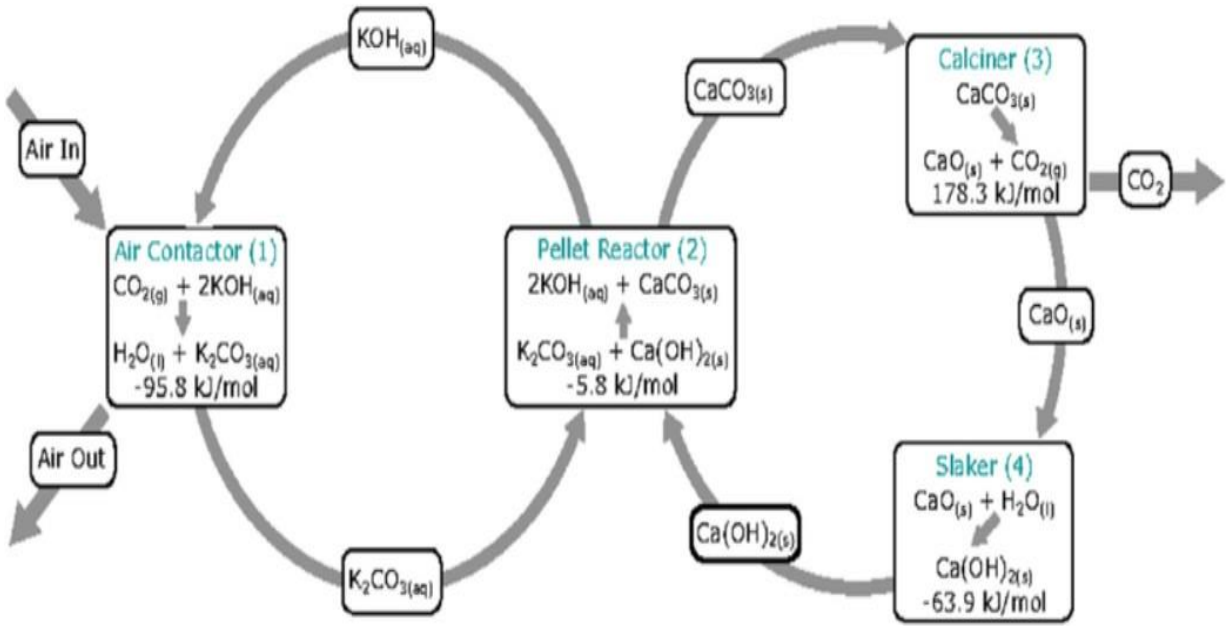


Figure 3 A Diagram Illustrating the Direct Air Capture [3]

2.3.1. Principles

Direct air capture includes the extraction of CO2 immediately from the ecosystem using sorbents or solvents. The captured CO2 can be saved underground or utilized for diverse applications.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

2.3.2. Advantages

Direct air seize has the ability to cope with emissions from dispersed assets and gain terrible emissions. It isn't always restrained to particular industrial facilities, offering flexibility in deployment.

2.3.3. Limitations

The technology is presently energy-in-depth, and the price of keeping tons of captured CO2 remains quite high. Scaling up direct air seize facilities to have a meaningful impact on international emissions poses logistical and economic demanding situations.

2.3.4. Recent Developments

Recent improvements in direct air seize consciousness on growing extra efficient sorbents and solvents, optimizing seize device designs, and exploring synergies with different industrial techniques. Pilot tasks and investments imply a developing hobby in advancing this technology [3]. By examining the standards, advantages, and boundaries of each carbon seize era and exploring current trends, this paper aims to provide a comprehensive review of the cutting-edge state and future potential of carbon seize methods. Each of these carbons seize technologies offers unique advantages and demanding situations, influencing their suitability for extraordinary business applications. The following sections will further discover the efficiencies, fees, and environmental concerns associated with these strategies, supplying a comprehensive expertise of the panorama of carbon seize technologies.

3. Conversion of Captured Carbon to Fuel

As we explore the conversion of captured carbon into usable fuels, it's extremely important to have a look at numerous strategies and their potential for transforming carbon emissions into valuable electricity assets. This segment will delve into various techniques, with a focal point on artificial fuels and hydrogen manufacturing.

3.1. Synthetic Fuels

Synthetic fuels, often called carbon-impartial or e-fuels, are generated through the conversion of captured carbon, typically carbon dioxide (CO2), into hydrocarbons or different electricity-rich compounds. This technique often includes the use of renewable power sources to electricity chemical reactions, ensuring a closed carbon cycle. Fischer-Tropsch synthesis and electrochemical methods are a number of the techniques employed to supply artificial fuels. This subsection will explore the ideas at the back of those strategies, emphasizing their ability to benefit in terms of energy density, compatibility with present infrastructure, and reduced dependence on fossil fuels in Figure 4[5].

Hydrogenation of CO2:

$CO^2 + 3H_2 \leftrightharpoons CH_2 + 2H_2O\ (\Delta H^0_{298} = 125\ kJ\ mol^{-1};$

The CWGS reaction:

$CO_2 + H_2 \leftrightharpoons CO + H_2O\ (\Delta H^0_{298} = 41\ kJ\ mol^{-1});$ and

The FTS reaction:

$CO + 2H_2 \leftrightharpoons -(CH_2)- + H_2O\ (\Delta H^0_{298} = -166\ kJ\ mol^{-1})$

Figure 4 Representation of Chemical Co2 Conversion into Different Fuels [5]

3.2. Hydrogen Production

The conversion of captured carbon to hydrogen represents another road for sustainable energy technology. Hydrogen is taken into consideration as a flexible and easy power provider, and its manufacturing from captured carbon can make a contribution to a decarbonized electricity panorama. Steam methane reforming (SMR) with carbon seize and garage (CCS) and electrolysis using renewable power are two essential techniques. This component will discuss the concepts, blessings, and obstacles of those methods, considering elements such as power performance, scalability, and the environmental footprint related to hydrogen manufacturing. Figure 5 A system for capturing carbon transforms carbon dioxide into both electricity and hydrogen fuel. closer to a low-carbon strength future [6].

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

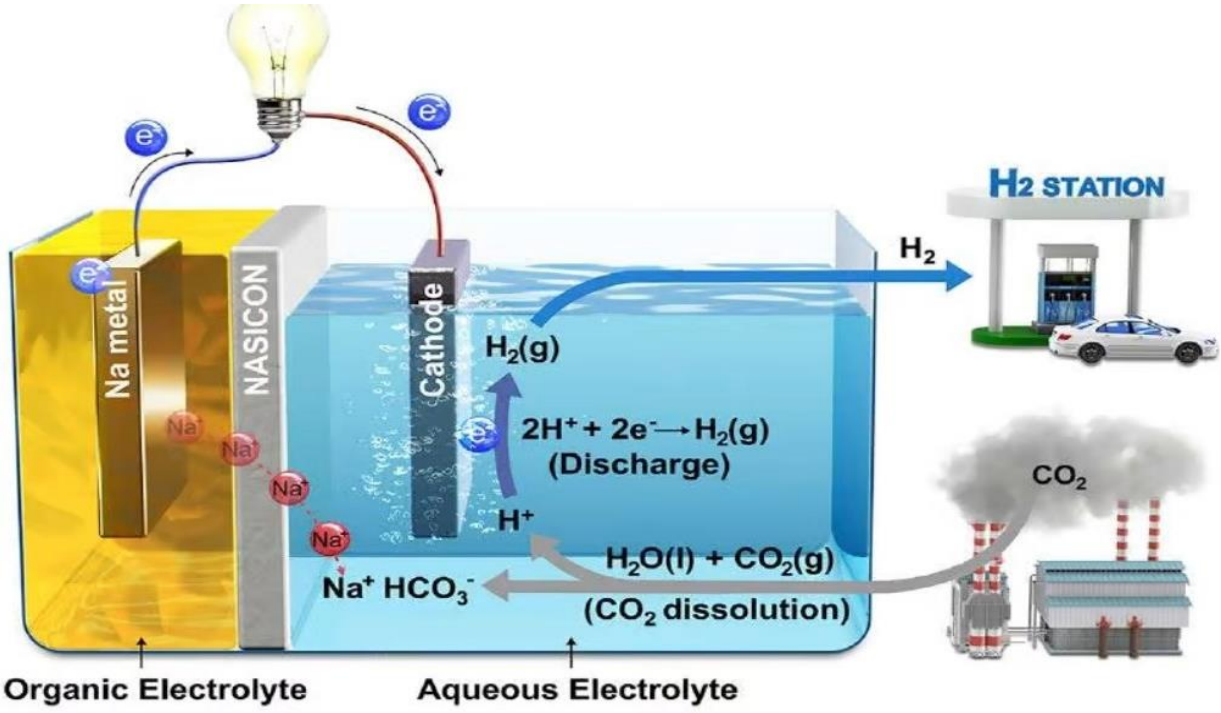


Figure 5 Carbon Transforms

By comprehensively exploring these methods, this segment aims to provide insights into the potential of changing captured carbon into fuels, addressing the growing call for sustainable electricity answers. Subsequent discussions will delve into the economic considerations, environmental influences, and actual-global applications of those conversion technologies, imparting a holistic knowledge of the position they play in the transition.

3.3. Chemical Processes and Catalysts for Conversion

The conversion of captured carbon into usable fuels involves problematic chemical strategies and often relies on catalysts to facilitate efficient reactions. Understanding those methods is critical for comparing the feasibility and effectiveness of carbon seize to fuel conversion.

3.3.1. Fischer-Tropsch Synthesis for Synthetic Fuels:

Fischer-Tropsch synthesis is a nicely-established chemical technique used inside the production of synthetic fuels. It involves the catalytic conversion of syngas (a combination of hydrogen and carbon monoxide, often derived from captured carbon) into liquid hydrocarbons. Transition metallic-primarily based catalysts, which includes iron or cobalt, play an essential position in selling the necessary reactions. This subsection will delve into the intricacies of Fischer-Tropsch synthesis, discussing reaction mechanisms, catalyst choice, and recent improvements in enhancing catalytic efficiency in Figure 5.

3.3.2. Electrolysis for Hydrogen Production

In the context of hydrogen manufacturing from captured carbon, electrolysis is a key procedure. This method includes passing an electric powered modern via water or different feedstocks containing captured carbon to cut up water molecules and generate hydrogen. Various electrolysis technologies, which include proton trade membrane (PEM) and alkaline electrolysis, employ particular catalysts like platinum or nickel to beautify response kinetics. This segment will explore the chemical concepts behind electrolysis, the function of catalysts, and current developments in electrolysis technology for efficient hydrogen production [6] in Figure 6.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

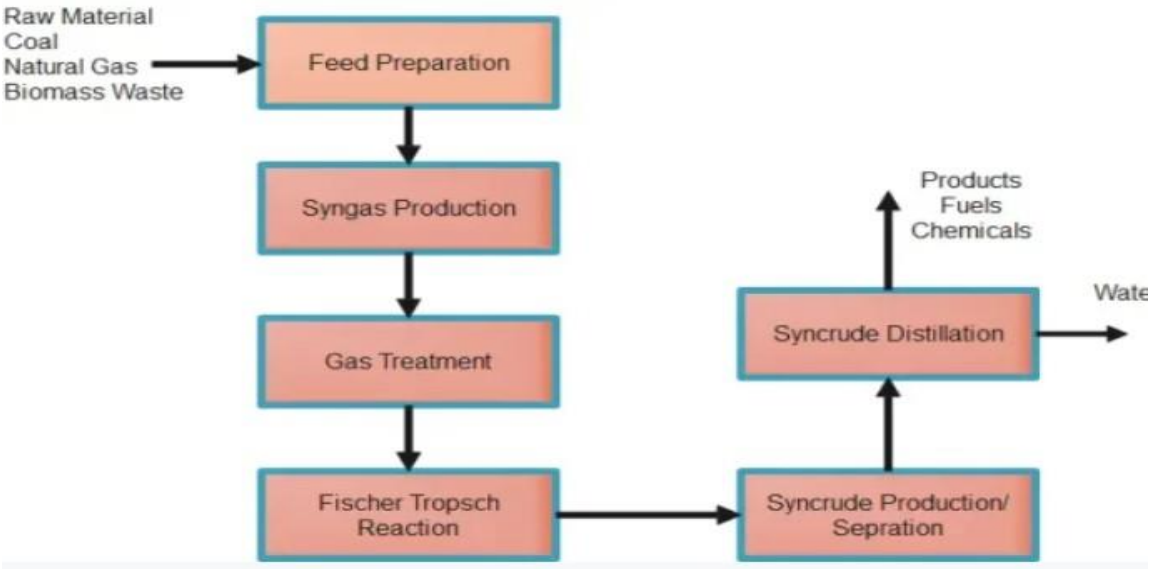


Figure 6 Steps Includes in Fischer- Tropsch [4]

3.3.3. Novel Catalysts and Emerging Technologies

Continual research efforts are centered on discovering novel catalysts and refining current ones to beautify the efficiency of carbon capture to gasoline conversion tactics. Advances in nanotechnology, catalytic substances, technological know-how, and computational modeling contribute to the identity of catalysts with stepped forward selectivity, sturdiness, and resistance to impurities. This subsection will spotlight current breakthroughs in catalyst improvement and emerging technologies that display promise in optimizing the energy efficiency and overall performance of carbon conversion methods [13].

3.4. Energy Requirements for Conversion

The conversion of captured carbon to fuels is inherently strength-extensive, and know-how the strength requirements is vital for assessing the overall sustainability of those strategies. This component will discuss the electricity inputs for diverse conversion methods, considering elements which include energy demand, thermal electricity, and standard power efficiency. It will also deal with techniques to optimize power use, integrate renewable energy assets, and beautify the overall energy balance of carbon capture to fuel conversion technology. By delving into the chemical processes, catalysts, and energy requirements related to the conversion of captured carbon to fuels, this phase aims to offer a complete information of the technical components underpinning those transformative technologies. Subsequent discussions will in addition explore the economic and environmental dimensions, supplying a holistic angle at the feasibility and ability implications of carbon seize to gasoline conversion [14].

3.5. Chemical Processes, Catalysts, and Energy Requirements

To recognize the efficiency and viability of carbon capture to fuel tactics, it's imperative to scrutinize the underlying chemical procedures, the position of catalysts, and the related power necessities. This phase will delve into the elaborate information of these factors, dropping mild on the important factors that influence the success of converting captured carbon into usable fuels.

3.5.1. Chemical Processes

The chemical methods used inside the conversion of captured carbon to fuel vary depending on the chosen technique, be it Fischer-Tropsch synthesis for synthetic fuels or steam methane reforming for

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

hydrogen production. A certain exam of those chemical pathways might be undertaken, elucidating the reactions, intermediates, and by means of-merchandise. Understanding the nuances of each procedure is important for evaluating their efficiency and ability for huge-scale implementation.

3.5.2. Catalysts

Catalysts play a pivotal role in facilitating and optimizing the chemical adjustments essential for converting captured carbon into fuels. This subsection will explore the styles of catalysts utilized in extraordinary techniques, ranging from heterogeneous catalysts in Fischer-Tropsch synthesis to electrocatalysts in electrolysis. The dialogue will contribute to the catalyst's impact on response fees, selectivity, and stability, presenting insights into the advancements and challenges in catalyst improvement for carbon seizure to fuel packages [7].

3.5.3. Energy Requirements

The electricity necessities of carbon capture to gasoline processes are paramount in figuring out their standard performance and environmental impact. Analyzing the strength inputs for every step, from carbon capture to gasoline synthesis, will be a focus. Consideration might be given to the assets of energy used, including renewable energy or excessive-temperature warmness, and the results for the general carbon footprint of the procedure. This scrutiny will make contributions to the assessment of the sustainability and feasibility of carbon capture to gas technology. By unraveling the intricacies of chemical methods, catalysts, and electricity needs, this section aims to offer a complete understanding of the factors influencing the efficiency and viability of converting captured carbon into fuels. Subsequent discussions will build upon this basis, addressing monetary considerations, environmental impacts, and the wider implications for sustainable power structures [15].

4. Economic Considerations

In comparing the transition from carbon capture to gasoline technologies from a monetary angle, this section will verify the financial feasibility via delving into the associated prices, potential revenue streams, and drawing insights from present economic models and case studies.

4.1. Implementation Costs

Understanding the economic viability of carbon capture to fuel technology necessitates a radical examination of implementation costs. This consists of charges related to the deployment of carbon capture structures, conversion techniques, and the essential infrastructure. Factors which include era adulthood, scale of deployment, and geographical concerns might be explored to offer a nuanced knowledge of the financial commitments involved [8].

4.2. Operational Costs

Beyond initial implementation, operational prices play a critical position in figuring out the economic feasibility of these technologies. This encompasses the everyday prices related to going for walks and maintaining carbon capture facilities, fuel conversion processes, and any ancillary systems. Analyzing those ongoing costs is important for assessing the sustainability of carbon seize to gasoline operations over the long term [9].

4.3. Revenue Streams

To balance the monetary equation, it's vital to pick out potential revenue streams associated with carbon capture to gas technologies. This ought to involve the sale of synthesized fuels, participation in carbon credit score markets, or leveraging government incentives and subsidies. The section will discover the numerous avenues via which sales can be generated, emphasizing the want for a comprehensive enterprise model to aid the monetary viability of these technologies [10].

4.4. Economic Models and Case Studies

Drawing insights from current economic fashions and case research is pivotal in providing actual-world context to the monetary concerns of carbon seize to gas technology. By examining successful projects or the ones dealing with demanding situations, this subsection will offer a comparative evaluation. Evaluation standards will encompass

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

return on funding, payback periods, and the effect of policy frameworks on challenge economics. Such an exam will make contributions to the development of robust monetary fashions for destiny ventures. Through a radical exploration of implementation expenses, operational costs, revenue streams, and actual-world case studies, this segment pursues a comprehensive financial assessment of carbon seize to gas technology. Subsequent discussions will amplify the analysis to encompass broader environmental effects, policy considerations, and capacity avenues for destiny studies and development [8][9].

5. Environmental Impact

Assessing the environmental effect of converting captured carbon to fuel is fundamental to knowledge of the wider implications of those technologies. This segment will comprehensively evaluate both the benefits and drawbacks, thinking about elements including greenhouse fuel emissions, resource utilization, and ecological considerations.

5.1. Environmental Benefits

Greenhouse Gas Emission Reduction: One of the number one environmental advantages of carbon seize to gasoline technologies lies in their potential to lessen net greenhouse fuel emissions. By utilizing captured carbon to supply artificial fuels or hydrogen, those tactics have the potential to offset emissions from traditional fossil fuel assets.

5.2. Resource Efficiency

The conversion of captured carbon into gasoline represents a shape of recycling carbon, imparting an alternative to extracting new carbon from fossil fuels. This aid efficiency thing can make a contribution to sustainable practices and a round carbon economic system.

5.3. Air Quality Improvement

Depending on the utility, the manufacturing and utilization of synthetic fuels or hydrogen may additionally result in lower air pollutant emissions compared to traditional combustion processes, contributing to advanced neighborhood air quality [11].

5.4. Environmental Drawbacks

Energy Intensity: The strength requirements for carbon seize, conversion, and gas synthesis methods may introduce a new set of environmental demanding situations. Depending on the electricity sources utilized, these procedures should make a contribution to improved demand for power and, in flip, environmental influences associated with strength manufacturing.

5.4.1. Resource Extraction:

The extraction and processing of raw materials required for the development and maintenance of carbon seize and fuel manufacturing infrastructure may also have associated environmental costs. These can encompass land use modifications, water intake, and different ecological concerns.

5.4.2. Emissions all through Production

The manufacturing of artificial fuels or hydrogen may also entail emissions, depending on the precise conversion methods hired. Assessing the lifestyle cycle emissions is important to know the general environmental impact of those technologies [12].

5.5. Ecological Considerations

Land Use and Biodiversity: The enlargement of carbon capture to gasoline infrastructure may additionally affect land use patterns, potentially leading to habitat disruption and biodiversity loss. Careful consideration of these ecological aspects is essential in ensuring the general sustainability of such tasks.

5.6. Water Usage

Depending on the approaches involved, the water requirements for carbon capture to fuel technologies ought to pose demanding situations, mainly in regions dealing with water scarcity. An evaluation of water usage and ability impacts on neighborhood ecosystems could be addressed in this segment. By analyzing both the tremendous and bad environmental elements, this segment aims to offer a complete knowledge of the environmental impact of converting captured carbon to gasoline. Subsequent discussions will discover capacity mitigation techniques,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

regulatory frameworks, and the function of those technologies in broader sustainability targets.

5.7. Greenhouse Gas Reduction

Discuss the Potential Reduction in Greenhouse Gas Emissions and Other Environmental Considerations. The ability to reduce greenhouse fuel (GHG) emissions sticks out as a key environmental benefit of changing captured carbon to fuel. This subsection will delve into the specifics of how those technologies contribute to mitigating climate trade and discover additional environmental concerns.

5.7.1. GHG Emission Mitigation

The core environmental advantage of carbon seize to gas technologies lies in their ability to mitigate GHG emissions. By capturing and making use of carbon dioxide (CO2) emissions that could in any other case be launched into the surroundings from business procedures or power technology, these technologies contribute to an enormous reduction in internet GHG emissions. The discussion will complicate at the mechanisms through which this mitigation happens, considering the existence cycle of gas manufacturing and consumption Figure 7.

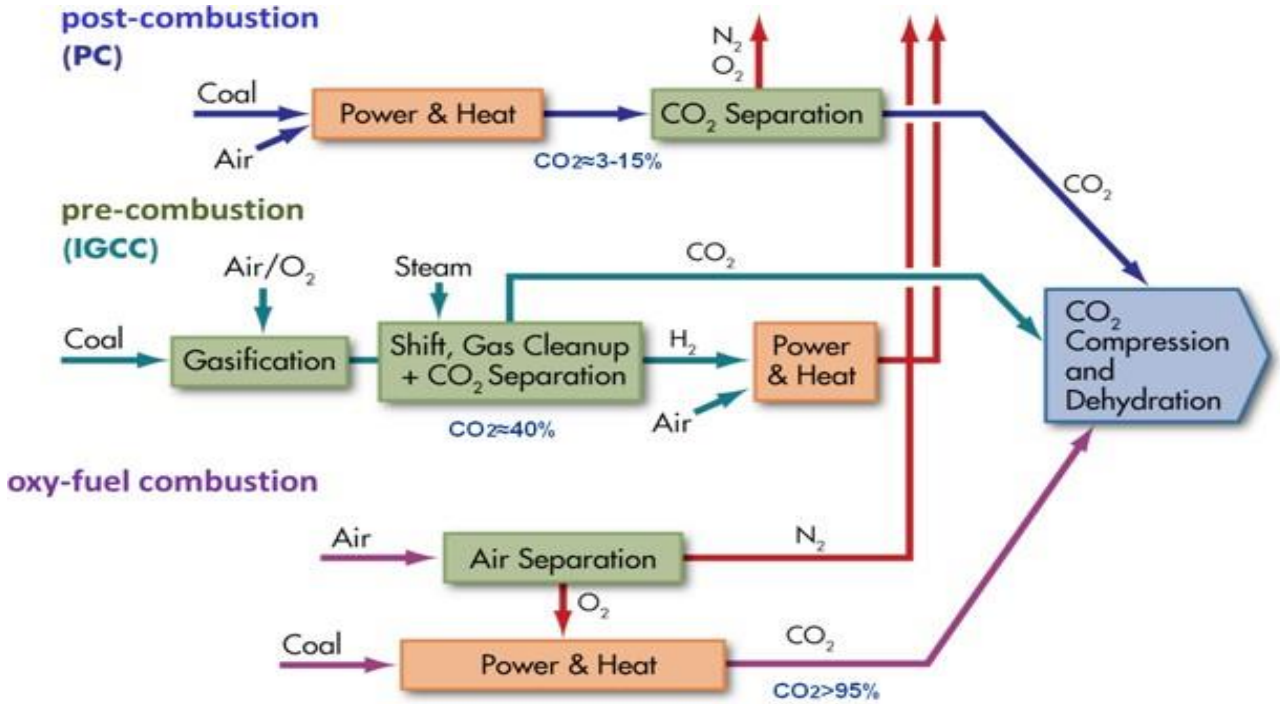


Figure 7 Greenhouse Gas Emission [12]

5.7.2. Carbon Circular Economy

Beyond emission reductions, carbon capture to fuel technologies contribute to the establishment of a carbon circular economy. This involves recycling captured carbon, minimizing reliance on fresh fossil fuel extraction. The utilization of captured CO2 to produce synthetic fuels or hydrogen fosters resource efficiency and aligns with sustainable principles.

5.7.3. Air Quality Improvement

In addition to GHG emissions, the production and use of synthetic fuels or hydrogen may lead to lower emissions of conventional air pollutants compared to traditional combustion processes. This improvement in air quality can have positive effects on human health and ecosystem well-being. The discussion will touch upon the specific pollutants affected and the potential benefits for local air quality.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

5.7.4. Contribution to Decentralized Energy

Carbon capture to fuel technologies can contribute to the decentralization of energy production, promoting distributed energy systems. This decentralized approach may lead to reduced environmental impacts associated with centralized energy production and distribution, such as transmission losses and infrastructure development.

5.7.5. Considerations for Other Greenhouse Gases

Beyond CO2, the discussion will touch upon considerations for other greenhouse gases, such as methane (CH4) or nitrous oxide (N2O), that may be associated with certain industrial processes. Assessing the overall impact on a broader spectrum of greenhouse gases is essential for a comprehensive understanding of the environmental benefits. By exploring these aspects, this subsection aims to provide a nuanced analysis of the potential reduction in GHG emissions and additional environmental considerations associated with converting captured carbon to fuel. Subsequent discussions will address challenges, regulatory frameworks, and the role of these technologies in advancing broader environmental sustainability goals.

6. Comparative Environmental Impact

Compare the Environmental Impact of Carbon Capture to Fuel with Other Carbon Capture Applications. To understand the relative environmental implications of carbon capture to fuel technologies, it is essential to compare them with other prevalent carbon capture applications. This section will assess how the environmental impact of converting captured carbon to fuel aligns with or differs from alternative carbon capture methods.

6.1. Comparison with Carbon Capture and Storage (CCS)

Carbon Capture and Storage (CCS) is a well-established method focused on capturing carbon dioxide (CO2) emissions from industrial processes and power plants and storing them underground. This subsection will delve into the environmental impact of CCS, emphasizing factors such as geological storage safety, potential leakage risks, and the long-term storage implications. A comparative analysis with carbon capture to fuel technologies will highlight the trade-offs and benefits associated with each approach.

6.2. Comparison with Enhanced Oil Recovery (EOR)

Enhanced Oil Recovery (EOR) involves injecting captured CO2 into depleted oil reservoirs to recover additional oil. This subsection will explore the environmental considerations associated with EOR, including the potential for induced seismicity, the long-term fate of stored CO2, and the overall sustainability of the process. A comparative assessment with carbon capture to fuel technologies will shed light on their divergent environmental impacts and associated challenges.

6.3. Comparison with Direct Air Capture (DAC)

Direct Air Capture (DAC) involves capturing CO2 directly from the ambient air. This subsection will discuss the environmental considerations of DAC, such as energy intensity, resource utilization, and the potential for negative emissions. A comparative analysis with carbon capture to fuel technologies will highlight the advantages and drawbacks of each method in the context of environmental impact.

6.4. Life Cycle Assessment (LCA)

A life cycle assessment will be employed to compare the environmental impact of carbon capture to fuel with other carbon capture applications comprehensively. This will involve evaluating the entire life cycle of each technology, from raw material extraction and manufacturing to operation and end-of-life considerations. The LCA approach will provide a holistic perspective on the environmental sustainability of different carbon capture methods.

6.5. Consideration of Co-Benefits

Apart from direct environmental impacts, this section will consider co-benefits associated with each carbon capture application. Co-benefits may include improved air quality, job creation, or

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

positive impacts on local ecosystems. An analysis of these co-benefits will contribute to a more comprehensive understanding of the overall environmental footprint of each technology. By comparing the environmental impact of carbon capture to fuel with other carbon capture applications, this section aims to provide a nuanced assessment of their relative sustainability. Subsequent discussions will explore potential synergies, policy implications, and the broader role of these technologies in addressing climate change and environmental challenges.

7. Case Studies: Opening an Overview of CCUS Technology (CCUS)

7.1. Methanol Examples: Following TEA and LCA Guidelines for CO2 Use

As we start the journey from TEA and LCA to methanol manufacturing, this situation takes a look as it opens a sensible guide. Following the tips of the Global CO2 Initiative, the complexities of changing hydrogen and captured carbon dioxide into methanol are explored. More than just TEA or LCA, it's a treasured lesson at the realistic implementation of the guidelines and provides perception into the challenges and opportunities.

7.2. Mineralization Work Cases: Building a Sustainable Future

In the sphere of creation, this observation makes use of TEA and LCA guidelines to take a look at the mineralization system that produces carbon-containing aggregates. The record introducing the implementation of the guide no longer best explains the environmental effects, but also highlights the training discovered. They observe how carbon capture can promote sustainable construction practices and emphasizes the importance of comprehensive tests.

7.3. OME Case Study: Towards Green Transport

Focusing on Oxyethylene Dimethyl Ether (OME) as a diesel alternative exhibits realistic insights from TEA pointers. In addition to its position as a helping report, this example takes a look at highlights training learned within the search for cleaner transportation solutions. It illustrates how TEA can lead the development of coal use technologies and make them financial and environmental profitability.

7.4. TEA Target Setting Guide: Determining Carbon Use Patterns

This case takes a look at isn't just about TEA goal setting; that is a comprehensive document on the house heating sector. The examination focuses on how carbon dioxide can reduce emissions and is going past the guidelines. Lessons found out encompass the tough balance needed to set targets, a sensitive technique that includes a radical region-precise assessment and an accurate expertise of the potential influences of coal use.

7.5. SNG in Action: Unlocking the Potential of Synthetic Natural Gas

This case is well-known, shows the complexities of synthetic natural fuel (SNG) manufacturing and applies TEA steering to sensible programs. This is a helping record that offers classes found in navigating the complexities of SNG carbon seize. Beyond the specifics of the generation, it sheds mild light at the broader implications and concerns while TEA steering is implemented to various coal usage technology.

7.6. Economic Feasibility of Carbon Dioxide Utilization: Visualizing a Sustainable Chemistry Pathway

U.S. The National Renewable Energy Laboratory's records visualization device isn't always just a presentation of financial analyses. It is a dynamic aid that gives perception into the monetary viability of various chemical compounds crafted from carbon dioxide. The instructions found out increase the capacity of effective communication through visualization that offers stakeholders a clear view of the monetary possibilities of sustainable chemistry.

7.7. Techno-Economic Modeling of Electrochemical Processes: A Blueprint for Future Innovation

This document courses us via electrochemical strategies inside the discipline of CCUS technologies. This file is more than a case look at, it is an outline. In addition to introducing TEA

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

applications, it affords steering and resources that emphasize the non-stop getting to know of techno-monetary modeling of electrochemical pathways. This is a testimony to the evolving panorama of carbon capture technologies and the want for adaptive modeling methods [16].

8. Challenges, Lessons Learned, and Future Directions

8.1. Current Challenges in the Implementation of Carbon Capture to Fuel Technologies

Implementing carbon capture to fuel technologies is not without its hurdles. This section outlines the current challenges and obstacles faced in the practical deployment of these innovative solutions.

8.1.1. Technological Maturity

Many carbons capture to fuel technologies are still in their infancy, lacking the maturity needed for large-scale deployment. The current technological landscape presents challenges related to efficiency, scalability, and reliability [14].

8.1.2. Economic Viability

The economic feasibility of these technologies remains a significant challenge. High initial costs, coupled with uncertainties in revenue streams, pose obstacles to widespread adoption. Bridging the gap between innovation and economic viability is a pressing concern [9]. Policy and Regulatory Frameworks: Inconsistent or insufficient policies and regulations can impede the implementation of carbon capture to fuel technologies. The absence of clear guidelines may hinder investment and industry commitment [13].

8.1.3. Public Acceptance

Public perception and acceptance play a crucial role in the success of any environmental technology. Addressing concerns, raising awareness, and fostering public trust are ongoing challenges in the deployment of carbon capture to fuel solutions.

8.2. Lessons Learned from Previous Initiatives

Drawing insights from past endeavors provides valuable lessons for navigating current challenges. Examining previous experiences in implementing carbon capture to fuel technologies offers a roadmap for future success.

8.2.1. Collaboration is Key

Successful initiatives often involve collaboration between government bodies, industries, and research institutions. Establishing effective partnerships can streamline the development and deployment processes. Early Stakeholder Engagement: Engaging stakeholders early in the development process is critical. Addressing concerns, incorporating feedback, and involving the community in decision-making contribute to smoother project execution. Flexibility in Technology Adoption: Recognizing the dynamic nature of technology, a flexible approach to adoption is essential. Adaptable strategies that accommodate evolving scientific and economic landscapes enhance resilience against unforeseen challenges.

8.3. Future Directions for Research and Development

To overcome current challenges and propel the implementation of carbon capture to fuel technologies, this section proposes future directions for research and development in the field.

8.3.1. Technological Advancements

Invest in research to enhance the efficiency, robustness, and scalability of carbon capture to fuel technologies. Innovations in materials, processes, and integration methods can accelerate technological maturity.

8.3.2. Economic Incentives and Policies

Advocate for the development and enhancement of supportive economic incentives and policies. Clear regulatory frameworks and financial support mechanisms can stimulate private and public investment.

8.3.3. Public Awareness Campaigns

Launch comprehensive public awareness campaigns to educate communities about the benefits and safety of carbon capture to fuel technologies. Building public support is crucial for overcoming social and regulatory challenges.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

8.3.4. Interdisciplinary Research Collaboration

Foster interdisciplinary collaboration between scientists, engineers, economists, and policymakers. Integrating diverse perspectives can lead to holistic solutions and expedite the development process. In navigating the challenges, learning from past experiences, and charting a course for future research, the implementation of carbon capture to fuel technologies can move closer to achieving its potential as a transformative solution in the fight against climate change.

9. Policy and Regulatory Landscape

9.1. Examination of Existing Policies and Regulations

A thorough examination of existing policies and regulations surrounding carbon capture and the production of fuels from captured carbon is crucial for understanding the current policy landscape.

9.1.1. Carbon Capture Policies

Assess the presence and effectiveness of policies specifically targeting carbon capture technologies. Identify regulatory frameworks that encourage or mandate the integration of carbon capture solutions across various industries.

9.1.2. Fuel Production Regulations

Examine regulations governing the production of fuels from captured carbon. This includes understanding standards, safety protocols, and environmental regulations applicable to synthetic fuel production. Environmental Compliance: Evaluate how existing environmental regulations address the potential environmental impacts of carbon capture to fuel technologies. Considerations may include air quality standards, water usage limits, and waste disposal guidelines.

9.2. Role of Government Incentives

Government incentives play a pivotal role in shaping the adoption and development of carbon capture to fuel technologies. This section explores the impact of such incentives on the industry.

9.2.1. Financial Support Programs

Analyze government-funded programs that provide financial support for research, development, and implementation of carbon capture to fuel technologies. Assess the effectiveness of grants, subsidies, and tax credits in stimulating industry participation.

9.2.2. Carbon Pricing Mechanisms

Investigate the role of carbon pricing mechanisms in incentivizing the adoption of carbon capture technologies. Examine how pricing signals influence investment decisions and contribute to the economic viability of carbon capture to fuel projects.

9.2.3. Technology-Specific Incentives

Identify incentives tailored to the unique characteristics of carbon capture to fuel technologies. This may include targeted funding for pilot projects, grants for technology demonstration, and rewards for achieving specific milestones in fuel production [17].

9.3. International Agreements and Collaborations

The global nature of climate change necessitates international collaboration. This section explores the role of international agreements and collaborations in promoting or hindering the advancement of carbon capture to fuel technologies.

9.3.1. Paris Agreement Commitments

Analyze how countries' commitments under the Paris Agreement influence their support for carbon capture technologies. Explore the alignment of national policies with international climate goals and targets.

9.3.2. Technology Transfer and Knowledge Sharing:

Investigate international collaborations focused on technology transfer and knowledge sharing in the field of carbon capture. Assess the impact of collaborative efforts on accelerating the global deployment of carbon capture to fuel technologies.

9.3.3. Trade Agreements

Explore how trade agreements address the exchange of carbon capture technologies and synthetic fuels between countries. Assess whether trade agreements facilitate or hinder the global

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

adoption of these technologies in Figure 8. By comprehensively examining existing policies, government incentives, and international collaborations, this section provides insights into the regulatory environment shaping the development and deployment of carbon capture to fuel technologies. Subsequent discussions will explore potential policy improvements, global cooperation strategies, and the role of these technologies in achieving broader climate objectives.

Element	Details
Policy Path	In the near term, CCUS policy should be carbon credits or incentives based, to seed and promote the CCUS sector in India through tax and cash credits. Over time (probably beyond 2050), the policy should transition to carbon taxes, so as to enable reaching India's net zero goals by 2070. The policy should establish early stage financing and funding mechanisms for CCUS projects.
Hub & cluster business model	Regional hub & cluster models need to be established to drive economies of scale. The role of emitters, aggregators, hub operator, disposers, and conversion agents needs to be defined.
Low carbon products	Preferential procurement in Government tenders for low carbon or carbon abated products. Incentives to foster innovation for low carbon products through schemes like PLI.
Environmental and social justice	Distribution of benefits of economic value added created to communities most affected by environmental and climate change. Protection of communities and jobs, especially in sectors affected by clean energy regulations.
Accounting and regulatory framework	Regulated emission levels and allowances for different sectors. Adoption of Life Cycle Analysis (LCA) framework to take into account Scope 2 and Scope 3 emissions and drive effective carbon abatement.
Risk mitigation	Limiting the CO_2 liability, and ownership of participants across the CCUS value chain. Monitoring, Verification, and Accounting (MVA) framework and monitoring for risk management.

Figure 8 Key Elements of A CCUS Policy Framework for India [18]

Conclusion

In conclusion, the exploration of carbon capture to fuel technologies unveils a promising frontier in the pursuit of sustainable energy solutions and climate change mitigation. The diverse facets covered in this research paper, ranging from technological intricacies to policy landscapes, collectively contribute to a nuanced understanding of the challenges and opportunities inherent in this field.

Technological Advancements and Challenges: The in-depth examination of various carbon capture technologies, their conversion into fuels, and the associated chemical processes has revealed the strides made in technological innovation. Yet, the journey is not without challenges. The sector grapples with the maturation of these technologies, economic viability concerns, and the need for scalable, efficient solutions.

Economic and Environmental Considerations: Assessing the economic feasibility and environmental impact of carbon capture to fuel technologies has underscored the importance of a balanced approach. The economic landscape poses hurdles, from high implementation costs to uncertainties in revenue streams, necessitating innovative solutions. Concurrently, the environmental benefits, including the potential reduction in greenhouse gas emissions and resource efficiency, emphasize the positive role these technologies can play in addressing climate change.

Policy Dynamics and International Collaboration: The examination of the policy and regulatory landscape has shed light on the

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

crucial role governments play through incentives and regulations. International collaborations and agreements further underscore the interconnectedness of global efforts in advancing carbon capture technologies. Striking a harmonious balance between national policies and international commitments emerges as a critical factor in fostering a conducive environment for the widespread adoption of carbon capture to fuel solutions.

Lessons Learned and Future Directions: The case studies presented, drawn from real-world applications, provide valuable insights and lessons for the future. Collaborative efforts, early stakeholder engagement, and flexibility in technology adoption have emerged as pivotal lessons. Looking ahead, a call for continued technological advancements, robust economic incentives, and increased public awareness becomes imperative to overcome current challenges and drive the industry toward maturity.

The field of carbon capture to fuel technologies stands at a crossroads, presenting both challenges and unprecedented opportunities. A concerted effort involving technological innovation, supportive policies, and international collaboration is vital to propel these technologies from theoretical promise to practical, scalable solutions. As we navigate this path, the collective commitment to a sustainable, low-carbon future remains the guiding beacon, encouraging researchers, policymakers, and industries to work collaboratively towards a resilient and cleaner energy landscape.

References

[1]. Liang, Zhiwu & Rongwong, Wichitpan & Liu, Helei & Fu, Kaiyun & Gao, Hongxia & Cao, Fan & Zhang, Rui & Sema, Teerawat & Henni, Amr & Sumon, Kazi & Nath, Devjyoti & Gelowitz, don & Srisang, Wayuta & Saiwan, Chintana & Benamor, Abdelbaki & Al-Marri, Mohammed & Shi, Huancong & Supap, Teeradet & Chan, Christine & Tontiwachwuthikul, Paitoon. (2015). Recent progress and new developments in post-combustion carbon-capture technology with amine-based solvents. International Journal of Greenhouse Gas Control. 40. 26-54.

[2]. Jansen, Daniel & Gazzani, Matteo & Manzolini, Giampaolo & van Dijk, Eric & Carbo, Michiel. (2015). Pre-combustion CO2 capture. International Journal of Greenhouse Gas Control. 40. 167-187. 10.1016/j.ijggc.2015.05.028.

[3]. Adamu, Abdullahi & Russo Abegão, Fernando & Boodhoo, Kamelia. (2020). Process intensification technologies for CO2 capture and conversion – a review. BMC Chemical Engineering. 2. 10.1186/s42480-019-0026-4.

[4]. Ketzer, J. M., Iglesias, R. S., & Einloft, S. (2012). Handbook of Climate Change Mitigation. doi=10.1007/978-1-4419-7991-9_37

[5]. Saleh, H. M., & Hassan, A. I. (2023). Green Conversion of Carbon Dioxide and Sustainable Fuel Synthesis. *Fire, 6*(3), 128. https://doi.org/10.3390/fire6030128

[6]. Article Efficient CO2 Utilization via a Hybrid Na-CO2 System Based on CO2 Dissolution Author links open overlay panelChangmin Kim 1 4, Jeongwon Kim 1 4, Sangwook Joo 1, Yunfei Bu 2, Meilin Liu 3, Jaephil Cho 1, Guntae Kim 1 5

[7]. Kevinjeorjios Pellumbi et al, Pushing the Ag-loading of CO2 electrolyzers to the minimum via molecularly tuned environments, Cell Reports Physical Science (2023). DOI: 10.1016/j.xcrp.2023.101746

[8]. Schmelz, W. J., Hochman, G., & Miller, K. G. (2020, August 14). Title of the Article. *Journal Name, volume number*(issue number), page range. https://doi.org/10.1098/rsfs.2019.0065

[9]. David, J., & Herzog, H. (2000, August). The cost of carbon capture. In fifth

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1387 - 1403
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0192

international conference on greenhouse gas control technologies, Cairns, Australia (pp. 13-16).

[10]. Roussanaly, S., Berghout, N., Fout, T., Garcia, M., Gardarsdottir, S., Nazir, S. M., ... & Rubin, E. S. (2021). Towards improved cost evaluation of Carbon Capture and Storage from industry. International Journal of Greenhouse Gas Control, 106, 103263.

[11]. https://www.azocleantech.com/article.aspx?ArticleID=1572

[12]. Llamas, B., Navarrete, B., Vega, F., Rodriguez, E., Mazadiego, L. F., Cámara, Á., & Otero, P. (2016). Greenhouse Gas Emissions – Carbon Capture, Storage and Utilisation. InTech. doi: 10.5772/63154

[13]. https://www.merckgroup.com/en/research/science-space/envisioning-tomorrow/scarcity-of-resources/co2-conversion.html

[14]. David.L.Chandler.https://news.mit.edu/2023/engineers-develop-efficient-fuel-process-carbon-dioxide-1030

[15]. Sifat, Najmus Saquib & Haseli, Yousef. (2019). A Critical Review of CO2 Capture Technologies and Prospects for Clean Power Generation. Energies. 12. 4143. 10.3390/en12214143.

[16]. https://assessccus.globalco2initiative.org/tea/case-studies/

[17]. Labanya Prakash and Shreyans Jain commentary titled 'What role should government play in the development of negative emission technologies?' https://www.lse.ac.uk/granthaminstitute/news/what-role-should-governments-play-in-the-development-of-negative-emission-technologies/

[18]. Niti Aayog Carbon Capture, Utilization and Storage (CCUS) https://www.niti.gov.in/sites/default/files/2022-11/CCUS-Report.pdf



Textual Context guided Vision Transformer with Rotated Multi-Head Attention for Sentiment Analysis

Chhavi Dhiman[1]
Department of Electronics and Communication Engineering, Delhi Technological University
New Delhi, India
chhavi.dhiman@dtu.ac.in

Gaurav Kumar[2]
Department of Electronics and Communication Engineering, Delhi Technological University
New Delhi, India
kr.zgaurav@gmail.com

ABSTRACT

Social media multimodal sentiment analysis has proliferated the research attention of the research community, as it opens up various paradigms for social issues such as cyberbullying, hate speech, healthcare, politics, business analysis and many more. At the same time, it is an open problem to learn the intrinsic representation of multiple modalities in order to identify correlated patterns. In the proposed work a sentiment analysis framework is presented that defines Textual Context guided Vision Transformer with Rotated Multi-Head Attention, in order to exploit correlation between image-text pair and mine rich discriminatory features for multimodal sentiment analysis. A novel Rotated-Multi-head attention mechanism is defined that translates the visual or text embeddings in distinct feature space resulting in Adaptively Rotated Refined Embedding (ARREmb). To exhibit the performance of the proposed work, extensive experiments are carried out on three publicly available datasets-BG, Twitter and MVSA-single dataset, in terms of Precision, Recall, F1-score and accuracy. The experiments support superior performance of the proposed approach by laying out comparison with SOTA followed by ablation study.

CCS CONCEPTS

• **Information systems** → **Multimedia information systems**; • **Computing methodologies** → **Computer vision**.

KEYWORDS

Vision Transformer, Language model, rotated embeddings, textual context guided attention

ACM Reference Format:
Chhavi Dhiman[1] and Gaurav Kumar[2]. 2024. Textual Context guided Vision Transformer with Rotated Multi-Head Attention for Sentiment Analysis. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24 Companion), May 13–17, 2024, Singapore, Singapore.* , 8 pages. https://doi.org/10.1145/3589335.3651968

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
WWW '24 Companion, May 13–17, 2024, Singapore, Singapore
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0172-6/24/05
https://doi.org/10.1145/3589335.3651968

1 INTRODUCTION

In present time, the major mode of communication is online social media platforms, which allow the common user to present their opinions or sentiments with text, images, emojis, videos etc. Sentiment analysis of content generated by common users at social platforms help to organize it and mine useful information in order to stay aligned with the exponentially growing pace of videos and images. It extends multiple solutions and advancements for various applications. It becomes essential that the performed sentiment analysis to be accurate for social good, such as identifying hate speeches made in public, misinformation propagation to harm or defame the personality of the person in the speeches, cyberbullying, mass notion about any new sensation in the market, impact of any political incident on public so on and so forth. Hence, sentiment analysis of multimodal data allows to understand the broader public consensus behind specific topics or social issues i.e., hate speech [8], cyberbullying [19], and misinformation across different online platforms, healthcare [6], politics [20], and business analysis [5] while ensuring unique reflection of the context and realities of the specific domains. Hence, multimodal sentiment analysis has emerged as a significant research problem with a potential to address a variety of societal challenges and promote positive social outcomes by leveraging insights gleaned from the integration of multiple modalities, i.e. image-text, speech-text, visual-text or video-text. Previous multimodal solutions attempted naïve fusion techniques-early fusion[10][17], late fusion [11], while utilising visual-text, speech-text, and visual-text modalities respectively in videos for sentiment score prediction. Such approaches fail to map the interactions between modalities. There exist few attempts [3] [28] that defined image and text interactions for sentiment score prediction using different attention mechanisms. With the recent proliferation in the performance of Transformer structures- BERT [10], Vision Transformers [7] for different applications, some researcher's [4][18] exploited the efficacy of transformers to learn image-text or visual-text pair interactions. Also, the fine-tuning of recent pretrained vision-language models defined with transformer-based encoders- ViLBERT [29], BLIP take much more complicated multimodal interactions into consideration, compared with late fusion or early fusion, have turned out to be a promising solution to learn complex visual-text interactions. While they offer significant advancements over BERT in terms of multimodal understanding, they also come with their own limitations, particularly in terms of model size and computational time, data dependencies, and interpretability. These limitations need to be carefully considered when choosing and deploying these models in real-world applications. Hence, multi-modal sentiment analysis is still an open challenge to present

each modality characteristics uniquely similar to human cognitive systems with practical constrains of implementations in real-time scenario. Therefore, effective text and visual interactions need to be apprehended to bridge this gap. In this direction, the proposed work deals with uniquely defining image-text interactions with a Textual Context guided Vision Transformer with Rotated Multi-Head Attention for Sentiment Analysis. The framework defines a novel adaptively rotated refined embeddings (ARREmb) that translate the text and visual embeddings in distinct feature space to distinctly perform attention on the correlated features of image and text, given by Rotated-Multi-Head Attention (R-MHA) in vision Transformer. Key contributions for the proposed framework are three folds:

- Proposes an efficient Textual Context guided Vision Transformer with Rotated Multi-Head Attention Multimodal Sentiment Analysis framework that learns the image-text corelation using textual context guided attention mechanism.
- Introduces a novel Adaptively Rotated Refined Embedding (ARREmb) representation of features-Mixed Text Embeddings (MTE) TF-IDF, POS, all-MiniLM-L12-v2 Embedding or visual embeddings. Where the amount of rotation is learnt adaptively with θ_t, θ_p, θ_{LM} parameters. The amount of rotation instils distinctness in each feature representations.
- A textual context guided VisTransformer encapsulates the visual scene region semantic features using ARREmb guided attention, that highlights relevant regions of the scene aligned with textual description for sentiment analysis.

Rest of the paper is organized as follows: Section 2 discusses recent existing state-of-the-art in three categories-fusion based approaches, attention-based approaches and transformer-based approaches. In section 2, proposed methodology is discussed in detail. In Section 4, the observed experimental results of the proposed work, are reported followed by section 5, in which conclusion of the proposed work is mentioned

2 RELATED WORKS

2.1 Fusion based approaches

In context of multimodal sentiment analysis, traditional fusion methods can be divided in three categories: early fusion, mid-level fusion and late fusion. Whereas a recent work [11] defined Multi-Model Fusion (MMF), that exploits a multichannel joint fusion model with self-attention to effectively capture the intrinsic visual and textual interactions. To model and learn end-to-end inter and intra-modality dynamics, the work [14], develops a Tensor Fusion Networks. It computes the outer product between unimodal representations. Another work [24] attempted to reduce the computational complexity of tensor-based approach by developing low-rank multimodal-fusion method yielding tensor representations. It decomposed the weights of high-dimensional fusion tensors, reducing redundant information, yet it was difficult to effectively encapsulate intermodal or specific modality dynamics. Another work [13] presented Quantum Interference (QIMF) guided Multimodal decision Fusion to address semantic gaps and model the correlation between the two modalities using density matrices. However, it has limited adaptability of feature differences and suffered with significant feature redundancy. Majumder et al. [26] defined hierarchical-fusion strategy that failed capture intramodal dynamics adequately.

2.2 Attention based approaches

The work [15] defined Attention-Based Modality-Gated Networks (AMGN)—to exploit the correlation between the modalities of images and texts and extract the discriminative features for multimodal sentiment analysis. Another work [9] developed deep multi-level attention network (DMLANet), which exploits the correlation between image and text modalities using bi-attentive visual map along the spatial and channel dimensions to magnify Convolutional neural network representation power.

Whereas, a bidirectional multi-level attention approach [23] mines complementary and comprehensive information between the image modality and text modality for joint visual-textual sentiment classification. Another work [3], utilised text guided regional attention and scene attention for attentive interactive learning of image-text features.

2.3 Transformer based approaches

These days researchers are readdressing their attention in the direction of pre-trained language models like Transformer, BERT, and ResNet as these language models present exceptional feature learning competence. A recent approach [22] defined self-adaptive representation transformer-based learning model (SRLM) using a gated network approach. Another work [4], recently, attempted transformer-based approach to exploit inner hierarchy of the documents for sentiment analysis followed by LIME analysis, while emphasizing the need for explainability of the model.

3 PROPOSED WORK

The proposed framework utilises both visual and text modality to analyse the sentiments $\{S\}$ as positive, negative, and neutral of a data given as $D \in \{V, T : S\}$, that includes visual information $\{V\}$, and its textual description $\{T\}$. Sentiment analysis problem can be defined as: For a given image and text pair $\{V_i, T_i\}$ and its associated sentiment label $\{S_i\}$; $i \in [1, n]$; n no. of samples, the objective is to learn the mapping function $\phi : (V, T) \rightarrow S$ using multimodal training sample set **{V, T: S}**. To learn the contextual semantics of the text and visuals together for efficient sentiment score prediction Multimodal Adaptive Attention based Sentiment Analysis Framework (see Figure 1) is proposed. It utilises TF-IDF embeddings, part-of-speech (POS) embeddings and all-MiniLM-L12-v2 language model-based embeddings of the given text to highlight both semantic and syntactic information in the text. The mixed text embeddings are translated in refined embedding space by adaptively rotating the embeddings by a factor of $\theta_t, \theta_p, \theta_{LM}$. This translation provides more relevant and distinct semantic and syntactic features present in the text, defined as ARREmb. Since, in Image-text pair it is important to have text aligned visual learning to ensure strong features are used for sentiment prediction. Therefore, a textual context guided vision transformer is defined, that utilises Rotary Multi-Head Attention (R-MHA). ARREmb are used for generating K and Q vectors for R-MHA block. The framework technical details are discussed in following subsections: 3.1 and 3.2.

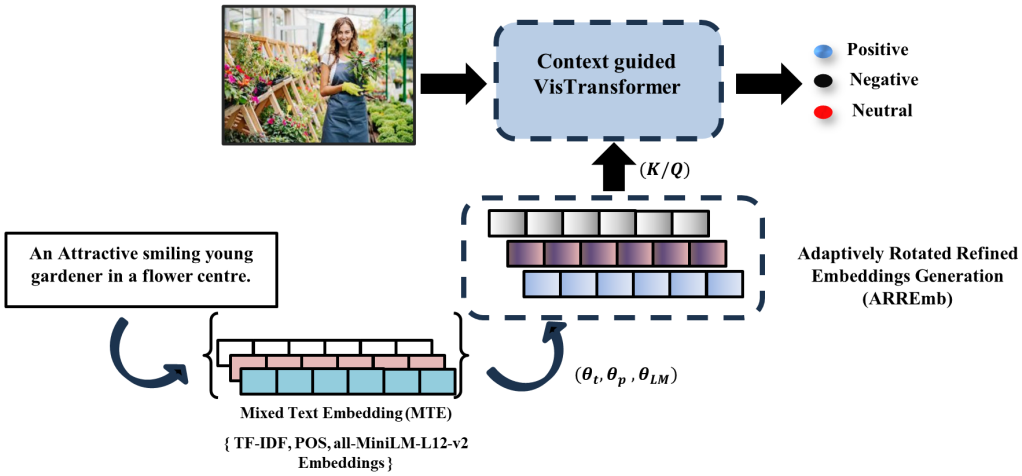


Figure 1: Proposed Multimodal Adaptive Attention based Sentiment Analysis Framework

3.1 Textual Feature Encoding

Before extracting the textual features, essential pre-processing is carried out. It includes case folding, followed by punctuation removal and stop word removal. It cleans the data and makes the features extraction more meaningful. Instead of relying on one type of textual feature representation richer textual features are learnt as a collection of word level embeddings, defined with TF-IDF and POS features- verb (VB), adverb (AVB), noun (NN), proper noun (PN) Adjectives (AD), Negation (NEG) and sentence level features, defined using all-MiniLM-L12-v2 language model. Where POS features presents structure of sentence, term frequency and inverse document frequency (TF-IDF) highlights relatively important words in a sentence, i.e. syntactic details. Language Model all-MiniLM-L12-v2 based features encrypt contextual meaning of the sentence, i.e. semantic details.

In order to refine semantic and syntactic details of textual input, mixed textual embeddings are rotated adaptively. The details are provided in following subsection.

3.1.1 Adaptively Rotated Refined Embeddings Generation (ARREmb). For the given text corpus $T = \{w_i\}, i \in no.of words(N)$, w_i represents i^{th} word in the sentence. It is translated in three types of embeddings space to ensure rich syntactic and semantic feature generation. Part-of-speech features (POS) emphasizes on word level syntactic representation of the given corpus T, such that

$$POS = \{NN, PN, VB, AVB, AD, NEG\} \tag{1}$$

where NN is a collection of set of nouns, i.e., $w_{i,k}^{NN}$, PN is a collection of set of pronouns, i.e., $w_{i,k}^{PN}$, VB is a collection of set of verbs, i.e., $w_{i,k}^{VB}$, AVB is a collection of set of adverbs, i.e., $w_{i,k}^{AVB}$, AD is a collection of set of adjectives, i.e., $w_{i,k}^{AD}$, NEG is a collection of set of negative words, i.e., $w_{i,k}^{NEG}$. Each collection is vectorized using *word-to-vec* embedding representation, given by POS_{Emb}. For sentence level T embedding to encapsulate semantic details a pre-trained all-MiniLM-L12-v2 language model is used to generate LM_{Emb}. . It maps each T_i sentence to a 384-dimensional dense vector space. Combining all three text embeddings a richer mixed text embedding is defined as shown in Eq. (3)

$$MTE = \{POS_{Emb}; TF-TDF_{Emb}; LM_{Emb}\} \tag{2}$$

The position of three embeddings in MTE is given as $m1, m2, m3$. The work [27] exhibited the impact of Rotary Position Embeddings to make the transformer learn better and converge fast. It motivated us to refine the MTE embeddings by applying rotation adaptively. Each embedding in MTE is rotated by an angle $\theta_p, \theta_t, \theta_{LM}$ respectively, learnt adaptively in such a way that in the rotated embedding space the respective embedding distinctly defines the text corpus T, resulting better text context understanding. Mathematically, embedding rotation function is defined as given in Eq. (4).

$$f_{ROT} : \{m1, m2, m3\} \rightarrow \{m1 * \theta_p, m2 * \theta_t, m3 * \theta_{LM}\} \tag{3}$$

$$ARREmb = f_{ROT}(MTE, m_i) = (W_{MTE} MTE_{mi}) e^{jm_i\theta}; \tag{4}$$

where $\theta \in \{\theta_p, \theta_t, \theta_{LM}\}; i \in [1, 3]$

Hence, f_{ROT} maps MTE to Adaptively Rotated Refined Embeddings ($ARREmb$). W_{MTE} are learnt adaptively in training phase.

3.2 Textual context guided Vision Transformer

Vision Transformers [7] learns the visual data by dividing the image in patches and performs multi-head attention layer in the encoding phase. The proposed multimodal sentiment analysis framework learns the visual analytics guided by contextual details of the text. It ensures the sentiment analysis is carried out in the alignment of both text and visual details. Hence, uniquely generated $ARREmb$ are used to guide the vision transformer with its text contextual details. The multi-head self-attention block in the vision transformer first integrates positional details with patch embeddings and converts them into queries, keys, and value representations, given as below:

$$Q_{mi} = K_{mi} = f_{ROT}(MTE, m_i) \tag{5}$$

$$V_j = f_{ROT}(\widetilde{\mu}, j); \tag{6}$$

$\forall j$:positional values of image embeddings $\widetilde{\mu}$

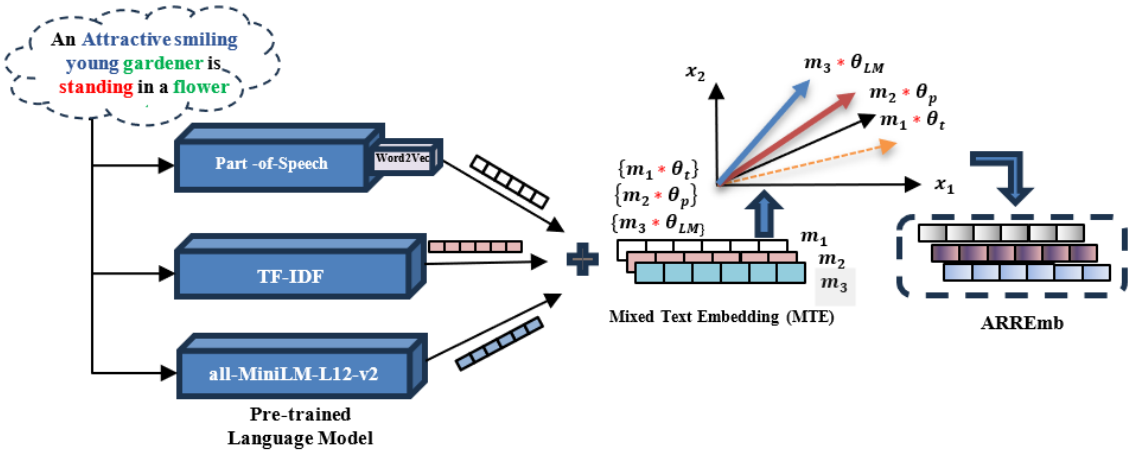


Figure 2: Illustration of Adaptively Rotated Refined Embeddings Generation (ARREmb)

$$V_j = f_{ROT}(\widetilde{\mu}, j) = (W_V, \widetilde{\mu}_j)e^{k(j_* \quad \theta_v)} \tag{7}$$

where θ_v defines adaptively learnt angle of rotation applied on values V_j.

$$ATTn_{mi,j} = \frac{e^{\frac{Q_{mi}^T K_{mi}}{\sqrt{d}}}}{\sum_{l=1}^{N} e^{Q_{mi}^T K_l}} \& O_{mi} = \sum_{j=1}^{L} ATTn_{mi,j} V_j \tag{8}$$

where $L = length\ \widetilde{\mu}$. In Eq. (9), query and key values are used to compute the attention weights $ATTn_{mi,j}$, and the output is computed as the weighted sum over the V_j value representation as O_{mi}. The positional rotated of key, query and values provide refined attention weights in the images.

4 EXPERIMENTAL RESULTS

To observe the performance of the proposed framework, experiments are conducted on three publicly available social media datasets: Getty Images dataset (BG), Twitter Dataset, Multi-View Sentiment Analysis (MVSA) dataset.

Getty Image dataset [21] is a large collection of photographs, videos, and audio from over 477 million resources. It is a user-friendly data resource that provides efficient query-based search engine. All the samples are paired with descriptive image descriptions. The dataset is named "Binary_Getty" (BG), which includes 10,098 positive images, 10,029 negative images with relevant textual explanations, and two labels.

Twitter dataset is a collection of English tweets paired with photos on twitter platform. The samples are collected using Twitter streaming Application Programming Interface (API) [1], using user-generated hashtags as keywords.Based on visual-text information, and the projected rotated sentiment polarity the tweets were manually categorized into negative, positive, and neutral sentiment polarities. The dataset contains 17,073 image-text paired tweets including 6075 positive, 5228 negative, and 5770 neutral samples.

The MVSA-Single dataset [2] includes 5129 image-text pairs extracted from Twitter. Each image-text pair is annotated with one of three polarities (neutral, negative, or positive). The contradicting textual and visual label samples have been deleted. For samples with one polarity as neutral and another as negative/positive, final polarity for the samples is considered as negative/positive. This procedure of filtering results in 4511 text-image pairs with 2683 positive, 1358 negative, and 470 neutral for MVSA-Single dataset. For experimentation the datasets are divides in train-valid-test split in 80-10-10 proportion. Since, MVSA-single dataset is imbalanced dataset, up-sampling is performed to minimize the effect of data imbalance on the smallest category. The architecture is implemented with Keras, and the experiment is performed on Google Colab pro with an Nvidia P100 GPU with 24 GB RAM. Adam optimizer is used for proposed transformer training, while learning rate is set at $10e^{-3}$ with 32 batch size for 80 epochs. Early stopping with patience value of 3 is used to control the training using binary cross-entropy loss. Xavier uniform initializer is used to set the rotation weights, i.e. $W_{MTE}, \theta_p, \theta_t, \theta_{LM}$, while defining *ARREmb*.

To measure the performance of the proposed work precision, recall, F1-score, and accuracy evaluation metrics are used.

The architecture is developed in five steps developing models Model 1 to Model 5, as described below:

- **Model 1 (Visual Data)** is a simple vision transformer pre-trained architecture with MHA. It utilises only images to predict the sentiments.
- **Model 2 (ViLBERT)** [29], the pre-trained model weights are used to learn the visual-text interactions
- **Model 3 (BLIP)** [12]: Bootstrapping Language-Image Pre-training model is used for unified vision-language understanding
- **Model 4 (Visual+MTE-Text)** is an extended version of model 1. Model 4 utilises Mixed Text Embeddings (MTE) to guide the MHA attention block in vision transformer.
- **Model 5 (Visual+ARREmb-Text)** is further extended version of Model 4. It utilises adaptively rotated refined embeddings (ARREmb) for generating K and Q vectors for textual context-based attention in vision transformer.

An ablation study is carried out, as reported in Table 1, to exhibit the effect of proposed text guided vision transformer with Rotated MHA over model 1, 2, 4 and 5. It is observed that the performance

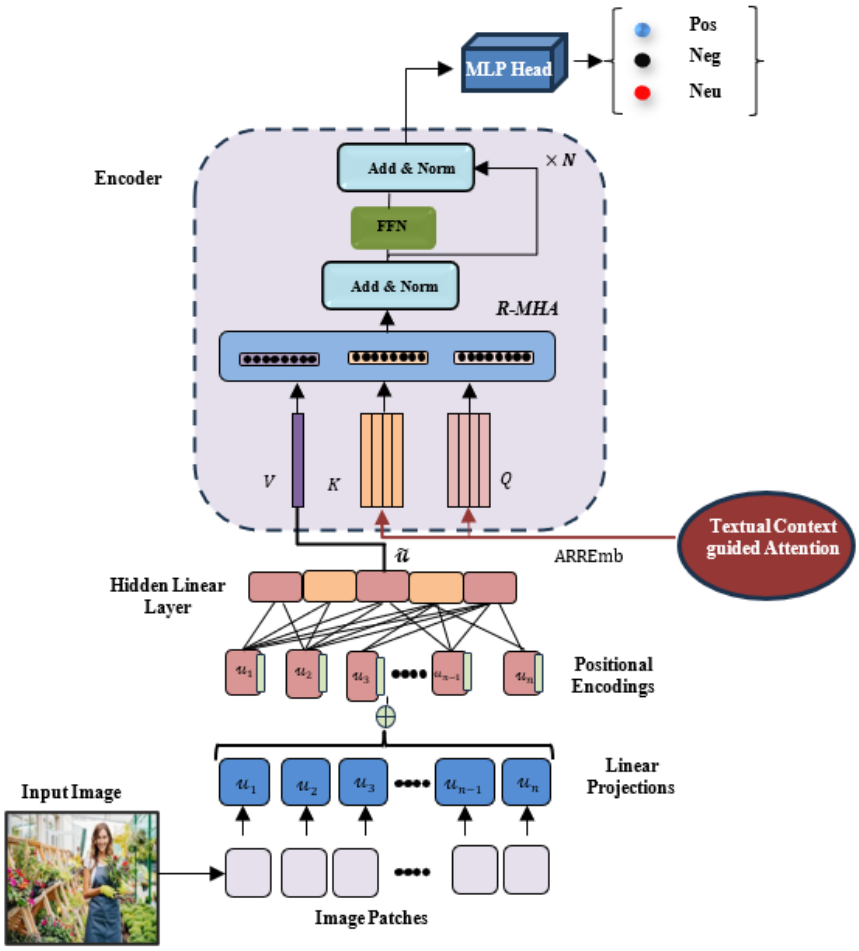


Figure 3: Textual context guided Vision Transformer architecture; R-MHA Rotated Multi-Head

is very naive, exhibiting limited sentiment prediction ability with Model 1 (only visual data). In case of Model 2 and 3, visual-language interaction is encapsulated using pretrained ViLBERT [29] with 110 epochs and BLIP [12] with 95 epochs models weights respectively. It is observed that Model 3 still performs better over Model 2, one of the reasons may be the lesser number of parameters that refrains it from overfitting the data. Whereas, In Model 4, Mixed Text Embeddings (MTE) in textual context guided Vision Transformer learns aligned visual and text interactions, reported as improved results in terms of precision, recall, F-Score and accuracy in Table 1. In case of Model 5, presented ARREmb generates K and Q vectors in multi-head attention, i.e. R-MHA, that projects the K, Q and V vectors in distinct embedding space resulting in refined attention in the visual regions and support enhanced sentiment prediction.

The observed accuracy of the proposed work for BG dataset is 99.80%, for Twitter dataset is 96.39%, and for MVSA dataset is 80.25%. The comparison of the proposed work performance is carried out, in Table 2, with following state-of-the-arts:

(i) MMF [11]: Multi-Model Fusion
(ii) DMVAN [3]: Deep Multi-View Attentive Network
(iii) HCIM [28]: Hierarchical Cross-Modality Interaction Model
(iv) AMGN [15]: Attention-Based Modality-Gated Networks
(v) DMLANet [9]: Deep Multi-Level Attentive network
(vi) HFN [16] :Hybrid Fusion Network
(vii) ITIN [12]: Image-Text Interaction Network
(viii) GLFN [25]: Global Local Fusion Neural Network

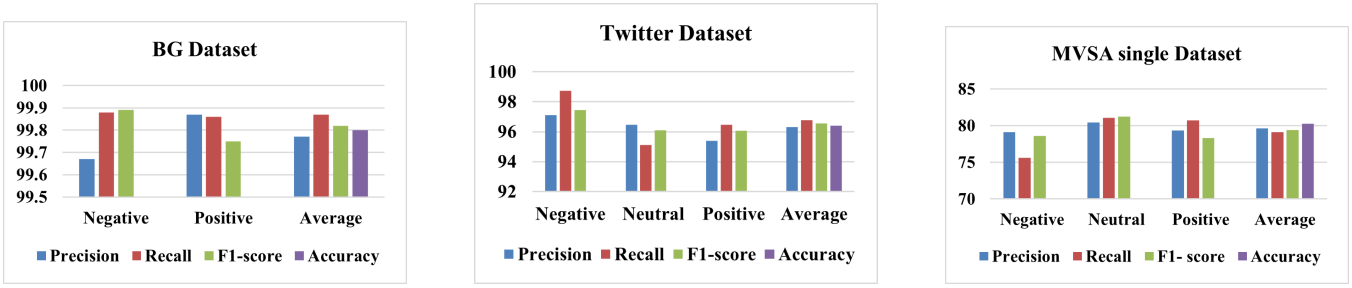
The reported results exhibit that the proposed framework that adaptively learns rotated feature embeddings followed by R-MHA outperforms MMF [11], HCIM [28], AMGN[15] with 0.02%, 6.2% and 11.6% accuracy, whereas DMVAN [3] exhibits 0.03% superior precision and equivalent accuracy for BG dataset. For Twitter dataset, multi-view attention strategy in DMVAN [3] allows it to produce marginally higher accuracy and precision over proposed work by 0.04% and 0.55% respectively. In case of MVSA-single dataset, proposed work exhibit 1.19%, 0.85%, 4.92%, 3.04%, higher accuracy from MMF [11], DMLANet [9], ITIN [12], GLFN [25]. The class wise performance of the proposed work for BG, Twitter and MVSA single dataset is presented in Figure4 respectively.

Table 1: Ablation Study Observations for BG, Twitter dataset and MVSA-single dataset Model 1: VisTranformer+ MHA; Model 2: ViLBERT Model 3: BLIP; Model 4: Text guided VisTranformer+ MHA; Model 5: Text guided VisTranformer+ R-MHA

Dataset		Model 1	Model 2	Model 3	Model 4	Model 5
BG Dataset	Precision	83.77	92.41	93.28	94.44	**99.77**
	Recall	82.59	93.18	94.53	93.51	**99.87**
	F-Score	83.17	92.79	93.90	93.97	**99.82**
	Accuracy	84.52	93.47	94.63	94.55	**99.80**
Twitter Dataset	Precision	78.94	89.45	90.47	91.51	**96.32**
	Recall	80.13	91.45	92.56	92.72	**96.77**
	F-Score	79.53	90.45	91.51	92.11	**96.54**
	Accuracy	80.25	90.41	91.55	92.03	**96.39**
MVSA Single Dataset	Precision	57.73	72.59	73.12	71.65	**79.65**
	Recall	58.31	71.64	72.96	72.91	**79.14**
	F-Score	58.01	72.43	73.15	72.27	**79.39**
	Accuracy	63.22	73.12	75.79	75.32	**80.25**

Table 2: Comparison of the proposed work with SOTA in terms of Precision, Recall, F-Score and Accuracy

Dataset	Methods	Precision	Recall	F-Score	Accuracy
BG Dataset	AMGN[15]	89.8	87.6	88.7	88.2
	HCIM[28]	92.8	93.6	93.2	93.6
	MMF[11]	99.78	99.77	99.78	99.78
	DMVAN[3]	**99.80**	99.80	99.80	99.80
	ours	99.77	**99.87**	**99.82**	**99.80**
Twitter Dataset	MMF[11]	95.70	35.76	95.73	95.70
	DMVAN[3]	**96.87**	**96.95**	**96.91**	**96.86**
	HFN[16]	-	-	68.52	71.35
	ours	96.32	96.77	96.54	96.39
MVSA Single Dataset	MMF[11]	78.10	77.26	77.62	79.06
	ITIN[12]	-	-	73.46	75.33
	GLFN[25]	-	-	76.42	77.21
	DMLANet[9]	**80.8**	78.6	**79.5**	79.4
	ours	79.65	**79.14**	79.39	**80.25**



(a) (b) (c)

Figure 4: Class wise performance of the Proposed work for (a) BG dataset, (b) Twitter dataset, and (c) MVSA single dataset

4.1 Error Analysis

The proposed work mitigates better image-text interactions over SOTA results except DMVAN [9] for BG and Twitter datasets. Whereas, for MVSA single dataset, the proposed framework observed least accuracy as 80.25% over other datasets. One of the reasons may be the annotation dispute on the image-text samples provided in MVSA single dataset, as shown in Figure5(a)-(c). In Figure 5a the caption includes negative terms 'paralyzed', 'rehab'. Whereas the sentence semantics point towards improved state of the baby, hence predicted as positive sentence. Similarly, in Figure

5b image strongly reflects positive sentiment with 'I miss you'. It is repeated two times in the caption. Whereas, the model predicts it as negative sample as the caption ends with '#depression' that echoes the negative sentiment. In Figure 5c the image stands out with neutral sentiment along with the caption, also predicted as neutral sentiment, whereas annotation of the sample is provided as positive. Further, it is found that there are 500 non-English tweets in MVSA-single dataset. Such tweets result in poor sentiment classification.

5 CONCLUSION

Like ViLBERT [29] and BLIP [12] models that learn visual-text interactions, proposed Textual Context guided Vision Transformer with Rotated Multi-Head Attention encapsulates the textual context guided visual sentiments that includes only single Vision Transformer. The proposed Rotated-MHA block projects the key (K), query (Q), and value (V) vectors in distinct embedding space, hence, defining refined attention maps. Textual context guided attention in vision transformer provides relevant text and visual mapping by detecting image regions aligned with text content. The observed results for three publicly available datasets exhibit superior performance of the proposed work, validated in terms of precision, recall, F1- score and accuracy evaluation metrics. The proposed sentiment analysis framework performance is also supported with ablation study and comparisons to other SOTA. This kind of structure may be used for general purposes to learn multi-modal interactions in an efficient way with limited computation demands, since it utilizes only single VisTransformer. Hence, it holds potential real-life applications beyond sentiment analysis such as visual question answering, visual grounding, multimodal document understanding, etc. by leveraging rich textual and visual interactions.

REFERENCES

[1] [n. d.]. Royalty-Free Stock Photos, Creative Images & Vectors | News, Fashion, and Entertainment imagery - Getty Images. https://www.gettyimages.in/

[2] [n. d.]. Use Cases, Tutorials, & Documentation. https://developer.twitter.com/en

[3] Israa Khalaf Salman Al-Tameemi, Mohammad-Reza Feizi-Derakhshi, Saeid Pashazadeh, and Mohammad Asadpour. 2023. Interpretable Multimodal Sentiment Classification Using Deep Multi-View Attentive Network of Image and Text Data. *IEEE Access* (2023), 91060–91081. https://doi.org/10.1109/ACCESS.2023.3307716

[4] Luca Bacco, Andrea Cimino, Felice Dell'Orletta, and Mario Merone. 2021. Explainable Sentiment Analysis: A Hierarchical Transformer-Based Extractive Summarization Approach. *Electronics* 18 (2021). https://doi.org/10.3390/electronics10182195

[5] Samuel W.K. Chan and Mickey W.C. Chong. 2017. Sentiment analysis in financial texts. *Decision Support Systems* (2017), 53–64. https://doi.org/10.1016/j.dss.2016.10.006

[6] Kerstin Denecke and Yihan Deng. 2015. Sentiment analysis in medical settings: New opportunities and challenges. *Artificial Intelligence in Medicine* 1 (2015), 17–27. https://doi.org/10.1016/j.artmed.2015.03.006

[7] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *CoRR* (2020). arXiv:2010.11929 https://arxiv.org/abs/2010.11929

[8] Soumitra Ghosh, Asif Ekbal, Pushpak Bhattacharyya, Tista Saha, Alka Kumar, and Shikha Srivastava. 2023. SEHC: A Benchmark Setup to Identify Online Hate Speech in English. *IEEE Transactions on Computational Social Systems* 2 (2023), 760–770. https://doi.org/10.1109/TCSS.2022.3157474

[9] Feiran Huang, Kaimin Wei, Jian Weng, and Zhoujun Li. 2020. Attention-Based Modality-Gated Networks for Image-Text Sentiment Analysis. *ACM Trans. Multimedia Comput. Commun. Appl.* 3, Article 79 (jul 2020), 19 pages. https://doi.org/10.1145/3388861

[10] Ju Huang, Pengtao Lu, Shuifa Sun, and Fangyi Wang. 2023. Multimodal Sentiment Analysis in Realistic Environments Based on Cross-Modal Hierarchical Fusion Network. *Electronics* 16 (2023). https://doi.org/10.3390/electronics12163504

[11] Saeed Pashazadeh Mohammad Asadpour Israa K. Salman Al-Tameemi, Mohammad-Reza Feizi-Derakhshi. 2023. Multi-Model Fusion Framework Using Deep Learning for Visual-Textual Sentiment Classification. *Computers, Materials & Continua* 2 (2023), 2145–2177. https://doi.org/10.32604/cmc.2023.040997

[12] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. In *Proceedings of the 39th Int'l Conf. on Machine Learning (Proceedings of Machine Learning Research)*, Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (Eds.). PMLR, 12888–12900. https://proceedings.mlr.press/v162/li22n.html

[13] Zhun Liu, Ying Shen, Varun Bharadhwaj Lakshminarasimhan, Paul Pu Liang, Amir Zadeh, and Louis-Philippe Morency. 2018. Efficient Low-rank Multimodal Fusion with Modality-Specific Factors. *CoRR* (2018). arXiv:1806.00064 http://arxiv.org/abs/1806.00064

[14] Jiasen Lu, Dhruv Batra, Devi Parikh, and Stefan Lee. 2019. ViLBERT: Pretraining Task-Agnostic Visiolinguistic Representations for Vision-and-Language Tasks. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2019/file/c74d97b01eae257e44aa9d5bade97baf-Paper.pdf

[15] N. Majumder, D. Hazarika, A. Gelbukh, E. Cambria, and S. Poria. 2018. Multimodal sentiment analysis using hierarchical fusion with context modeling. *Knowledge-Based Systems* (2018), 124–133. https://doi.org/10.1016/j.knosys.2018.07.041

[16] Teng Niu, Shiai Zhu, Lei Pang, and Abdulmotaleb El Saddik. 2016. Sentiment Analysis on Multi-View Social Data. In *MultiMedia Modeling*, Qi Tian, Nicu Sebe, Guo-Jun Qi, Benoit Huet, Richang Hong, and Xueliang Liu (Eds.). Springer Int'l Publishing, Cham, 15–27.

[17] Soujanya Poria, Iti Chaturvedi, Erik Cambria, and Amir Hussain. 2016. Convolutional MKL Based Multimodal Emotion Recognition and Sentiment Analysis. In *2016 IEEE 16th Int'l Conf. on Data Mining (ICDM)*. 439–448. https://doi.org/10.1109/ICDM.2016.0055

[18] Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, and Rada Mihalcea. 2023. Beneath the Tip of the Iceberg: Current Challenges and New Directions in Sentiment Analysis Research. *IEEE Transactions on Affective Computing* 1 (2023), 108–132. https://doi.org/10.1109/TAFFC.2020.3038167

[19] Semiu Salawu, Yulan He, and Joanna Lumsden. 2020. Approaches to Automated Detection of Cyberbullying: A Survey. *IEEE Transactions on Affective Computing* 1 (2020), 3–24. https://doi.org/10.1109/TAFFC.2017.2761757

[20] Ankita Sharma and Udayan Ghose. 2020. Sentimental Analysis of Twitter Data with respect to General Elections in India. *Procedia Computer Science* (2020), 325–334. https://doi.org/10.1016/j.procs.2020.06.038 Int'l Conf. on Smart Sustainable Intelligent Computing and Applications under ICITETM2020.

[21] Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. 2024. RoFormer: Enhanced transformer with Rotary Position Embedding. *Neurocomputing* (2024), 127063. https://doi.org/10.1016/j.neucom.2023.127063

[22] Jie Xu, Feiran Huang, Xiaoming Zhang, Senzhang Wang, Chaozhuo Li, Zhoujun Li, and Yueying He. 2019. Visual-textual sentiment classification with bi-directional multi-level attention networks. *Knowledge-Based Systems* (2019), 61–73. https://doi.org/10.1016/j.knosys.2019.04.018

[23] Ashima Yadav and Dinesh Kumar Vishwakarma. 2023. A Deep Multi-level Attentive Network for Multimodal Sentiment Analysis. *ACM Trans. Multimedia Comput. Commun. Appl.* 1, Article 15 (jan 2023), 19 pages. https://doi.org/10.1145/3517139

[24] Amir Zadeh, Minghai Chen, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. 2017. Tensor Fusion Network for Multimodal Sentiment Analysis. *CoRR* (2017). arXiv:1707.07250 http://arxiv.org/abs/1707.07250

[25] Sun Zhang, Bo Li, and Chunyong Yin. 2022. Cross-Modal Sentiment Sensing with Visual-Augmented Representation and Diverse Decision Fusion. *Sensors* 1 (2022). https://doi.org/10.3390/s22010074

[26] Yazhou Zhang, Dawei Song, Peng Zhang, Panpan Wang, Jingfei Li, Xiang Li, and Benyou Wang. 2018. A quantum-inspired multimodal sentiment analysis framework. *Theoretical Computer Science* (2018), 21–40. https://doi.org/10.1016/j.tcs.2018.04.029 Quantum structures in computer science: language, semantics, retrieval.

[27] Yazhou Zhang, Yang Yu, Mengyao Wang, Min Huang, and M. Shamim Hossain. 2024. Self-Adaptive Representation Learning Model for Multi-Modal Sentiment and Sarcasm Joint Analysis. *ACM Trans. Multimedia Comput. Commun. Appl.* 5, Article 125 (jan 2024), 17 pages. https://doi.org/10.1145/3635311

[28] Tao Zhou, Jiuxin Cao, Xuelin Zhu, Bo Liu, and Shancang Li. 2021. Visual-Textual Sentiment Analysis Enhanced by Hierarchical Cross-Modality Interaction. *IEEE Systems Journal* 3 (2021), 4303–4314. https://doi.org/10.1109/JSYST.2020.3026879

[29] Tong Zhu, Leida Li, Jufeng Yang, Sicheng Zhao, Hantao Liu, and Jiansheng Qian. 2023. Multimodal Sentiment Analysis With Image-Text Interaction Network. *IEEE Transactions on Multimedia* (2023), 3375–3385. https://doi.org/10.1109/TMM.2022.3160060



(a) (b) (c)

Figure 5: Sample images of MVSA-single dataset showing errors

arXiv:2405.15883v1 [gr-qc] 24 May 2024

The estimation of parameters of generalized cosmic Chaplygin gas and viscous modified Chaplygin gas and Accretions around Black Hole in the background of Einstein-Aether gravity

Puja Mukherjee,[1,*] Ujjal Debnath,[1,†] Himanshu Chaudhary,[2,3,4,‡] and G. Mustafa[5,§]

[1] *Department of Mathematics, Indian Institute of Engineering Science and Technology, Shibpur, Howrah-711 103, India.*
[2] *Department of Applied Mathematics, Delhi Technological University, Delhi-110042, India*
[3] *Pacif Institute of Cosmology and Selfology (PICS), Sagara, Sambalpur 768224, Odisha, India*
[4] *Department of Mathematics, Shyamlal College, University of Delhi, Delhi-110032, India.*
[5] *Department of Physics, Zhejiang Normal University, Jinhua 321004, People's Republic of China,*

In this paper, we have investigated the phenomenon of accelerated cosmic expansion in the late universe and the mass accretion process of a 4-dimensional Einstein-Aether black hole. Starting with the basics of Einstein-Aether gravity theory, we have first considered the field equations and two eminent models of Chaplygin gas, viz. generalized cosmic Chaplygin gas model and viscous modified Chaplygin gas model. Then, we obtained the energy density and Hubble parameters equations for these models in terms of some dimensionless density parameters and some unknown parameters. After finding the required parameters, we proceeded with the mass accretion process. For both models, we obtained the equation of mass in terms of the redshift function and represented the change of mass of the black hole graphically with redshift. At the same time, we have made a graphical comparison between the above-mentioned models and the ΛCDM model of the universe. Eventually, we have concluded that the mass of a 4-dimensional black hole will increase along the universe's evolution in the backdrop of Einstein-Aether gravity.

CONTENTS

I. Introduction 1

II. Basics of Einstein-Aether gravity: 4
A. Generalized Cosmic Chaplygin Gas (GCCG) Model: 5
B. Viscous Modified Chaplygin Gas (VMCG) Model: 5

III. Methodology 6
A. Cosmological data sets 6
1. Cosmic chronometers (CC) 6
2. Type Ia supernova (SNIa) 6

IV. Accretions of GCCG and VMCG around Black Hole in Einstein-Aether Gravity: 7
A. Generalized Cosmic Chaplygin Gas (GCCG) Model: 10
B. Viscous Modified Chaplygin Gas (VMCG) Model: 11

V. Results 11

VI. Conclusions 11

VII. ACKNOWLEDGEMENTS 12

References 12

* pmukherjee967@gmail.com
† ujjaldebnath@gmail.com
‡ himanshuch1729@gmail.com
§ gmustafa3828@gmail.com

I. INTRODUCTION

Our universe is such an enigma, whose most of the secrets are still hidden from us. In the last century one of such concealed mystery came into light when with the help of evidences like Chandra X-ray observations [1], Planck observations [2], Wilkinson Microwave Anisotropy Probe (WMAP) prediction [3–5], Baryon acoustic oscillation (BAO) [6, 7], observation of Ia-type supernovae [8–12], Sloan Digital Sky Survey (SDSS) [13, 14], large-scale structure theory [12, 13, 15–18] and observation of cosmic microwave background (CMB) which shows a peak while measuring angular power spectrum on degree scales [19–22] firmly indicated towards the current accelerated rate of expansion of our universe. Among all of the possibilities for such kind of accelerated expansion rate, most of the cosmologist agreed upon an exotic kind of energy named as "dark energy" which possess quite an unique type of properties such as negative pressure and positive energy density. Dark energy has several forms, like a static form commonly known as the cosmological constant [23, 24]. Then comes the ones that changes over time called as a scalar field or quintessence [25–30]. As the time progresses researchers came across many other candidates of dark energy named as phantom energy [31, 32], k-essence [33, 34], quintom [35, 36], dilaton [37], tachyon field [38–40], hessence [41], DBI-essence [42, 43] etc. Some interacting models like holographic dark energy model [44–46] and braneworld model of Universe [47] have been investigated, besides considering the theory of quantum gravity, models like the agegraphic dark energy (ADE) model and the new agegraphic dark energy (NADE) model have been introduced [48, 49].

Chaplygin gas(CG) models are another prominent candidate of dark energy, which is basically a combination of dark matter and dark energy, constructed to measure the acceleration rate of our universe. It has different types such as; pure Chaplygin gas [50–52], more general form named generalized Chaplygin gas (GCG) [53–55], modified version commonly popular as modified Chaplygin gas (MCG) [56, 57], which propagates between dust (or radiation) stage to ΛCDM stage. In 2003, a new model of CG, named as generalized cosmic Chaplygin gas (GCCG) model has been introduced [58] with some striking features like, this model can be made stable and without any unphysical behaviour, even though the vacuum fluid satisfies the condition of phantom energy. For such unique properties, this model can easily overcome the concept that the phantom energy era will finally lead us to Big-Rip as, in finite time, derivative of scale-factor with respect to time gives infinite value. The GCCG model has fewer constraints than GCG and MCG models, so it can adapt to any cosmological framework with just some appropriate choices of parameters. Thus it is more universal in nature as well as in this model black holes does not evaporate to zero masses and Big-Rip singularity can be avoided quite effortlessly. Intriguing by such flexible properties of this model several researches have been conducted on this every now and then [59–66]. Another addition to this list of dark energy models is viscous modified Chaplygin gas (VMCG) model, which have the property of bulk viscosity.Its cosmological modeling has been done in classical and loop quantum cosmology [67], the dynamical behaviour of VMCG has been studied in classical and loop quantum cosmology with the help of some recent observational data [68], the evolution of VMCG in $f(R,T)$ gravity with respect to flat FLRW universe has been studied [69], and it's interaction effects on f(R, $\mathcal{G}$, $\mathcal{T}$) gravity has been investigated [70].

Apart from this, scientists came across another very aspiring theory to explain the current accelerated expansion phase of our universe known to be the modified gravity theory, which stands on the concept that although general relativity works fine for small scales but in cases of larger cosmological distances it needs modification. This theory gives priority to the arise of cosmic acceleration due to the dynamics of modified gravity framework. It has quite a good dynamical approach regarding the alternate of ΛCDM model of the Universe. Most common type of modified gravity is basically the DGP brane-world model of gravity [71], in an alternate way we can change the law of gravity by considering the $f(R)$ theory of action in the place of usual Einstein-Hilbert action [72, 73]. Then there are some other models to explain the concept of the present time acceleration rate of our Universe [74]. This new approach opened several ways for researchers owing to which numerous researches have been carried out on this topic [75–82]. There are other forms of modified gravities including $F(T)$ gravity [83], $F(G)$ gravity [84], Hořava-Lifshitz gravity [78], Brans-Dicke gravity [85], Gauss-Bonnet gravity [81] etc.

Einstein-Aether theory is also among one of these modified gravity theories which has now been a very well known topic as it deals with the violation of Lorentz symmetry principle [76, 86–89].At first Jacobson and Mattingly developed this theory [76, 87] which later has been generalized by Zlosnik et al. [90, 91]. It works on each point of space in preferable timelike direction [92].Nature of singularities of a vacuum black hole solution has been found in this theory [93].Being verified by many observational evidences [94–97], this theory has become quite suitable for further exploration regarding it's other aspects, be it cosmological perturbations [98, 99], it's effects on the gravitational waves [100, 101], analyzing a positive energy theorem [102], finding Aetherizing Lambda by using barotropic fluid as dark energy [103] or some study related to very early universe [104]. Recently analytical static aether solutions and possible range for stability of constant density stars have been obtained [105], paving the way for researchers to study more about black holes in Einstein-Aether gravity framework. Subsequently, studies on spherically symmetric black holes in the non-reduced Einstein–Aether scenarios have been conducted [106], charged black holes in Einstein-Maxwell-aether theory have been investigated [107], n-dimensional charged black holes in Einstein–Aether gravity have been discussed in detail [108], cosmic censorship in Einstein-Aether gravity and as a result the possibility of a collapse in its dynamics leading to unstable behavior around spherical solution has been analyzed [109]. Barausse et al. have examined slowly rotating black holes in the Einstein–Aether gravity scenario, which shows that their solutions do not have any naked singularities [110]. Also, shadows of charged and slowly rotating black holes have been studied in [111]. Many other black hole related researches in Einstein–Aether gravity framework have been conducted on regular intervals [112–123].

On the other hand, in current research world, parameter constraining and observational data analysis have gain immense popularity due to it's close to reality implications. Researchers have observed the process of tallying theoretical methods with various observational databases such as cosmological supernova type Ia data, TORNY and Gold sample databases etc. measured in surveys like- GOODS ACS Treasury program, IfA Deep Survey and several other related research programs [124–127].To continue this legacy, constraining of parameter space for generalized Chaplygin gas(GCG) model taking into consideration the location of CMB radiation peak [128] and using WMAP data [129] have been performed, placement of constraints on GCG model according to type Ia supernovae (SNIa) observations have been

studied in [130]. Several other works on observational constraining of GCG model have been performed as well [131–137]. Also, applying constraining on cosmic microwave background radiation (CMBR) spectrum the range for the value of the parameter in the case of modified Chaplygin gas model has been found in [138] and parameter constraining on variable Chaplygin gas (VCG) model using gold sample database of type Ia supernova has been discussed in [139]. Cosmologists have always been keen about the concept of parameter constraining on different type of Chaplygin gas models. Hence there has been generous amount of such works in the research field of cosmology [140–156].Moving a step ahead researchers took interest on constraining of parameters in different gravities; like Brans-Dicke gravity [157–173], loop quantum cosmology [174, 175], Hořava-Lifshitz gravity [176–178], f(T) gravity [179], Chern-Simons gravity [180], Galileon gravity [181], RS II braneworld gravity [182], Einstein-Aether gravity [183] etc. Some other remarkable works on this topic has been discussed in [184–186]. Motivated by these EoS parameterizations, researchers have also extensively discussed and scrutinized the parametrization of the deceleration parameter, Hubble parameter, and a few other geometrical and physical parameters in the literature. A brief summary can be found in [187–208]. As the Universe evolves, it shifts from a deceleration phase to a late-time acceleration phase. Consequently, any cosmological model must include a transition from deceleration to acceleration to explain the Universe's evolution completely. The deceleration parameter, denoted as $q = -\frac{a\ddot{a}}{\dot{a}^2}$, where $a(t)$ represents the customary scale factor, plays a pivotal role in determining whether the Universe is undergoing acceleration ($q < 0$) or deceleration ($q > 0$). In this paper, we are particularly interested in generalized cosmic Chaplygin gas (GCCG) and viscous modified Chaplygin gas (VMCG) models, regarding which some amazing works have already been conducted in [67, 68, 144, 170, 177, 209].

Mass accretion phenomena have been quite a significant topic for researchers devoted to the studies of black holes. This concept first emerged through the work of Bondi [210]. Then Michel followed this up by considering some polytropic gases and formulating the equations of motion of spherical symmetric steady-state flow of matter to and fro some condensed objects like black holes, neutron stars, etc. [211]. This accretion process gives encouragement to the concept of existence of supermassive black holes at the center of most of the galaxies [212] and it can contribute to some of the energy portion of galaxies and quasars [213]. Accretion of phantom energy onto Schwarzschild black hole which ultimately leads to the Big Rip has been discussed in [214, 215]. Phantom energy onto black holes in oscillating or cyclic universe has been studied in [216]. One by one, accretion of modified variable Chaplygin gas and viscous generalized Chaplygin gas onto black holes [217], accretion of phantom energy onto a stringy magnetically charged black hole [218], accretion process of modified Hayward black hole as well as it's evaporation [219], dark energy accretion onto Kerr–Newman black holes [220], accretion of generalized cosmic Chaplygin gas (GCCG) and new variable modified Chaplygin gas (NVMCG) onto Schwarzschild and Kerr–Newman black holes [61], black holes in expanding universe [221] have been studied. Besides, accretion process of black holes and wormholes have been discussed simultaneously in [222] and accretion of generalized cosmic Chaplygin gas and variable modified Chaplygin gas onto Morris–Thorne wormhole have been illustrated in [223].

Afterward, many pieces of research on mass accretion process of higher dimensional black holes have been evolved in [224–229] along with there have been several remarkable works on mass accretion process of various black holes in the framework of different gravities and models of the universe [229–237]. Greatly inspired by all of these researches, we have analyzed about the generalized cosmic Chaplygin gas (GCCG) and viscous modified Chaplygin gas (VMCG) models on the backdrop of Einstein-Aether gravity. Even though we are familiar with the concept that modified gravity is an alternative to dark energy, and that is the reason for not considering any other dark energies in the case of Einstein-Aether gravity as it may itself initiate dark energy [80, 238]. However in our work we have exempted this hypothesis and consider the above mentioned dark energies from our side. In Sec.II, we have discussed the basic idea of Einstein-Aether gravity. In Sec.II A, we have considered the GCCG model of dark energy and obtained its energy density equation as well as the equation of the Hubble parameter associated with it in the form of some dimensionless density parameters. In Sec.II B we have done the same for the VMCG model of dark energy. In Sec.III, we have conducted the data analysis for the above two discussed models and elaborated the outcomes graphically. In Sec.IV, we have studied the accretion process of a 4-dimensional Einstein-Aether black hole in the case of GCCG and VMCG models of dark energy and portrayed the changes of mass of the black hole with respect to redshift function graphically first in the case of GCCG induced universe then in the case of VMCG induced universe and lastly shown the comparison of the same graphically for GCCG filled universe, VMCG filled universe and ΛCDM model of the universe side by side. Then, we discussed the results of data analysis and mass accretion process and explained the corresponding graphs in a detailed manner in Sec.V. Finally we have summarized all of our findings in the Sec.VI as a conclusion of this work.

II. BASICS OF EINSTEIN-AETHER GRAVITY:

The equation for describing the action of the Einstein-Aether gravity theory coupled with the usual Einstein-Hilbert part action is given by the following form [80, 90, 183]:

$$S = \int d^4x\sqrt{-g}\left[\frac{R}{16\pi G} + \mathbb{L}_V + \mathbb{L}\right], \tag{1}$$

where $\mathbb{L}_V$ is the Lagrangian density of the vector field and $\mathbb{L}$ represents the same for all of the other matter fields. The Lagrangian density related to the vector field is given by [80, 90, 183]:

$$\mathbb{L}_V = \frac{m^2}{16\pi G}E(K) + \frac{1}{16\pi G}\lambda(B^iB_i + 1), \tag{2}$$

$$K = m^{-2}K^{ij}_{kl}\nabla_iB^k\nabla_jB^l, \tag{3}$$

$$K^{ij}_{kl} = e_1g^{ij}g_{kl} + e_2\delta^i_k\delta^j_l + e_3\delta^i_l\delta^j_k, \tag{4}$$

Here, e_i are dimensionless constants, and m is a coupling constant with mass dimension. λ serves as a Lagrange multiplier enforcing the unit constraint for the time-like vector field. B^i represents a contravariant vector, g_{ij} is the metric tensor, and $E(K)$ denotes an arbitrary function of K. Eqn (1) yields the field equations [183]:

$$G_{_{ij}} = T^V_{ij} + 8\pi GT_{ij}, \tag{5}$$

and

$$\nabla_i(E'Q^i_j) = 2\lambda B_j, \tag{6}$$

Where,

$$E' = \frac{dE}{dK} \quad and \quad Q^i_j = 2K^{il}_{jk}\nabla_lB^k \tag{7}$$

Where, T_{ij} is given by [80, 183]:

$$T_{_{ij}} = (\rho + p)\mathfrak{u}_i\mathfrak{u}_j + pg_{_{ij}}, \tag{8}$$

which gives us the expression of energy-momentum tensor for all matter fluids. Here, p and ρ stand for all matter fluids' respective pressure and density. Also, $\mathfrak{u}_i = (1, 0, 0, 0)$ denotes the vector of fluid 4-velocity. Again, T^V_{ij} stands for the energy-momentum tensor of vector field whose expression is given as follows [80, 183]:

$$\begin{aligned} T^V_{ij} =& \frac{1}{2}\nabla_l\left[\left(Q_{(i}{}^lB_{j)} - Q^l{}_{(i}B_{j)} - Q_{(ij)}B^l\right)E'\right] \\ &- C_{(ij)}E' + \frac{1}{2}g_{ij}m^2E + \lambda B_iB_j, \end{aligned} \tag{9}$$

together with

$$C_{ij} = -e_1[(\nabla_lB_i)(\nabla^lB_j) - (\nabla_iB_l)(\nabla_jB^l)], \tag{10}$$

Here, (ij) subscript shows the symmetricity property of the indices involved. Also, $B^i = (1, 0, 0, 0)$ represents nonvanishing time-like unit vectors that follow the relation: $B^iB_i = -1$. Now, the Friedmann-Robertson-Walker (FRW) metric of the universe given by:

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + sin^2\theta d\phi^2\right)\right], \tag{11}$$

Here, $a(t)$ denotes the scale factor and k is called the curvature scalar, which can take values $-1, 0, 1$. Again, from the above two Eqn.(3) and Eqn.(4) we will have the following one [183]:

$$K = m^{-2}\left(e_1g^{ij}g_{kl} + e_2\delta^i_k\delta^j_l + e_3\delta^i_l\delta^j_k\right) = \frac{3\beta H^2}{m^2}, \tag{12}$$

Here, β is a constant given by the expression: $\beta = e_1 + 3e_2 + e_3$. Also, from the above Eqn.(5) we will have a modified version of Friedmann equations in the background of Einstein-Aether gravity as follows [80, 90, 183]:

$$\beta\left(-E' + \frac{E}{2K}\right)H^2 + \left(H^2 + \frac{k}{a^2}\right) = \frac{8\pi G}{3}\rho \tag{13}$$

and

$$\beta\frac{d}{dt}(HE') + \left(-2\dot{H} + \frac{2k}{a^2}\right) = 8\pi G(\rho + p), \tag{14}$$

Where, H is the Hubble parameter, given by: H ($\equiv \frac{\dot{a}}{a}$). It is quite clear from the above two Eqn.(12) and Eqn.(13) that whenever we put the L.H.S. of the first expressions of both of the Eqn.(12) and Eqn.(13) equals to zero, then we will get the general field equations for Einstein's gravity. Hence we can conclude that the first expression appears due to the Einstein-Aether gravity. Again, the conservation equation is given as follows:

$$\dot{\rho} + 3H(\rho + p) = 0 \tag{15}$$

If we consider that the matter field of our universe is an amalgamation of dark matter and dark energy. Then, the pressure and energy density will be of the forms: $\rho = \rho_{_{DM}} + \rho_{_{DE}}$ and $p = p_{_{DM}} + p_{_{DE}}$ respectively. Again, separate applications of conservation equations on dark matter and dark energy will lead us to the following equations:

$$\dot{\rho}_{_{DM}} + 3H(\rho_{_{DM}} + p_{_{DM}}) = 0 \tag{16}$$

and

$$\dot{\rho}_{_{DE}} + 3H(\rho_{_{DE}} + p_{_{DE}}) = 0. \tag{17}$$

Since dark matter has nearly no pressure so, we can take $p_{_{DM}} = 0$, also by considering the redshift equation: $(1 + z) = \frac{1}{a}$, finally the Eqn.(16) will reduces to the form, given as: $\rho_{_{DM}} = \rho_{_{DM0}}a^{-3}$ which in terms of redshift function z will be of the following form:

$$\rho_{_{DM}} = \rho_{_{DM0}}(1 + z)^3. \tag{18}$$

The term $\rho_{_{DM0}}$ gives the present energy density value of dark matter.

A. Generalized Cosmic Chaplygin Gas (GCCG) Model:

Let us consider GCCG to be a candidate of dark energy in this Einstein-Aether gravity framework.The equation of state for GCCG model is given by [58]:

$$p_{DE} = -\rho_{DE}^{-\alpha}\left[\mathcal{C} + (\rho_{DE}^{1+\alpha} - \mathcal{C})^{-\omega}\right], \tag{19}$$

Here, $\mathcal{C} = \frac{\mathcal{B}}{1+\omega} - 1$ and $-l < \omega < 0$ with l being a constant having value greater than one. Also, α and $\mathcal{B}$ are any two constants. It is quite clear from the above Eqn.(19) that whenever $\omega \to 0$, it will reduces to the equation of state for generalized Chaplygin gas(GCG). Again, whenever $\omega \to -1$, it will give us the de Sitter fluid. Now, using Eqn.(17) and Eqn.(19) together, we will have the energy density equation of GCCG as follows [59, 170]:

$$\rho_{DE} = \left\{\mathcal{C} + \left(1 + \mathcal{D}\, a^{-3(1+\alpha)(1+\omega)}\right)^{\frac{1}{1+\omega}}\right\}^{\frac{1}{1+\alpha}}, \tag{20}$$

where, $\mathcal{D}$ represents a constant. We can write the above Eqn.(20) in the following form [170]:

$$\rho_{DE} = \rho_{DE0}\left\{\mathcal{A} + (1-\mathcal{A})\left(\mathcal{A}_s + (1-\mathcal{A}_s)(1+z)^{3(1+\alpha)(1+\omega)}\right)^{\frac{1}{1+\omega}}\right\}^{\frac{1}{1+\alpha}}, \tag{21}$$

Here, $\mathcal{A}_s = \frac{1}{1+\mathcal{D}}$, $\mathcal{A} = \left(1 + \mathcal{C}^{-1}\mathcal{A}_s^{-\frac{1}{1+\omega}}\right)^{-1}$, and $\rho_{DE0}^{1+\alpha} = \mathcal{C} + \mathcal{A}_s^{-\frac{1}{1+\omega}}$. Also, z symbolizes the redshift function and ρ_{DE0} gives us the present energy density of DE. Now, we can choose different forms of the function $E(K)$ since it is a free function of K. According to [90, 239]; $E(K)$ can be taken as: $E(K) = \gamma(-K)^n$. On the other hand in [80, 238]; $E(K)$ has been taken as: $E(K) = \gamma\sqrt{-K} + \sqrt{\frac{3K}{\beta}}\ \ln(-K)$. But for easier calculation we will take $E(K)$ as [183]: $E(K) = (\frac{2}{\beta})K(1-\epsilon K)$, where ϵ is a constant. So, if we solve Eqn.(13) considering the above form of $E(K)$ then we will have the expression for Hubble parameter in terms of redshift function z as follows:

$$H^2(z) = \frac{m}{3\sqrt{3\epsilon\beta}}\left[8\pi G\rho_{DM0}(1+z)^3 - 3k(1+z)^2 + 8\pi G\rho_{DE0}\left\{\mathcal{A} + (1-\mathcal{A})\left(\mathcal{A}_s + (1-\mathcal{A}_s)(1+z)^{3(1+\alpha)(1+\omega)}\right)^{\frac{1}{1+\omega}}\right\}^{\frac{1}{1+\alpha}}\right]^{\frac{1}{2}}, \tag{22}$$

Again, if we define the dimensionless parameters in the following forms: $\Omega_{EA} = \frac{m}{3H_0\sqrt{\epsilon\beta}}$, $\Omega_{k0} = \frac{k}{H_0^2}$, $\Omega_{m0} = \frac{8\pi G\rho_{DM0}}{3H_0^2}$, $\Omega_{DE0} = \frac{8\pi G\rho_{DE0}}{3H_0^2}$, the above Eqn.(22) will reduces to the form given as follows:

$$H(z) = H_0\sqrt{\Omega_{EA}}\left[\Omega_{m0}(1+z)^3 - \Omega_{k0}(1+z)^2 + \Omega_{DE0}\left\{\mathcal{A} + (1-\mathcal{A})\left(\mathcal{A}_s + (1-\mathcal{A}_s)(1+z)^{3(1+\alpha)(1+\omega)}\right)^{\frac{1}{1+\omega}}\right\}^{\frac{1}{1+\alpha}}\right]^{\frac{1}{4}}, \tag{23}$$

Also, the above Eqn.(23) and Eqn.(13) will together leads to the following relation:

$$\sqrt{\Omega_{EA}}\ \left[\Omega_{m0} + \Omega_{DE0} - \Omega_{k0}\right]^{\frac{1}{4}} = 1. \tag{24}$$

Combining the above equations, we can obtain the following:

$$H(z) = H_0\left[\left\{\Omega_{m0}(1+z)^3 - \Omega_{k0}(1+z)^2\right\}\Omega_{EA}^2 + \left\{1 - \Omega_{EA}^2(\Omega_{m0} - \Omega_{k0})\right\}\left\{\mathcal{A} + (1-\mathcal{A})\left(\mathcal{A}_s + (1-\mathcal{A}_s)(1+z)^{3(1+\alpha)(1+\omega)}\right)^{\frac{1}{1+\omega}}\right\}^{\frac{1}{1+\alpha}}\right]^{\frac{1}{4}}. \tag{25}$$

B. Viscous Modified Chaplygin Gas (VMCG) Model:

If we appraised viscous modified Chaplygin gas (VMCG), a generalization of modified Chaplygin gas(MCG), considering the expansion process as of the cumulation of states that are out of thermal equilibrium resulting the bulk viscosity in the Einstein-Aether gravity context.The expression of pressure for VMCG is given by [56, 67, 68]:

$$p_{DE} = A\rho_{DE} - \frac{B}{\rho_{DE}^{\xi}} - 3\zeta_0\sqrt{\rho_{DE}}H, \tag{26}$$

where, A and B are constants.Also, ζ_0 stands for the positive coefficient of bulk viscosity and ξ is any positive valued constant. Substituting p_{DE} from Eqn.(26) into the Eqn.(18), we get:

$$\rho_{DE} = \left(\frac{B}{1 + A - \sqrt{3}\zeta_0} + \frac{C_1}{a^{3(1+\xi)(1+A-\sqrt{3}\zeta_0)}}\right)^{\frac{1}{1+\xi}}, \tag{27}$$

where C_1 represents the constant of integration. The above expression can be further re-written as:

$$\rho_{DE} = \rho_{DE0}\left\{B_s + (1-B_s)(1+z)^{3(1+\xi)(1+A-\sqrt{3}\zeta_0)}\right\}^{\frac{1}{1+\xi}}. \tag{28}$$

where ρ_{DE0} being the energy density value of the dark energy at the present epoch, $B_s = \frac{B}{(1+A-\sqrt{3}\zeta_0)C_1+B}$ satisfying the conditions $0 < B_s < 1$ and $1+A-\sqrt{3}\zeta_0 > 0$, and $\rho_{DE0}^{1+\xi} = \frac{(1+A-\sqrt{3}\zeta_0)C_1+B}{1+A-\sqrt{3}\zeta_0}$. From the above equations, we can obtain the following:

$$H(z) = H_0\Big[\left\{\Omega_{m0}(1+z)^3 - \Omega_{k0}(1+z)^2\right\}\Omega_{EA}^2 + \left\{1-\Omega_{EA}^2(\Omega_{m0}-\Omega_{k0})\right\}\left\{B_s + (1-B_s)(1+z)^{3(1+\xi)(1+A-\sqrt{3}\zeta_0)}\right\}^{\frac{1}{1+\xi}}\Big]^{\frac{1}{4}}. \tag{29}$$

III. METHODOLOGY

In cosmology, parameter estimation typically follows a Bayesian framework, where the goal is to compute the posterior distribution of parameters θ given observed data D. This is mathematically expressed as:

$$P(\theta \mid D) = \frac{\mathcal{L}(D \mid \theta)P(\theta)}{P(D)} \tag{30}$$

Here, $\mathcal{L}(D \mid \theta)$ represents the likelihood function, $P(\theta)$ is the prior distribution of the parameters, and $P(D)$ is the marginal likelihood. The posterior distribution is explored through the parameter space θ, often using algorithms such as Metropolis-Hastings [240]. This algorithm helps a random walker navigate the parameter space, giving preference to regions with higher likelihood values. To estimate the parameters, we determine the mean and error of the distribution by analyzing the regions where the walker frequently visits and its deviations within the parameter space. By allowing the walker to take numerous steps, we can approximate the shape of the posterior distribution. For Bayesian model selection, we calculate and compare the evidence for each model. This process typically demands significant computational resources, often utilizing nested sampling methods. However, when the posteriors resemble a Gaussian distribution, simpler tools like information criteria can offer reliable model preferences [241]. Our study explores new parameterizations for the dark energy equation of state (EoS) within both Einstein's gravity and Horava-Lifshitz gravity frameworks, determining the optimal values of the free parameters in each. Utilizing Markov Chain Monte Carlo (MCMC) analysis with the Polychord package facilitates efficient exploration of the parameter space and provides reliable estimates [242].

A. Cosmological data sets

For our Markov Chain Monte Carlo (MCMC) simulations, we utilize a diverse set of cosmological data. This includes data from Cosmic Chronometers [243–248] and Supernovae [249]

1. *Cosmic chronometers (CC)*

Cosmic chronometers are valuable tools for probing cosmic expansion, providing a model-independent method for determining the Hubble constant and the expansion rate. By analyzing the age and metallicity of nearby passive galaxies, we can estimate the expansion rate $H_{\rm CC}(z)$ at a given redshift $z_{\rm CC}$. This estimation is based on the observation that the expansion rate can be approximated by the ratio of the redshift difference to the time difference, adjusted for the redshift: $H_{\rm CC}(z) \approx -(\Delta z_{\rm CC}/\Delta t)/(1+z_{\rm CC})$. We collect cosmic chronometer data from various studies [243–248], covering the redshift range $0.07 \lesssim z \lesssim 1.97$, as compiled in [241]. These data provide direct constraints on the universe's expansion history. To evaluate how well theoretical predictions align with cosmic chronometer measurements at different redshifts, we calculate the chi-squared statistic $\chi^2_{\rm CC}$:

$$\chi^2_{\rm CC} = \sum_{i=1}^{31}\left(\frac{H(z_{\rm CC}) - H_{\rm CC}(z_{\rm CC})}{\sigma_{\rm CC}(z_{\rm CC})}\right)^2 \tag{31}$$

In this equation, $H_{\rm CC}(z_{\rm CC}) \pm \sigma_{\rm CC}(z_{\rm CC})$ represents the measured expansion rate at redshift $z_{\rm CC}$ along with its associated uncertainty. We assume that the cosmic chronometer observations are uncorrelated.

2. *Type Ia supernova (SNIa)*

We utilize the Pantheon+ compilation supernovae dataset [250] as our primary observational data source. This dataset encompasses 1701 data points collected from 1550 type Ia supernovae, spanning a redshift range of $0.001 \leq z \leq 2.3$. To assess the concordance between our theoretical model and the observed supernovae data, we

employ the χ^2 function, defined as:

$$\chi^2_{\text{Pantheon+}}(\Theta) = \mathbf{F}^T \cdot \mathbf{C}^{-1}_{\text{Pantheon+}} \cdot \mathbf{F} \tag{32}$$

Here, $\mathbf{C}_{\text{Pantheon+}}$ denotes the covariance matrix derived from the Pantheon+ dataset, accounting for both statistical and systematic uncertainties. The vector $\mathbf{F}$ comprises discrepancies between the apparent magnitudes $m_{B,i}$ and the predicted distance moduli μ_{model}, where the latter is determined by a chosen cosmological model:

$$\mu_{\text{model}}(z_i, \Theta) = 5\log_{10} D_L(z_i, \Theta) + 25 \tag{33}$$

The luminosity distance $D_L(z_i, \Theta)$ is computed as an integral of the Hubble parameter $H(z, \Theta)$ over redshift z, where Θ represents the set of cosmological parameters. Additionally, we adjust the vector $\mathbf{F}$ to accommodate the influence of Cepheid hosts on the absolute magnitude M, resulting in the modified vector $\mathbf{F}'$:

$$F'_i = \begin{cases} m_{B,i} - M - \mu_{\text{Ceph},i} & \text{if } i \text{ is in a Cepheid host} \\ m_{B,i} - M - \mu_{\text{model}}(z_i) & \text{otherwise} \end{cases} \tag{34}$$

Here, $\mu_{\text{Ceph},i}$ denotes the distance modulus associated with the Cepheid host of the i-th supernova, independently measured by Cepheid calibrators. Consequently, the modified χ^2 function is expressed as:

$$\chi^2_{\text{SNIa}} = \mathbf{F}'^T \cdot \mathbf{C}^{-1}_{\text{Pantheon+}} \cdot \mathbf{F}' \tag{35}$$

MCMC Results			
Model	Priors	Parameter	Value
ΛCDM Model	H_0	[50.,100.]	$69.854848^{\pm 2.386935}_{\pm 1.259100}$
	Ω_{m0}	[0.,1.]	$0.318654^{\pm 0.028134}_{\pm 0.012822}$
	Ω_{k0}	[-0.1,0.1]	$-0.072537^{\pm 0.025674}_{\pm 0.019007}$
GCCG Model	H_0	[50.,100.]	$69.765455^{\pm 1.517539}_{\pm 0.788664}$
	Ω_{m0}	[0.,0.6]	$0.295293^{\pm 0.371892}_{\pm 0.284314}$
	Ω_{k0}	[0.,0.1]	$0.061630^{\pm 0.056411}_{\pm 0.036983}$
	Ω_{EA}	[0.,1.]	$0.572881^{\pm 0.519862}_{\pm 0.383366}$
	A	[0.1,0.5]	$0.315725^{\pm 0.207903}_{\pm 0.128765}$
	A_s	[0.,0.6]	$0.257420^{\pm 0.240137}_{\pm 0.135803}$
	α	[1.7,2.]	$1.883296^{\pm 0.163223}_{\pm 0.087537}$
	ω	[-0.2,0.]	$-0.082886^{\pm 0.105872}_{\pm 0.064216}$
	M	[-19.2334,0.0404]	$-19.268738^{\pm 0.047949}_{\pm 0.023370}$
VMCG Model	H_0	[50.,100.]	$69.859683^{\pm 1.369140}_{\pm 0.859873}$
	Ω_{m0}	[0.,0.6]	$0.298908^{\pm 0.376904}_{\pm 0.278864}$
	Ω_{k0}	[-0.1,0.1]	$0.004638^{\pm 0.100714}_{\pm 0.064960}$
	Ω_{EA}	[0.,1.]	$0.429626^{\pm 0.401530}_{\pm 0.287205}$
	A	[0.,1.]	$0.778558^{\pm 0.084330}_{\pm 0.043710}$
	B_s	[0.,1.]	$0.485455^{\pm 0.458603}_{\pm 0.282278}$
	ξ	[0.5,1.5]	$1.036391^{\pm 0.463270}_{\pm 0.322349}$
	ζ_0	[0.,1.]	$0.263673^{\pm 0.247089}_{\pm 0.181835}$
	M	[-19.2334,0.0404]	$-19.312621^{\pm 0.039679}_{\pm 0.026045}$

TABLE I: Summary of the MCMC Results.

IV. ACCRETIONS OF GCCG AND VMCG AROUND BLACK HOLE IN EINSTEIN-AETHER GRAVITY:

In this section, we are going to discuss about the mass accretion process of a 4-dimensional Einstein-Aether black hole.In this regard, let us consider a spherically symmetric and static black hole solution which have the following form:

$$ds^2 = -g(r)dt^2 + g^{-1}(r)dr^2 + r^2(d\Theta^2 + sin^2\Theta d\Phi^2). \tag{36}$$

The Einstein-Aether vector is given by [114]:

$$u^\mu(r) = \left(\delta(r) - \frac{\eta(r)}{g(r)}, \eta(r), 0, 0\right). \tag{37}$$

It would be in correspondence with the boundary condition that is the metric given by Eqn.(36) will reduces to that of the Minkowski metric only when as $r \to +\infty$, $u^\mu = (1, 0, 0, 0)$. Depending upon the coupling constants, we can have two types of exact black hole solutions from Einstein-Aether field of equations [92, 251, 252]. Among them the solution which is consistent with Eqn.(1) and represents a non-rotating black hole, is commonly known as first Einstein-Aether black hole solution.In this solution the metric function $g(r)$ is given as follows [107, 112, 114]:

$$g(r) = 1 - \frac{2M}{r} - \left(\frac{27e_{_{13}}}{256(1-e_{_{13}})}\right)\left(\frac{2M}{r}\right)^4, \tag{38}$$

Where, $e_{_{13}} = e_1 + e_3$ and the coupling constants satisfies the following conditions: $e_{_{14}} \equiv e_1 + e_4 = 0$, $e_{_{123}} \equiv e_1 + e_2 + e_3 \neq 0$. Also, the functions $\delta(r)$ and $\eta(r)$ of the Einstein-Aether framework are given by [114, 252]:

$$\delta(r) = \left(\frac{3\sqrt{3}}{16\sqrt{1-e_{_{13}}}}\left(\frac{2M}{r}\right)^2 + \sqrt{1 - \frac{2M}{r} + \frac{27}{256}\left(\frac{2M}{r}\right)^4}\right)^{-1}, \tag{39}$$

and

$$\eta(r) = -\frac{3\sqrt{3}}{16\sqrt{1-e_{_{13}}}}\left(\frac{2M}{r}\right)^2. \tag{40}$$

We will get the Schwarzschild black hole solution whenever $e_{_{13}} \to 0$. Now, to investigate the accretion process of the above type of black hole in Einstein-Aether gravity background, we need to assume the energy-momentum

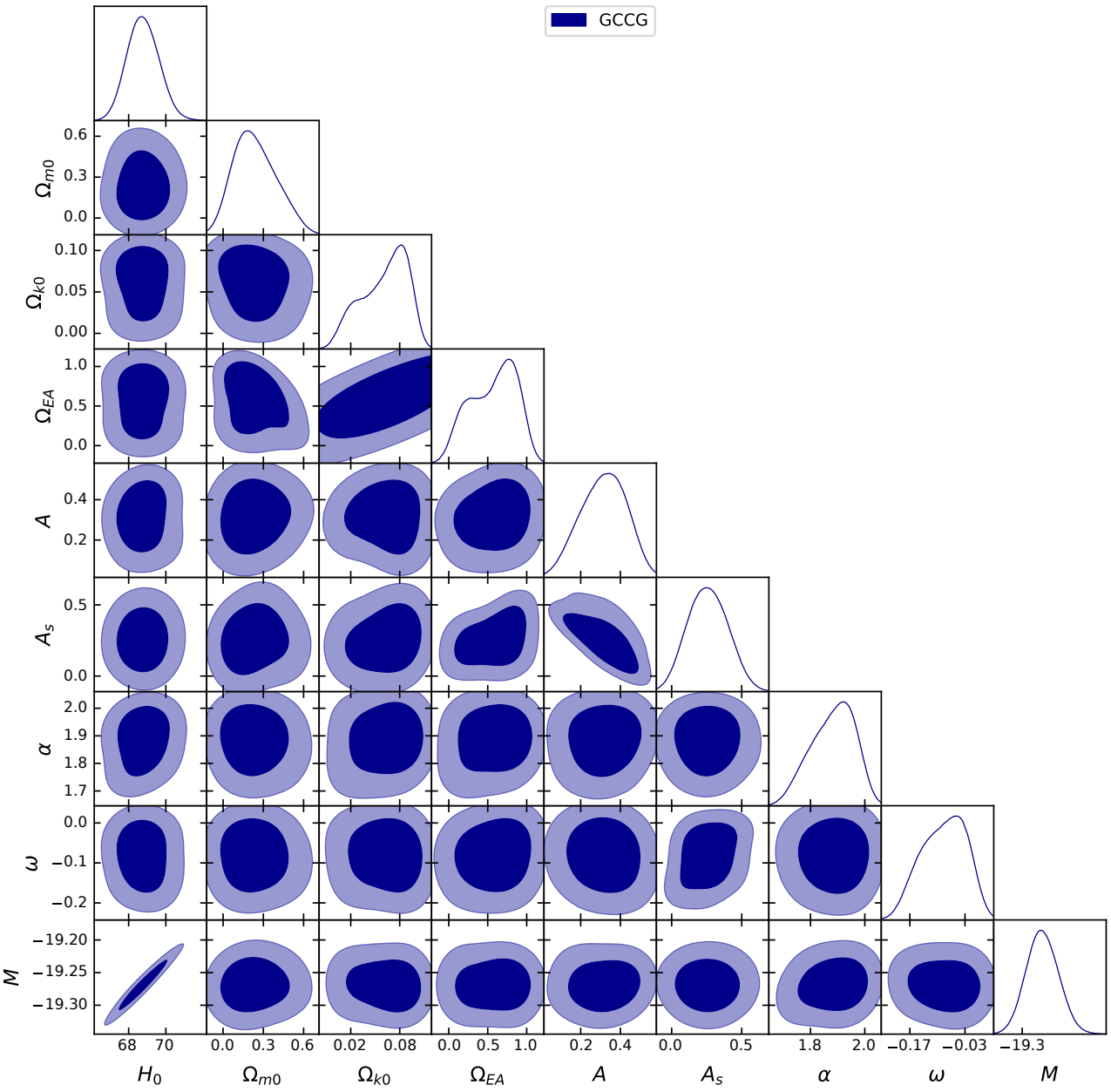


FIG. 1: The figure displays the posterior distribution for various observational data measurements using the Generalized Cosmic Chaplygin Gas Model, highlighting the 1σ and 2σ confidence levels.

tensor for the dark energy in the form of a perfect fluid whose equation of state is $p = p(\rho)$ given by:

$$T_{ij} = (\rho + p)\mathfrak{u}_i\mathfrak{u}_j + pg_{ij} \tag{41}$$

where, ρ and p are as usual the energy density and pressure of the dark energy.The fluid four-velocity vector taken as $\mathfrak{u}^i = \frac{dx^i}{ds}$ has the form $(\mathfrak{u}^0, \mathfrak{u}^1, 0, 0)$ and also satisfies the equation $\mathfrak{u}^i\mathfrak{u}_i = -1$, here only $\mathfrak{u}^0$ and $\mathfrak{u}^1$ are the two non-zero terms for this vector.Also, $\mathfrak{u} < 0$ for the fluid's inward flow towards the black hole. Considering the fact that the spherical symmetry of the black hole does not have any kind of changes due to the in-falling of dark energy fluid, we are going to advance in this accretion process. The time component of the relativistic Bernoulli's equation for the conservation of T_{ij} leads us to the following:

$$\mathfrak{u}r^2M^{-2}\left(\rho + p\right)\left(\sqrt{g(r) + \mathfrak{u}^2}\right) = I_0, \tag{42}$$

This is the first integral of motion, where I_0 stands for an integration constant having the same dimension as of the energy density. Again to obtain the second integration of motion, mostly known as the energy flux equation, we

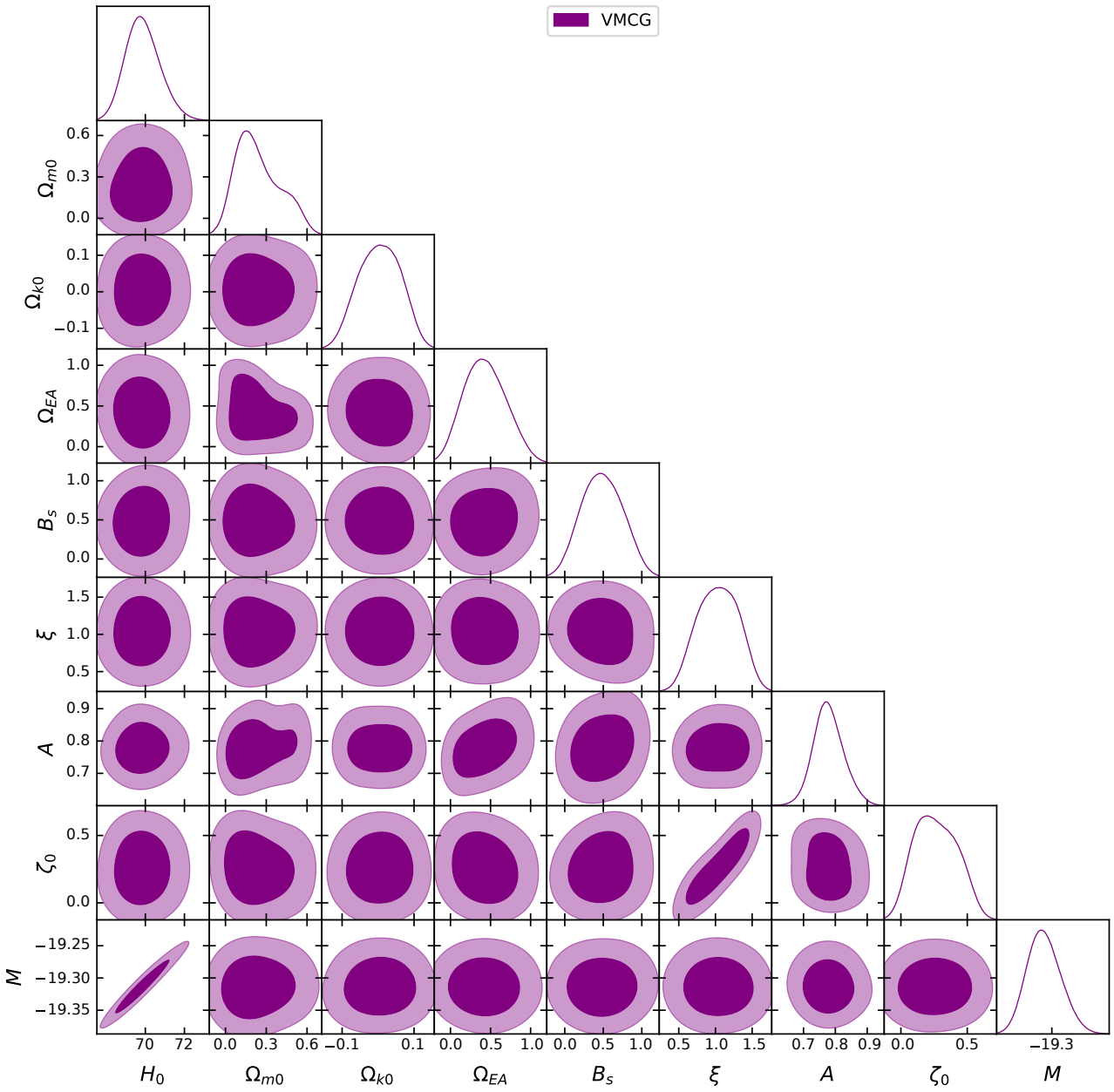


FIG. 2: The figure displays the posterior distribution for various observational data measurements using the Viscous Modified Chaplygin Gas Model

need to project T_{ij} onto $\mathfrak{u}^i = \frac{dx^i}{ds}$, which is given by the equation $\mathfrak{u}_i T^{ij}_{;j} = 0$ and can be written in simplified form as $\mathfrak{u}^i \rho_{,i} + (\rho + p)\mathfrak{u}^i_{;i} = 0$ will lead us to the following:

$$\mathfrak{u} r^2 M^{-2} exp\left[\int_{\rho_\infty}^{\rho} \frac{d\tilde{\rho}}{\tilde{\rho} + p(\tilde{\rho})}\right] = -I_1, \tag{43}$$

where I_1 is another integration constant, predominantly related to the energy flux onto the black hole, given by [214, 215, 253]. The related negative sign has been used for the convenience of the calculation. Energy densities of the black hole at horizon and infinity are given by ρ and ρ_∞, respectively. Now, by dividing the above two equations, we will have the following equation:

$$(\rho + p)\left(\sqrt{g(r) + \mathfrak{u}^2}\right) exp\left[-\int_{\rho_\infty}^{\rho} \frac{d\tilde{\rho}}{\tilde{\rho} + p(\tilde{\rho})}\right] = I_2, \tag{44}$$

Here, the constant I_2, used in the RHS of the above equation, is given by: $I_2 = -\frac{I_0}{I_1} = \rho_\infty + p(\rho_\infty)$. To compute the rate of change of mass of the black hole, i.e. $\dot{M}$ we need to integrate the flux of the dark energy over the 4-dimensional volume of the black hole [226, 254]:

$$\dot{M} = -4\pi I_0 M^2, \tag{45}$$

Now, considering the value of I_0, the above Eqn.(45) will become:

$$\dot{M} = 4\pi I_1 M^2 \left[\rho_\infty + p(\rho_\infty)\right], \tag{46}$$

From the Eqn.(46), we will get the rate of change of mass of a 4-dimensional black hole. Interestingly, for all the values of ρ and p which are in contrast with dominant energy condition, the above equation performs quite efficiently. So, in general, the above equation can be written as follows:

$$\dot{M} = 4\pi I_1 M^2 (\rho + p). \tag{47}$$

The above equation clearly implies that the changes in the mass of the black hole solely depends upon the term $\rho + p$. So, in the case of phantom energy dominance, universe will contract as $(\rho + p) < 0$ in this case. On the other hand in the case of quintessence kind of dark energy dominance, where $(\rho + p) > 0$ the universe will expand. Again, envisaging the Einstein-Aether gravity framework, we will get the following equation by using the Eqn.(13) and Eqn.(15) as follows:

$$(\rho + p) = -\frac{\dot{\rho}}{3}\left[\frac{\beta(-E' + \frac{E}{2K} + 1)}{\frac{8\pi G}{3}\rho - \frac{k}{a^2}}\right]^{\frac{1}{2}}. \tag{48}$$

Now, making use of Eqn.(48) into the Eqn.(47) and taking integration on both sides, we will obtain the required equation of mass in the following form:

$$M = \frac{M_0}{1 + \frac{4\pi I_1 M_0}{3}\int_{\rho_0}^{\rho} \frac{d\rho}{\sqrt{\frac{\left(\frac{8\pi G\rho}{3} - \frac{k}{a^2}\right)}{\beta\left(1 - E' + \frac{E}{2K}\right)}}}} \tag{49}$$

where M_0 represents the current mass of the black hole and ρ_0 represents the current energy density which can be written as $\rho_0 = \rho_{DM0} + \rho_{DE0}$.

The above equation can be written in a more simplified manner as follows:

$$M = \frac{M_0}{1 + \frac{4\pi I_1 M_0}{3}\int_{\rho_0}^{\rho} \frac{d\rho}{\sqrt{\frac{\left(\frac{8\pi G\rho}{3} - \frac{k}{a^2}\right)}{\left(\beta + \frac{15\epsilon\beta H^2}{m^2} - 1\right)}}}}. \tag{50}$$

The above Eqn.(50) is our desired equation of mass of a 4-dimensional black hole in Einstein-Aether gravity framework. Now, let us consider some Chaplygin gas models and analyzed their effects onto the mass of the Black hole during the evolution of the Universe.

A. Generalized Cosmic Chaplygin Gas (GCCG) Model:

We have already discussed this particular kind of Chaplygin gas in Sec.II A. Now, here we are going to investigate the accretion process of GCCG onto a 4-dimensional Einstein-Aether black hole. In this regard, considering Eqn.(18), Eqn.(21), Eqn.(24) and Eqn.(25) into Eqn.(50) all at once, we will get the desired equation of mass as follows:

$$M = \frac{M_0}{1 + \frac{4\pi I_1 M_0}{3}\int_{\rho_0}^{\rho} \frac{d\left(\frac{3H_0^2}{8\pi G}\left[\Omega_{m0}(1+z)^3 + \left(\Omega_{EA}^{-2} + \Omega_{k0} - \Omega_{m0}\right)\left\{\mathcal{A} + (1-\mathcal{A})\left(\mathcal{A}_s + (1-\mathcal{A}_s)(1+z)^{3(1+\alpha)(1+\omega)}\right)^{\frac{1}{1+\omega}}\right\}^{\frac{1}{1+\alpha}}\right]\right)}{\sqrt{\left(\frac{8\pi G\left(\frac{3H_0^2}{8\pi G}\left[\Omega_{m0}(1+z)^3 + \left(\Omega_{EA}^{-2} + \Omega_{k0} - \Omega_{m0}\right)\left\{\mathcal{A} + (1-\mathcal{A})\left(\mathcal{A}_s + (1-\mathcal{A}_s)(1+z)^{3(1+\alpha)(1+\omega)}\right)^{\frac{1}{1+\omega}}\right\}^{\frac{1}{1+\alpha}}\right]\right)}{3} - \frac{k}{a^2}\right)}{\left(\beta + \frac{15\epsilon\beta\left(H_0^2\left[\left\{\Omega_{m0}(1+z)^3 - \Omega_{k0}(1+z)^2\right\}\Omega_{EA}^2 + \left\{1 - \Omega_{EA}^2(\Omega_{m0} - \Omega_{k0})\right\}\left\{\mathcal{A} + (1-\mathcal{A})\left(\mathcal{A}_s + (1-\mathcal{A}_s)(1+z)^{3(1+\alpha)(1+\omega)}\right)^{\frac{1}{1+\omega}}\right\}^{\frac{1}{1+\alpha}}\right]^{\frac{1}{2}}\right)}{m^2} - 1\right)}}}. \tag{51}$$

The above Eqn.(51) stands for the equation of mass of a 4-dimensional Einstein-Aether black hole due to the ac-

cretion of GCCG. The effects of GCCG accretion onto the black hole mass M have been portrayed by the Fig.3 in terms of redshift function z, which clearly shows the increase in black hole mass with the evolution of our universe caused by the accretion of GCCG.

B. Viscous Modified Chaplygin Gas (VMCG) Model:

In Sec.II B we have briefly talked about VMCG model, hence here we are going to explore its accretion process onto a 4-dimension Einstein-Aether black hole. If we consider Eqn.(18), Eqn.(24), Eqn.(28) and Eqn.(29) altogether in Eqn.(50), we will have the desired equation of mass as follows:

$$M=\frac{M_0}{1+\frac{4\pi I_1 M_0}{3}\int_{\rho_0}^{\rho}\frac{d\left(\frac{3H_0^2}{8\pi G}\left[\Omega_{m0}(1+z)^3+\left(\Omega_{EA}^{-2}+\Omega_{k0}-\Omega_{m0}\right)\left\{B_s+(1-B_s)(1+z)^{3(1+\xi)(1+A-\sqrt{3}\zeta_0)}\right\}^{\frac{1}{1+\xi}}\right]\right)}{\sqrt{\left(\frac{8\pi G\left(\frac{3H_0^2}{8\pi G}\left[\Omega_{m0}(1+z)^3+\left(\Omega_{EA}^{-2}+\Omega_{k0}-\Omega_{m0}\right)\left\{B_s+(1-B_s)(1+z)^{3(1+\xi)(1+A-\sqrt{3}\zeta_0)}\right\}^{\frac{1}{1+\xi}}\right]\right)}{3}-\frac{k}{a^2}\right)\left(\beta+\frac{15\epsilon\beta\left(H_0^2\left[\left\{\Omega_{m0}(1+z)^3-\Omega_{k0}(1+z)^2\right\}\Omega_{EA}^2+\left\{1-\Omega_{EA}^2(\Omega_{m0}-\Omega_{k0})\right\}\left\{B_s+(1-B_s)(1+z)^{3(1+\xi)(1+A-\sqrt{3}\zeta_0)}\right\}^{\frac{1}{1+\xi}}\right]^{\frac{1}{2}}\right)}{m^2}-1\right)}}}. \tag{52}$$

The above Eqn.(52) stands for the equation of mass of a 4-dimensional Einstein-Aether black hole due to the accretion of VMCG. The effects of VMCG accretion onto the black hole mass M have been portrayed by the Fig.4 in terms of redshift function z which clearly shows the increase in black hole mass with the evolution of our universe caused by the accretion of VMCG. In Fig.5 the comparison of mass change with redshift has been depicted for GCCG model, VMCG model as well as ΛCDM model of the universe alongside eachother.

V. RESULTS

Figures 1 and 2 illustrate the 68% and 95% confidence intervals for key cosmological parameters in the Generalized Cosmic Chaplygin Gas (GCCG) Model and the Viscous Modified Chaplygin Gas (VMCG) Model, respectively. Table I summarizes the MCMC results for three cosmological models: the ΛCDM model, the GCCG Model, and the VMCG Model. All three models yield consistent H_0 values around 69.85, aligning with the findings in [255]. The ΛCDM model and the VMCG model exhibit tighter constraints compared to the GCCG model. However, the matter density parameter (Ω_{m0}) is lower than the values reported in [255] ($\Omega_{m0}=0.315\pm0.007$) for all three cases. Then we have proceeded to observe the mass accretion process, with the help of these estimated values of parameters. The outcomes have been more than satisfactory as we have obtained the mass equation of a 4-dimensional Einstein-Aether black hole in terms of some dimensionless density parameters and redshift function z. After plotting the results as black hole mass M vs redshift z graphs in Fig.3 and Fig.4, respectively, we can distinctly observe that with time, black hole mass increases for both the generalized cosmic Chaplygin gas (GCCG) model and viscous modified Chaplygin gas (VMCG) model. In Fig.5, we can observe separately the pattern of black hole mass M vs redshift z graphs in the cases of generalized cosmic Chaplygin gas (GCCG), viscous modified Chaplygin gas (VMCG) as well as ΛCDM model of the universe. All of the above-mentioned models show an increase in black hole mass with the expansion of the Universe, just as predicted in the case of the dark energy-induced Universe.

VI. CONCLUSIONS

Black Holes have always been an object of fascination for cosmologists. Several studies have been conducted on them time and again. In this study, we have investigated Einstein-Aether black holes, basically about their mass accretion process. Firstly, we have considered the non-flat FRW model of the universe, which combines dark matter and dark energy. Then in the framework of Einstein-Aether gravity, we have taken the modified Friedmann equations. Applying the conservation equation on dark matter and dark energy separately has led us to energy densities of the dark matter and dark energy, respectively. Also, for the simplicity of the calculation, we have taken $E(K)$ as a quadratic function of K. Afterward, we contemplated two different kinds of dark energy models: the generalized cosmic

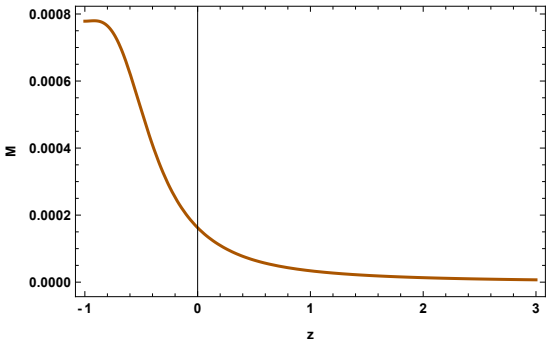


FIG. 3: The changes in BH mass M against redshift function z on the account of GCCG accretion.

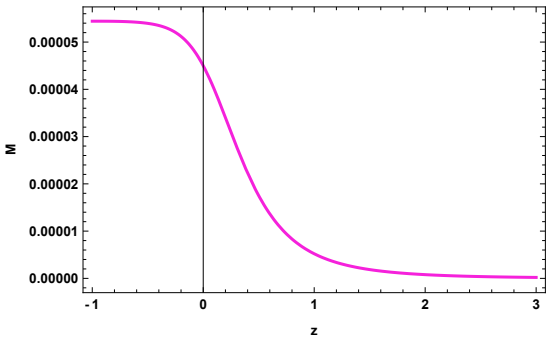


FIG. 4: The changes in BH mass M against redshift function z on the account of VMCG accretion.

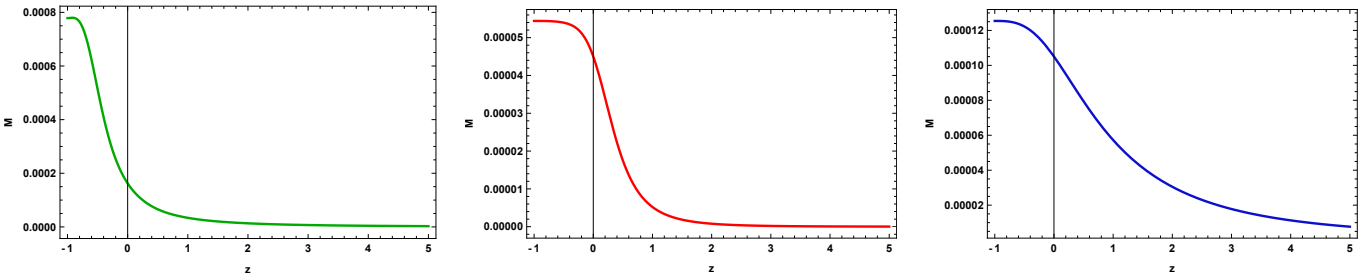
Chaplygin gas (GCCG) model and the viscous modified Chaplygin gas (VMCG) model as the candidates for dark energy. In both cases, we have calculated the energy density and Hubble parameter equation in terms of some dimensionless density parameters and some unknown parameters one by one. Next, we selected 31 uncorrelated data points from Cosmic Chronometers and 1701 data points from the Type Ia Supernova dataset to obtain the best-fit values for the model parameters using MCMC. Finally, after estimating the required parameters, we analyzed the mass accretion process of a 4-dimensional Einstein-Aether black hole. Primarily we have obtained the mass equation for both the generalized cosmic Chaplygin gas (GCCG) model and viscous modified Chaplygin gas (VMCG) model one after another. Then we have picturized the result through the black hole mass M vs. redshift function z graph in Fig.3 for the case of GCCG accretion and in Fig.4 for the case of VMCG accretion respectively. Both Fig.3 and Fig.4 clearly show that the black hole mass will only grow larger with time. Again, in Fig.5, we have picturized the comparison of the black hole mass M vs. redshift function z graph for the GCCG model, VMCG model as well as ΛCDM model of the universe in the case of Einstein-Aether gravity. So, on the whole, we can conclude that with the expansion of the universe, a 4-dimensional Einstein-Aether black hole mass will increase for accretion of dark energies like generalized cosmic Chaplygin gas and viscous modified Chaplygin gas as well as in the case of ΛCDM model of the Universe.

VII. ACKNOWLEDGEMENTS

PM is thankful to IIEST, Shibpur, India, for providing Institute Fellowship (JRF).

[1] S. Allen, R. W. Schmidt, H. Ebeling, A. Fabian, L. Van Speybroeck, Constraints on dark energy from Chandra observations of the largest relaxed galaxy clusters, Monthly Notices of the Royal Astronomical Society 353 (2) (2004) 457–467.

[2] R. Adam, P. A. Ade, N. Aghanim, Y. Akrami, M. Alves, F. Argüeso, M. Arnaud, F. Arroja, M. Ashdown, J. Aumont, et al., Planck 2015 results-I. Overview of products and scientific results, Astronomy & Astrophysics 594 (2016) A1.



(a) M vs z in GCCG Model (b) M vs z in VMCG Model (c) M vs z in ΛCDM Model

FIG. 5: The changes in BH mass M with redshift function z for GCCG model, VMCG model and ΛCDM model of the universe in Einstein-Aether gravity framework.

[3] C. L. Bennett, M. Bay, M. Halpern, G. Hinshaw, C. Jackson, N. Jarosik, A. Kogut, M. Limon, S. Meyer, L. Page, et al., The microwave anisotropy probe* mission, The Astrophysical Journal 583 (1) (2003) 1.

[4] D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, M. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, et al., First-year Wilkinson Microwave Anisotropy Probe (WMAP)* observations: determination of cosmological parameters, The Astrophysical Journal Supplement Series 148 (1) (2003) 175.

[5] L. Verde, A. F. Heavens, W. J. Percival, S. Matarrese, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, S. Cole, M. Colless, et al., The 2dF Galaxy Redshift Survey: the bias of galaxies and the density of the Universe, Monthly Notices of the Royal Astronomical Society 335 (2) (2002) 432–440.

[6] D. J. Eisenstein, W. Hu, Baryonic features in the matter transfer function, The Astrophysical Journal 496 (2) (1998) 605.

[7] D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H.-J. Seo, M. Tegmark, Z. Zheng, et al., Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies, The Astrophysical Journal 633 (2) (2005) 560.

[8] S. Perlmutter, G. Aldering, S. Deustua, S. Fabbro, G. Goldhaber, D. E. Groom, A. Kim, M. Kim, R. Knop, P. Nugent, et al., Cosmology from type Ia supernovae, arXiv preprint astro-ph/9812473 (1998).

[9] P. M. Garnavich, R. P. Kirshner, P. Challis, J. Tonry, R. L. Gilliland, R. C. Smith, A. Clocchiatti, A. Diercks, A. V. Filippenko, M. Hamuy, et al., Constraints on cosmological models from Hubble Space Telescope observations of high-z supernovae, The Astrophysical Journal 493 (2) (1998) L53.

[10] N. A. Bahcall, J. P. Ostriker, S. Perlmutter, P. J. Steinhardt, The cosmic triangle: Revealing the state of the universe, Science 284 (5419) (1999) 1481–1488.

[11] S. Perlmutter, et al., Supernova cosmology project collaboration, Astrophys. J 517 (2) (1999) 565.

[12] A. V. Filippenko, A. G. Riess, Results from the high-z supernova search team, Physics Reports 307 (1-4) (1998) 31–44.

[13] K. Abazajian, J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, K. S. Anderson, S. F. Anderson, J. Annis, N. A. Bahcall, I. K. Baldry, S. Bastian, et al., The second data release of the sloan digital sky survey, The Astronomical Journal 128 (1) (2004) 502.

[14] J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, C. A. Prieto, K. S. Anderson, S. F. Anderson, J. Annis, N. A. Bahcall, C. Bailer-Jones, I. K. Baldry, et al., The sixth data release of the sloan digital sky survey, The Astrophysical Journal Supplement Series 175 (2) (2008) 297.

[15] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, et al., Measurements of Ω and Λ from 42 high-redshift supernovae, The Astrophysical Journal 517 (2) (1999) 565.

[16] S. Perlmutter, G. Aldering, M. D. Valle, S. Deustua, R. Ellis, S. Fabbro, A. Fruchter, G. Goldhaber, D. Groom, I. Hook, et al., Discovery of a supernova explosion at half the age of the Universe, Nature 391 (6662) (1998) 51–54.

[17] M. Tegmark, M. A. Strauss, M. R. Blanton, K. Abazajian, S. Dodelson, H. Sandvik, X. Wang, D. H. Weinberg, I. Zehavi, N. A. Bahcall, et al., Cosmological parameters from SDSS and WMAP, Physical review D 69 (10) (2004) 103501.

[18] S. Perlmutter, et al., Measurements of the cosmological parameters Omega and Lambda from the first 7 supernovae at z>=0.35, Astrophys. J. 483 (1997) 565. arXiv:astro-ph/9608192, doi:10.1086/304265.

[19] P. de Bernardis, P. A. Ade, J. J. Bock, J. Bond, J. Borrill, A. Boscaleri, K. Coble, B. Crill, G. De Gasperis, P. Farese, et al., A flat Universe from high-resolution maps of the cosmic microwave background radiation, Nature 404 (6781) (2000) 955–959.

[20] A. E. Lange, P. A. Ade, J. Bock, J. Bond, J. Borrill, A. Boscaleri, K. Coble, B. Crill, P. de Bernardis, P. Farese, et al., Cosmological parameters from the first results of Boomerang, Physical Review D 63 (4) (2001) 042001.

[21] A. Balbi, P. Ade, J. Bock, J. Borrill, A. Boscaleri, P. De Bernardis, P. G. Ferreira, S. Hanany, V. Hristov, A. H. Jaffe, et al., Constraints on cosmological parameters from MAXIMA-1, The Astrophysical Journal 545 (1) (2000) L1.

[22] D. N. Spergel, R. Bean, O. Doré, M. Nolta, C. Bennett, J. Dunkley, G. Hinshaw, N. e. Jarosik, E. Komatsu, L. Page, et al., Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: implications for cosmology, The Astrophysical Journal Supplement Series 170 (2) (2007) 377.

[23] T. Padmanabhan, Cosmological constant—the weight of the vacuum, Physics Reports 380 (5-6) (2003) 235–320.

[24] S. M. Carroll, W. H. Press, E. L. Turner, The cosmological constant, Annual Review of Astronomy and Astrophysics. Vol. 30 (A93-25826 09-90), p. 499-542. 30 (1992) 499–542.

[25] R. R. Caldwell, R. Dave, P. J. Steinhardt, Cosmological imprint of an energy component with general equation of state, Physical Review Letters 80 (8) (1998) 1582.

[26] P. Peebles, B. Ratra, Cosmology with a time-variable cosmological'constant', Astrophysical Journal, Part 2-Letters to the Editor (ISSN 0004-637X), vol. 325, Feb. 15, 1988, p. L17-L20. NSF-supported research. 325 (1988) L17–L20.

[27] K. Bamba, S. Capozziello, S. Nojiri, S. D. Odintsov, Dark energy cosmology: the equivalent description via different theoretical models and cosmography tests, Astrophysics and Space Science 342 (2012) 155–228.

[28] Y. Chen, C.-Q. Geng, S. Cao, Y.-M. Huang, Z.-H. Zhu, Constraints on a ϕCDM model from strong gravitational lensing and updated Hubble parameter measurements, Journal of Cosmology and Astroparticle Physics 2015 (02) (2015) 010.

[29] V. Smer-Barreto, A. R. Liddle, Planck satellite constraints on pseudo-Nambu-Goldstone boson quintessence, Journal of Cosmology and Astroparticle Physics 2017 (01) (2017) 023.

[30] C. Wetterich, Cosmology and the fate of dilatation symmetry, Nuclear Physics B 302 (4) (1988) 668–696.

[31] R. R. Caldwell, A phantom menace? Cosmological consequences of a dark energy component with super-negative equation of state, Physics Letters B 545 (1-2) (2002) 23–29.

[32] E. Elizalde, S. Nojiri, S. D. Odintsov, Late-time cosmology in a (phantom) scalar-tensor theory: dark energy and the cosmic speed-up, Physical Review D 70 (4) (2004) 043539.

[33] C. Armendariz-Picon, V. Mukhanov, P. J. Steinhardt, Dynamical solution to the problem of a small cosmological constant and late-time cosmic acceleration, Physical Review Letters 85 (21) (2000) 4438.

[34] C. Armendariz-Picon, V. Mukhanov, P. J. Steinhardt, Essentials of k-essence, Physical Review D 63 (10) (2001) 103510.

[35] B. Feng, X. Wang, X. Zhang, Dark energy constraints from the cosmic age and supernova, Physics Letters B 607 (1-2) (2005) 35–41.

[36] Z.-K. Guo, Y.-S. Piao, X. Zhang, Y.-Z. Zhang, Cosmological evolution of a quintom model of dark energy, Physics Letters B 608 (3-4) (2005) 177–182.

[37] M. Gasperini, F. Piazza, G. Veneziano, Quintessence as a runaway dilaton, Physical Review D 65 (2) (2001) 023508.

[38] A. Sen, Rolling tachyon, Journal of High Energy Physics 2002 (04) (2002) 048.

[39] A. Sen, Tachyon matter, Journal of High Energy Physics 2002 (07) (2002) 065.

[40] G. Gibbons, Cosmological evolution of the rolling tachyon, Physics Letters B 537 (1-2) (2002) 1–4.

[41] H. Wei, R.-G. Cai, D.-F. Zeng, Hessence: a new view of quintom dark energy, Classical and Quantum Gravity 22 (16) (2005) 3189.

[42] B. Gumjudpai, J. Ward, Generalized DBI quintessence, Physical Review D 80 (2) (2009) 023528.

[43] J. Martin, M. Yamaguchi, DBI-essence, Physical Review D 77 (12) (2008) 123508.

[44] M. Li, A model of holographic dark energy, Physics Letters B 603 (1-2) (2004) 1–5.

[45] S. D. Hsu, Entropy bounds and dark energy, Physics Letters B 594 (1-2) (2004) 13–16.

[46] A. G. Cohen, D. B. Kaplan, A. E. Nelson, Effective field theory, black holes, and the cosmological constant, Physical Review Letters 82 (25) (1999) 4971.

[47] V. Sahni, Y. Shtanov, Braneworld models of dark energy, Journal of Cosmology and Astroparticle Physics 2003 (11) (2003) 014.

[48] R.-G. Cai, A dark energy model characterized by the age of the universe, Physics Letters B 657 (4-5) (2007) 228–231.

[49] H. Wei, R.-G. Cai, A new model of agegraphic dark energy, Physics Letters B 660 (3) (2008) 113–117.

[50] A. Kamenshchik, U. Moschella, V. Pasquier, An alternative to quintessence, Physics Letters B 511 (2-4) (2001) 265–268.

[51] N. Bilić, G. B. Tupper, R. D. Viollier, Unification of dark matter and dark energy: the inhomogeneous Chaplygin gas, Physics Letters B 535 (1-4) (2002) 17–21.

[52] V. Gorini, A. Kamenshchik, U. Moschella, V. Pasquier, The Chaplygin gas as a model for dark energy, in: The Tenth Marcel Grossmann Meeting: On Recent Developments in Theoretical and Experimental General Relativity, Gravitation and Relativistic Field Theories (In 3 Volumes), World Scientific, 2005, pp. 840–859.

[53] U. Alam, V. Sahni, T. Deep Saini, A. Starobinsky, Exploring the expanding universe and dark energy using the Statefinder diagnostic, Monthly Notices of the Royal Astronomical Society 344 (4) (2003) 1057–1074.

[54] V. Gorini, A. Kamenshchik, U. Moschella, Can the Chaplygin gas be a plausible model for dark energy?, Physical Review D 67 (6) (2003) 063509.

[55] M. Bento, O. Bertolami, A. A. Sen, Generalized Chaplygin gas, accelerated expansion, and dark-energy-matter unification, Physical Review D 66 (4) (2002) 043507.

[56] H. B. Benaoum, et al., Modified Chaplygin gas cosmology, Advances in High Energy Physics 2012 (2012).

[57] U. Debnath, A. Banerjee, S. Chakraborty, Role of modified Chaplygin gas in accelerated universe, Classical and Quantum Gravity 21 (23) (2004) 5609.

[58] P. F. González-Díaz, You need not be afraid of phantom energy, Physical Review D 68 (2) (2003) 021303.

[59] W. Chakraborty, U. Debnath, S. Chakraborty, Generalized cosmic Chaplygin gas model with or without interaction, arXiv preprint arXiv:0711.0079 (2007).

[60] P. Rudra, Role of generalized cosmic Chaplygin gas in accelerating universe: a field theoretical prescription, Modern Physics Letters A 28 (22) (2013) 1350102.

[61] J. Bhadra, U. Debnath, Accretion of new variable modified Chaplygin gas and generalized cosmic Chaplygin gas onto Schwarzschild and Kerr–Newman black holes, The European Physical Journal C 72 (2012) 1–9.

[62] M. Sharif, R. Saleem, Effects of viscous pressure on warm inflationary generalized cosmic Chaplygin gas model, Journal of Cosmology and Astroparticle Physics 2014 (12) (2014) 038.

[63] P. Rudra, Dynamics of interacting generalized cosmic Chaplygin gas in brane-world scenario, Astrophysics

and Space Science 342 (2012) 579–599.
[64] M. Sharif, A. Jawad, Phantom-like generalized cosmic chaplygin gas and traversable wormhole solutions, The European Physical Journal Plus 129 (2014) 1–13.
[65] A. Eid, Schwarzschild–De Sitter thin shell wormholes supported by a generalized cosmic Chaplygin gas, Gravitation and Cosmology 24 (4) (2018) 378–383.
[66] M. Sharif, M. Azam, Reissner–Nordström thin-shell wormholes with generalized cosmic Chaplygin gas, The European Physical Journal C 73 (2013) 1–6.
[67] D. Aberkane, N. Mebarki, S. Benchikh, Viscous modified Chaplygin gas in classical and loop quantum cosmology, Chinese Physics Letters 34 (6) (2017) 069801.
[68] N. Mebarki, D. Aberkane, Dynamical study of viscous modified Chaplygin gas and confrontation with recent observational data, in: Journal of Physics: Conference Series, Vol. 1258, IOP Publishing, 2019, p. 012025.
[69] M. Sharif, A. Siddiqa, Interaction of viscous modified Chaplygin gas with f (R, T) gravity, Modern Physics Letters A 32 (28) (2017) 1750151.
[70] F. Mavoa, H. Hova, G. Ganiou, S. Houndjo, F. Ahloui, E. Baffou, A. Kanfon, Effect of viscous generalized and modified Chaplygin gas on f (R,$\mathcal{G}$,$\mathcal{T}$) gravity, Indian Journal of Physics 97 (7) (2023) 2247–2260.
[71] G. Dvali, G. Gabadadze, M. Porrati, Metastable gravitons and infinite volume extra dimensions, Physics Letters B 484 (1-2) (2000) 112–118.
[72] A. De Felice, S. Tsujikawa, f (R) theories, Living Reviews in Relativity 13 (1) (2010) 1–161.
[73] S. Nojiri, S. D. Odintsov, Unified cosmic history in modified gravity: from F (R) theory to Lorentz non-invariant models, Physics Reports 505 (2-4) (2011) 59–144.
[74] T. Clifton, J. D. Barrow, The power of general relativity, Physical Review D 72 (10) (2005) 103005.
[75] G. Dvali, G. Gabadadze, M. Porrati, 4D gravity on a brane in 5D Minkowski space, Physics Letters B 485 (1-3) (2000) 208–214.
[76] T. Jacobson, D. Mattingly, Gravity with a dynamical preferred frame, Physical Review D 64 (2) (2001) 024028.
[77] M. C. B. Abdalla, S. Nojiri, S. D. Odintsov, Consistent modified gravity: dark energy, acceleration and the absence of cosmic doomsday, Classical and Quantum Gravity 22 (5) (2005) L35.
[78] P. Hořava, Membranes at quantum criticality, Journal of High Energy Physics 2009 (03) (2009) 020.
[79] J. D. Barrow, Some inflationary Einstein-aether cosmologies, Physical Review D 85 (4) (2012) 047503.
[80] X. Meng, X. Du, Einstein-aether theory as an alternative to dark energy model?, Physics Letters B 710 (4-5) (2012) 493–499.
[81] S. Nojiri, S. D. Odintsov, Modified Gauss–Bonnet theory as gravitational alternative for dark energy, Physics Letters B 631 (1-2) (2005) 1–6.
[82] S. Tsujikawa, Modified gravity models of dark energy, in: Lectures on cosmology: Accelerated Expansion of the Universe, Springer, 2010, pp. 99–145.
[83] K. Yerzhanov, S. R. Myrzakul, I. Kulnazarov, R. Myrzakulov, Accelerating cosmology in F (T) gravity with scalar field, arXiv preprint arXiv:1006.3879 (2010).
[84] I. Antoniadis, J. Rizos, K. Tamvakis, Singularity-free cosmological solutions of the superstring effective action, Nuclear Physics B 415 (2) (1994) 497–514.
[85] C. Brans, R. H. Dicke, Mach's principle and a relativistic theory of gravitation, Physical Review 124 (3) (1961) 925.
[86] C. Eling, T. Jacobson, Static post-Newtonian equivalence of general relativity and gravity with a dynamical preferred frame, Physical Review D 69 (6) (2004) 064005.
[87] T. Jacobson, D. Mattingly, Einstein-aether waves, Physical Review D 70 (2) (2004) 024003.
[88] C. Eling, T. Jacobson, D. Mattingly, Einstein-aether theory, in: Deserfest, World Scientific, 2006, pp. 163–179.
[89] B. Z. Foster, T. Jacobson, Post-Newtonian parameters and constraints on Einstein-aether theory, Physical Review D 73 (6) (2006) 064015.
[90] T. G. Zlosnik, P. G. Ferreira, G. D. Starkman, Modifying gravity with the aether: An alternative to dark matter, Physical Review D 75 (4) (2007) 044017.
[91] T. Zlosnik, P. Ferreira, G. Starkman, Growth of structure in theories with a dynamical preferred frame, Physical Review D 77 (8) (2008) 084010.
[92] C. Eling, T. Jacobson, Black holes in Einstein-aether theory, Classical and Quantum Gravity 23 (18) (2006) 5643.
[93] V. Satheeshkumar, Nature of singularities in vector-tensor theories of gravity, in: The Sixteenth Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories: Proceedings of the MG16 Meeting on General Relativity; 5–10 July 2021, World Scientific, 2023, pp. 1095–1103.
[94] K. Yagi, D. Blas, E. Barausse, N. Yunes, Constraints on Einstein-Æther theory and Hořava gravity from binary pulsar observations, Physical Review D 89 (8) (2014) 084067.
[95] Y. Gong, S. Hou, D. Liang, E. Papantonopoulos, et al., Gravitational waves in Einstein-æther and generalized TeVeS theory after GW170817, Physical Review D 97 (8) (2018) 084040.
[96] J. Oost, S. Mukohyama, A. Wang, Constraints on Einstein-aether theory after GW170817, Physical Review D 97 (12) (2018) 124023.
[97] M. Khodadi, E. N. Saridakis, Einstein-Æther gravity in the light of event horizon telescope observations of M87, Physics of the Dark Universe 32 (2021) 100835.
[98] B. Li, D. F. Mota, J. D. Barrow, Detecting a Lorentz-violating field in cosmology, Physical Review D 77 (2) (2008) 024032.
[99] R. A. Battye, F. Pace, D. Trinh, Cosmological perturbation theory in generalized Einstein-Aether models, Physical Review D 96 (6) (2017) 064041.
[100] W.-b. Han, The Gravitational Waves from Test Particles around Black Holes Immersed in a Strong Magnetic Field, International Journal of Theoretical Physics 48 (2009) 621–629.
[101] C. Zhang, X. Zhao, A. Wang, B. Wang, K. Yagi, N. Yunes, W. Zhao, T. Zhu, Gravitational waves from the quasicircular inspiral of compact binaries in Einstein-aether theory, Physical Review D 101 (4) (2020) 044002.
[102] D. Garfinkle, T. Jacobson, A positive-energy theorem for einstein-aether and hořava gravity, Physical Review Letters 107 (19) (2011) 191102.

[103] E. V. Linder, R. J. Scherrer, Aetherizing Lambda: Barotropic fluids as dark energy, Physical Review D 80 (2) (2009) 023008.

[104] M. Gasperini, Repulsive gravity in the very early universe, General Relativity and Gravitation 30 (12) (1998) 1703–1709.

[105] C. Eling, T. Jacobson, Spherical solutions in Einstein-aether theory: Static aether and stars, Classical and Quantum Gravity 23 (18) (2006) 5625.

[106] R. Konoplya, A. Zhidenko, Gravitational spectrum of black holes in the Einstein–Aether theory, Physics Letters B 648 (2-3) (2007) 236–239.

[107] C. Ding, A. Wang, X. Wang, Charged Einstein-aether black holes and Smarr formula, Physical Review D 92 (8) (2015) 084055.

[108] K. Lin, F.-H. Ho, W.-L. Qian, Charged Einstein-æther black holes in n-dimensional spacetime, International Journal of Modern Physics D 28 (03) (2019) 1950049.

[109] M. Meiers, M. Saravani, N. Afshordi, Cosmic censorship in Lorentz-violating theories of gravity, Physical Review D 93 (10) (2016) 104008.

[110] E. Barausse, T. P. Sotiriou, I. Vega, Slowly rotating black holes in Einstein-æther theory, Physical Review D 93 (4) (2016) 044044.

[111] T. Zhu, Q. Wu, M. Jamil, K. Jusufi, Shadows and deflection angle of charged and slowly rotating black holes in Einstein-Æther theory, Physical Review D 100 (4) (2019) 044055.

[112] J. Rayimbaev, A. Abdujabbarov, M. Jamil, W.-B. Han, Dynamics of magnetized particles around Einstein-Æther black hole with uniform magnetic field, Nuclear Physics B 966 (2021) 115364.

[113] R. Chan, M. da Silva, V. Satheeshkumar, Thermodynamics of Einstein-Aether black holes, The European Physical Journal C 82 (10) (2022) 943.

[114] H.-M. Wang, Z.-C. Lin, S.-W. Wei, Optical appearance of Einstein-Æther black hole surrounded by thin disk, Nuclear Physics B 985 (2022) 116026.

[115] C. Zhang, X. Zhao, K. Lin, S. Zhang, W. Zhao, A. Wang, Spherically symmetric static black holes in Einstein-aether theory, Physical Review D 102 (6) (2020) 064043.

[116] C. Liu, S. Yang, Q. Wu, T. Zhu, Thin accretion disk onto slowly rotating black holes in Einstein-Æther theory, Journal of Cosmology and Astroparticle Physics 2022 (02) (2022) 034.

[117] A. Adam, P. Figueras, T. Jacobson, T. Wiseman, Rotating black holes in Einstein-aether theory, Classical and Quantum Gravity 39 (12) (2022) 125001.

[118] B. Z. Foster, Noether charges and black hole mechanics in Einstein-aether theory, Physical Review D 73 (2) (2006) 024005.

[119] D. Garfinkle, C. Eling, T. Jacobson, Numerical simulations of gravitational collapse in Einstein-aether theory, Physical Review D 76 (2) (2007) 024003.

[120] T. Tamaki, U. Miyamoto, Generic features of Einstein-Aether black holes, Physical Review D 77 (2) (2008) 024026.

[121] C. Ding, A. Wang, X. Wang, T. Zhu, Hawking radiation of charged Einstein-aether black holes at both Killing and universal horizons, Nuclear Physics B 913 (2016) 694–715.

[122] C. Ding, Quasinormal ringing of black holes in Einstein-aether theory, Physical Review D 96 (10) (2017) 104021.

[123] K. Lin, X. Zhao, C. Zhang, T. Liu, B. Wang, S. Zhang, X. Zhang, W. Zhao, T. Zhu, A. Wang, Gravitational waveforms, polarizations, response functions, and energy losses of triple systems in Einstein-aether theory, Physical Review D 99 (2) (2019) 023010.

[124] A. G. Riess, L.-G. Strolger, J. Tonry, S. Casertano, H. C. Ferguson, B. Mobasher, P. Challis, A. V. Filippenko, S. Jha, W. Li, et al., Type Ia supernova discoveries at z¿ 1 from the Hubble Space Telescope: Evidence for past deceleration and constraints on dark energy evolution, The Astrophysical Journal 607 (2) (2004) 665.

[125] T. R. Choudhury, T. Padmanabhan, Cosmological parameters from supernova observations: A critical comparison of three data sets, Astronomy & Astrophysics 429 (3) (2005) 807–818.

[126] J. L. Tonry, B. P. Schmidt, B. Barris, P. Candia, P. Challis, A. Clocchiatti, A. L. Coil, A. V. Filippenko, P. Garnavich, C. Hogan, et al., Cosmological results from high-z supernovae, The Astrophysical Journal 594 (1) (2003) 1.

[127] B. J. Barris, J. L. Tonry, S. Blondin, P. Challis, R. Chornock, A. Clocchiatti, A. V. Filippenko, P. Garnavich, S. T. Holland, S. Jha, et al., Twenty-three high-redshift supernovae from the Institute for Astronomy Deep Survey: Doubling the supernova sample at z¿ 0.7, The Astrophysical Journal 602 (2) (2004) 571.

[128] M. d. C. Bento, O. Bertolami, A. Sen, Generalized Chaplygin gas and cosmic microwave background radiation constraints, Physical Review D 67 (6) (2003) 063003.

[129] M. d. C. Bento, O. Bertolami, A. Sen, WMAP constraints on the generalized Chaplygin gas model, Physics Letters B 575 (3-4) (2003) 172–180.

[130] M. Makler, S. Q. de Oliveira, I. Waga, Constraints on the generalized Chaplygin gas from supernovae observations, Physics Letters B 555 (1-2) (2003) 1–6.

[131] Z.-H. Zhu, Generalized Chaplygin gas as a unified scenario of dark matter/energy: observational constraints, Astronomy & Astrophysics 423 (2) (2004) 421–426.

[132] O. Bertolami, A. A. Sen, S. Sen, P. Silva, Latest supernova data in the framework of the generalized Chaplygin gas model, Monthly Notices of the Royal Astronomical Society 353 (1) (2004) 329–337.

[133] Y. Gong, Observational constraints on the generalized Chaplygin gas model, Journal of Cosmology and Astroparticle Physics 2005 (03) (2005) 007.

[134] D. Su-Mei, W. Pu-Xun, Observation constraints on the simplified GCG model, Chinese Physics Letters 24 (6) (2007) 1782.

[135] A. Sen, R. J. Scherrer, Generalizing the generalized Chaplygin gas, Physical Review D 72 (6) (2005) 063511.

[136] J. Lu, Y. Gui, L. X. Xu, Observational constraint on generalized Chaplygin gas model, The European Physical Journal C 63 (2009) 349–354.

[137] C.-G. Park, J.-c. Hwang, J. Park, H. Noh, Observational constraints on a unified dark matter and dark energy model based on generalized Chaplygin gas, Physical Review D 81 (6) (2010) 063532.

[138] L. Dao-Jun, L. Xin-Zhou, Cosmic microwave background radiation constraints on a modified Chaplygin gas model, Chinese Physics Letters 22 (7) (2005) 1600.

[139] Z.-K. Guo, Y.-Z. Zhang, Observational constraints on variable chaplygin gas, arXiv preprint astro-ph/0509790 (2005).

[140] K. Liao, Y. Pan, Z.-H. Zhu, Observational constraints on the new generalized Chaplygin gas model, Research in Astronomy and Astrophysics 13 (2) (2013) 159.

[141] S. Carneiro, C. Pigozzo, Observational tests of non-adiabatic Chaplygin gas, Journal of Cosmology and Astroparticle Physics 2014 (10) (2014) 060.

[142] E. Kahya, B. Pourhassan, Observational constraints on the extended Chaplygin gas inflation, Astrophysics and Space Science 353 (2014) 677–682.

[143] B. R. Dinda, S. Kumar, A. A. Sen, Inflationary generalized Chaplygin gas and dark energy in light of the Planck and BICEP2 experiments, Physical Review D 90 (8) (2014) 083515.

[144] C. Ranjit, U. Debnath, Constraining parameters of generalized cosmic Chaplygin gas in loop quantum cosmology, Astrophysics and Space Science 354 (2014) 651–665.

[145] W.-T. Zhu, P.-X. Wu, H.-W. Yu, Constraining the Generalized and Superfluid Chaplygin Gas Models with the Sandage-Loeb Test, Chinese Physics Letters 32 (5) (2015) 059501.

[146] G. Sharov, Observational constraints on cosmological models with Chaplygin gas and quadratic equation of state, Journal of Cosmology and Astroparticle Physics 2016 (06) (2016) 023.

[147] L. G. Collodel, G. M. Kremer, Chaplygin gas of Tachyon Nature Imposed by Noether Symmetry and constrained via H (z) data, Research in Astronomy and Astrophysics 16 (4) (2016) 014.

[148] F. Salamate, I. Khay, A. Safsafi, H. Chakir, M. Bennai, Observational Constraints on the Chaplygin Gas with Inverse Power Law Potential in Braneworld Inflation, Moscow University Physics Bulletin 73 (2018) 405–412.

[149] F. Salahedin, R. Pazhouhesh, M. Malekjani, Cosmological constrains on new generalized Chaplygin gas model, The European Physical Journal Plus 135 (5) (2020) 429.

[150] U. Debnath, Gravitational waves for some dark energy models in FRW Universe, Physics of the Dark Universe 32 (2021) 100832.

[151] U. Debnath, Roles of modified Chaplygin–Jacobi and Chaplygin–Abel gases in FRW universe, International Journal of Modern Physics A 36 (33) (2021) 2150245.

[152] J. Lu, L. Xu, J. Li, B. Chang, Y. Gui, H. Liu, Constraints on modified Chaplygin gas from recent observations and a comparison of its status with other models, Physics Letters B 662 (2) (2008) 87–91.

[153] L. Xu, Y. Wang, H. Noh, Modified Chaplygin gas as a unified dark matter and dark energy model and cosmic constraints, The European Physical Journal C 72 (3) (2012) 1931.

[154] J. Bhadra, U. Debnath, Constraining the Parameters of New Variable Modified Chaplygin Gas Model, International Journal of Theoretical Physics 53 (6) (2014) 1821–1831.

[155] U. Debnath, Gravitational waves for variable modified Chaplygin gas and some parametrizations of dark energy in the background of FRW universe, The European Physical Journal Plus 135 (2) (2020) 135.

[156] S. Sahlu, H. Sami, R. Hough, M. Elmardi, A.-M. Swart, A. Abebe, Confronting the chaplygin gas with data: background and perturbed cosmic dynamics, arXiv preprint arXiv:2112.11695 (2021).

[157] F.-Q. Wu, X. Chen, Cosmic microwave background with Brans-Dicke gravity. II. Constraints with the WMAP and SDSS data, Physical Review D 82 (8) (2010) 083003.

[158] Y.-C. Li, F.-Q. Wu, X. Chen, Constraints on the Brans-Dicke gravity theory with the Planck data, Physical Review D 88 (8) (2013) 084053.

[159] R. Morganstern, Observational constraints imposed by Brans-Dicke cosmologies, Physical Review D 7 (6) (1973) 1570.

[160] D. F. Torres, Quintessence, superquintessence, and observable quantities in Brans-Dicke and nonminimally coupled theories, Physical Review D 66 (4) (2002) 043522.

[161] R. Nakamura, M. A. Hashimoto, K. Arai, Constraints on Brans-Dicke cosmology with a varying Λ term due to the big-bang nucleosynthesis and WMAP, in: ORIGIN OF MATTER AND EVOLUTION OF GALAXIES: International Symposium on Origin of Matter and Evolution of Galaxies 2005: New Horizon of Nuclear Astrophysics and Cosmology, 2006, pp. 449–451.

[162] U. Debnath, Constraining the parameters of modified Chaplygin gas in Brans–Dicke theory, Physics of the Dark Universe 31 (2021) 100764.

[163] E. Thushari, R. Nakamura, M. Hashimoto, K. Arai, Brans-Dicke model constrained from the Big Bang nucleosynthesis and magnitude redshift relations of supernovae, Astronomy & Astrophysics 521 (2010) A52.

[164] A. Cid, S. del Campo, Constraints from CMB in the intermediate Brans-Dicke inflation, Journal of Cosmology and Astroparticle Physics 2011 (01) (2011) 013.

[165] R. Freitas, S. Gonçalves, Observational constraints of the gravitational waves in the Brans–Dicke theory: Einstein frame and Jordan–Brans–Dicke frame, Physics Letters B 710 (4-5) (2012) 504–515.

[166] H. Farajollahi, A. Salehi, Stability analysis and observational measurement in chameleonic generalised Brans-Dicke cosmology, Journal of Cosmology and Astroparticle Physics 2011 (07) (2011) 036.

[167] H. Farajollahi, A. Salehi, Stability analysis and observational measurement in 5-D Brans–Dicke cosmology, Physics Letters B 718 (3) (2013) 709–717.

[168] B. Tahmasebzadeh, K. Rezazadeh, K. Karami, Brans-Dicke inflation in light of the Planck 2015 data, Journal of Cosmology and Astroparticle Physics 2016 (07) (2016) 006.

[169] H. Amirhashchi, A. K. Yadav, Constraining an exact Brans–Dicke gravity theory with recent observations, Physics of the Dark Universe 30 (2020) 100711.

[170] U. Debnath, Observational data analysis for generalized cosmic Chaplygin gas in the background of Brans–Dicke theory, International Journal of Modern Physics A 36 (21) (2021) 2150157.

[171] Ö. Akarsu, A. Chopovsky, V. Shulga, E. Yalçınkaya, A. Zhuk, Weak field limit of higher dimensional massive Brans-Dicke gravity: Observational constraints, Physical Review D 101 (2) (2020) 024004.

[172] R. Prasad, A. K. Yadav, A. K. Yadav, Constraining Bianchi type V universe with recent H (z) and BAO observations in Brans–Dicke theory of gravitation, The European Physical Journal Plus 135 (3) (2020) 1–16.

[173] R. Biswas, U. Debnath, Constraining red-shift parametrization parameters in Brans-Dicke theory: evolution of open confidence contours, Astrophysics and Space Science 353 (2014) 721–730.

[174] S. Chakraborty, U. Debnath, C. Ranjit, Observational constraints of modified Chaplygin gas in loop quantum cosmology, The European Physical Journal C 72 (2012) 1–8.

[175] R. Biswas, U. Debnath, Constraining redshift parametrization parameters of dark energy: loop quantum gravity as back ground, The European Physical Journal C 73 (2013) 1–9.

[176] B. Paul, P. Thakur, M. Verma, Observational constraints on modified Chaplygin gas in Horava–Lifshitz gravity with dark radiation, Pramana 81 (2013) 691–718.

[177] C. Ranjit, P. Rudra, Observational data fitting to constrain Variable Modified Chaplygin Gas in the background of Horava-Lifshitz Gravity, International Journal of Theoretical Physics 55 (2016) 636–647.

[178] R. Biswas, U. Debnath, Observational constraints of red-shift parametrization parameters of dark energy in horava-lifshitz gravity, International Journal of Theoretical Physics 54 (2015) 341–357.

[179] H. Chaudhary, U. Debnath, T. Roy, S. Maity, G. Mustafa, Constraints on the Parameters of Modified Chaplygin-Jacobi and Modified Chaplygin-Abel Gases in $f(T)$ Gravity Model, arXiv preprint arXiv:2307.14691 (2023).

[180] U. Debnath, Observational Constraints of Modified Chaplygin Gas in Chern-Simons Gravity, International Journal of Theoretical Physics 54 (2015) 22–35.

[181] C. Ranjit, P. Rudra, U. Debnath, Study of some parameters of modified Chaplygin gas in Galileon gravity theory from observational perspective, Canadian Journal of Physics 92 (12) (2014) 1667–1675.

[182] C. Ranjit, P. Rudra, S. Kundu, Observational constraints of modified Chaplygin gas in RS II brane, Astrophysics and Space Science 347 (2013) 423–431.

[183] U. Debnath, et al., Constraining the parameters of modified chaplygin gas in Einstein-Aether gravity, Advances In High Energy Physics 2014 (2014).

[184] W. Yang, E. Di Valentino, S. Pan, S. Basilakos, A. Paliathanasis, Metastable dark energy models in light of Planck 2018: Alleviating the H0 tension, arXiv preprint arXiv:2001.04307 (1880).

[185] W. Yang, S. Pan, E. Di Valentino, A. Paliathanasis, J. Lu, Challenging bulk viscous unified scenarios with cosmological observations, Physical Review D 100 (10) (2019) 103518.

[186] W. Yang, S. Pan, A. Paliathanasis, S. Ghosh, Y. Wu, Observational constraints of a new unified dark fluid and the H 0 tension, Monthly Notices of the Royal Astronomical Society 490 (2) (2019) 2071–2085.

[187] H. Chaudhary, A. Bouali, U. Debnath, T. Roy, G. Mustafa, Constraints on the parameterized deceleration parameter in frw universe, Physica Scripta 98 (9) (2023) 095006.

[188] A. Bouali, H. Chaudhary, U. Debnath, A. Sardar, G. Mustafa, Data analysis of three parameter models of deceleration parameter in flrw universe, The European Physical Journal Plus 138 (9) (2023) 816.

[189] A. Bouali, B. Shukla, H. Chaudhary, R. K. Tiwari, M. Samar, G. Mustafa, Cosmological tests of parametrization q= α- β h in f (q) flrw cosmology, International Journal of Geometric Methods in Modern Physics 20 (09) (2023) 2350152.

[190] A. Bouali, H. Chaudhary, S. Mumtaz, G. Mustafa, S. Maurya, Observational constraining study of new deceleration parameters in frw universe, Fortschritte der Physik 71 (10-11) (2023) 2300033.

[191] B. K. Shukla, A. Bouali, H. Chaudhary, R. K. Tiwari, M. S. Martín, Cosmographic studies of q= α- β h parametrization in $f(t)$ framework, International Journal of Geometric Methods in Modern Physics 20 (14) (2023) 2450007–8.

[192] H. Chaudhary, A. Kaushik, A. Kohli, Cosmological test of $\sigma\theta$ as function of scale factor in f (r, t) framework, New Astronomy 103 (2023) 102044.

[193] D. Arora, H. Chaudhary, S. K. J. Pacif, G. Mustafa, Diagnostic and comparative analysis of dark energy models with q (z) parametrizations, The European Physical Journal Plus 139 (5) (2024) 1–20.

[194] B. Shukla, D. Sofuoğlu, H. Chaudhary, F. Atamurotov, G. Mustafa, Cosmic evolution in $f(q,t)$ gravity with observational constraints: A comparative analysis with λcdm, Journal of High Energy Astrophysics (2024).

[195] H. Chaudhary, A. Bouali, H. Duru, E. Gudekli, G. Mustafa, Exploring the deceleration parameter in $f(t)$ gravity: A comprehensive analysis using parametrization techniques and observational data, International Journal of Geometric Methods in Modern Physics (2024).

[196] H. Chaudhary, S. Mumtaz, A. Bouali, U. Debnath, G. Mustafa, Parametrization of the deceleration parameter in a flat flrw universe: constraints and comparative analysis with the λ cdm paradigm, General Relativity and Gravitation 55 (11) (2023) 133.

[197] H. Chaudhary, A. Bouali, N. U. Molla, U. Debnath, G. Mustafa, Cosmological tests of $f(r,g,t)$ dark energy model in frw universe, The European Physical Journal C 83 (10) (2023) 918.

[198] B. K. Shukla, A. Bouali, H. Chaudhary, R. K. Tiwari, M. S. Martín, Cosmographic studies of q= α- β h parametrization in f (t) framework, International Journal of Geometric Methods in Modern Physics 20 (14) (2023) 2450007–8.

[199] M. Khurana, H. Chaudhary, S. Mumtaz, S. Pacif, G. Mustafa, Cosmic evolution in f (q, t) gravity: Exploring a higher-order time-dependent function of deceleration parameter with observational constraints, Physics of the Dark Universe 43 (2024) 101408.

[200] A. Bouali, H. Chaudhary, A. Mehrotra, S. K. J. Pacif, Model-independent study for a quintessence model of dark energy: Analysis and observational constraints, Fortschritte der Physik 71 (12) (2023) 2300086.

[201] R. Kundu, U. Debnath, H. Chaudhary, G. Mustafa, Gravitational lensing and clues of $r_\{d\}$ and $h_\{0\}$ tension: Viscous modified chaplygin gas and variable modified chaplygin gas in loop quantum cosmology, arXiv preprint arXiv:2402.07515 (2024).

[202] G. N. Gadbail, H. Chaudhary, A. Bouali, P. K. Sahoo, Statistical and Observation Comparison of Weyl-Type $f(Q,T)$ Models with the ΛCDM Paradigm (5 2023). arXiv:2305.11190.

[203] S. Dahmani, H. Chaudhary, A. Bouali, S. K. J. Pacif, T. Ouali, Polytropic gas cosmology and late-time acceleration, Available at SSRN 4697513.

[204] H. Chaudhary, U. Debnath, T. Roy, S. Maity, G. Mustafa, Constraints on the parameters of modified chaplygin-jacobi and modified chaplygin-abel gases

in $f(t)$ gravity model, arXiv preprint arXiv:2307.14691 (2023).
[205] S. Shekh, H. Chaudhary, A. Bouali, A. Dixit, Observational constraints on teleparallel effective equation of state, General Relativity and Gravitation 55 (8) (2023) 95.
[206] H. Chaudhary, N. U. Molla, M. Khurana, U. Debnath, G. Mustafa, Cosmological test of dark energy parameterizations in hořava–lifshitz gravity, The European Physical Journal C 84 (3) (2024) 223.
[207] H. Chaudhary, N. U. Molla, M. Khurana, U. Debnath, G. Mustafa, Cosmological test of dark energy parameterizations in hořava–lifshitz gravity, The European Physical Journal C 84 (3) (2024) 223.
[208] H. Chaudhary, U. Debnath, S. K. J. Pacif, N. U. Molla, G. Mustafa, S. Maurya, Investigating $\omega(z)$ parametrizations in horava-lifshitz gravity: Observational constraints with covariance matrix simulation, arXiv preprint arXiv:2402.12324 (2024).
[209] S. Maity, H. Chaudhary, U. Debnath, S. Maurya, G. Mustafa, Constraining Cosmological Parameters with Viscous Modified Chaplygin Gas and Generalized Cosmic Chaplygin Gas Models in Horava-Lifshitz Gravity: Utilizing Late-time Datasets, arXiv preprint arXiv:2403.06286 (2024).
[210] H. Bondi, On spherically symmetrical accretion, Monthly Notices of the Royal Astronomical Society 112 (2) (1952) 195–204.
[211] F. C. Michel, Accretion of matter by condensed objects, Astrophysics and Space Science 15 (1972) 153–160.
[212] J. Moffat, Do supermassive black holes exist at the center of galaxies?, arXiv preprint astro-ph/9704232 (1997).
[213] D. Merritt, L. Ferrarese, Relationship of black holes to bulges, arXiv preprint astro-ph/0107134 (2001).
[214] E. Babichev, V. Dokuchaev, Y. Eroshenko, Black hole mass decreasing due to phantom energy accretion, Physical Review Letters 93 (2) (2004) 021102.
[215] E. Babichev, V. Dokuchaev, Y. N. Eroshenko, The accretion of dark energy onto a black hole, Journal of Experimental and Theoretical Physics 100 (2005) 528–538.
[216] C.-Y. Sun, Phantom energy accretion onto black holes in a cyclic universe, Physical Review D 78 (6) (2008) 064060.
[217] M. Jamil, Evolution of a Schwarzschild black hole in phantom-like Chaplygin gas cosmologies, The European Physical Journal C 62 (2009) 609–614.
[218] M. Sharif, G. Abbas, Phantom energy accretion by a stringy charged black hole, Chinese Physics Letters 29 (1) (2012) 010401.
[219] U. Debnath, Accretion and evaporation of modified Hayward black hole, The European Physical Journal C 75 (3) (2015) 129.
[220] J. A. Jiménez Madrid, P. F. Gonzalez-Díaz, Evolution of a Kerr-Newman black hole in a dark energy universe, Gravitation and Cosmology 14 (2008) 213–225.
[221] S.-W. Kim, Y. Kang, Dark energy and dark matter accretion onto a black hole in expanding universe, in: International Journal of Modern Physics: Conference Series, Vol. 12, World Scientific, 2012, pp. 320–329.
[222] J. A. J. Madrid, P. Martín-Moruno, On accretion of dark energy onto black-and worm-holes, arXiv preprint arXiv:1004.1428 (2010).
[223] U. Debnath, Accretions of various types of dark energies onto Morris–Thorne wormhole, The European Physical Journal C 74 (2014) 1–8.
[224] S. P. Kim, S. K. Kim, K.-S. Soh, J. H. Yee, Renormalized thermodynamic entropy of black holes in higher dimensions, Physical Review D 55 (4) (1997) 2159.
[225] S. G. Ghosh, Nonstatic charged BTZ-like black holes in N+ 1 dimensions, International Journal of Modern Physics D 21 (03) (2012) 1250022.
[226] A. J. John, S. G. Ghosh, S. D. Maharaj, Accretion onto a higher dimensional black hole, Physical Review D 88 (10) (2013) 104005.
[227] S. Hendi, Charged BTZ-like black holes in higher dimensions, The European Physical Journal C 71 (2) (2011) 1551.
[228] P. Mukherjee, U. Debnath, A. Pradhan, Accretion of modified Chaplygin-Jacobi gas and modified Chaplygin-Abel gas onto Schwarzschild black hole, International Journal of Geometric Methods in Modern Physics 20 (12) (2023) 2350218–443.
[229] G. Abbas, R. Ramzan, Thermodynamics of Phantom Energy Accreting onto a Black Hole in Einstein—Power—Maxwell Gravity, Chinese Physics Letters 30 (10) (2013) 100403.
[230] B. Nayak, M. Jamil, Effect of vacuum energy on evolution of primordial black holes in Einstein gravity, Physics Letters B 709 (3) (2012) 118–122.
[231] D. Dwivedee, B. Nayak, M. Jamil, L. P. Singh, R. Myrzakulov, Evolution of primordial black holes in loop quantum cosmology, Journal of Astrophysics and Astronomy 35 (2014) 97–106.
[232] J. A. S. d. Lima, D. C. Guariento, J. E. Horvath, Analytical solutions of accreting black holes immersed in a ΛCDM model, Physics Letters B 693 (3) (2010) 218–220.
[233] M. Sharif, G. Abbas, Phantom accretion onto the Schwarzschild de-Sitter black hole, Chinese Physics Letters 28 (9) (2011) 090402.
[234] P. Martín-Moruno, A.-E. L. Marrakchi, S. Robles-Pérez, P. F. González-Díaz, Dark energy accretion onto black holes in a cosmic scenario, General Relativity and Gravitation 41 (2009) 2797–2811.
[235] M. G. Rodrigues, A. E. Bernardini, Accretion of Non-minimally Coupled Generalized Chaplygin Gas Into Black Holes, International Journal of Modern Physics D 21 (09) (2012) 1250075.
[236] P. Martín-Moruno, J. A. J. Madrid, P. F. González-Díaz, Will black holes eventually engulf the Universe?, Physics Letters B 640 (4) (2006) 117–120.
[237] G. Abbas, Phantom energy accretion onto a black hole in Hořava-Lifshitz gravity, Science China Physics, Mechanics and Astronomy 57 (2014) 604–607.
[238] X.-H. Meng, X.-l. Du, A specific case of generalized Einstein-Aether theories, Communications in Theoretical Physics 57 (2) (2012) 227.
[239] J. Zuntz, T. Zlosnik, F. Bourliot, P. Ferreira, G. Starkman, Vector field models of modified gravity and the dark sector, Physical Review D 81 (10) (2010) 104015.
[240] B. Calderhead, A general construction for parallelizing metropolis- hastings algorithms, Proceedings of the National Academy of Sciences 111 (49) (2014) 17408–17413.
[241] R. C. Bernardo, D. Grandón, J. L. Said, V. H. Cárdenas, Parametric and nonparametric methods hint dark en-

ergy evolution, Physics of the Dark Universe 36 (2022) 101017.

[242] J. Dunkley, M. Bucher, P. G. Ferreira, K. Moodley, C. Skordis, Fast and reliable markov chain monte carlo technique for cosmological parameter estimation, Monthly Notices of the Royal Astronomical Society 356 (3) (2005) 925–936.

[243] D. Stern, R. Jimenez, L. Verde, M. Kamionkowski, S. A. Stanford, Cosmic chronometers: constraining the equation of state of dark energy. i: H (z) measurements, Journal of Cosmology and Astroparticle Physics 2010 (02) (2010) 008.

[244] M. Moresco, A. Cimatti, R. Jimenez, L. Pozzetti, G. Zamorani, M. Bolzonella, J. Dunlop, F. Lamareille, M. Mignoli, H. Pearce, et al., Improved constraints on the expansion rate of the universe up to z 1.1 from the spectroscopic evolution of cosmic chronometers, Journal of Cosmology and Astroparticle Physics 2012 (08) (2012) 006.

[245] C. Zhang, H. Zhang, S. Yuan, S. Liu, T.-J. Zhang, Y.-C. Sun, Four new observational h (z) data from luminous red galaxies in the sloan digital sky survey data release seven, Research in Astronomy and Astrophysics 14 (10) (2014) 1221.

[246] M. Moresco, Raising the bar: new constraints on the hubble parameter with cosmic chronometers at z 2, Monthly Notices of the Royal Astronomical Society: Letters 450 (1) (2015) L16–L20.

[247] M. Moresco, L. Pozzetti, A. Cimatti, R. Jimenez, C. Maraston, L. Verde, D. Thomas, A. Citro, R. Tojeiro, D. Wilkinson, A 6% measurement of the hubble parameter at z 0.45: direct evidence of the epoch of cosmic re-acceleration, Journal of Cosmology and Astroparticle Physics 2016 (05) (2016) 014.

[248] A. Ratsimbazafy, S. Loubser, S. Crawford, C. Cress, B. Bassett, R. Nichol, P. Väisänen, Age-dating luminous red galaxies observed with the southern african large telescope, Monthly Notices of the Royal Astronomical Society 467 (3) (2017) 3239–3254.

[249] M. Smith, C. B. D'Andrea, M. Sullivan, A. Möller, R. Nichol, R. Thomas, A. Kim, M. Sako, F. Castander, A. Filippenko, et al., First cosmology results using supernovae ia from the dark energy survey: Survey overview, performance, and supernova spectroscopy, The Astronomical Journal 160 (6) (2020) 267.

[250] D. Brout, et al., The Pantheon+ Analysis: Cosmological Constraints, Astrophys. J. 938 (2) (2022) 110. arXiv:2202.04077, doi:10.3847/1538-4357/ac8e04.

[251] E. Barausse, T. Jacobson, T. P. Sotiriou, Black holes in Einstein-aether and Hořava-Lifshitz gravity, Physical Review D 83 (12) (2011) 124043.

[252] P. Berglund, J. Bhattacharyya, D. Mattingly, Towards thermodynamics of universal horizons in Einstein-aether theory, Physical Review Letters 110 (7) (2013) 071301.

[253] E. O. Babichev, V. I. Dokuchaev, Y. N. Eroshenko, Black holes in the presence of dark energy, Physics-Uspekhi 56 (12) (2013) 1155.

[254] U. Debnath, Accretions of dark matter and dark energy onto (n+ 2)-dimensional Schwarzschild black hole and Morris-Thorne wormhole, Astrophysics and Space Science 360 (2) (2015) 40.

[255] P. Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. Banday, R. Barreiro, N. Bartolo, et al., Planck 2018 results. vi. cosmological parameters (2020).

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Toward Multimodal Complaint Severity Detection From Social Media

Apoorva Singh, Prince Jha, Souryadip Das, Raghav Jain, and Sriparna Saha, *Senior Member, IEEE*

***Abstract*—The prevalence of complaints submitted online and the sheer volume of information made available by social media platforms highlight the need for automated complaint analysis tools. In linguistic studies, complaints have been classified according to how much of personal risk the complainant is willing to take. This is crucial information for understanding the motivations of complainants and how people come up with reasonable means of reparation. Few attempts have been made to use the existing multimodal complaint model, which focuses on improving the textual mode with the help of images, to find specific visual features that help identify complaints. Our aim is to find a solution to this problem. In order to detect complaints and the severity level associated with them in a multitask setting, we propose Multimodal framEwork for Complaint and Severity-level detectIon (MECSI), a novel multimodal framework that uses local and global attributes (in both modalities) and relates them to the textual context. To do this, we add severity-level annotation to the newly released CESAMARD dataset, which is a compilation of reviews and images of products listed on the Amazon website. The experimental findings confirmed the superiority of our proposed model over the state-of-the-art model and other strong rival baselines, proving the efficacy of our proposed framework.**

***Index Terms*—Multimodal complaint detection, multimodal dataset, multitask learning, severity-level classification, social media mining.**

I. Introduction

In linguistics, complaints are categorized based on emotional depth through severity levels, the level of face threat involved, and the intentions behind them [1]. Complaints serve to express dissatisfaction, seek clarification, or achieve both objectives. Recognizing and evaluating complaints and their severity levels in natural language holds crucial importance for businesses aiming to enhance customer relations by addressing and resolving issues. Additionally, it provides linguists with a broader scope for analyzing complaint speech acts across varying levels of granularity [2].

Manuscript received 25 July 2023; revised 27 December 2023 and 28 March 2024; accepted 11 April 2024. *(Corresponding author: Apoorva Singh.)*

Apoorva Singh, Prince Jha, Raghav Jain, and Sriparna Saha are with the Department of Computer Science and Engineering, Indian Institute of Technology, Patna 801103, India (e-mail: apoorva_1921cs19@iitp.ac.in).

Souryadip Das was with the University of Kalyani, West Bengal 741235, India. He is now with Hungerbox, Bangalore 560102, India.

Digital Object Identifier 10.1109/TCSS.2024.3393729

When making a purchasing decision, we pay attention to the associated texts and images in a review post. To ascertain the user's legitimate interest in the product, information from several modalities must be analyzed together. Multimodal inputs contribute to the development of effective classification models, which enable the identification of the customer's affective state and sentiment as well as the detection of genuine complaints and their associated severity levels.

Recent work on multimodal CD [3] has shown that off-the-shelf vision models, which work well on visual-linguistic tasks, do not help much when adapted to product images signifying complaints. Some challenges include: First, user-uploaded images can be chaotic, so relying solely on them without product and category context is often insufficient. Second, in product reviews, visual and textual information are sometimes uncorrelated and the entire semantics is provided in entirety. Fig. 1 illustrates two different complaint scenarios communicated via both modalities. The first image shows a low-quality page. The textual modality lacks explicit criticism. Therefore, relying solely on textual analysis does not reliably predict a violation of expectation. In one scenario, the text implies a neutral review, while the visuals suggest a negative allegation. These differing modalities in both cases highlight the importance of multiple evidence sources for detecting complaints.

Research objectives: The current study aims to achieve the following.

1) We intend to investigate how multimodal cues help identify complaints and associated severity levels from social media content.
2) We aim to understand the effect of polarity and emotion recognition (ER) tasks individually and in different combinations in the proposed framework.
3) This work aims to study multimodal complaint and severity-level prediction using local information and background knowledge (global attributes) for text and image modalities.

Contributions: The main contributions of the current work include the following.

1) We manually annotate the recently released multimodal complaint dataset, CESAMARD, with complaint severity labels for aiding fine-grained analysis of multimodal complaint severity-level prediction research.
2) We propose Multimodal framEwork for Complaint and Severity-level detectIon (MECSI), a novel multimodal framework that uses local and global attributes and relates

2329-924X © 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 June 05,2024 at 06:53:02 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.

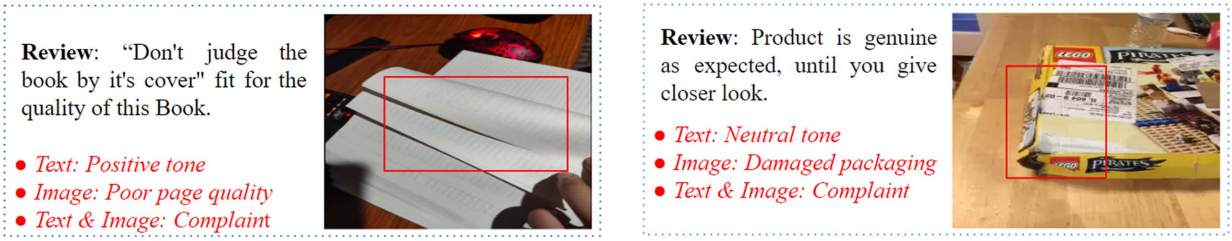


Fig. 1. Some sample instances highlight the influence of multimodal inputs in complaint detection (CD).

them to the textual context to detect complaints and associated severity-level in a multitask setting. CD and complaint severity-level classification (CS) are the main tasks and ER and polarity classification (PC) are considered additional tasks in the multitask framework.

3) We propose the benchmark model for complaint and severity-level prediction in a multimodal multitask setting.

The remainder of the manuscript follows this structure: First, we provide a concise summary of the current literature in Section II. Next, Section III delves into the details of the expanded *CESAMARD* dataset. Subsequently, in Section IV, we explain our novel multimodal multitask approach to detect complaints and severity from social media content. Section V is dedicated to discussing the experimental setup and analyzing the results. Finally, in Section VI, we present a summary and conclusion, along with a glimpse of potential future research directions.

II. Related Works

Earlier works in pragmatics interpreted complaints based on their intensity and straightforwardness [4]. Complaints were categorized into four fine-grained severity classifications, according to [1]: 1) no specific reproach; 2) disapproval; 3) accusation; and 4) blame. Prior studies in complaint diagnosis have primarily focused on distinguishing between complaints and noncomplaints. They employed various approaches including feature-based machine learning models [5], deep transformer-network-based models [6], and text-based multitask modeling techniques that incorporated polarity and emotion information for complaint mining [7], [8]. In [5], the authors used logistic regression with several different features to detect Twitter-based complaints. In a recent work on complaint mining, [9] trained multiple transformer-based networks using linguistic information to make predictions about the severity level of complaints from social media.

A. Multimodal and Multitask Learning

Recent studies such as [7], [10], [11], [12], and [13] have shown the efficacy of multitask learning where multiple related tasks are learned simultaneously by a single model. Although polarity and emotion are two separate concepts [14], they are strongly related. This is also because polarity and emotion are mostly concerned with perceptions that occur as a result of a combination of biological, psychological, and social conditions [15]. Numerous works in the literature examine the simultaneous analysis of PC and ER [10], [16], and others, along with their usage in various natural language processing tasks [17] and others. Multimodality has emerged as a major research subject in the field of emotion and polarity detection in human interactions [18], [19], [20].

B. Complaint Identification

The abundance of online information has made it increasingly difficult for companies and merchants to effectively identify and resolve complaints. This is particularly challenging when screening complaints on social networking sites, where the text fragments are often chaotic and limited by character count, and may include the use of aliases, satire, and allegations [21]. To address this, a new approach to CD has emerged that involves a multimodal approach instead of relying solely on text-based classification. Earlier research has applied methods such as semisupervised techniques, hand-crafted feature-based machine learning techniques, and deep learning models to recognize text-based complaints [5], [8], [22]. For example, logistic regression with customized features was used in [5], while a variety of neural language models supplemented by transformer networks were used in [6]. Complaints have been sorted according to the accountable department, urgency level, delivery issues, and the hazards and risks associated with the product [23], [24], [25]. Some studies have also focused on the propensity for complaints to escalate, in workplace emails, boosting ensemble models with linguistic style parameters have been employed to identify complaints [26].

Even though it has been hypothesized that multimodal sources of information (such as images along with text) might offer additional relevant data in detecting complaints, only the work by [3] has looked into this. However, while they worked to improve the textual modality, they instead used the generic ResNet-152 image classification model for the visual modality, which does not account for any of the unique characteristics of images that might be used to identify product-related complaints. As a result, we propose MECSI, a deep learning-based multimodal framework that detects complaints and correlates severity levels in a multitask setup using local and global attributes (for both modalities).

III. Corpus Description

To date, available complaint datasets have only covered textual complaints [5], [9]. We use the CESAMARD dataset[1] published in [3]. The CESAMARD dataset is a compilation

[1]https://www.iitp.ac.in/ai-nlp-ml/resources.html#CESAMARD

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 05,2024 at 06:53:02 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
COMPARATIVE STUDY OF EXISTING COMPLAINT CORPORA

Dataset	Labels				
	Complaint	Severity	Polarity	Emotion	Images
Complaints [5]	✓	×	×	×	×
Complaints [9]	✓	✓	×	×	×
Complaints [7]	✓	×	✓	✓	×
CESAMARD [3]	✓	×	✓	✓	✓
CESAMARD v2 (current work)	✓	✓	✓	✓	✓

of reviews and images of products bought from Amazon.[2] The dataset contains 3962 instances, 2641 of which are noncomplaints, and 1321 of which are complaints. Since it is a publicly available, high-quality multimodal complaint dataset, we opted to use it for the current study. According to the linguistic definition provided by [27], if a given instance of a CESAMARD review contains at least one complaint speech act, the complete review is interpreted as a complaint. There are three types of polarity to consider: negative, neutral, and positive. Anger, disgust, fear, happiness, sadness, and surprise are the six key emotions used for emotion annotation based on the work of psychologist Paul Ekman. In this article, we delve into the finer details of the CESAMARD dataset's extension (*CESAMARD v2*). The expanded *CESAMARD v2* dataset is compared to other complaint datasets in Table I.

A. Task Definition

The primary objective of this work is to extend the CESAMARD dataset with severity-level annotation. We utilize the severity classes defined in the benchmark study[3] [1]. Following are the severity classes.

1) No explicit reproach: The complaint is of a neutral tone, and the reason for the complaint is not expressed explicitly.
2) Disapproval: Complaint explicitly shows negative emotions such as annoyance or discontent. The claim may imply that the complainant holds the complainee accountable without saying so.
3) Accusation: Complainant directly charges someone of wrongdoing.
4) Blame: Complainant holds someone responsible for wrongdoing.

B. Annotation

The three annotators (one doctoral and two undergraduate computer science students) with sufficient domain knowledge and experience in developing supervised corpora annotated the severity classes for each instance in the dataset. The selection process for the three students underwent a rigorous competition, allowing participation exclusively from the computer science department with expertise in natural language processing. Initially, an adept doctoral student having experience in the CD domain annotated 50 examples for each severity class, establishing gold standard references for annotation and ambiguity resolution. Students were provided with the annotation guidelines outlined in Table II. Following this, they annotated 50 samples, and based on the quality, accuracy, and semantic coherence of their annotations, three students were selected to proceed with annotating the CESAMARD dataset samples.

TABLE II
SEVERITY LEVEL ANNOTATION INSTRUCTIONS

S.No.	Annotation Guidelines
1	Severity levels classify complaints by nature rather than intensity.
2	"No explicit reproach" is an indicative approach in which the complainee is typically left out of the statement.
3	"Disapproval" conveys only an undesirable feeling or situation.
4	In "Blame," the complainer believes the complainee is directly responsible for the crime.
5	In case of ambiguity, annotators, and authors collaborate to discuss and clarify the instance.

Majority voting determined the final severity classes. In cases where the annotations differed among the three annotators, the doctoral student expert endeavored to resolve any ambiguity or uncertainty. Instances where a consensus was not reached among annotators were excluded from consideration. To measure the interrater agreement, we computed Fleiss' kappa score and achieved an agreement score of 0.67 (k) for the severity task, indicating high-quality annotations and significant agreement [28]. In the work [3], the authors state the interrater agreement scores of 0.83, 0.69, and 0.81 on the complaint, emotion, and polarity task, respectively, depicting good quality annotation. The distribution of tweets across complaint severity classes is as follows: *Accusation* (254), *Blame* (238), *Disapproval* (433), and *No explicit reproach* (396). Table III shows examples from the *CESAMARD v2* dataset.[4]

IV. METHODOLOGY

This section details the specifics of the proposed architecture. Fig. 2 depicts the proposed MECSI model's architecture.

Problem definition: We aim to detect complaints and associated severity levels from product reviews in a multimodal, multitask setting. The training data consists of R review instances, and is represented as $\{[T, I, W, c, s, e, p]_i\}_{i=1}^{N}$, where T denotes

[2]https://www.amazon.in/

[3]According to [1], classes are disconnected.

[4]The dataset is available here: https://github.com/appy1608/IEEETCSS_Multimodal-Complaint-Severity-Detection

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 05,2024 at 06:53:02 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE III
FEW EXAMPLES FROM THE *CESAMARD v2* CORPUS. COMP: COMPLAINT

Title	Review	Label	Severity	Emotion	Polarity
Poor printing quality of the book	Received the book. Many pages of the book are blank, not readable due to poor printing. Disappointed!	Comp	Disapproval	Sadness	Negative
Rotten product	The package when opened was full of worms, the lentil was foul smelling. Picture attached. Coming from Tata's surprised; disappointed.	Comp	Accusation	Disgust	Negative
Defective product	Seems such as a quality defective product is being sold online... without even checking! There is a button missing in the shirt I received!!	Comp	Blame	Anger	Negative

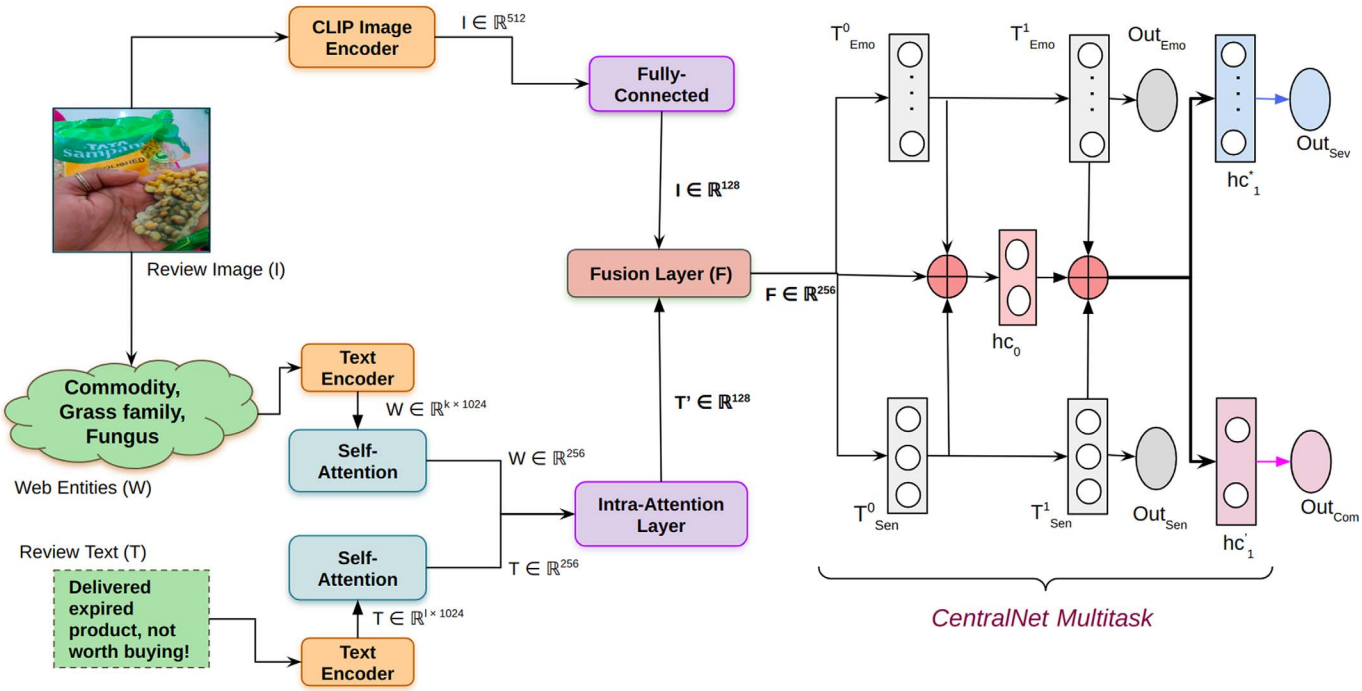


Fig. 2. Architectural diagram of the proposed model, MECSI. T^0_{Emo} is the hidden representation of Emotion task-specific network at 0th layer. T^0_{Sen} is the hidden representation of sentiment task-specific network at 0th layer.

the review text, I is the review image, and W is the web entity tag, c denotes the complaint label, s is the severity label, e is the emotion class label, and p is the polarity class label.

A. Feature Composition Module

This section discusses the feature extraction techniques used for different modules.

1) *CLIP*: For image encoding we utilize contrastive language-image pretraining (CLIP) [29], a pretrained visual-linguistic model, exploiting its encoding abilities to preserve the whole semantic. CLIP was trained using 400 million image-text pairs obtained from the Internet. CLIP provides excellent zero-shot capabilities due to the large number of images it has encountered and the natural language monitoring. We retrieved the CLIP image embedding, F_I which is a 512-dimensional vector.
2) *BERT*: For the review-text, the wordwise textual features are obtained by averaging the last four hidden states of a pretrained bidirectional encoder representations from transformers model [30]. We use *bert-large-uncased* model[5] (uncased: 12-layer, 1024-hidden, 12-heads, 110 M parameters).
3) *Web entities*: Inspired by [31], we introduce web entities[6] as a more advanced image concept for our model. The review images displayed in Fig. 3 manifest entities such as fabric, juicer, grinder-mixer, outerwear, etc., presenting contextual information corresponding to each image, aiding in complaint identification. Leveraging the Google Cloud Vision API,[7] we foreground various web entities.

[5]https://huggingface.co/bert-large-uncased
[6]Objects of interest are highlighted within rectangular bounding boxes.
[7]https://cloud.google.com/vision/docs/detecting-web

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 05,2024 at 06:53:02 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. 3. Identified web entities in a selection of images sourced from the *CESAMARD v2* dataset.

Let us assume that for a given review image input, the entities are denoted as $P_1, P_2, \ldots, P_k$. Each web entity is encoded using a $\text{BERT}_{\text{large}}$ encoder to create 1024-dimensional features. These features are represented as $O_f = f_1, f_2, \ldots, f_k$, where $O_f \in R^{k\times 1024}$.

B. Attention Technique

We compute attention by associating a query with key-value pairs to produce an output. This output is determined by the weighted sum of values, with weights assigned based on the compatibility function between a query and its respective key.

1) *Self attention*: We utilize self-attention across l review texts and k web entities to allocate weights to the most pertinent entities based on their relevance or importance within the given context. These weights are calculated by assessing the compatibility between each entity and every other entity in the respective set, individually for both the review texts and the web entities. Consequently, the resulting weights signify the significance of each entity relative to the others, enabling the model to prioritize the most relevant entities during processing. The resultant self-attended representations can be expressed as: $\text{SA}_i = \text{softmax}(Q_i K^T i)Vi$, where i belongs to the set of web entities, review text.
2) *Intramodality attention*: Review text is significant for some product reviews meanwhile web entities provide background context that acts as a higher-level concept regarding the review image. Hence both the modalities (text and web entity) play crucial roles in their own way. To merge both textual modalities, intramodality attention [32] has been employed. Both the self-attended representations of review text and web entities are concatenated into a vector and passed through dense layers. Finally, a sigmoid activation function is used to calculate attention scores. These attention values are the weights of each modality, and the resulting modality representations T', which is further fed into the fusion layer.

C. Fusion Layer

Both textual and visual modalities may not be equally significant for identifying the complaint. This encourages the development of a fusion network that takes into account each modality independently, similar to the work of [33].

The fusion layer takes in the textual modality $M_{T'}$ and visual modality M_I. These modalities are combined by concatenating them to create a vector V. This vector then undergoes a dense layer followed by a sigmoid activation function, computing attention scores. These scores determine the importance of each modality, and their representations are combined to form the fusion vector F. The mathematical specifics are detailed in the following equation:

$$\begin{aligned} V &= M_{T'} \bigoplus M_I \\ \text{Att}_{T'} &= \sigma(W_{T'}^T V + b_{T'}) \\ \text{Att}_I &= \sigma(W_I^T V + b_I) \\ F &= \text{Att}_{T'}.M_{T'} \bigoplus \text{Att}_I.M_I. \end{aligned} \tag{1}$$

F is then passed through CentralNet Multitask Network for the decision process.[8]

D. CentralNet Multitask

Inspired by the line of works made by [33], we introduce CentralNet Multitask. It is a neural network architecture with n independent subnetworks, where $n-1$ subnetworks are devoted to auxiliary tasks and 1 is for the main task. Complaints can be identified and their severity can be predicted at a standard layer known as the central network. The central network accumulates features learned for individual tasks by considering a weighted summation of task-specific networks and their previous layers (polarity, emotion). The multitask layers can be defined as follows:

$$h_{c_{i+1}} = \alpha_{c_i} h_{c_i} + \sum_{k=1}^{n-1} \alpha_{T_i^k} T_i^k$$

where $n-1$ denotes number of task-specific networks, α's are scalar trainable weights, h_{c_i} is the central hidden representation of the main task and T_k^i is the hidden representation of kth task-specific network at ith layer. $h_{c_{i+1}}$ is the ultimate layer, which passed to a fully connected layer accompanied by an activation function. The initial layer of CentralNet Multitask is a shared representation obtained from the fusion layer.

Loss function: The categorical-cross entropy (J_{CE}) losses are calculated for the complaint (CD), severity (CS), emotion (ER),

[8] $\{W_{T'}^T, W_I^T, b_{T'}, b_I\}$ are learnable parameters.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 05,2024 at 06:53:02 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 polarity (PC) tasks. The integrated loss function $(J(\theta))$ is realized as follows:

$$J(\theta) = J_{\text{CE}}^{\text{CD}}(\theta) + J_{\text{CE}}^{\text{CS}}(\theta) + J_{\text{CE}}^{\text{ER}}(\theta) + J_{\text{CE}}^{\text{PC}}(\theta)$$

All the model parameters to be optimized are denoted by θ, which ranges from 0 to 1, and decides the per task loss-share to the total loss.

V. Experiments and Results

A. Baselines

We evaluate the proposed MECSI model with several strong baselines, divided into three classes: multimodal methods, multitask systems, and ablation models.

1) Multimodal methods: These baselines include various multimodal methods used to evaluate the visual modality, and textual and visual modality together.

a) *EfficientNet [34]:* We utilize EfficientNet for extracting visual features instead of CLIP model in the proposed framework. EfficientNet is a family of models obtained by effectively scaling up MobileNet and ResNet.

b) *ViLBERT [35]:* To acquire a visual foundation for paired visual-linguistic data, we use a collaborative technique known as vision and language BERT (ViLBERT) [35]. It has separate pathways for understanding language and visual, which communicate with one another through coattentional transformer layers. In order to work together, different modalities can share attention at separate representational levels, which is made possible by coattentional layers. Compared to the single-stream unified model, it performs better on a wider range of activities.

c) *UNITER [36]:* As one of the multimodal baselines, we additionally employ the state-of-the-art (SOTA) UNiversal Image-TExt Representation (UNITER) model [36]. UNITER is a pretrained large-scale model, which leverages its elegant self-attention mechanism built for learning contextualized representations by using transformers as the core of its model. UNITER is pretrained on four pretraining tasks relating to language modeling, masked region modeling, word-region alignment, and image-text matching, as inspired by BERT [30].

d) *VGG-19:* For the visual modality we also used a pretrained VGG-19[9] model to extract the image features. Each bounding box is clipped and resized before being passed into VGG-19, encoding it into a 4096-dimensional representation.

2) Multitask systems: The *Baseline$_1$* model (text only) is a unimodal multitask baseline developed inspired by the recent work in CD in multitask framework [7]. It is an adversarial multitask framework that uses BERT to categorize complaints, emotions, polarity, and sarcasm. By retaining the same experimental setting as in our current work, we reimplement the Baseline$_1$ model for the simultaneous learning of CD, CS, ER, and PC tasks.

3) Ablation models: To gauge the individual effects of ER and PC tasks on the primary tasks CD and CS, we create several ablation models while keeping CD and CS tasks consistent across all cases. The architecture remains unchanged except for these specific tasks.

B. Experimental Setup

We used Python-based libraries: Pytorch[10] and Tensorflow.[11] Our experiments were conducted on a hybrid cluster equipped with multiple GPUs, featuring RTX 2080Ti. We reported the weighted average F1 value along with the model accuracy. In all our experimental models, 80% of the *CESAMARD v2* dataset served as training data (10% for validation), while the remaining 20% was allocated for testing. The *K*-fold cross-validation technique, employing seven folds, was used to train the model. In the dense layers, ReLU activation was utilized to prevent overfitting, followed by dropout rates of 30%, with dropout ϵ values ranging from 20% to 50% after the dense layers. For severity, emotion, and polarity detection tasks, output layers utilized Softmax activation with 5, 6, and 3 neurons, respectively. The Sigmoid activation function was applied for complaint classification. The framework was trained with a batch size of 32 and learning rate settings of lr = 1e$-$4, where $lr\epsilon$ 1e$-$1, 1e$-$2, 1e$-$3, 1e$-$4, $\beta_1 = 0.9$, and $\beta_2 = 0.999$, utilizing the Adam optimizer.

C. Results and Discussion

The current study aims to improve CD and CS tasks using multimodal information such as texts, images, and web entities, as well as information from additional tasks (ER and PC). Consequently, we state the findings and related discussion with CD and CS as key tasks in all combinations.

Table IV depicts the classification results from the various experiments for the main tasks of CD and CS on accuracy and weighted-average F1 score metrics. As shown in Table IV, our model outperforms all the other models. Compared to baselines without web entity attributes, variants with web entity information perform well for all the multimodal baselines and the proposed MECSI$_{\text{All}}$ model. Overall the multimodal frameworks outperform the unimodal models. Moreover, the SOTA vision-text models with underlying transformer architecture such as ViLBERT and UNITER show lower performances than the MECSI (BERT + CLIP-Image, BERT + CLIP-Image + Web Entity) variants, illustrating the efficient usage of self-attention, intraattention and fusion layer in the proposed framework. In terms of tasks, MECSI$_{\text{All}}$ which includes all four tasks (CD, CS, ER, and PC), outperforms all the other variants (MECSI$_{\text{CD + CS}}$, MECSI$_{\text{CD + CS + ER}}$, MECSI$_{\text{CD + CS + PC}}$). Besides, the auxiliary task PC variant of the proposed model (MECSI$_{\text{CD + CS + PC}}$) outperforms the ER variant (MECSI$_{\text{CD + CS + ER}}$). This could be motivated by the fact that emotion, as a fine-grained task, is particularly context-dependent and therefore unable to add significantly to the multitask model by itself. It is also worth

[9] https://keras.io/api/applications/vgg/

[10] https://pytorch.org

[11] https://www.tensorflow.org/

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 05,2024 at 06:53:02 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE IV
EXPERIMENTAL RESULTS IN TERMS OF WEIGHTED-AVERAGE F1 SCORE (F1) AND ACCURACY (A) VALUE FOR THE TWO MAIN TASKS OF COMPLAINT AND SEVERITY DETECTION

		Complaint (CD)		Severity (CS)	
Type	**Model**	**F1**	**A**	**F1**	**A**
Text-only	**SOTA$_{CS}$** [9]	-	-	57.67	58.28
Text and image	**SOTA$_{CD}$** [3]	89.07	89.64	-	-
Text-only	**Baseline$_1$** [7]	85.04	86.48	64.47	65.27
Multimodal methods	**[BERT + VGG19 + WebEntity]$_{All}$**	88.44	88.67	72.54	69.38
	[BERT + EfficientNet]$_{All}$	87.94	87.75	69.92	72.88
	[BERT + EfficientNet + WebEntity]$_{All}$	88.84	89.04	69.65	73.16
	VilBERT$_{All}$	77.25	77.42	67.59	69.74
	[VilBERT + Web Entity]$_{All}$	79.92	79.87	67.38	70.56
	UNITER$_{All}$	75.97	75.72	60.56	66.88
	[UNITER + Web Entity]$_{All}$	78.97	78.81	63.86	68.52
	[BERT + CLIP-Image]$_{All}$	90.72	90.33	70.4	73.35
Proposed model and ablations	**MECSI$_{CD+CS}$**	88.13	88.89	70.66	74.05
	MECSI$_{CD+CS+ER}$	90.53	90.18	71.16	74.43
	MECSI$_{CD+CS+PC}$	90.68	90.65	71.06	74.64
	MECSI$_{All}$	**91.47**†	**91.38**†	**71.26**†	**74.87**†

Note: F1, A metrics are given in %. All: Signifies all the four tasks (CD, CS, ER, and PC) have been considered for the experiment. Bold-faced values represent the maximum scores achieved.
†: Signifies statistically significant findings.

TABLE V
EXPERIMENTAL RESULTS IN TERMS OF WEIGHTED-AVERAGE F1 SCORE (F1) AND ACCURACY (A) VALUE FOR SENTIMENT CLASSIFICATION AND EMOTION DETECTION TASKS

		Polarity (PC)		Emotion (ER)	
Type	**Model**	**F1**	**A**	**F1**	**A**
Text and image	**SOTA$_{CD}$** [3]	56.18	56.87	45.71	46.25
Text-only	**Baseline$_1$** [7]	52.87	54.05	41.33	42.89
Multimodal methods	**[BERT + VGG19 + WebEntity]$_{All}$**	60.08	61.37	50.14	50.78
	[BERT + EfficientNet]$_{All}$	58.94	59.65	50.32	51.78
	[BERT + EfficientNet + WebEntity]$_{All}$	60.14	60.59	51.11	51.54
	VilBERT$_{All}$	56.38	56.90	44.29	45.34
	[VilBERT + Web Entity]$_{All}$	57.76	58.44	45.08	45.83
	UNITER$_{All}$	54.62	55.22	43.56	44.75
	[UNITER + Web Entity]$_{All}$	55.57	56.16	45.49	46.52
	[BERT + CLIP-Image]$_{All}$	61.61	62.83	47.22	47.95
Proposed model and ablations	**MECSI$_{CD+PC}$**	61.07	61.91	-	-
	MECSI$_{CD+ER}$	-	-	46.92	47.53
	MECSI$_{CS+PC}$	60.11	60.21	-	-
	MECSI$_{CS+ER}$	-	-	45.66	46.48
	MECSI$_{CD+CS+PC}$	62.17	63.29	-	-
	MECSI$_{CD+CS+ER}$	-	-	50.08	52.13
	MECSI$_{All}$	**63.66**†	**64.34**†	**51.85**†	**52.87**†

Note: All: Signifies all the four tasks (CD, CS, ER, and PC) have been considered for the experiment. Bold-faced values represent the maximum scores achieved.
†: Signifies statistically significant findings.

noting that, among the three tasks (CS, ER, and SA), the CS task enhances the CI task's performance the most. We additionally notice that the text-only model achieves the best accuracy and weighted-average F1 score metrics, with 86.48% and 85.04%, respectively; this suggests that textual information is essential. In Table V, the results of the auxiliary tasks for sentiment

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 05,2024 at 06:53:02 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 emotion detection are presented. It is evident that the *[BERT + CLIP-Image]$_{All}$* baseline stands as the second-best performing model for the sentiment task, achieving an accuracy of 62.83%. For the emotion task, the *[BERT + EfficientNet]$_{All}$* model performs well, boasting an overall accuracy of 51.78%. The results shown in this work are statistically significant.[12]

1) Comparison With SOTA: We also compare our proposed approach with the existing SOTA techniques for CD and CS tasks. For multimodal CD, the SOTA approach (SOTA$_{CD}$) [3] employs an attention-based adversarial multitask model. Table IV demonstrates that the proposed approach, MECSI, along with its majority of variants surpasses the results of the SOTA$_{CD}$ model. Furthermore, for the CS task, we compare our proposed MECSI framework with the SOTA framework proposed by [9]. Kindly note the SOTA model is a text-based model for CS tasks; we chose this model for comparison as we are unaware of any other multimodal complaint severity prediction framework. SOTA$_{CS}$ utilizes an array of transformer networks with additional linguistic features. The proposed MECSI model outperforms the SOTA$_{CS}$ technique for the primary task of complaint severity prediction.

D. Error Analysis

Expressions of disapproval or accusation-related terms within reviews are seldom misclassified as noncomplaints. Here are some reasons for the proposed model's occasional failure to detect complaints and severity.

1) Disguised complaints: In scenarios where the overall sentiment appears neutral but the content is a complaint, the MECSI$_{All}$ model mistakenly categorizes the instance as noncomplaint. For instance, consider the following text: "Family fun pack! They supposedly forgot to mention its as ant family. There were live ants in it." While the actual label for this example is "complaint," the model classifies it as "noncomplaint," potentially due to the absence of explicit terms indicating disapproval.
2) Proximate severity-level errors: Reviews that belong to adjoining severity levels are similar in semantics, leading to incorrect predictions. For example, "Country of Origin is <USER>, but in product description over <USER> is mentioned <USER>. Disappointing with false details mentioned in the description." The severity level predicted is "disapproval," but the actual label is "accusation." One of the possible reasons for misclassification could be the presence of negative words such as *disappointing*, which is generally associated with the disapproval severity class.

E. Limitations

Although our approach outperforms the baselines in a variety of single-task and multitask scenarios, there are certain drawbacks to our approach, which are addressed as follows.

1) While expressing complaints, certain users employ an implied tone that includes sarcasm or irony. The proposed system, however, does not specifically detect sarcasm or irony as distinct tasks. Consequently, the approach might not effectively identify complaints that contain subtle forms of sarcasm.
2) When voicing complaints, some users prefer sharing only text-based reviews, and some just share the image of the product. In such cases, if only one modality is present, the proposed model will not be able to identify the complaint. In the current work, both modalities are needed for the model to correctly proceed further with predictions.
3) The model's training and evaluation were solely based on reviews composed in English. To extend its applicability to other languages or scenarios involving code-mixed language interactions, further training in diverse languages would be necessary.

VI. Conclusion

In this work, we have explored the tasks of complaint and severity-level detection from social media. The framework is constructed based on a multimodal multitask setting such that complaint and severity-level predictions are determined based on product reviews consisting of images and associated texts. Previous work on multimodal CD focused on enhancing the textual modality; without paying heed to specific image features, which could be beneficial in identifying complaints. We propose MECSI, a novel multimodal framework that uses local and global attributes (for both image and text) to detect complaints and associated severity-level in a multitask setting. Extensive studies were conducted to illustrate the efficacy of our proposed model.

Future research will focus on recognizing multimodal complaints on a fine-grained aspect level, in addition to finding the cause of complaints.

Acknowledgment

This publication is an outcome of the R&D work undertaken in the project under the Visvesvaraya Ph.D. Scheme of Ministry of Electronics & Information Technology, Government of India, being implemented by Digital India Corporation (Formerly Media Lab Asia).

References

[1] A. Trosborg, *Interlanguage Pragmatics: Requests, Complaints, and Apologies*, vol. 7. Berlin, Germany: Walter de Gruyter, 2011.
[2] C. Vásquez, "Complaints online: The case of tripadvisor," *J. Pragmatics*, vol. 43, no. 6, pp. 1707–1717, 2011.
[3] A. Singh, S. Dey, A. Singha, and S. Saha, "Sentiment and emotion-aware multi-modal complaint identification," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 11, 2022, pp. 12163–12171.
[4] E. Olshtain and L. Weinbach, "Complaints: A study of speech act behavior among native and nonnative speakers of Hebrew. The pragmatic perspective," pp. 1–195, 1987.
[5] D. Preoţiuc-Pietro, M. Gaman, and N. Aletras, "Automatically identifying complaints in social media," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 5008–5019.
[6] M. Jin and N. Aletras, "Complaint identification in social media with transformer networks," 2020, *arXiv:2010.10910*.

[12]The results are statistically significant when testing the null hypothesis (p-value < 0.05).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 05,2024 at 06:53:02 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.

[7] A. Singh, A. Nazir, and S. Saha, "Adversarial multi-task model for emotion, sentiment, and sarcasm aided complaint detection," in *Proc. Adv. Inf. Retrieval 44th Eur. Conf. IR Res. (ECIR)*, Stavanger, Norway, M. Hagen et al., Eds., vol. 13185. New York, USA: Springer, 2022, pp. 428–442.

[8] A. Singh, R. Bhatia, and S. Saha, "Complaint and severity identification from online financial content," *IEEE Trans. Comput. Social Syst.*, vol. 11, no. 1, pp. 660–670, Feb. 2024.

[9] M. Jin and N. Aletras, "Modeling the severity of complaints in social media," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol. (NAACL-HLT)*, K. Toutanova et al., Eds., Pennsylvania, USA: Association for Computational Linguistics, 2021, pp. 2264–2274.

[10] T. Saha, A. Upadhyaya, S. Saha, and P. Bhattacharyya, "Towards sentiment and emotion aided multi-modal speech act classification in twitter," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol., (NAACL-HLT)*, K. Toutanova et al., Eds., Pennsylvania, USA: Association for Computational Linguistics, 2021, pp. 5727–5737.

[11] N. Majumder, S. Poria, H. Peng, N. Chhaya, E. Cambria, and A. Gelbukh, "Sentiment and sarcasm classification with multitask learning," *IEEE Intell. Syst.*, vol. 34, no. 3, pp. 38–43, May/Jun. 2019.

[12] A. Singh, S. Saha, M. Hasanuzzaman, and K. Dey, "Multitask learning for complaint identification and sentiment analysis," *Cogn. Comput.*, vol. 14, no. 1, pp. 212–227, 2022

[13] A. Singh and S. Saha, "Are you really complaining? A multi-task framework for complaint identification, emotion, and sentiment classification," in *Proc. Int. Conf. Document Anal. Recognit.*, New York, USA: Springer, 2021, pp. 715–731.

[14] N. Majumder, S. Poria, D. Hazarika, R. Mihalcea, A. Gelbukh, and E. Cambria, "DialogueRNN: An attentive RNN for emotion detection in conversations," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 6818–6825.

[15] G. E. Lindzey and E. E. Aronson, "The handbook of social psychology," vol. 1, 1998.

[16] S. Poria, I. Chaturvedi, E. Cambria, and A. Hussain, "Convolutional MKL based multimodal emotion recognition and sentiment analysis," in *Proc. 16th Int. Conf. Data Mining (ICDM),* Barcelona, Spain, F. Bonchi, J. Domingo-Ferrer, R. Baeza-Yates, Z. Zhou, and X. Wu, Eds., IEEE Computer Society, 2016, pp. 439–448.

[17] H. Zhou, M. Huang, T. Zhang, X. Zhu, and B. Liu, "Emotional chatting machine: Emotional conversation generation with internal and external memory," in *Proc. 32nd AAAI Conf. Artif. Intell., (AAAI), 30th Innov. Appl. Artif. Intell. (IAAI), 8th AAAI Symp. Educ. Adv. Artif. Intell. (EAAI)*, New Orleans, LO, USA, S. A. McIlraith and K. Q. Weinberger, Eds., Washington D.C, USA: AAAI Press, 2018, pp. 730–739.

[18] T. Saha, A. P. Patra, S. Saha, and P. Bhattacharyya, "Towards emotion-aided multi-modal dialogue act classification," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics, (ACL)*, D. Jurafsky, J. Chai, N. Schluter, and J. R. Tetreault, Eds., Pennsylvania, USA: Association for Computational Linguistics, 2020, pp. 4361–4372.

[19] S. Poria, D. Hazarika, N. Majumder, G. Naik, E. Cambria, and R. Mihalcea, "MELD: A multimodal multi-party dataset for emotion recognition in conversations," in *Proc. 57th Conf. Assoc. Comput. Linguistics (ACL)*, Florence, Italy, A. Korhonen, D. R. Traum, and L. Marquez, Eds., Pennsylvania, USA: Association for Computational Linguistics, 2019, pp. 527–536.

[20] K. Maity, P. Jha, S. Saha, and P. Bhattacharyya, "A multitask framework for sentiment, emotion and sarcasm aware cyberbullying detection from multi-modal code-mixed memes," in *Proc. 45th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2022, pp. 1739–1749.

[21] S. Pawar, N. Ramrakhiyani, G. K. Palshikar, and S. Hingmire, "Deciphering review comments: Identifying suggestions, appreciations and complaints," in *Proc. Int. Conf. Appl. Natural Lang. Inf. Syst.*, New York, USA: Springer, 2015, pp. 204–211.

[22] A. Singh, S. Saha, M. Hasanuzzaman, and A. Jangra, "Identifying complaints based on semi-supervised mincuts," *Expert Syst. Appl.*, vol. 186, Dec. 2021, Art. no. 115668.

[23] W. Yang et al., "Detecting customer complaint escalation with recurrent neural networks and manually-engineered features," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol., (NAACL-HLT)*, Minneapolis, MN, USA, A. Loukina, M. Morales, and R. Kumar, Eds., Pennsylvania, USA: Association for Computational Linguistics, 2019, pp. 56–63.

[24] D. Law, R. Gruss, and A. S. Abrahams, "Automated defect discovery for dishwasher appliances from online consumer reviews," *Expert Syst. Appl.*, vol. 67, pp. 84–94, Jan. 2017.

[25] A. Chopra, M. Arora, and S. Pandey, "Delivery issues identification from customer feedback data," 2021, *arXiv:2112.13372*.

[26] K. Coussement and D. Van den Poel, "Improving customer complaint management by automatic email classification using linguistic style features as predictors," *Decis. Support Syst.*, vol. 44, no. 4, pp. 870–882, 2008.

[27] A. D. Cohen and E. Olshtain, "The production of speech acts by EFL learners," *Tesol Quart.*, vol. 27, no. 1, pp. 33–56, 1993.

[28] R. Artstein and M. Poesio, "Inter-coder agreement for computational linguistics," *Comput. Linguistics*, vol. 34, no. 4, pp. 555–596, 2008.

[29] A. Radford et al., "Learning transferable visual models from natural language supervision," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2021, pp. 8748–8763.

[30] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[31] Y. Cai, H. Cai, and X. Wan, "Multi-modal sarcasm detection in Twitter with hierarchical fusion model," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 2506–2515.

[32] T. Saha, A. Upadhyaya, S. Saha, and P. Bhattacharyya, "Towards sentiment and emotion aided multi-modal speech act classification in Twitter," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol.*, 2021, pp. 5727–5737.

[33] P. Jha et al., "Combining vision and language representations for patch-based identification of Lexico-semantic relations," in *Proc. 30th ACM Int. Conf. Multimedia (ACM MM)*, 2022, pp. 4406–4415.

[34] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2019, pp. 6105–6114.

[35] J. Lu, D. Batra, D. Parikh, and S. Lee, "Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 13–23.

[36] Y.-C. Chen et al., "Uniter: Universal image-text representation learning," in *Proc. Eur. Conf. Comput. Vis.*, New York, USA: Springer, 2020, pp. 104–120.



Apoorva Singh received the M.tech. degree in computer science from Birla Institute of Technology, Mesra, India, in 2016.

She is a Research Scholar with the Department of Computer Science and Engineering, Indian Institute of Technology Patna, Patna, India. She has been working toward the research in text mining, information retrieval and applications of federated learning.



Prince Jha is an undergraduate student majoring in B.tech. computer science and engineering with the Indian Institute of Technology Patna, Patna, India, in 2023.

His area of research includes social media, computational linguistics, and lexical semantics. He has authored research articles in top-tier conferences such as *ACM SIGIR*, *ACM MultiMedia*, and *ECIR*.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 05,2024 at 06:53:02 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.



Souryadip Das received the M.S. degree in data science from the University of Kalyani, West Bengal, India, in 2022.

He is working as an SDE/Data Scientist with Hungerbox.



Raghav Jain received the B.tech. degree in software engineering from Delhi Technological University, New Delhi, India, in 2021.

He is currently working as a Research Fellow with Indian Institute of Technology Patna, Patna, India. His research interests include natural language generation and multimodal learning.



Sriparna Saha (Senior Member, IEEE) is currently an Associate Professor with the Department of Computer Science and Engineering, Indian Institute of Technology Patna, Patna, India. She received the M.tech. and Ph.D. degrees in computer science from the Indian Statistical Institute, Kolkata, India, in 2005 and 2011, respectively. She is the author of a book published by Springer-Verlag. She has authored or coauthored more than 290 papers. Her research interests include deep learning, natural language processing, machine learning, information extraction, text mining, bioinformatics and multiobjective optimization. She is involved with different sponsored research projects in the broad areas of artificial intelligence and machine learning technologies, funded by different government and private agencies.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 05,2024 at 06:53:02 UTC from IEEE Xplore. Restrictions apply.

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.2024.3397890

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Digital Object Identifier

Towards Energy-Efficiency: Integrating DRL and ze-RIS for Task Offloading in UAV-MEC Environments

MUHAMMAD NAQQASH TARIQ [1], JINGYU WANG [1], (Senior Member, IEEE), SAIFULLAH MEMON [1], MOHAMMAD SIRAJ [2], (Senior Member, IEEE), MAJID ALTAMIMI [2], (Member, IEEE), and MUHAMMAD AYZED MIRZA[3].

[1]State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications.
[2]Department of Electrical Engineering, College of Engineering, King Saud University, Riyadh 11543, Saudi Arabia.
[3]School of Computer Science and Information Engineering, Qilu Institute of Technology, Jinan, Shandong 250200 P.R, China.

The authors are grateful to the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunication (BUPT), Beijing, China, for providing the Laboratory Services. 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 Unmanned aerial vehicles (UAVs) play an important role within mobile edge computing (MEC) networks in improving communications for ground users during emergency situations. However, sustaining high-quality service for extended periods is challenging because of constraints on battery capacity and computing capabilities of UAVs. To address this issue, we leverage zero-energy reconfigurable intelligent surfaces (ze-RIS) within UAV-MEC networks and introduce a comprehensive strategy that combines task offloading and resource sharing. A deep reinforcement learning (DRL) driven energy efficient task offloading (DEETO) scheme is presented. The primary objective is to minimize UAV energy ingestion. DEETO aims to enhance task offloading decision mechanism, computing and communication resource allocation, while adopting hybrid task offloading mechanism with intelligent RIS phase-shift control. We begin by modeling it as a DRL problem, structuring it as a Markov decision process (MDP), and subsequently resolving it effectively through the use of the advantage actor-critic (A2C) algorithm. Our simulation results highlight the superiority of the DEETO scheme compared to alternative approaches. DEETO excelled by achieving a notable energy savings of 16.98% from the allocated energy resources, coupled with the highest task turnover rate of 94.12%, all achieved within a shorter learning time frames per second (TFPS) and yielding higher rewards.

INDEX TERMS DRL, MEC, Task offloading, UAV, ze-RIS.

I. INTRODUCTION

The mobile edge computing (MEC) networks have evolved as a very potential technological boost the quality of service (QoS) for tasks that demand significant computational power and low latency [1]. When MEC networks incorporate unmanned aerial vehicles (UAVs) into their infrastructure, they can harness the flexibility of UAV deployment [2]. This means that they can quickly expand the communication coverage and also guarantee the connectivity of the mobile users, particularly in cases of the destruction of the terrestrial communication infrastructures or a sudden influx of large numbers of users [3].

However, deploying additional UAVs in MEC networks comes with substantial costs and increased power consumption caused by the constraints in UAVs' energy and computational resources [4]. Moreover, UAV-assisted MEC systems encounter several hurdles. For instance, in urban settings, direct links between UAVs and IoT devices may encounter obstacles such as buildings, leading to significant deterioration in channel conditions. Similarly, since UAV is energy limited, it becomes critical to co-optimize UAV trajectory and computation energy to improve the overall energy efficiency. To take full advantage of UAV integration benefits in MEC and to further enhance the IoT task offloading performance, there is increasing interest in exploiting a new paradigm which is referred to as reconfigurable intelligent surfaces (RIS). RIS has emerged as an energy-efficient alternative capable of enhancing network capacity [5].

With the integration of a energy-efficient, cost-effective and high-gain metasurface, wireless propagation environ-

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.2024.3397890

ment can be modified by RIS. This reconfiguration can minimize the energy consumption of the MEC networks and enhances communication capacity [6]. Recent research, such as studies in [7] through [8], has explored the applicability of RIS in MEC systems. Lately, numerous studies have delved into the theoretical exploration of leveraging RIS for network optimization. For instance, a notable work [9] focuses on integrating RIS into MEC systems to facilitate device task offloading. It seeks to enhance computational performance by maximizing the summation of computational bits. This is achieved through a joint optimization approach involving time allocation, transmit power and CPU frequency for computational offloading, while adjusting phase shifts of the RIS. The RIS beamforming and sensing mechanism for near field communications is investigated in [10]. Additionally, research by [11] delves into joint resource allocation and user grouping aided by IRS. Furthermore, [12] offers valuable insights into the resource trading mechanisms within MEC enabled UAV networks. Subsequently, in [13], the focus shifts on enhancing the total completed task-input bits within RIS-assisted MEC networks, a problem initially tackled by using an algorithm called block coordinate descending (BCD). A framework of deep learning is devised which can support the BCD algorithm and its online implementation in order to mitigate its computational complexity. Moreover, work in [14] explores the integration of wireless energy transfer (WET) into RIS-assisted MEC networks to meet both computation and energy supply demands of IoT devices.

In the context of RIS-assisted UAV networks, RIS deployment serves to mitigate blockages and improve achievable rates. In [15], the focus lies on maximizing the sum rate of users in the down-link through a joint optimization approach. This involves optimizing UAV trajectory, power control, THz sub-band allocation and RIS phase shifts. To ensure secure communication between the UAV and legitimate users, [16] employs RIS deployment to enhance legitimate transmission quality while compromising eavesdropping. To achieve the maximum average secrecy rate an iterative algorithm was proposed. Likewise, [17] explores secure transmission in UAV and RIS-assisted mmWave networks, achieving near-optimal RIS and UAV positions through exhaustive search methods. Moreover, [18] introduces a decaying deep Q-network (D-DQN) algorithm to address the dynamic stochastic environments in RIS-assisted UAV networks. This algorithm aims to minimize UAV energy consumption by optimizing RIS phase shift, power allocation, UAV trajectory and decoding order. Results of the simulation demonstrate that D-DQN algorithm effectively balances training speed acceleration, convergence to local optima, and oscillation avoidance. In [19], an active RIS-aided MEC system was the subject of attention. Acknowledging the fact that there are some tasks that cannot be partitioned, the author presented a multi-user MEC system with RIS assistance, which operates based on both binary and partial offloading policy [20]. Meanwhile, in [21], the author proposed an RIS-assisted iterative algorithm for MEC downlink communication system, aiming to minimize transmission power through alternating optimization (AO) and penalty-based optimization.

In studies [22] and [23], an innovative communication concept was introduced, featuring a UAV-based base station transmitting data to the users on ground with the simultaneous support of a reflecting and transmitting reflecting RIS (STAR-RIS). In [24], an approach was developed to enhance the energy efficacy of the system by jointly augmenting the UAV's flight trajectory in three-dimensional space and designing RIS phase shifts. Authors in [25] introduced an innovative MEC framework powered by RIS and UAV relays, specifically targeting a maximum-minimum computation capability issue.

The previously discussed networks utilizing RIS predominantly focus on two areas: RIS-assisted MEC networks, where user equipment (UE) computational capabilities benefit from access point (AP) resources, and RIS-assisted UAV architectures. This enhances the UE communication rates via UAV trajectory adaptability. However, scant attention has been given, to our knowledge, to the performance enhancement potential of RIS in UAV-enabled MEC environments. In studies [26], despite the use of a UAV-mounted RIS to aid ground user communication with an MEC server, the UAV's computational capacity remains unaddressed. Task-input data offloading by users to the MEC server via orthogonal multiple access (OMA) protocol is employed, which fails to fully exploit time and frequency resources. Subsequently, the authors in [27] introduces a non-orthogonal multiple access (NOMA) into the RIS-assisted UAV-MEC system, aiming to optimize UAV computational capacity. Simulation outcomes demonstrate the superior performance of the NOMA scheme compared to OMA. In association to it, a waveform design and detection mechanism in NOMA systems is also presented in [28]. Work presented in [29] also considered a IRS based NOMA assisted MEC mechanism for energy efficiency with queue stability option. A Lyapunov-function-based mixed integer DDPG (LMIDDPG) algorithm is proposed for centralized learning and a heterogeneous MA-LMIDDPG algorithm is proposed for distributed learning, considering system throughput, power consumption, and queue stability for computing decisions, power allocation, and phase shifts optimizations.

In an other work [30] a DDQN based resource allocation scheme, DTORA, is introduced. DTORA optimizes task offloading and resource allocation decision using double DQN algorithm. These decisions target to mitigate UAV energy consumption through optimizing computing and communication resources and RIS phase shifts. Additionally, work presented in [31] utilized another concept of RIS-equipped UAVs to facilitate MEC. A DDPG based energy-efficient solution mechanism is proposed for MEC network with RIS-equipped UAVs, considering energy conservation for ground users and UAVs. The optimization problem maximizes system energy efficiency by jointly optimizing ground user transmission power, UAV trajectory, and RIS phase shifts. The works presented in [32] explored a multi-UAV assisted

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.2024.3397890

TABLE 1: Contrast of our work and related studies, emphasizing disparities and distinctive contributions.

	Offloading Mechanism		RAT		RIS Usage	ZED Association	Relaying		Objective (Minimization)		Solution Type	
	Binary	Partial	Cellular	NR			UAV	RIS	Energy	Delay	Other	DRL
This Work	✓	✓	–	✓	✓	✓	✓	✓	✓	✓	–	A2C
[20]	✓	✓	–	✓	✓	✓	✗	✓	✓	✓	–	A2C
[21]	✗	✓	✓	–	✓	✗	✗	✓	✓	✗	Optimization	–
[24]	✓	✗	✓	–	✓	✗	✗	✓	✓	✗	–	DQN
[29]	✗	✓	✓	–	✓	✗	✗	✓	✓	✓	–	LMIDDPG
[30]	✓	✗	✓	–	✓	✗	✓	✓	✓	✓	–	DDQN
[31]	✓	✗	✓	–	✓	✗	✓	✓	✓	✗	–	DDPG
[32]	✗	✓	✓	–	✓	✗	✗	✓	✓	✓	–	MA-TD3

MEC system with RIS. A multi agent DRL based MA-TD3 algorithm is proposed while aiming the optimization of computation offloading, UAV trajectories, and RIS phase shifts.

Motivated by the promising features of RIS and the performance limitations encountered by UAVs operating in MEC based networks, we intend to harness RIS to enhance UAV communications within MEC based networks. The RISs offer opportunities to reduce energy expenditures, boost UAVs' channel capacity, improve transmission reliability in MEC based networks, and expand wireless coverage. However, there is a notable scarcity of research concerning RIS-assisted UAVs in the field of mobile edge computing networks. While previously discussed studies have laid the groundwork for a novel system framework that incorporates RIS-assisted UAV-MEC, limited consideration has been devoted to investigating the energy ingestion aspect of RIS-assisted UAV-MEC systems.

Besides, while the offloading users intend to use UAV's resources, but given the limited resources and UAV's energy constraints, it's infeasible for users to execute their computation tasks concurrently on the UAV. Hence, there's a need to mutually augment task offloading and resource allocation to efficiently manage UAV resources and decrease energy consumption [13], [14], [21], [24]–[26]. However, the challenge lies in the complex, non-convex nature of this joint optimization problem, making it impractical to swiftly attain the optimal solution using conventional methods. Providentially, DRL has recently emerged as a powerful instrument for tackling numerous optimization problems within wireless communication networks. DRL offers the ability to derive optimal solutions for wide-reaching, intricate, and non-convex problems without necessitating complete and precise network information [18], [20], [24], [33].

The study outlined in [20] focused on minimizing energy and delay with ze-RIS considerations, albeit limited to vehicular networks and lacking the UAV relaying mechanism. Similarly, the research in [21] solely addressed partial offloading mechanism without considering ze-RIS or UAV relaying, while also being a non-DRL solution, potentially increasing complexity overhead. Additionally, [24] exclusively explored binary offloading mechanism without incorporating ze-RIS and only focused on energy optimizations. Moreover, [29] and [30] examined partial offloading mechanisms without considering UAV relaying or ze-RIS. Both [30] and [31] only delved into binary offloading mechanisms without any association with ze-RIS, with the latter not considering delay optimizations. The work in [32] is limited to partial offloading mechanisms without a UAV relaying mechanism.

Taking these factors into account, this paper introduces a hybrid resource allocation and task offloading scheme tailored for ze-RIS-assisted UAVs operating within MEC networks. ze-RIS offers energy efficiency, making it suitable for power-constrained environments [34]. The ze-RIS and conventional passive RIS differs in their power requirements. ze-RIS operates without external power sources, relying on energy harvesting or wireless power transfer. In contrast, conventional passive RIS requires a continuous power supply. While both enhance wireless communication by reflecting and manipulating radio waves [35], [36].

The scheme formalizes the objective of minimizing energy ingesting and leverages the DRL-based A2C algorithm to efficiently address the optimization problem. A specified comparison of related studies and our work is presented in the Table 1, and the key contributions of this work are as follows:

- The proposed DRL driven energy efficient task offloading (DEETO) scheme is based on B5G technological infrastructure. It incorporates an innovative RIS-assisted UAV-MEC system that not only equips the UAV with computation capabilities but also empowers it for relay transmissions. Within the proposed framework, we conduct a comprehensive joint optimization of several critical parameters, including the allocation of the UAV's computational resources, working environment communication resource allocation, task offloading decisions, and phase shifts for the RIS. The primary objective of this optimization is to minimize the energy consumption of the UAV.
- Given the non-convex nature of the objective function, our approach involves an initial problem formulation using a Markov decision process (MDP). In this context, it's important to highlight that both of the MDP spaces, the state space and the action space are continuous. To efficiently tackle this challenge, we employ the A2C algorithm.
- Our simulation outcomes demonstrate the notable energy efficiency improvements achieved by our RIS-

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.2024.3397890

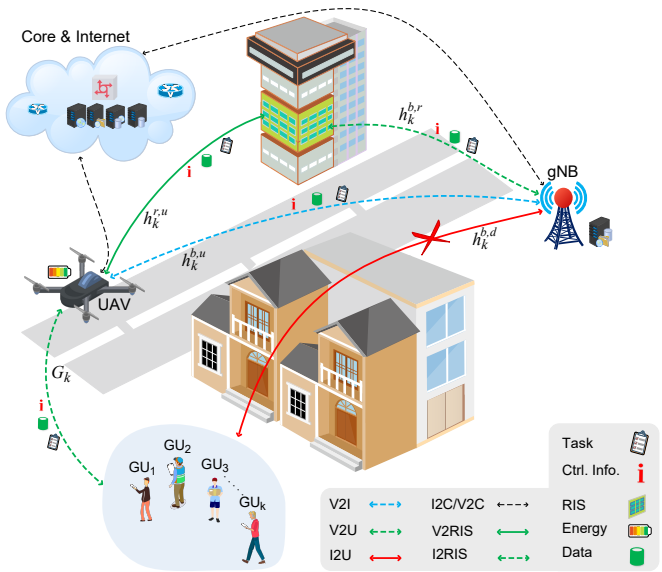


FIGURE 1: System model, demonstrating the conceptual framework.

aided UAV-MEC system, which employs the A2C algorithm. When compared to alternative benchmark schemes, our approach substantially reduces the overall energy consumption of UAVs.

II. SYSTEM AND OFFLOADING MODEL

A. SYSTEM MODEL

Recognizing the capabilities and advancements in 5G-NR communications, we have explored an RIS-enhanced UAV-MEC system, as depicted in Fig. 1. The UAV serves as a spatial base station, equipped with computational capabilities and the ability to function as a relay station. Thus, enabling the transmission of ground user's computational tasks to the next generation evolved node B (gNB) type base transceiver station (BTS). The BTS is supposed to be provided with multi CPU MEC server. We have integrated zero-energy reconfigurable intelligent surface (ze-RIS) elements into the architecture, thoughtfully positioned on the building's facade. We have followed the work proposed in [34], [35], and [36] for ze-RIS, and the zero energy device (ZED) based network configurations. The ze-RIS elements play a pivotal role in redirecting signals when needed. When users need to offload their computational tasks, they can choose between transmitting them directly to the UAV or using the UAV as a relay to send them to a distant small cell BTS for processing. The remarkable contribution of ze-RIS is its capacity to assist in signal relay from the UAV to the BTS, particularly in some situations where a direct communication is created amongst users and the BTS is hindered by physical barriers or obstructions.

Within the scheme, we have gu_m ground users, a UAV av_i, an RIS r_j equipped with k reflecting elements and a 5G small cell gNB type BTS b_l connected to MEC infrastructure. Here $i \in \mathcal{I}, j \in \mathcal{J}, k \in \mathcal{K}, l \in \mathcal{L}, m \in \mathcal{M}$ takes on values from the set $\{1, 2, \ldots, \mathcal{N}\}$, and $\mathcal{N}$ belongs to a natural number. To manage the RIS's reflection capabilities, we employ a reflection factor matrix denoted as $\Theta = \text{diag}[\theta_1(n), \theta_2(n), \ldots, \theta_k(n)]^T$, which embodies the reflecting beam-forming functionalities of the RIS. Here, $\theta_k(n)$ signifies the reflection factor of the k-th reflecting component during time slot n. The phase shift, $\phi_k(n)$, is contained within the interval $[0, 2\pi)$, and λ_k is the amplitude reflection coefficient, residing within the range $[0, 1]$. To optimize signal reflection, we typically set $\lambda_k = 1$, as previously done in [33]. We then divide total time period, T, into n distinct time periods. The variable n belongs to the set $\mathcal{N}$, and each of these time slots has a duration of δ. Users, during each of these time slots, offload their intended tasks to the in-range UAV following orthogonal frequency-division multiple access (OFDMA).

The channel gains for the ground user gu_m to UAV (gu_m–av_i link), from the UAV to BS (av_i–b_l link), UAV to RIS (av_i–r_j link), and from RIS to BS (r_j–b_l link) during time period n can be mathematically signified as $G_m(n)$, $h_m^{i,l}(n)$, $\mathbf{h}_m^{i,j}(n)$, and $\mathbf{h}_m^{j,l}(n)$, respectively. The gu_m–av_i link, av_i–b_l link, and av_i–r_j link are categorized as line-of-sight (LoS) links, and modeled using the free-space path loss model, and expressed as:

$$G_m(n) = \sqrt{\beta_0 {d_{m,i}}^{-2}(n)} \;:\; G_m(n) \in gu_m - av_i - \text{link}, \tag{1}$$

$$h_m^{i,l}(n) = \sqrt{\beta_0 {d_{i,l}}^{-2}(n)} \;:\; h_m^{i,l}(n) \in av_i - b_l - \text{link}, \tag{2}$$

$$\mathbf{h}_m^{i,j}(n) = \sqrt{\beta_0 d_{i,j}^{-2}(n)} \left[1, e^{-j\frac{2\pi}{\lambda} d\varphi_{i,j}(n)}, \ldots, e^{-j\frac{2\pi}{\lambda} d(K-1)\varphi_{i,j}(n)}\right] \;:\; \mathbf{h}_m^{i,j}(n) \in av_i - r_j - \text{link}, \tag{3}$$

In the given context, $d_{m,i}(n)$, $d_{i,l}(n)$, and $d_{i,j}(n)$ signify the distances for the gu_m–av_i, av_i–b_l, and av_i–r_j links in time slot n, respectively. The path-loss at reference distance of one meter is represented by β_0, d represents the inter RIS elements distance, and $\varphi_{i,j}(n) = (x_j - x_i)/(d_{i,j})$ denotes the cosine of the angle of arrival (AoA) of the signal. It's important to note that, for simplicity, the channel from RIS to BTS is considered to be mainly influenced by the Rician fading channel, expressed as:

$$\mathbf{h}_k^{j,l}(n) = \sqrt{\beta_0 d_{j,l}^{-2}(n)} \left[\sqrt{\frac{\beta_1}{1+\beta_1}} \mathbf{h}_{j,l}^{LoS}(n) + \sqrt{\frac{1}{\beta_1 + 1}} h_{j,l}^{NLoS}(n)\right] \;:\; \mathbf{h}_m^{j,l}(n) \in r_j - b_l - \text{link}, \tag{4}$$

$$\mathbf{h}_{j,l}^{LoS}(n) = \left[1, e^{-j\frac{2\pi}{\lambda} d\varphi_{j,l}(n)}, \ldots, e^{-j\frac{2\pi}{\lambda} d(K-1)\varphi_{j,l}(n)}\right] \tag{5}$$

where β_1 represents the Rician factor, $d_{j,l}(n)$ corresponds to the distance for the r_j–b_l link., and $\mathbf{h}_{j,l}^{LoS}(n)$ is the LOS

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.2024.3397890

component. $\varphi_{j,l}(n) = (x_l - x_j)/(d_{j,l})$ denotes the AoD signal's cosine. For the NLoS component, $\mathbf{h}_{j,l}^{NLoS}(n)$ follows a Gaussian distribution with zero mean and unit variance, i.e., $\mathbf{h}_{j,l}^{NLoS}(n) \sim \mathcal{CN}(0,1)$.

B. OFFLOADING AND PROCESSING MODEL

In our study, we categorize the processing of user computation tasks into three specific phases. Initially, there is the task offloading from the ground users to the UAV, followed by the computation processing at the UAV, and finally, the relaying of tasks to the BTS. It is important to note that the BTS, being grounded and potentially equipped with multi-CPU MEC servers. These servers are expected to have a considerably higher computation rate compared to the UAV. Therefore, we can safely disregard the computing latency of BTS within all time slot.

1) UAV offloading

Given the constraints of finite computational power then the battery size for the ground user, the decision is made to transfer all of the computation tasks to UAV. In this context, the computation tasks of the m-th user in the n-th time slot are denoted as $t_k(n)$. Each $tk(n)$ is characterized as a 3-tuple: $tk_m(n) = \{s_m, c_m, t_m, e_m\}$, where s_m represents the task size in bits, c_m indicates the required CPU cycles for 1 bit computation, t_m, and e_m are the time and energy threshold units allocated for task completion. Since all the tasks are pre-decided to be offloaded, the transmission rate $R_m^{ofld}(n)$ of the m-th user in the n-th time slot while offloading task $tk_m(n)$ is given by:

$$R_m^{ofld}(n) = B_0 \log_2\left(1 + \frac{p_m(n)\ ||G_m(n)||^2}{\sigma^2}\right), \quad \text{and} \quad B_0 = B/m, \tag{6}$$

Here, B is the total bandwidth, distributed into m even channels as. $p_m(n)$ is the transmission power of m-th user in the n-th time slot. σ^2 denotes the power of additive Gaussian white noise. The offloading delay for the m-th user, when offloading their computation extensive tasks to the in range UAV, is calculated while following Eq. (7). Additionally, offloading a given task must follow the constraint $t_m^{ofld} < t_m$.

$$t_m^{ofld}(n) = s_m(n)/R_m^{ofld}(n). \tag{7}$$

2) UAV Computation

Since all of the user's computational tasks are transferred to UAV. The UAV dynamically determines the decisions of offloading depending on the incoming task count from the ground users. By employing a partial task offloading strategy, the UAV handles the ζ portion of the task, while the β segment is relayed to the BTS. The decision-making process involves the UAV determining the allocation of the m-th user's computation tasks, deciding how much of the task should be processed on the UAV and how much should be relayed to the BTS. These decisions adhere to the constraint $\zeta + \beta = 1$. The decision parameters ζ_m and β_m are binary in nature. In the case of adopting a partial task processing model, both of these parameters are set to 1. When exclusively UAV processing is preferred, ζ_m is set to 1, and β_m is set to 0. Conversely, when only relaying is favored, β_m is set to 1, and ζ_m is set to 0. The time t_m^{av} and energy e_m^{av} consumed in processing task $t_k(n)$ during the n-th time slot at UAV av_i can be expressed as follows:

$$t_m^{av}(n) = \Big((1-\zeta_m(n)(1-\beta_m(n))\Big)\left\{\frac{\zeta s_m(n)\ c_m(n)}{f_m^{av}(n)}\right\} \tag{8}$$

$$e_m^{av}(n) = (1 - d_m(n))\Big\{\zeta s_m(n)\ c_m(n)\ \varphi^{av}\ \left(f_m^{av}(n)\right)^2\Big\} \tag{9}$$

where $f_m^{av}(n)$ in Eq. (8) signifies the CPU resources allocated to the task t_k in n-th time slot, and φ^{av} in Eq. (9) is the UAV chip specific effective capacitance coefficient, determined by architecture of the chip.

3) UAV–BTS Relaying

For the decisions $(\zeta_m = \beta_m = 1)$ or $(\zeta_m = 0 \wedge \beta_m - 1)$ the task is to be relayed to the BTS for processing, either in parts or as a whole. Therefore, the transmission rate and time for the UAV–BTS link can be calculated while following Eq. (10), and Eq. (11), while the consumed energy for relaying can be calculated while following Eq. (12).

$$R_m^{rel} = B_0 \log_2\left(1 + \frac{p_m(n)\ ||h(n)||^2}{\sigma^2}\right), \quad \text{where}\ \ h(n) = h_m^{i,l}(n) + \mathbf{h}_m^{i,j}(n)\ \Theta \mathbf{h}_k^{j,l}(n) \tag{10}$$

$$t_m^{rel}(n) = \Big((1 - \zeta_m(n)(1 - \beta_m(n))\Big)\left\{\frac{\beta\ s_m(n)}{R_m^{rel}}\right\} \tag{11}$$

$$e_m^{rel}(n) = \Big((1-\zeta_m(n)(1-\beta_m(n))\Big)\left\{\frac{P_m^{rel}\beta\ s_m(n)}{R_m^{rel}}\right\} \tag{12}$$

III. OBJECTIVE AND SOLUTION METHODOLOGY

In this segment, we present our approach to addressing the challenge of reducing energy ingesting while enhancing the computational efficacy of the UAV based MEC network. The key objective is to fine-tune the use of computation resources, while aiming to maximize the proportion of total processed bits to the energy consumed by UAVs. In line with our goal, we can articulate the average energy ingesting and time consumed while offloading, computing, and relaying the task m is expressed in Eq. (13) and Eq. (14).

$$\mathcal{T}_m = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \left(t_m^{ofld}(n) + t_m^{av}(n) + t_m^{rel}(n)\right) \tag{13}$$

$$\mathcal{E}_m = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \left(e_m^{av}(n) + e_m^{rel}(n)\right) \tag{14}$$

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.2024.3397890

A. DEXTERITY FACTOR

To ensure fair and impartial assessments of efficiency, we have introduced a parameter known as the dexterity factor. This factor is designed to incorporate standard benchmarks for time and energy consumption into the evaluation process. Specifically, we establish t_m^m as the average of the maximum and minimum time thresholds assigned to tasks, representing the benchmark for delay, and e_m^m as the average of the maximum and minimum energy thresholds allocated to tasks, serving as the benchmark for energy. These benchmarks, derived from the range of allowable time and energy values, provide a reference point against which the actual performance of the UAV can be compared.

In a dynamic RIS-assisted MEC network, the performance of a UAV in terms of time and energy can vary significantly across different time slots. To ensure unbiased evaluations of efficiency, we have introduced a dexterity factor. This factor is designed to incorporate standard benchmarks for time and energy consumption into the evaluation process. We define $t_m^m = \text{mean}(t_m^{max}, t_m^{min})$ as the benchmark for delay and $e_m^m = \text{mean}(e_m^{max}, e_m^{min})$ as the benchmark for energy. Here, t_m^{max} and t_m^{min} signify the maximum and minimum threshold times allocated to the tasks, while e_m^{max} and e_m^{min} signify the maximum and minimum threshold energy units allocated to the tasks, respectively. These benchmarks provide reference points against which the actual performance of the UAV can be compared, allowing for a more accurate evaluation of its efficiency. The dexterity factor is then formulated as:

$$\mathcal{D}_m = \{((t_m + t_m^m) - \mathcal{T}_m) + ((e_m + e_m^m) - \mathcal{E}_m)\} \quad (15)$$

It is worth noting that a larger value of $\mathcal{D}_m$ corresponds to a greater amount of saved energy and reduced delay. Since the dexterity factor $\mathcal{D}_m$ is associated with single task tk_m , the average dexterity factor for all M tasks can be expressed as:

$$\mathcal{D}_a = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \{((t_m + t_m^m) - \mathcal{T}_m) + ((e_m + e_m^m) - \mathcal{E}_m)\} \quad (16)$$

Building upon these foundational concepts, we establish our joint objective to optimize $\mathcal{D}_m$. This encompasses decisions associated with offloading, UAV processing and execution, and the relaying mechanism.

$$\underset{\zeta,\, \beta,\, \Theta,\, P^{rel},\, f^{av}}{\text{maximize}} \quad \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \left(\frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \mathcal{D}_m \right)(n),$$

s.t.

$$C1 : \mathcal{E}_m(n) \leq e_m(n), \quad (17)$$

$$C2 : \mathcal{T}_m(n) \leq t_m(n), \quad (18)$$

$$C3 : 0 \leq f_m^{av}, \text{ and}$$
$$\sum_{m=1}^{\mathcal{M}} (1 - \zeta_m(n)(1 - \beta_m(n)) f_m^{av}(n) \leq F_{max}(n), \quad (19)$$

$$C4 : 0 \leq P_m^{rel}, \text{ and}$$
$$\sum_{m=1}^{\mathcal{M}} (1 - \zeta_m(n)(1 - \beta_m(n))\, P_m^{rel}(n) \leq P_{max}^{rel}(n), \quad (20)$$

$$C5 : 0 \leq \phi_k < 2\pi, \quad (21)$$

$$C6 : \forall\, \zeta_m \in \{0, 1\}, \text{ and } \forall\, \beta_m \in \{0, 1\}. \quad (22)$$

Constraint C1 is directly associated with energy and stipulates that the energy consumed during task processing and relaying must be less than or equal to the allocated energy. Constraint C2 pertains to user experience, ensuring that the time taken for task offloading, processing, and relaying must be less than or equal to the allocated threshold time. Constraints C3 and C4 signify the limitations on UAV computation and energy resources, respectively. Constraint C5 denotes the phase shift constraints for RIS elements. Constraint C6 signifies the parameters governing task offloading decisions.

The objective of the optimization problem described in Eq. (17), i.e., to maximize the D_m, corresponds to minimize the overall energy consumption of the UAV during its operational tasks. This involves a comprehensive approach to optimizing various aspects, including the task offloading decisions ζ_m and β_m, the phase shift Θ of RIS elements, the power consumption P^{rel} during relay transmissions, and the allocation of UAV computational resources f^{av}.

Given the dynamic nature of problem Eq. (17) and its inherent involvement in a multidimensional spaciotemporal environment, employing traditional optimization approaches proves challenging due to significant difficulty and high-order complexity. To address these challenges, we have chosen to employ a DRL based methodology. This decision is driven by the adaptive and learning-centric nature of DRL, which is particularly well-suited for handling dynamic and complex optimization problems in dynamic environments.

B. MDP FORMULATION

In the scope of this paper, the UAV plays the role of an agent that actively interact with the environment, with the primary objective of collecting comprehensive state information and formulating an optimum policy to exploit the cumulative reward. The designations for the MDP components, including the state, action spaces, and reward function, are explicitly outlined as follows:

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.2024.3397890

1) State Space

The state space s_t in our MDP at the t time slot is characterized by various UAV states, including UAV position state s_u^p, UAV velocity state s_u^v, UAV energy state s_u^e, and UAV processing capacity state s_u^c. Additionally, the state space also incorporates RIS configuration state s_r, BS status state s_b, MEC workload state s_m^w, MEC capacity state s_m^c, and task-related states such as task size state s_{tk}^s, task CPU cycles requirement state s_{tk}^c, task threshold time state s_{tk}^t, and task energy threshold state s_{tk}^e. The comprehensive definition of the state space is as follows:

$$S \triangleq \{s|s = (s_u^p,\ s_u^v,\ s_u^e,\ s_u^c,\ s_r,\ s_b,\ s_m^w,\ s_m^c,\ s_{tk}^s, s_{tk}^c,\ s_{tk}^t,\ s_{tk}^e)\} \tag{23}$$

2) Action Space

The action state a_t in our MDP at t time slot encompasses UAV flight related actions including UAV speed a_u^s, UAV direction a_u^d, and UAV altitude a_u^a. Additionally, it also includes actions related to RIS phase shifts adjustments a_r, local task processing actions a_{tk}^l, task offloading decision action a_{tk}^o, and task relaying actions a_{tk}^r. In a comprehensive view, the action space can be articulated as follows:

$$A \triangleq \{a|a = (a_u^s,\ a_u^d,\ a_u^a,\ a_r,\ a_{tk}^l,\ a_{tk}^o,\ a_{tk}^r)\} \tag{24}$$

3) Reward Function

The reward function $R(s_t, a_t, s_{t+1})$ in our MDP encapsulates the goal of minimizing UAV energy consumption, ensuring task completion, and promoting efficient utilization of the ze-RIS. We have formulated the reward function as specified in Eq. (25), incorporating various factors to effectively train the DRL agent for optimal alignment with our objectives.

$$R(s_t, a_t, s_{t+1}) \triangleq \{r|r = (r_{df} + r_{ee} + r_{tc} + r_{ru})\} \tag{25}$$

The rewards r_{df}, r_{ee}, r_{tc}, and r_{ru} in Eq. (25) are directly related to the objective function. r_{df}, dexterity factor reward, this factor represents the agility and flexibility of the system in handling tasks. In the objective function, this could be associated with constraints related to task completion C2 and energy consumption C1. Maximizing r_{df} encourages the system to efficiently complete tasks while managing energy resources effectively. r_{ee}, the energy efficiency factor reward, this factor focuses on minimizing energy consumption, which directly aligns with constraint C1 in the objective function. By maximizing r_{ee}, the system aims to optimize energy usage while achieving its objectives. r_{tc}, the task completion reward, this factor is binary, indicating whether tasks are successfully completed $r_{tc} = 1$ or not $r_{tc} = 0$. In the objective function, this corresponds to the task completion constraint C2. Maximizing r_{tc} encourages the system to prioritize completing tasks efficiently. r_{ru}, the RIS utilization reward, this factor promotes the optimal utilization of RIS elements to enhance communication and energy consumption, aligning with constraints C4 and C5 in the objective function. Maximizing r_{ru} encourages the system to leverage RIS effectively to achieve its goals.

In order to train the DRL agent and guiding that towards higher rewards we have added weights with these factors as these can be observed from Eq. (26)

$$R(s_t, a_t, s_{t+1}) \triangleq \{r|r = (r_{df} + (-w_1 \cdot r_{ee}) + (w_2 \cdot r_{tc}) + (w_3 \cdot r_{ru}))\} \tag{26}$$

Here, w_1 weight is used to prioritize the importance of UAV energy consumption. The weight w_2 is associated with task completion reward, and w_3 is a measure indicating how effectively the RIS is utilized.

C. THE A2C BASED DEETO ALGORITHM

The DEETO scheme includes the collaboration among the DRL agent and the RIS-assisted MEC based UAV task offloading environment. The algorithm 1 integrates the Advantage Actor-Critic (A2C) algorithm with the UAV task offloading and relaying scenario for obtaining optimal task offloading decisions policy. The DEETO algorithm starts by initializing the policy network π_θ and the value function V_ϕ, along with the learning rate α and discount factor γ hyperparameters. Then algorithm iterates through episodes, representing sequences of interactions between the UAV and the environment. Each episode captures a series of time steps. At each time step within an episode, the algorithm the policy network π_θ samples an action a_t based on the current state s_t. Later, the sampled action a_t is executed in the environment, resulting in a reward r_t and a new state s_{t+1}. Then, based on the current state and connectivity information, the algorithm determines the communication scenario. This scenario can be a direct link $gu_m - av_i$ between the mobile user and the UAV or relay link $gu_m - av_i - b_l$ from mobile user to UAV to BS, or an RIS assisted link $gu_m - av_i - r_j - b_l$ an RIS-assisted link. Depending on the communication scenario, the state of the mobile user, BS, RIS, and UAV is updated to reflect changes in parameters such as task status, bandwidth, energy levels, delays, phase shifts, channel conditions, and resource availability. After updating, the transition tuple (s_t, a_t, r_t, s_{t+1}) is stored in a trajectory buffer for later use in policy and value function updates. Then the advantage A_t is calculated using the reward r_t and the value functions V_ϕ for the current and next states s_t and s_{t+1}, respectively. Next the policy parameters θ and value function parameters ϕ are updated using the A2C algorithm. The policy parameters are updated in the direction that increases the likelihood of selecting actions that yield higher advantages, while the value function parameters are updated to reduce the difference between the estimated value of the current state and the estimated value of the next state. After the specified number of episodes, the algorithm outputs the optimal policy π_θ and value function V_ϕ, which can then be used to make task offloading and relaying decisions in the UAV environment.

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.2024.3397890

Algorithm 1 A2C assisted DEETO Algorithm

Input: Initial state s_0, policy network π_θ, value function V_ϕ, gu_m, av_i, r_j, b_l parameters
Initialize policy network π_θ and value function V_ϕ, learning rate α, discount factor γ
for each episode **do**
  Initialize environment initial state s_0
  for each time step t **do**
    Take sample offloading action $a_t \sim \pi_\theta(\cdot|s_t)$
    Execute offloading action a_t, observe reward r_t and new environment state s_{t+1}
    Determine communicating node locations and connectivity
    Select communication link ($gu_m - av_i$, $gu_m - av_i - b_l$, $gu_m - av_i - r_j - b_l$)
    if $gu_m - av_i$ **then**
      Update location, trajectory, computation and energy states of av_i
    else if $gu_m - av_i - b_l$ **then**
      Update location, trajectory, computation and energy states of av_i,
      Update b_l states
    else if $gu_m - av_i - r_j - b_l$ **then**
      Update location, trajectory, computation and energy states of av_i,
      Update phase shifts, channel conditions, energy states of r_j,
      Update b_l states
    end if
    Update gu_m, av_i, r_j, and b_l state parameters
    Update communication scenario used,
    Update energy consumed,
    Update delay incurred,
    Update available resources
    Store transition (s_t, a_t, r_t, s_{t+1}) in trajectory buffer
    Calculate advantage $A_t = r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t)$
    Update policy parameters θ and value function parameters ϕ using A2C:

$$\theta_{\text{new}} = \theta + \alpha \nabla_\theta \log \pi_\theta(a_t|s_t) \cdot A_t$$

$$\phi_{\text{new}} = \phi + \alpha \nabla_\phi \left(0.5(V_\phi(s_{t+1}) - V_\phi(s_t))^2\right)$$

  end for
end for
Output: Optimal offloading policy π_θ and value function V_ϕ

Throughout the algorithm, mathematical notations such as π_θ, V_ϕ, α, γ, ∇_θ, and ∇_ϕ are used to represent the policy network, value function, hyperparameters, and gradients for policy and value function updates. The algorithm provides a rigorous framework for delay and energy optimizing task offloading and relaying decisions in the UAV environment.

TABLE 2: Simulation Parameters

Simulation Parameter	*Value*
CPU frequency of av_i, and MEC	2MHz–2GHz, 2GHz–5GHz
b_l antennas, RIS elements	4, 50
Maximum p_m, P_m^{rel}, β_0 and σ^2	0.5W, 1W, -50dB, and -90dBm
$av_i - r_j$ link Rician factor	10dB
$r_j - b_l$ link Rician factor	10dB
$gu_m - av_i$ link Rician factor	3dB
$r_j - b_l$ link path loss exponent	2dB
$gu_m - av_i$ link path loss exponent	3.5dB
$av_i - r_j$ link path loss exponent	2.8dB
Cellular Bandwidth	10MHz
A2C Algorithm Parameter	
Time steps	100k
Update window size per step	100
γ, GAE λ, and Learning rate	0.99,1.0, 0.0005
Value function coefficient	0.5

IV. SIMULATION SETUP, RESULTS AND DISCUSSION

In this section, we present the experimental setup and outcomes of our study on optimizing task offloading and relaying in a UAV-assisted MEC system while minimizing delays and energy consumption. The underlying subsection contains the details regarding simulation parameters, and configurations. The subsequent section contains the achieved results, discussing the DEETO algorithm's performance.

A. SIMULATION SETUP

We consider a UAV-assisted MEC based communication environment with RIS enhancement. The network is equipped with BTS having the communication range radius of 200m with ze-RISs support. Similarly, the UAV have a communication range radius of 500m. The positions of gu_m users are circularly distributed in an area around the UAV with 100m radius, randomly. We configure the channel parameters by equipping the ze-RIS with a uniform planar array (UPA), organized in an $M_x \times M_y$ grid layout. Employing the Rician fading model, our channel model encompasses both line-of-sight (LoS) and non-LoS components for the links between the $gu_m - av_i$, $av_i - b_l$, $av_i - r_j$, and $r_j - b_l$. The path loss is characterized by $PL(\alpha, d) = \beta_0(d/d_0) - \alpha$, where d represents the transmission distance, β_0 denotes the path loss at the reference distance $d_0 = 1$m, and is set to $-40dB$. The α signifies the path loss exponent. The task attributes, including size and processing time thresholds, are selected at random from predefined intervals. Task sizes vary from 1 to 50Mb, while processing thresholds range from 200 to 500ms, and 1–50 joules of energy units to process it, accordingly. Additionally, our system incorporates a sub-task feature, enabling each task to comprise a flexible number of sub-tasks, ranging from 1 to 100. The simulations were executed on a M1-MacBook, using Python 3.7, PyTorch 2.0, Spyder, and Jupyter Notebook IDEs were employed for implementation. The DRL algorithms adhered to OpenAI standards. Other simulation parametric values and hyperparameters are given in Table 2.

For thorough analysis, we conducted convergence tests on

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.2024.3397890

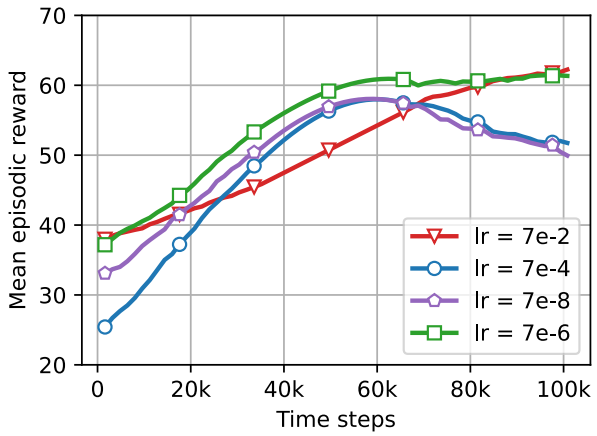


FIGURE 2: Convergence analysis of A2C based DEETO scheme for different learning rates.

the proposed DEETO scheme using various A2C DRL agent learning rates. Specifically, we evaluated the A2C algorithm's performance with learning rates set at $7e^{-2}$, $7e^{-4}$, $7e^{-6}$, and $7e^{-8}$. The convergence and reward collection analysis, depicted in Fig. 2, revealed distinct patterns. Notably, results indicated that the $7e^{-4}$ and $7e^{-8}$ learning rates achieved earlier convergence compared to $7e^{-2}$ and $7e^{-6}$, albeit experiencing declines in performance over additional time steps. While the convergence pattern for $7e^{-2}$ appeared superior to $7e^{-4}$ and $7e^{-8}$, it exhibited lower reward collection than $7e^{-6}$. Remarkably, the A2C algorithm demonstrated notable efficacy with DEETO at a learning rate of $7e^{-6}$, exhibiting enhanced convergence and higher rewards. Consequently, we have selected $7e^{-6}$ as the learning rate for subsequent experimental analyses.

B. RESULTS AND DISCUSSION

The efficiency of the DEETO scheme is systematically examined in dynamical scenarios, comparing it with other candidate DRL-based offloading techniques. we have evaluated and analyzed these schemes in terms of rewards, time frames per second (TFPS), task completion rate, energy consumption rate, and dexterity factors. Confirming an adequate comparison, all schemes are assessed using consistent simulation parameters. we have compared the DREEO scheme with the following contender schemes:

- *DEETO*: This is the proposed scheme.
- *DREEO*: This is a DRL based energy optimizing task offloading scheme. DREEO is an RIS assisted scheme, comprising hybrid task offloading scheme including binary and partial offloading mechanisms [20].
- *DTORA*: This is an RIS enabled UAV based MEC task offloading scheme. The DTORA optimizes task offloading and resource allocation decision using double DQN algorithm. These decisions target to mitigate UAV energy consumption through optimizing computing and communication resources and RIS phase shifts [30].
- *DQN*: This is an other DRL driven scheme to optimize task offloading delays and energy consumption. The DQN algorithm is adopted to optimizing task offloading decisions.

Several assessments have been made to verify and affirm the efficacy of our proposed DEETO scheme in a dynamic UAV-centric edge computing network environment. Fig. 3 provides a comprehensive comparison between DEETO and other schemes concerning rewards, a key metric in DRL-based approaches. Fig. 3(a) illustrates raw episodic rewards, while Fig. 3(b) depicts smoothed episodic rewards across time steps. Here, the DRL agent interacts with the environment, receiving rewards or penalties based on its actions. As the iteration with the environment progresses, the DRL agent adapts its policy to optimize its reward value. Favorable actions corresponding to the current environmental state yield higher rewards, whereas unfavorable actions result in reduced or negative rewards (penalties).

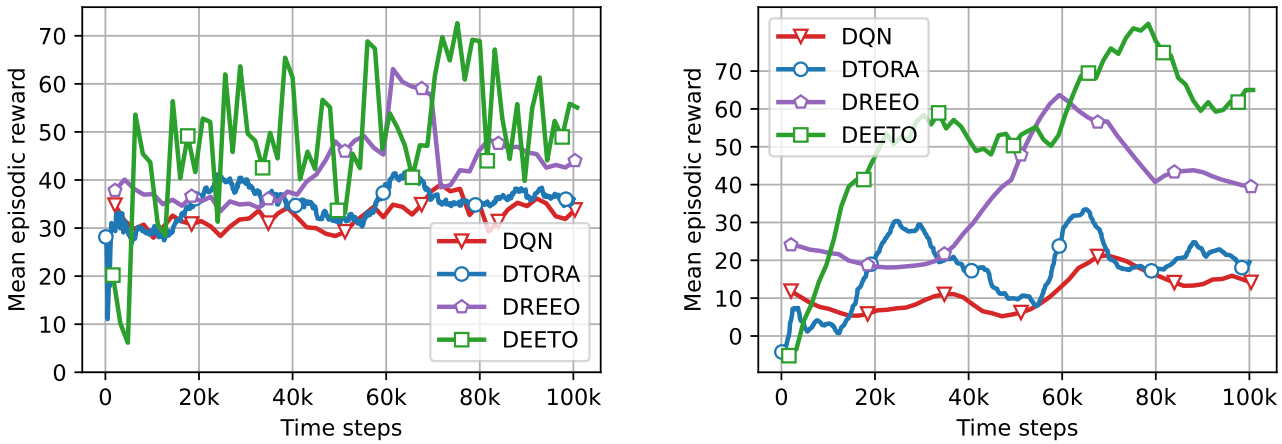
Besides, in the context of evaluating DRL based algorithms, apart from the comparison of rewards, time frames per seconds (TFPS) serves as a crucial performance metric. TFPS refers to the rate at which an algorithm processes the environment state steps within a given timeframe. A lower TFPS indicates that the algorithm can process each time frame more efficiently, meaning it can accomplish its tasks or decision-making processes in less time. Therefore, achieving a lower TFPS over time signifies improved performance and efficiency. Fig. 4 provides an illustration of the TFPS for all the contender approaches.

The observations indicate that the proposed DEETO scheme initially shows a higher TFPS, which gradually decreases over time and eventually converges to the lowest rate compared to all competing schemes. DEETO scheme achieves the lowest average TFPS rate of 870 among all other schemes. The DREEO scheme closely follows the characteristics of the DEETO scheme in terms of TFPS rate, yet it concludes with a slightly higher average TFPS of 910. Contrarily, the DTORA scheme starts with a higher TFPS, experiences a temporary decrease, but eventually shows an increasing trend over time, and finally converging at an average TFPS of 992. In contrast, the DQN-based scheme performs the poorest in terms of TFPS, obtaining an average TFPS rate of 1048.

The DRL characteristics shown in Fig. 3 and Fig. 4 illustrate the performance of each scheme in terms of rewards and TFPS, respectively. However, evaluating both features simultaneously unveils optimal algorithmic performance when higher rewards are achieved at lower TFPS rates collectively. It is understandable that the proposed DEETO scheme outperforms all contestant schemes while adhering to this measure. This performance gap between DEETO and other schemes can be accredited to several advantages of employing the A2C algorithm. A2C's simplicity, combining policy gradients and value-based methods, along with lower variance, renders it more robust. These attributes make it

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.2024.3397890



(a) Reward Raw

(b) Reward Smoothed

FIGURE 3: Evaluation of mean episodic rewards across different DRL based offloading schemes.

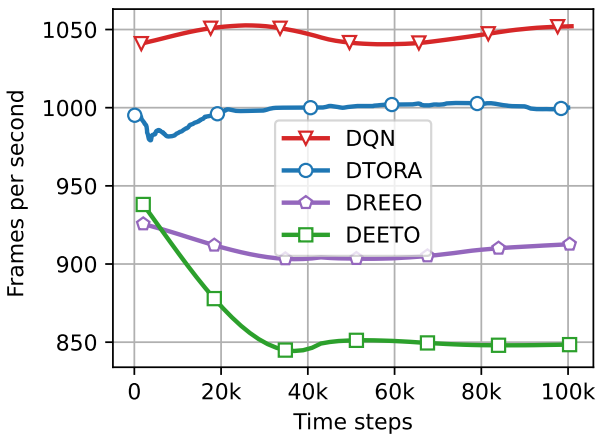


FIGURE 4: Evaluation of time frames per second across completing DRL based offloading schemes.

predominantly suitable for real-time learning scenarios, enabling firmer convergence and compliance to the dynamic communication environments.

Apart from the DRL characteristics, it is necessary to consider the interacting environmental factors in which the DRL agent operates. Therefore, to assess the effectiveness of these task offloading schemes we have analyzed the task turnover rate and energy consumption rate, as shown in Fig. 5(a) and Fig. 5(b). The analysis illustrates how accurately and effectively task offloading decisions are made, balancing task completion and energy optimization considerations. This decision involves determining whether to execute tasks locally, on the UAV, or to relay them to the edge server. For the DEETO and DREEO schemes, there exists an additional dimension of decision-making, i.e., whether task processing/offloading occurs in a binary or partial manner. Fig. 5(a) illustrates the analysis, representing that all schemes performed worthily in terms of task success and drop rates, although with slight variations. However, the DEETO scheme achieved the highest task success rate of 94.12% and the lowest drop rate of 05.87%. The DREEO scheme closely followed, achieving a 93.42% task success rate and a 06.58% drop rate. Conversely, the DTORA and DQN schemes achieved task success rates of 83.04% and 80.13%, with corresponding task drop rates of 16.96% and 19.87%, respectively.

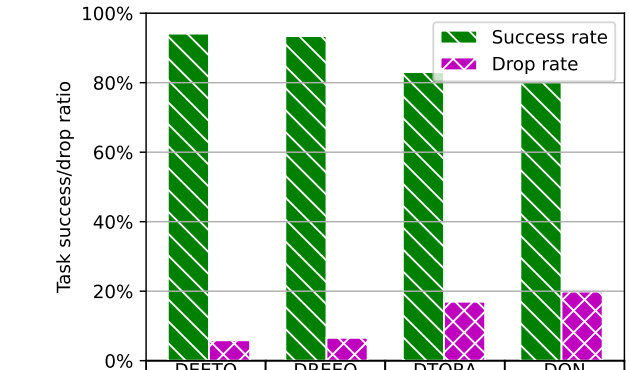
Similarly, Fig. 5(b) depicts the percentage of energy consumed and saved during task processing and relaying. All schemes successfully conserved energy within the allocated resources while observing to delay and task completion constraints. The DREEO scheme demonstrated the highest energy savings, conserving 16.98% of the allocated energy resources. Moreover, DREEO remained a close competitor to the DEETO scheme, achieving energy savings of 12.51%. In comparison, the DTORA and DQN schemes saved 12.22% and 8.32% of the allocated energy resources, respectively.

In addition to the aforementioned analyses, we conducted a comparative investigation involving the dexterity factor across varying numbers of tasks, task threshold times, and task sizes. The dexterity factor is composed as the sum of differences between consumed time and benchmark time, as well as consumed energy and benchmark energy.

Fig. 6a illustrates the comparative performance of all offloading schemes as the number of tasks increases. It is evident that an increase in the number of tasks leads to higher delay and energy consumption. However, the dexterity factor also increases concurrently, as it represents the average sum of saved energy and time. Notably, both the DREEO and DEETO schemes exhibit a similar pattern of energy and time savings, although with differing scales. This similarity can be attributed to the shared utilization of a hybrid task offloading mechanism. However, DEETO achieves the highest average dexterity factor, reaching 239.07 against 209.79 of

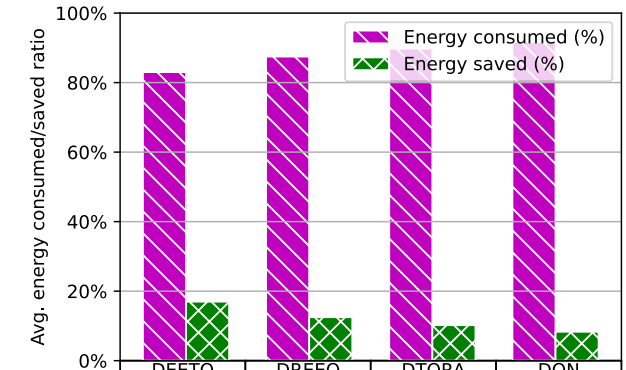
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

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.2024.3397890



	DEETO	DREEO	DTORA	DQN
Success rate	94.12 %	93.42 %	83.04 %	80.13 %
Drop rate	05.87 %	06.58 %	16.96 %	19.87 %

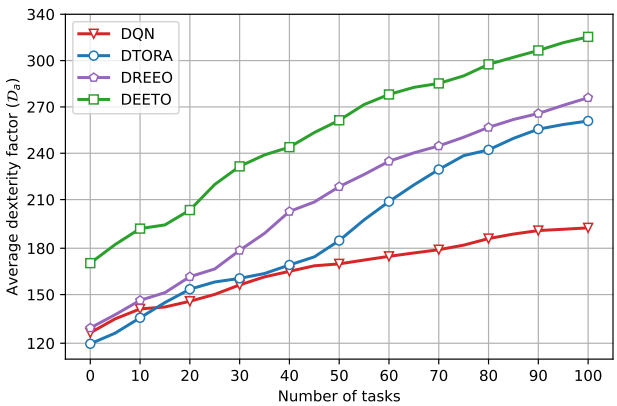
(a) task rate



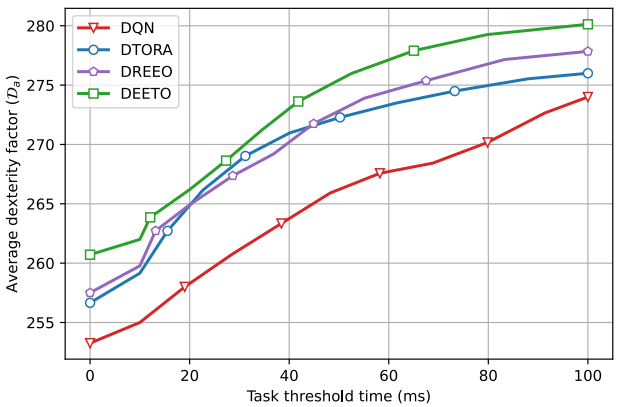
	DEETO	DREEO	DTORA	DQN
Consumed	83.02 %	87.49 %	89.78 %	91.68 %
Saved	16.98 %	12.51 %	10.22 %	08.32 %

(b) energy rate

FIGURE 5: Task completion and energy consumption of various DRL based schemes.



(a) Dexterity factor vs increasing number of tasks



(b) Dexterity factor vs increasing threshold time (t_{max})

FIGURE 6: Evaluation of dexterity factor against increasing number of tasks and the task threshold time

DREEO, 195.55 of DTORA, and 178.79 of DQN scheme. The superior decision-making mechanism of DEETO results in a higher dexterity factor, reflecting the cumulative savings in energy and time. This further validates the efficacy of relaying through ze-RIS via efficient phase shift mechanisms in reducing UAV energy consumption.

Expanding the analysis, we examined the performance of the DEETO scheme against the other schemes while increasing the threshold time, as depicted in Fig. 6b. It is evident that at lower threshold times, DEETO, DREEO, and DTORA exhibit similar performance levels, but as the threshold time increases, DEETO emerges as the superior scheme. However, the DQN scheme demonstrates the least efficiency. The proposed DEETO scheme achieved an average dexterity factor of 270.89, closely followed by DREEO with an average dexterity factor of 268.90. DTORA also closely trails DREEO with an average dexterity factor of 266.81, while the DQN scheme attained an average dexterity factor of 264.45.

The Fig. 7 illustrates the relationship of dexterity factor against increasing task size. it is obvious that as the task size increases the computation and communication over head also increases. the results show that DEETO and the DREEO schemes again has the same pattern of achieving dexterity factor, but the DEETO scheme has a smooth non linear trend as the task size increases and tends to convergence. However all the other schemes has a near linear sharp decreasing trend of dexterity factor with the increasing task size. The DEETO scheme achieved the highest average dexterity factor of 282.32, while DREEO, DTORA, and DQN achieved average dexterity factor of 277.65, 273.26, and 271.84, respectively.

The DREEO scheme emerged as the decent performer among all contender schemes due to its intelligent optimization of ze-RIS phase shifts coupled with a strategic hybrid task offloading mechanism. DREEO consistently remained a close competitor of DEETO across all evaluation results, leveraging the A2C algorithm alongside its hybrid task offloading approach, however with a random ze-RIS phase shift mechanism. In contrast, the DTORA scheme utilized the DDQN algorithm along with an intelligent RIS phase shift

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.2024.3397890

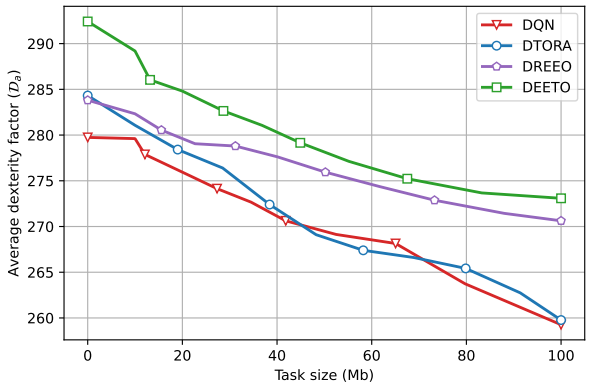


FIGURE 7: Dexterity factor vs increasing task size

mechanism but lacked a hybrid task offloading mechanism, resulting in slightly inferior performance. On the other hand, the DQN scheme, employing a deep Q-Network and employing a binary task offloading mechanism with random RIS phase shifts put it poorer scheme against all the competitor task offloading schemes.

V. CONCLUSIONS

This study addresses the challenge of sustaining high-quality communication services over extended periods in mobile edge computing (MEC) networks, particularly during emergencies, by leveraging UAVs and RIS. We introduced a comprehensive strategy that integrates ze-RIS into UAV-MEC networks, incorporating task offloading and resource allocation mechanisms. Our proposed DRL-driven energy-efficient task offloading (DEETO) scheme minimizes UAV energy consumption by optimizing task offloading decisions, UAV computing resource allocation, communication resource allocation, and RIS phase shift control, employing a hybrid task offloading mechanism. We modeled this problem as a DRL problem and structured it as an MDP and employed A2C algorithm due to its simplicity and less complexity. Simulation results demonstrate that the DREEO scheme emerged as a strong performer, utilizing intelligent ze-RIS phase shift optimization and a strategic hybrid task offloading mechanism. DEETO scheme showcased a significant energy savings of 16.98% from allocated resources, alongside the highest task turnover rate of 94.12%. Notably, DEETO scheme achieved these outcomes within shorter average learning time frames per second (TFPS) of 870 while yielding higher rewards of 57.38, underscoring its effectiveness over alternative approaches.

REFERENCES

[1] M. Y. Akhlaqi and Z. B. M. Hanapi, "Task offloading paradigm in mobile edge computing-current issues, adopted approaches, and future directions," Journal of Network and Computer Applications, vol. 212, p. 103568, 2023.

[2] Y. Bai, H. Zhao, X. Zhang, Z. Chang, R. Jäntti, and K. Yang, "Towards autonomous multi-uav wireless network: A survey of reinforcement learning-based approaches," IEEE Communications Surveys & Tutorials, 2023.

[3] M. N. Kharil, M. S. Johal, F. Idris, and N. Hashim, "Uav-enabled communications using noma for 5g and beyond: research challenges and opportunities," Indonesian Journal of Electrical Engineering and Computer Science, vol. 31, no. 3, pp. 1420–1432, 2023.

[4] F. Pervez, A. Sultana, C. Yang, and L. Zhao, "Energy and latency efficient joint communication and computation optimization in a multi-uav assisted mec network," IEEE Transactions on Wireless Communications, 2023.

[5] T. Sharma, A. Chehri, and P. Fortier, "Reconfigurable intelligent surfaces for 5g and beyond wireless communications: A comprehensive survey," Energies, vol. 14, no. 24, p. 8219, 2021.

[6] F. Naaz, A. Nauman, T. Khurshaid, and S.-W. Kim, "Empowering the vehicular network with ris technology: A state-of-the-art review," Sensors, vol. 24, no. 2, p. 337, 2024.

[7] M. Ahmed, A. Wahid, S. S. Laique, W. U. Khan, A. Ihsan, F. Xu, S. Chatzinotas, and Z. Han, "A survey on star-ris: Use cases, recent advances, and future research challenges," IEEE Internet of Things Journal, 2023.

[8] F. Xu, T. Hussain, M. Ahmed, K. Ali, M. A. Mirza, W. U. Khan, A. Ihsan, and Z. Han, "The state of ai-empowered backscatter communications: A comprehensive survey," IEEE Internet of Things Journal, 2023.

[9] Z. Chu, P. Xiao, M. Shojafar, D. Mi, J. Mao, and W. Hao, "Intelligent reflecting surface assisted mobile edge computing for internet of things," IEEE Wireless Communications Letters, vol. 10, no. 3, pp. 619–623, 2020.

[10] Y. Jiang, F. Gao, M. Jian, S. Zhang, and W. Zhang, "Reconfigurable intelligent surface for near field communications: Beamforming and sensing," IEEE Transactions on Wireless Communications, vol. 22, no. 5, pp. 3447–3459, 2023.

[11] Y. Gao, Q. Wu, W. Chen, Y. Liu, M. Li, and D. B. Da Costa, "Irs-aided overloaded multi-antenna systems: Joint user grouping and resource allocation," IEEE Transactions on Wireless Communications, pp. 1–1, 2024.

[12] M. Liwang, Z. Gao, and X. Wang, "Let's trade in the future! a futures-enabled fast resource trading mechanism in edge computing-assisted uav networks," IEEE Journal on Selected Areas in Communications, vol. 39, no. 11, pp. 3252–3270, 2021.

[13] X. Hu, C. Masouros, and K.-K. Wong, "Reconfigurable intelligent surface aided mobile edge computing: From optimization-based to location-only learning-based solutions," IEEE Transactions on Communications, vol. 69, no. 6, pp. 3709–3725, 2021.

[14] S. Mao, N. Zhang, L. Liu, J. Wu, M. Dong, K. Ota, T. Liu, and D. Wu, "Computation rate maximization for intelligent reflecting surface enhanced wireless powered mobile edge computing networks," IEEE Transactions on Vehicular Technology, vol. 70, no. 10, pp. 10 820–10 831, 2021.

[15] Y. Pan, K. Wang, C. Pan, H. Zhu, and J. Wang, "Uav-assisted and intelligent reflecting surfaces-supported terahertz communications," IEEE Wireless Communications Letters, vol. 10, no. 6, pp. 1256–1260, 2021.

[16] X. Pang, N. Zhao, J. Tang, C. Wu, D. Niyato, and K.-K. Wong, "Irs-assisted secure uav transmission via joint trajectory and beamforming design," IEEE Transactions on Communications, vol. 70, no. 2, pp. 1140–1152, 2021.

[17] G. Sun, X. Tao, N. Li, and J. Xu, "Intelligent reflecting surface and uav assisted secrecy communication in millimeter-wave networks," IEEE Transactions on Vehicular Technology, vol. 70, no. 11, pp. 11 949–11 961, 2021.

[18] X. Liu, Y. Liu, and Y. Chen, "Machine learning empowered trajectory and passive beamforming design in uav-ris wireless networks," IEEE Journal on Selected Areas in Communications, vol. 39, no. 7, pp. 2042–2055, 2020.

[19] Y. Li, Y. Zou, H. Hui, J. Zhu, and B. Ning, "Improving computing capability for active ris assisted noma-mec networks," IEEE Wireless Communications Letters, 2024.

[20] M. A. Mirza, J. Yu, M. Ahmed, S. Raza, W. U. Khan, F. Xu, and A. Nauman, "Drl-driven zero-ris assisted energy-efficient task offloading in vehicular edge computing networks," Journal of King Saud University-Computer and Information Sciences, vol. 35, no. 10, p. 101837, 2023.

[21] X. Qin, Z. Song, T. Hou, W. Yu, J. Wang, and X. Sun, "Joint resource allocation and configuration design for star-ris-enhanced wireless-powered mec," IEEE Transactions on Communications, vol. 71, no. 4, pp. 2381–2395, 2023.

[22] J. Zhao, Y. Zhu, X. Mu, K. Cai, Y. Liu, and L. Hanzo, "Simultaneously transmitting and reflecting reconfigurable intelligent surface (star-ris) assisted uav communications," IEEE Journal on Selected Areas in Communications, vol. 40, no. 10, pp. 3041–3056, 2022.

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.2024.3397890

[23] A. V. Savkin, C. Huang, and W. Ni, "Joint multi-uav path planning and los communication for mobile-edge computing in iot networks with riss," IEEE Internet of Things Journal, vol. 10, no. 3, pp. 2720–2727, 2022.

[24] L. Wang, K. Wang, C. Pan, and N. Aslam, "Joint trajectory and passive beamforming design for intelligent reflecting surface-aided uav communications: A deep reinforcement learning approach," IEEE Transactions on Mobile Computing, 2022.

[25] Y. Xu, T. Zhang, Y. Liu, D. Yang, L. Xiao, and M. Tao, "Computation capacity enhancement by joint uav and ris design in iot," IEEE Internet of Things Journal, vol. 9, no. 20, pp. 20 590–20 603, 2022.

[26] Z. Zhai, X. Dai, B. Duo, X. Wang, and X. Yuan, "Energy-efficient uav-mounted ris assisted mobile edge computing," IEEE Wireless Communications Letters, vol. 11, no. 12, pp. 2507–2511, 2022.

[27] Y. Xu, T. Zhang, Y. Zou, and Y. Liu, "Reconfigurable intelligence surface aided uav-mec systems with noma," IEEE Communications Letters, vol. 26, no. 9, pp. 2121–2125, 2022.

[28] C. Wang, M. Pang, G. Cui, X. Chang, F. Jiang, Y. Yao, and W. Wang, "Joint waveform design and detection in symbiotic ambient backscatter noma systems," IEEE Internet of Things Journal, vol. 10, no. 22, pp. 19 507–19 517, 2023.

[29] J. Yu, Y. Li, X. Liu, B. Sun, Y. Wu, and D. Hin-Kwok Tsang, "Irs assisted noma aided mobile edge computing with queue stability: Heterogeneous multi-agent reinforcement learning," IEEE Transactions on Wireless Communications, vol. 22, no. 7, pp. 4296–4312, 2023.

[30] Q. Liu, J. Han, and Q. Liu, "Joint task offloading and resource allocation for ris-assisted uav for mobile edge computing networks," in 2023 IEEE/CIC International Conference on Communications in China (ICCC). IEEE, 2023, pp. 1–6.

[31] L. Linpei, Z. Chuan, S. Yu, H. Jiahao, H. Yao, and L. Haojin, "Energy-efficient computation offloading assisted by ris-based uav," The Journal of China Universities of Posts and Telecommunications, vol. 31, no. 1, p. 37, 2024.

[32] S. Wang, X. Song, T. Song, and Y. Yang, "Fairness-aware computation offloading with trajectory optimization and phase-shift design in ris-assisted multi-uav mec network," IEEE Internet of Things Journal, 2024.

[33] H. Mei, K. Yang, Q. Liu, and K. Wang, "3d-trajectory and phase-shift design for ris-assisted uav systems using deep reinforcement learning," IEEE Transactions on Vehicular Technology, vol. 71, no. 3, pp. 3020–3029, 2022.

[34] D. Tyrovolas, S. A. Tegos, V. K. Papanikolaou, Y. Xiao, P.-V. Mekikis, P. D. Diamantoulakis, S. Ioannidis, C. K. Liaskos, and G. K. Karagiannidis, "Zero-energy reconfigurable intelligent surfaces (zeris)," arXiv preprint arXiv:2305.07686, 2023.

[35] Y. Zheng, S. A. Tegos, Y. Xiao, P. D. Diamantoulakis, Z. Ma, and G. K. Karagiannidis, "Zero-energy device networks with wireless-powered riss," IEEE Transactions on Vehicular Technology, 2023.

[36] S. Naser, L. Bariah, S. Muhaidat, and E. Basar, "Zero-energy devices empowered 6g networks: Opportunities, key technologies, and challenges," TechRxiv, 2023.



MUHAMMAD NAQQASH TARIQ obtained his Bachelor's degree (BEE) from The University of Faisalabad (TUF), Faisalabad in 2012, and his Master's degree in Electronics & Communication Engineering from the Beijing University of Posts and Telecommunications, Beijing, China in 2017. Currently, he is a Ph.D. candidate at the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China. His research interests include Unmanned Ariel vehicular communication networks, Software Defined Networking, fog and edge computing, internet of things, machine learning and deep reinforcement learning, and big data analytics.



JINGYU WANG (Senior Member, IEEE) received the Ph.D. degree from the Beijing University of Posts and Telecommunications, in 2008. He is currently a Professor with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. He has published more than 100 articles in international journals, including IEEE Communications Magazine, IEEE TRANSACTIONS ON CLOUD COMPUTING, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON MULTIMEDIA, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. His research interests include the IoV and AIoT, SDN, overlay networks, and traffic engineering.



DR. SAIFULLAH MEMON received a master's degree from Quaid-e-awam University of Engineering, Science and Technology, (QUEST) in 2015. Onward he acquired his Ph.D. degree from Beijing University of Posts and Telecommunications (BUPT), China in 2023 in the field of communication engineering. He is the author of more than 30 research articles and is currently working as an Assistant Professor at QUEST. His research interests include wireless communication, sensors, wireless body area networks (WBAN), QoS issues in sensor and ad-hoc networks, routing algorithms, and network security.



MOHAMMAD SIRAJ (Senior Member IEEE) 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. He has worked as a Scientist in Defense Research and Development Organization, India. 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, University 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 computing architecture, integrating mobile computing with cloud computing, and studying and designing green ICT solutions.

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.2024.3397890



MUHAMMAD AYZED MIRZA received his BSCS degree from Govt. College University, Faisalabad in 2008 and his MSCS degree in Computer Science from the National Textile University, Faisalabad in 2016. He worked as a lecturer at Punjab College University Campus Faisalabad for 4 years and at National Textile University, Faisalabad from 2016 to 2018. He completed his Ph.D. in 2023 from the School of Electronic Engineering at Beijing University of Posts and Telecommunications, Beijing, China. Currently, he is working as Associate Professor at the School of Computer Science and Information Engineering, Qilu Institute of Technology, Jinan, Shandong 250200 P.R, China. His research focuses on vehicular communications and networks, the IoT/IoV, fog and edge computing, and task/data offloading techniques.

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

International Journal of Environmental Science and Technology
https://doi.org/10.1007/s13762-024-05702-9

ORIGINAL PAPER

Trihalomethanes monitoring and their seasonal variation in urban municipal water supply system in North India

M. A. Mazhar[1] · S. Ahmed[1] · G. Singh[2] · A. Husain[1] · Rahisuddin[3]

Received: 23 June 2023 / Revised: 22 December 2023 / Accepted: 30 April 2024
© The Author(s) under exclusive licence to Iranian Society of Environmentalists (IRSEN) and Science and Research Branch, Islamic Azad University 2024

Abstract

This study reports the trihalomethanes generated during disinfection within the water distribution network of Delhi, India. For 12 months study period tap water samples were collected (n = 216) from the command areas of nine water treatment plants to provide a comprehensive picture of spatial and seasonal variation of trihalomethanes in the city. The mean concentrations of chloroform, bromodichloromethane, dibromochloromethane, bromoform were 34.62 ± 13.09 µg/l, 25.39 ± 11.75 µg/l, 15.83 ± 9.66 µg/l, 1.74 ± 1.51 µg/l respectively, the total trihalomethanes concentration ranged between 11.41–175.54 µg/l. Chloroform was prevalent accounting 45% in total trihalomethanes followed by bromodichloromethane, dibromochloromethane, bromoform accounting 33%, 20%, and 2% respectively. Trihalomethanes concentrations were found to be increasing during summer season and got decreased in the winter and post-monsoon season. Total trihalomethanes levels were found to be highest at Chandrawal water treatment plant command area (85.35–148.38 µg/l) and lowest at Okhla water treatment plant command area (11.41–63.38 µg/l). In many samples trihalomethanes concentrations were discovered to be higher than the allowable limit specified by Indian Standards. A strong relationship between trihalomethanes formation with total organic carbon (r = 0.934) and residual chlorine (r = 0.801) accentuate its capability for monitoring disinfection by-products in the distribution network. Results showed that considerable levels of trihalomethanes are generated in the study area, therefore the competent authorities must take the appropriate steps to regularly monitor the trihalomethanes in drinking water. The findings of this study may help to provide operators with invaluable information on drinking water quality and open up several chances to enhance water quality management.

Keywords Disinfection by-products · Chlorination · Water distribution system · Spatial variability · Seasonal variability

Introduction

For decades chlorine is being used for disinfection purpose in municipal drinking water for controlling the waterborne diseases. Chlorination kills the microorganisms and makes sure in diminishing the waterborne diseases like cholera, typhoid, infectious hepatitis, giardiasis, amoebiasis, bacillary dysentery, enteroviral diarrhoea etc. which results in decreasing death rate (Pan et al. 2014; WHO 2018; Zhang et al. 2010). Chlorine as a disinfectant has higher oxidation potential, high efficiency and low cost when compared to other disinfectants like chlorine dioxide, ozone, ultraviolet radiation, etc. It can also safeguard against regrowth of microorganism and inline biological contamination by leaky water distribution network (Elshorbagy et al. 2000; Chu et al. 2010). All forms of chlorine whether solid, liquid or gas are easily available worldwide and it is primarily used in India. Furthermore, it has previously been discovered that the reaction between chlorine and natural organic matter (NOM) during disinfection produces harmful disinfection by-products (DBPs) like trihalomethanes (THMs), haloacetic acids (HAAs), haloacetonitriles (HANs), etc., (Mazhar et al. 2020; Li et al. 2017; Richardson and Postigo 2015; States et al. 2013; Chowdhury 2013; Zamyadi et al. 2012). During chlorination process THMs and HAAs

Editorial responsibility: S. Mirkia.

✉ M. A. Mazhar
er.aamirmazhar@gmail.com

[1] Department of Civil Engineering, Jamia Millia Islamia, New Delhi, India

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

[3] Department of Chemistry, Jamia Millia Islamia, New Delhi, India

Published online: 11 May 2024

are mostly formed DBPs. The NOM is formed by natural degradation of some organic elements and are not easily eliminated by conventional water treatment processes like coagulation further entering the water treatment system (Grenoble et al. 2007). The following can be said about how DBPs form: (Krasner 2009).

NOM + Free chlorine → THMs + HAAs + HANs + Cyanogen-Halides + Other DBPs.

Since the discovery of THMs the first group of DBPs in early 1970s (Rook 1974), the number of identified species are more than 600 till date (DeMarini 2021). A lot of species is known today still THMs and HAAs are considered as decent indicators of the DBPs formed during chlorination (Selvam et al. 2018; Richardson and Postigo 2016; Serrano et al. 2015; Zhai et al. 2014; Shah and Mitch 2012). During the past decade's researchers have found DPBs to be potential carcinogens which are of great concerned (Pan et al. 2014). Various epidemiological studies showed that DPBs are potential carcinogens to human beings and causes bladder cancer, colorectal cancer, adverse reproductive outcomes like still birth and retardation in growth (Jones et al. 2019; Quist et al. 2018; Font-Ribera et al. 2018; Barul et al. 2017; Beane Freeman et al. 2017; Kumar et al. 2014; Bove et al. 2007). Therefore, these THMs are being regulated globally by various agencies and are categorised as possible or probable carcinogens.

The measurement of total trihalomethanes (TTHMs) involves adding together four different species: chloroform (CF), bromodichloromethane (BDCM), dibromochloromethane (DBCM), and bromoform (BF). The four THMs species are considered as a group and guideline values has been set up by many organizations. Due to its carcinogenic nature various international organizations like USEPA, World Health Organization (WHO), European Union (EU) etc., establish the TTHM maximum contaminant levels (MCL) as shown in Table 1.

The presence of NOM in surface water varies at different locations due to difference in environment and different activities by humans. Therefore, the DBPs formed are spatially variable depending upon the concentration and constituents of NOM present in raw water (Brown et al. 2011; Chang et al. 2010a, b; Charisiadis et al. 2015; Delpla et al. 2018; Feungpean et al. 2015; Fooladvand et al. 2011; Guilherme and Rodriguez 2015; Valdivia-Garcia et al. 2016). Numerous studies have demonstrated its variability in relation to a number of variables, including the water supply, pH, temperature, residence time, chlorine dose, type of disinfectant, treatment plant operating parameters, concentration of bromide, applied chlorine dose, residual chlorine, etc., (Kalankesh et al. 2021; Kumari and Gupta 2015; Brown et al. 2011; Legay et al. 2011; Fooladvand et al. 2011; Chang et al. 2010a, b; Chowdhury et al. 2009; Chaib and Moschandreas 2006; Richardson et al. 2003). Owing to climate change or human caused discharges there is temporal difference in source water which are of great concern (States et al. 2013; Wang et al. 2017). The THMs formed in drinking water should be monitored effectively to protect human health. Various lab scale studies have been done resulting in higher concentration of DBPs formed (Leavey-Roback et al., 2016; Linge et al. 2017; Rodriguez and Sérodes 2001). These monitoring program would be more effective when they are studied statistically by modelling to find predictive values. Also, health risk linked with DBPs in water can be optimised by these modelling approach.

Given that chlorine is India's most common disinfectant, there is a considerable risk of THMs exposure in Delhi. The MCL for THMs was established by the Indian government in 2012 to adhere to WHO guidelines (IS:10500 2012). There have been concerns that it was crucial to quantify the THMs because very few studies in India are looking at THM levels in drinking water. The first and second objective was to analyse DBPs spatially i.e., THMs in the municipal supply system of Delhi and to examine the variation of THMs seasonally respectively. The primary goal was to utilise this database to analyse the variation of THMs spatially and seasonally. There is lack of monitoring data nationwide to explain the DBPs formation and the results obtained will help in future water

Table 1 Drinking water guidelines for THMs

DBPs	India	WHO	USEPA (µg/l)	EU	China
CF	200	300	80	–	60
BDCM	60	60	80	–	60
DBCM	100	100	80	–	100
BF	100	100	80	–	100
Total THM	≤1*	≤1*	80	100	≤1*
References	IS:10500 (2012)	WHO (2017)	USEPA (2018)	EUDWD (2019)	CDWQ (2006)

*$\frac{C.CF}{G.V.CF} + \frac{C.BDCM}{G.V.BDCM} + \frac{C.DBCM}{G.V.DBCM} + \frac{C.BF}{G.V.BF} \leq 1$ where C = concentration of species; G.V. = guideline value

management for the populace receiving treated drinking water.

Materials and methods

Study area

Delhi is one of the important metropolitan cities located in Northern India having an area of 1483 km^2. The population of the city is 19.43 million (Census of India 2020). There are nine major water treatment plants (WTP) supplying drinking water in Delhi, namely Chandrawal WTP, Wazirabad WTP, Haiderpur WTP, Bhagirathi WTP, Bawana WTP, Nangloi WTP, Sonia Vihar WTP, Okhla WTP, Dwarka WTP abbreviated as CWTP, WWTP, HWTP, BWTP, BAWTP, NWTP, SVWTP, OWTP, DWTP respectively as shown in Fig. 1 and the command areas of the nine WTPs are referred as CCA, WCA, HCA, BCA, BACA, NCA, SVCA, OCA, DCA respectively. One minor WTP was built at Commonwealth Games Village (CGWTP). The major resource of water in Delhi is Ganga River and Yamuna River which is conventionally treated by processes i.e., coagulation, flocculation, sedimentation, filtration, followed by disinfection before supply. The atmosphere of the city is humid subtropical with high variation of temperature in summers, winters and rainy season which effects the availability of raw water.

Water sampling program and sample collection

From January 2019 to December 2019, a year-long sampling programme was implemented to cover all four seasons i.e., summer (April, May, June), monsoon (July, August September), post monsoon (October, November, December) and winter (January, February, March). Two tap water samples from the treated/processed water after chlorination were taken at the user's end at two locations (L1 and L2) as illustrated in Fig. 1 from the nine Delhi WTPs command areas, with the exception of CGWTP because of its small catchment area. Grab sampling was done at all sampling points in a 40 mL Teflon lined screw cap glass vials washed with cleansing agent, then rinsed with deionized water and dried at 400 °C for 30 min (APHA 2017). The addition of sodium thiosulfate came before sampling to quench the formation of THMs and eradicate chlorine reaction. To prevent THM volatilization the sampling vials were filled carefully eliminating air bubbles inside the sample. The samples were transferred to the laboratory after collection for extraction at

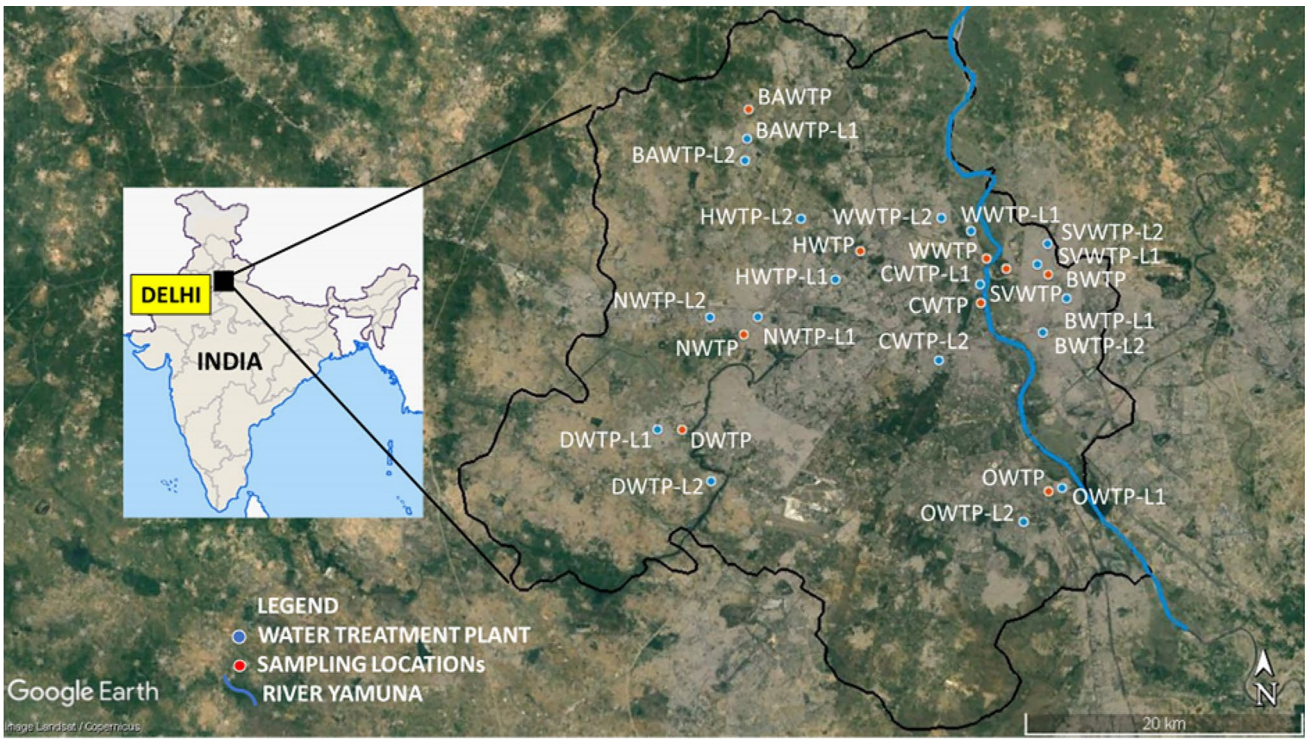


Fig. 1 Sampling locations in command areas of nine WTP's of Delhi

Table 2 Methods, equipment's and parameters analysed in this study

Parameters	Abbreviations	Equipment used	Analytical method
Chloroform	CF	Agilent 7890B & 7000B GC–MS/MS	EPA 551.1 APHA 6232
Bromodichloromethane	BDCM		
Dibromochloromethane	DBCM		
Bromoform	BF		
Temperature	Temp	8603 Mettler Toledo	IS:3025 (Part 9): 2002
pH	pH		IS:3025 (Part 11): 2002
Total organic carbon	TOC	TOC analyzer (TOC-5000, Shimadzu, Japan)	APHA 5310C
Residual chlorine	Res. Cl	Titrimetric	IS:3025 (Part 26): 2002 APHA 4500-Cl-G

4 °C in an ice box away from light. Separate samples were collected for physicochemical parameters.

Analytical methods

The investigation procedures and equipment used in this study are shown in Table 2. A pH meter was used to measure the temperature and pH. The samples were filtered by a glass microfiber (2-µm) then Total Organic Carbon (TOC) was analysed by TOC analyzer in accordance with the 5310C Standard Method (APHA 2017) using combustion techniques. Residual chlorine (in mg/l) was determined by the N,N-diethyl-p-phenylenediamine (DPD) titrimetric method 4500-Cl-G given in Standard method (APHA 2017). Four THMs (CF, BDCM, DBCM and BF) were analysed in according to United States Environmental Protection Agency (USEPA) 551.1 Method (USEPA 1995). Samples were prepared and extracted using liquid–liquid extraction (LLE) method with n-pentane for injection according to 6232-B Standard Method (APHA 2017). The THMs concentration in aqueous samples were determined by gas chromatography coupled with mass spectrophotometer (GC–MS) system, while chromatographic separation of the analyte was carried out with a HP-5ms column, 30 m × 0.25 mm × 0.25 µm (Agilent Technologies). All chemicals were of the analytical grade from Merck (India), solvents and standards were from Sigma-Aldrich. High purity gas Helium, and de-ionized water was used in the study.

Results and discussion

Water quality parameters

The important water quality parameters measured in the command areas of nine WTPs of Delhi are shown in Table 3. All the statistical analyses were done using Microsoft Office Excel for calculating the measured parameters as mean ± standard deviation (SD). The pH measured in the samples showed very little variation in the pH values having no significant difference in the studied region. The maximum pH value in command areas of nine WTPs were found to be highest at HWTP (7.85) followed by BAWTP (7.84), CWTP (7.82), NWTP (7.81), BWTP (7.7), WWTP (7.68), SVWTP (7.66), DWTP (7.62) and lowest at OWTP (7.61). The pH measured in the distribution system was in the range of 6.65–7.85 with an average of 7.29 ± 0.23 for the study period. All the water samples revealed a normal range of pH values. In this study the temperature pattern is also

Table 3 THMs and water quality parameters concentrations in the study area

Parameter	CCA	WCA	HCA	BCA	BACA	NCA	SVCA	OCA	DCA
	Average value ± SD (standard deviation), n = 24 for each WTP								
pH	7.26 ± 0.21	7.29 ± 0.22	7.29 ± 0.26	7.33 ± 0.19	7.35 ± 0.25	7.38 ± 0.24	7.30 ± 0.24	7.20 ± 0.20	7.21 ± 0.20
Temp (°C)	21.46 ± 5.92	21.62 ± 6.04	21.43 ± 6.03	21.45 ± 6.00	21.63 ± 6.18	21.47 ± 6.14	21.53 ± 6.08	21.43 ± 6.13	21.57 ± 6.08
TOC (mg/l)	8.07 ± 2.35	7.27 ± 2.55	6.75 ± 2.94	6.44 ± 2.76	3.79 ± 1.44	5.52 ± 1.94	6.48 ± 2.12	2.49 ± 1.10	2.54 ± 0.91
Res. Cl (mg/l)	1.41 ± 0.25	1.25 ± 0.25	1.17 ± 0.35	1.17 ± 0.38	0.60 ± 0.14	0.94 ± 0.20	1.10 ± 0.23	0.40 ± 0.16	0.42 ± 0.12
CF (µg/l)	49.89 ± 7.65	43.06 ± 8.43	41.31 ± 11.22	39.58 ± 10.87	24.26 ± 4.57	35.03 ± 7.41	40.38 ± 6.26	18.52 ± 6.49	19.58 ± 5.33
BDCM (µg/l)	39.87 ± 5.70	33.94 ± 6.64	30.39 ± 9.61	30.27 ± 10.07	16.14 ± 3.81	25.82 ± 6.18	30.35 ± 5.80	10.63 ± 4.32	11.09 ± 3.66
DBCM (µg/l)	26.27 ± 6.88	23.09 ± 7.06	20.95 ± 7.89	20.33 ± 8.53	9.26 ± 2.56	16.09 ± 4.63	19.06 ± 4.97	3.85 ± 1.98	3.56 ± 1.82
BF (µg/l)	3.51 ± 1.54	2.48 ± 0.94	2.14 ± 1.21	2.38 ± 1.52	0.99 ± 0.52	1.61 ± 0.84	2.38 ± 1.40	0.06 ± 0.18	0.10 ± 0.28
TTHM (µg/l)	119.53 ± 20.51	102.57 ± 21.94	94.80 ± 28.92	92.56 ± 30.26	50.65 ± 10.39	78.55 ± 17.86	92.16 ± 16.59	33.05 ± 12.58	34.34 ± 10.01

similar for all the command areas of nine WTPs. The measured temperature was in the range of 12.60–31.60 °C for the duration of the research with an average of 21.51 ± 5.95 °C.

The maximum TOC value in the command areas of nine WTPs were found to be highest at BWTP (12.92 mg/l) followed by CWTP (12.61 mg/l), HWTP (12.41 mg/l), WWTP (12.38 mg/l), SVWTP (11.41 mg/l), NWTP (9.88 mg/l), BAWTP (7.19 mg/l), OWTP (4.75 mg/l) and lowest at DWTP (4.28 mg/l). The TOC ranged between 0.57 and 12.92 mg/l with an average of 5.48 ± 2.85 mg/l for the study period. TOC concentrations at the command areas of CWTP, WWTP, HWTP, BWTP, SVWTP were found to be higher than BAWTP, NWTP, OWTP and DWTP. Variability was seen in the levels of TOC in the collected tap water reflecting some infestation of wastewater into the distribution network.

Residual chlorine works as a disinfectant in the distribution system. The maximum residual Cl value measured in nine WTPs command areas were found to be highest at BWTP (2.17 mg/l) followed by CWTP (1.83 mg/l), WWTP (1.75 mg/l), HWTP (1.74 mg/l), SVWTP (1.66 mg/l), NWTP (1.32 mg/l), BAWTP (0.97 mg/l), OWTP (0.78 mg/l) and lowest at DWTP (0.63 mg/l). The residual Cl in the water samples ranged from 0.14 to 2.17 mg/l with an average of 0.94 ± 0.43 mg/l for the study period, found to be much higher than the permissible limit, in treated water the free residual Cl ranges between 0.2 and 1 mg/l (IS:10500 2012). Free residual chlorine concentrations were found on the higher side at the command areas of CWTP, WWTP, HWTP, BWTP, NWTP, SVWTP and were found within the permissible limit only at BAWTP, OWTP and DWTP command area. This can be due to an error in estimating the post-chlorination dose or it can be due to some infestation of wastewater into the distribution network. Higher values of residual chlorine suggest more formation of trihalomethanes in the water samples.

Trihalomethanes

The command areas of nine WTPs in Delhi were found to have different concentrations of four THM species. The maximum CF value quantified in the command areas of nine WTPs were found to be highest at BWTP (66.80 µg/l) followed by CWTP (62.34 µg/l), HWTP (57.85 µg/l), SVWTP (57.45 µg/l), WWTP (56.88 µg/l), NWTP (48.68 µg/l), BAWTP (36.37 µg/l), OWTP (32.14 µg/l) and the lowest at DWTP (28.08 µg/l). The CF ranges between 7.53 and 66.80 µg/l with an average of 34.62 ± 13.09 µg/l. The maximum BDCM value quantified in the command areas of nine WTPs were found to be highest at BWTP (56.46 µg/l) followed by CWTP (49.37 µg/l), WWTP (46.26 µg/l), HWTP (45.86 µg/l), SVWTP (43.52 µg/l), NWTP (34.57 µg/l), BAWTP (25.37 µg/l), OWTP (21.37 µg/l) and lowest at DWTP (18.98 µg/l). The BDCM ranges between 3.88 and 56.46 µg/l with an average of 25.39 ± 11.75 µg/l.

The maximum DBCM value quantified in the command areas of nine WTPs were found to be highest at BWTP (45.68 µg/l) followed by WWTP (36.85 µg/l), CWTP (36.54 µg/l), HWTP (33.65 µg/l), SVWTP (28.63 µg/l), NWTP (23.59 µg/l), BAWTP (16.83 µg/l), OWTP (9.22 µg/l) and lowest at DWTP (7.19 µg/l). The DBCM ranges between 0 and 45.68 µg/l with an average of 15.83 ± 9.66 µg/l. Sometimes DBCM concentration levels were not detected (ND) in the water samples. The maximum BF value quantified in the command areas of nine WTPs were found to be highest at BWTP (6.60 µg/l) followed by CWTP (6.26 µg/l), SVWTP (5.67 µg/l), HWTP (4.68 µg/l), WWTP (3.85 µg/l), NWTP (3.68 µg/l), BAWTP (1.68 µg/l), DWTP (1.24 µg/l) and lowest at OWTP (0.65 µg/l). The BF ranges between 0 and 6.60 µg/l with an average of 1.74 ± 1.51 µg/l. BF concentrations were often below the limit of detection in these samples. When BF is present, it means that raw water contains bromide. A summary of the THM concentrations in tap water samples is shown in Table 3.

The variation of TTHMs at all the locations are shown in Fig. 2. The maximum TTHM value measured in the command areas of nine WTPs were found to be highest at BWTP (175.54 µg/l) followed by CWTP (148.38 µg/l), WWTP (142.10 µg/l), HWTP (141.21 µg/l), SVWTP (134.18 µg/l), NWTP (107.19 µg/l), BAWTP (78.93 µg/l), OWTP (63.38 µg/l) and lowest at DWTP (50.74 µg/l). The average TTHM concentration for the study period was found to be 77.58 ± 35.30 µg/l and it ranges from 11.41 µg/l to 175.54 µg/l. WHO has adopted a fractional approach to account the total toxicity of trihalomethanes and the results obtained were checked as per WHO's formula. The TTHM values were found to be higher as per WHO G.V. in the command area of CWTP, WWTP, HWTP, SVWTP and NWTP as shown in Fig. 2e. The variance in TTHM concentration may be a result of the treated water's varied TOC and residual chlorine content as well as its geographical distribution.

For the study period the average TTHM concentrations of the 24 samples taken from each of the command area of nine WTP's ranged from 85.35 to 148.38 µg/l at CWTP, 54.36 to 142.10 µg/l at WWTP, 46.42 to 141.21 µg/l at HWTP, 37.50 to 175.54 µg/l at BWTP, 33.13 to 78.93 µg/l at BAWTP, 21.48 to 107.19 µg/l at NWTP, 57.50 to 134.18 µg/l at SVWTP, 11.41 to 63.38 µg/l at OWTP and 15 to 50.74 µg/l at DWTP. The concentration of THMs measured were lower than those measured in Jharkhand, India (i.e. 231 to 484 µg/l), (Kumari and Gupta 2015). The spatial variation in the average trihalomethanes concentration for the study period is shown in Fig. 3.

Among the THMs formed, CF had the maximum concentration (45% of TTHM) trailed by BDCM (33% of TTHM), DBCM (20% of TTHM) and BF (2% of TTHM).

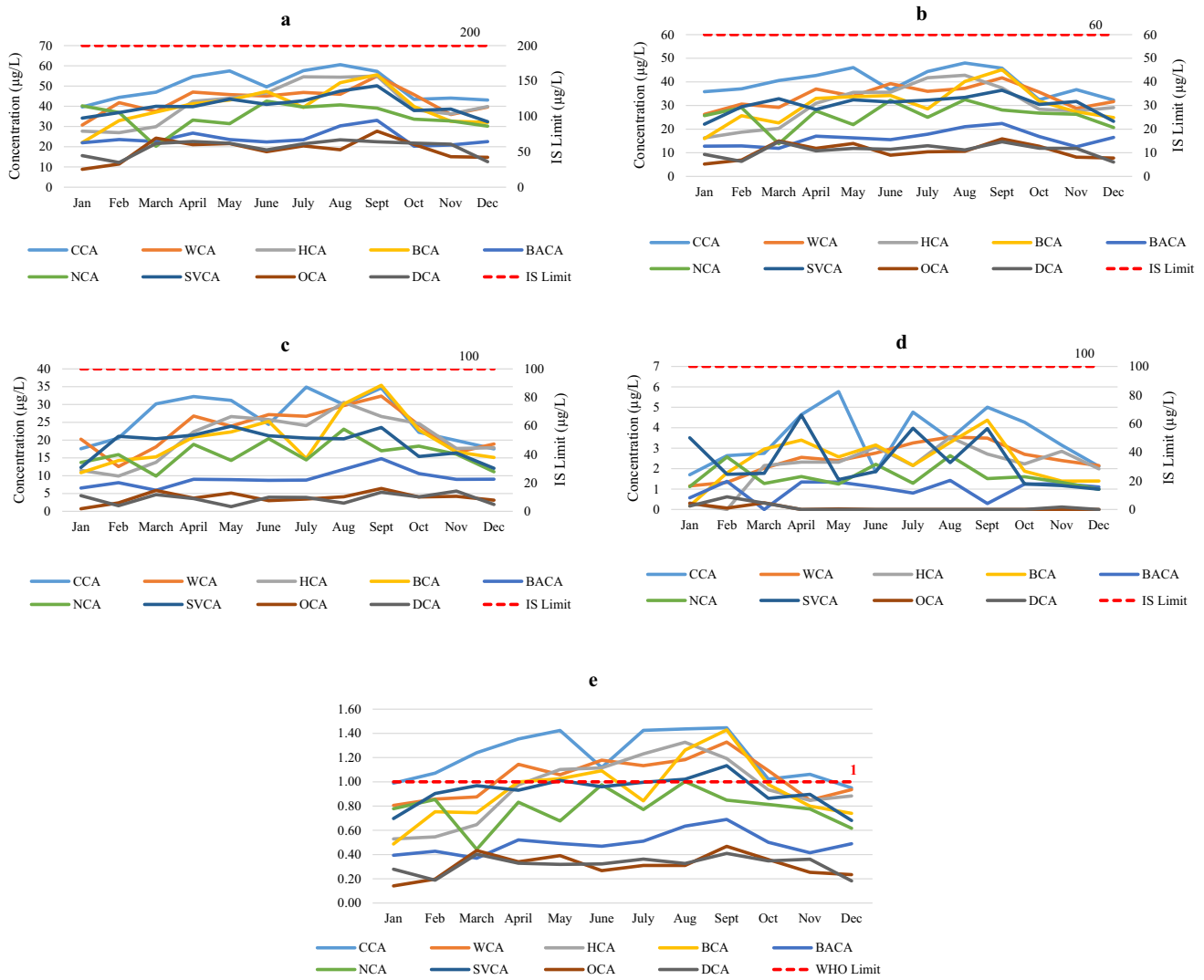


Fig. 2 THMs variation in the distribution system of Delhi **a** CF **b** BDCM **c** DBCM **d** BF **e** TTHM based on WHO's G.V

Mean speciation at all the locations is shown in Fig. 4. The order of contribution in TTHM by each THM species is CF > BDCM > DBCM > BF. Earlier findings have confirmed that the highest concentration levels among the THMs species is of CF (Domínguez-Tello et al. 2015; Uyak et al. 2014). The presence of significant levels of organic matter in the source water may be the reason chloroform typically makes up the highest share of DBPs in chlorinated drinking water. Furthermore, surface water sources are more likely than ground water sources to have greater TOC levels. The above observation is supported by the high concentration of TOC discovered in this investigation. The average THM concentrations of each species measured were mostly within the acceptable limits by Indian standards for drinking water, but they were not within the permissible limit corresponding to WHO and USEPA guidelines for Total THMs. A comparison of THMs concentrations found in drinking water detected in India and around the world is shown in Table 4.

Variation in THMs were analyzed using box and whiskers plot representing the minimum, maximum, mean and median values of THMs for the study period. The median value is shown by the lines inside the boxes, the 75th and 25th percentile values are shown by the top and bottom of the boxes, and the mean value is represented by the cross (x). Outlying points are indicated by the dot (.). The maximum and minimum values are displayed at the top and bottom

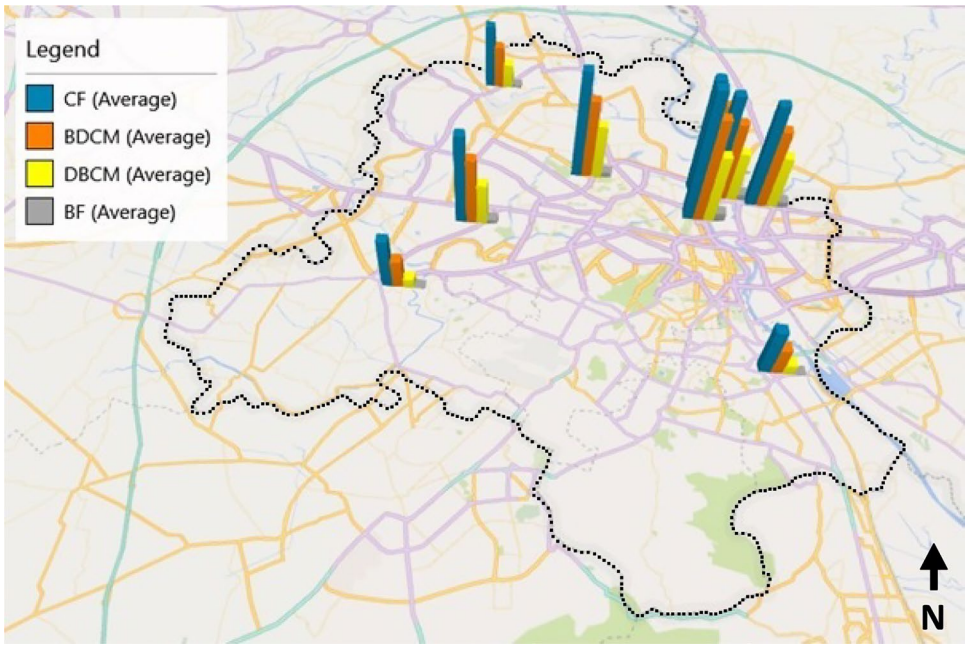


Fig. 3 Spatial variation of THMs concentrations in municipal supply system of Delhi

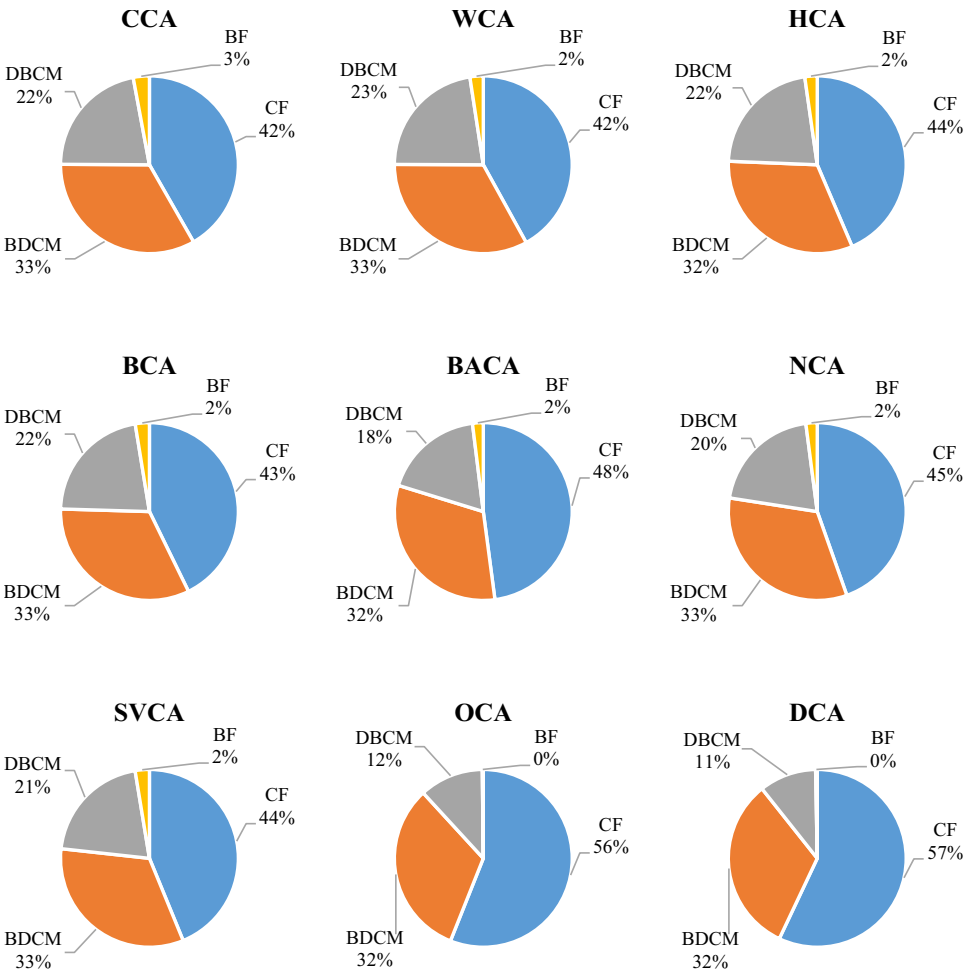


Fig. 4 Mean concentration (%) of THMs in municipal supply system of Delhi

Table 4 DBPs concentrations found in drinking water supplies in India and worldwide

Sample type/ location	Location/ country	DBPs detected	DBPs concentration (µg/L)	Time period	References
DWTP	Thessaloniki, Greece	TCM BDCM DBCM TBM	TCM = 1.33 ± 0.68 BDCM = 1.74 ± 0.76 DBCM = 1.69 ± 0.79 TBM = BDL TTHMs = 4.77 ± 1.93		Albanakis et al. (2021)
DWTP DWDS	China	TCM BDCM DBCM TBM	TCM = 13.2 ± 13.1 (DWTP) 18.0 ± 13.3 (DWDS) BDCM = 9.0 ± 6.5 (DWTP) 9.7 ± 4.1 (DWDS) DBCM = 4.2 ± 4.9 (DWTP) 2.8 ± 1.3 (DWDS) TBM = 1.7 ± 2.7 (DWTP) 1.1 ± 1.2 (DWDS) TTHMs = 24.45 ± 2.39 (DWTP) 32.89 ± 2.93 (DWDS)	2013–2019	Du et al. (2021)
DWDS	Jazan province, KSA	TCM BDCM DBCM TBM	TTHMs = 61.43 ± 2.71	2019–2020	Reyad et al. (2021)
DWDS	Zhejiang Province, China	TCM BDCM DBCM TBM	TTHMs = 10.42–42.25	2018–2020	Dong et al. (2022)
DWDS	Barcelona, Spain	TCM BDCM DBCM TBM	TCM = 17 ± 14 BDCM = 4.6 ± 3.7 DBCM = 8.5 ± 5.7 TBM = 23 ± 15 TTHMs = 45 ± 13	2020	Redondo-Hasselerharm et al. (2022)
DWDS	Egypt	TCM BDCM DBCM TBM	TTHMs = 29.07 to 86.01		Mishaqa et al. (2022)
DWTP DWDS	Kampala, Uganda	TCM BDCM	TCM = 19.86 ± 2.34 (DWTP) 22.39 ± 2.05 (DWDS) BDCM = 4.59 ± 0.59 (DWTP) 10.5 ± 1 (DWDS) TTHMs = 24.45 ± 2.39 (DWTP) 32.89 ± 2.93 (DWDS)	2021	Nshemereirwe et al. (2022)
DWDS	Z City, China	TCM BDCM DBCM TBM	TTHMs = 3.67 to 30.30		Pang et al. (2022)
DWDS	Kharagpur, West Bengal, India	TCM BDCM DBCM TBM	TCM = 5.34 ± 3.34 BDCM = 8.415 ± 5.332 DBCM = 10.76 ± 6.106 TBM = 3.243 ± 1.662 TTHMs = 0.51 to 68.91	2019	Parveen et al. (2023)
DWDS	Huzhou City, Zhejiang Province, China	TCM BDCM DBCM TBM	TCM = 12.532 ± 10.703 BDCM = 3.187 ± 3.435 DBCM = 2.330 ± 4.780 TBM = 0.923 ± 2.813	2019–2020	Liu et al. (2023)
DWDS	Suzhou, China	TCM BDCM DBCM TBM	TCM = 1.8 to 3.2 BDCM = 4.9 to 8.4 DBCM = 4.4 to 8.1 TBM = 1.2 to 2		Li et al. (2023)

Table 4 (continued)

Sample type/ location	Location/ country	DBPs detected	DBPs concentration (µg/L)	Time period	References
DWTP DWDS	Spain	TCM BDCM DBCM TBM	TCM = 0 (DWTP), 2.8 (DWDS) BDCM = 0 (DWTP), 10 (DWDS) DBCM = 3.2 (DWTP), 27.5 (DWDS) TBM = 4.9 (DWTP), 32 (DWDS) TTHMs = 8.2 ± 1.4 (DWTP) 72.3 ± 12.7 (DWDS)	2020–2021	Ramírez et al. (2023)
DWDS	Addis Ababa, Ethiopia	TCM BDCM DBCM TBM	TCM = 4.3 to 79.40 BDCM = 2.42 to 19 DBCM = 2.22 to 13.1 TBM = 2.7 to 3.21 TTHMs = 76.31	2022	Tafesse et al. (2023)
DWDS	Ho Chi Minh City, Vietnam	TCM BDCM DBCM TBM	TCM = 63.9 ± 48 BDCM = 8.95 ± 3.8 DBCM = 3.11 ± 1.0 TBM = 0.446 ± 0.11 TTHMs = 82.7 ± 47	2022	Dat et al. (2023)
DWDS	Delhi, India	TCM BDCM DBCM TBM	TCM = 34.62 ± 13.09 BDCM = 25.39 ± 11.75 DBCM = 15.83 ± 9.66 TBM = 1.74 ± 1.51 TTHMs = 11.41 to 175.54	This study	

DBPs disinfection by products, *TTHMs* total trihalomethanes *TCM* chloroform, *BDCM* bromodichloromethane, *DBCM* dibromochloromethane, *TBM* bromoform, *DWTP* drinking water treatment plant, *DWDS* drinking water distribution system, *BDL* below detection level

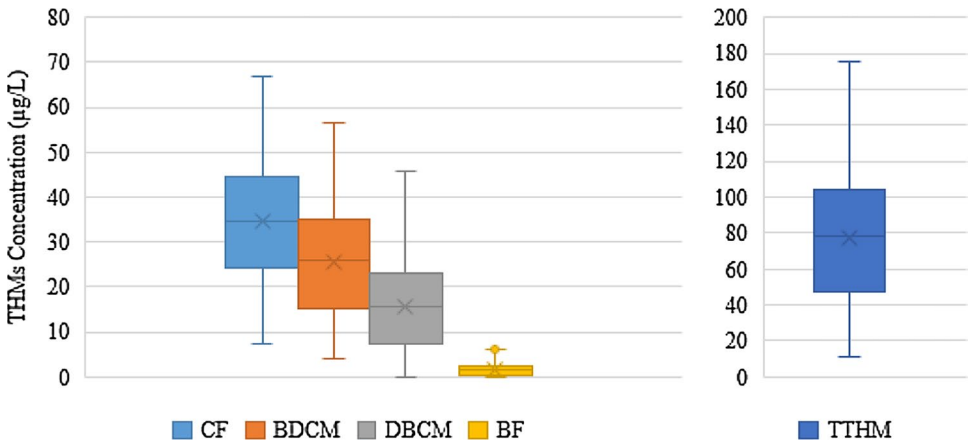


Fig. 5 Box and whiskers plot of THMs

of the whiskers. The median values are displayed as a line across the box. The box and whiskers plot are shown in Fig. 5. CF shows the maximum mean concentration value of 34.62 µg/l followed by BDCM (25.39 µg/l), DBCM (15.83 µg/l) and BF (1.74 µg/l). The median concentrations of CF, BDCM, DBCM and BF were 36.57, 27.51, 15.99 and 1.56 µg/l respectively. The 25th and 75th percentile concentration of CF, BDCM, DBCM and BF were 22.56, 14.93, 8.55, 0.32 µg/l respectively and 43.45, 33.52, 22.49, 2.64 µg/l respectively. The minimum, maximum, mean and median concentration levels of TTHMs was found to be 11.41, 175.54, 77.58 and 78.57 µg/l respectively. The 25th

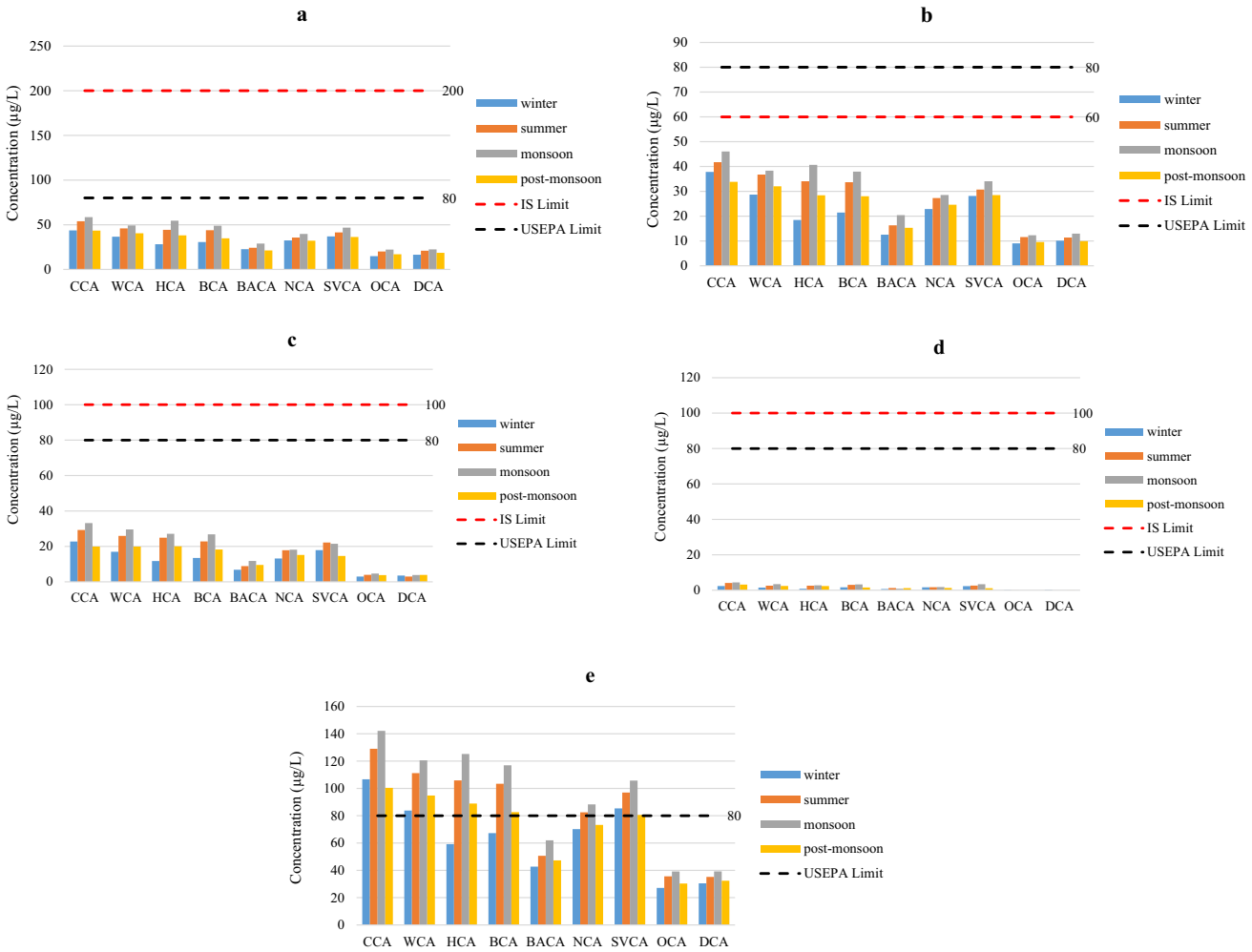


Fig. 6 Seasonal variation of THMs in the municipal supply system of Delhi **a** CF **b** BDCM **c** DBCM **d** BF **e** TTHM

and the 75th percentile concentration of the TTHMs were 45.79 and 101.83 µg/l respectively.

Seasonal variation of THMs

THMs formation significantly varied in the four seasons and their seasonal variation in the municipal water supply system are presented in Fig. 6. The results showed the concentration of CF formed is higher than the other THM species at all locations. The maximum THMs formation was observed at CWTP command area: 142.12 µg/l in the monsoon season. The minimum THMs formation was observed in OWTP command area: 27.11 µg/l in the winter season. Higher THMs concentrations were observed in the monsoon followed by summer and post-monsoon season, while the lowest THMs concentrations were observed in winters. In summers higher temperature generally increased the THM levels and the variations were consistent with other studies reported (Delpla et al. 2018; Domínguez-Tello et al. 2015; Kalankesh et al. 2021; Uyak et al. 2014; Wang et al. 2017). During all the seasons the THMs measured in the water samples were below the MCLs in drinking water according to Indian standards (IS:10500, 2012). In summer and monsoon season the populace residing in the command area of CWTP, WWTP, HWTP, BWTP were exposed to THMs greater than 100 µg/l. Variation of TTHM were much higher at command areas of four WTP i.e., CWTP, WWTP, HWTP, BWTP in summer and monsoon. The lowest levels of TTHM were observed at command areas of OWTP and DWTP in winter and post monsoon season. The THMs measured in the water samples similarly varied at all locations as the source of water is same.

Generally, the variation in THM levels were higher in the order monsoon > summer > post-monsoon > winter. It's possible that this phenomenon is connected to high water temperatures that drop in the post-monsoon and winter months. The mean water temperature was 28.77 °C in the summer, 24.16 °C in the monsoon, 17.03 °C

in the winter, and 16.08 °C in the post-monsoon. Since high temperatures typically speed up reactions, there was a strong correlation found between water temperature and THM production (Scheili et al. 2015; Rahman et al. 2014; Chowdhury et al. 2011). The summertime had greater THM levels than the winter, according to numerous studies (Ahmed et al. 2019; Guilherme and Rodriguez 2014; Lee et al. 2013; Mercier Shanks et al. 2013; O'Driscoll et al. 2018; Sadeghi et al. 2019; Scheili et al. 2015).

THMs in the water could be detected by an organic precursor like TOC, which measures the NOM (Guilherme and Rodriguez 2014; Serrano et al. 2015; Zhou et al. 2019). The average THM levels in our study were greater in the monsoon than they were in the summer. In Delhi's subtropical monsoon climate, the monsoon brings exceptionally high rainfall (around 380.1 mm precipitation) despite the city's maximum average temperature of 28.77 °C during the summer. Water resources NOM concentration is increased by surface runoff caused by rain, raising the TOC level in drinking water. Its average concentrations in the samples were found to be 3.57 mg/l in winter, 5.81 mg/l in summer, 7.82 mg/l in monsoon and 4.73 mg/l in post-monsoon, so much greater during the monsoon. Compared to winter, the monsoon had 2.2 times greater TOC levels. Warm weather and corresponding high TOC values are likely to blame for the greater average THM concentration during the monsoon.

The monsoon season brings with it greater surface runoff, algal development, and biomass production than the winter months, which transferred more NOM and other contaminants from the regions of water catchment to the sources of raw water. Due to higher TOC and higher temperatures recorded, it is possible for bacteria to grow more quickly in the distribution system. Additionally, stronger chlorine dosages are typically used during this season to prevent infections and regulate the growth of algae, which raises the content of THMs. Furthermore, a greater residual chlorine content could potentially encourage the formation of THMs (Gallard and Von Gunten 2002). Average concentration of residual chlorine in the samples were found to be 0.79 mg/l in winter, 0.97 mg/l in summer, 1.14 mg/l in monsoon and 0.86 mg/l in post-monsoon, so markedly higher in the monsoon. Moreover, increased disinfectant dosages and organic precursors that react with chlorine may be the cause of summer and monsoon THM concentrations (Serrano et al. 2015).

Winter and the post-monsoon seasons have the minimal THM levels, It could be connected to a decreased rate of reaction that reduced THM generation. At lower temperatures, the rate at which chlorine and NOM react to generate THMs is slower (Valdivia-Garcia et al. 2016). In contrast, the average temperature in post-monsoon (16.08 °C) and in winter (17.03 °C) is notably lower than during the summer (28.77 °C). Additionally, there was significantly less rainfall in post-monsoon (about 41.6 mm precipitation) and in winter (about 58 mm precipitation). Numerous causes could be responsible for the seasonal variations in THM concentrations found in Delhi's drinking water. These discrepancies could be caused by seasonal changes in the characteristics and makeup of organic precursors, the source water's quality, the water's temperature, the quantity of chlorine supplied, and the residual chlorine (Brown et al. 2011; Pieri et al. 2014; Won et al. 2013). Determining the effect of each operational and water quality characteristic on the occurrence of THM in water distribution networks is difficult, though, due to the interconnections between parameters (Rodriguez and Sérodes 2001).

The emergence of THMs, a type of chlorinated DBP, in drinking water has highlighted the need to investigate alternative disinfectants and novel treatment methods. Other disinfectants that can be used alone or in conjunction with chlorine include ozone, chloramines, chlorine dioxide, and, more recently, ultraviolet (UV) light (Liu et al. 2021, 2020; Luo et al. 2020; Yang et al. 2013). It is necessary to consider the usage of different disinfection alternatives to chlorination, even though they can result in non-chlorinated DBPs. In general, the health risk associated with DBP exposure in drinking water can be effectively controlled by creating alternate disinfection techniques for low DBPs (Kimura et al. 2019; Mazhar et al. 2022). The removal of organic and inorganic substances before disinfection has been found to be an efficient way to reduce the likelihood for the development of chlorinated DBPs, as these substances act as precursors for DBPs (Matilainen et al. 2010). The NOM can be largely eradicated using a normal treatment by combining or replacing its components with more efficient methods such enhanced coagulation, membrane filtration, and granular activated carbon (GAC) filtering (coagulation, flocculation, sedimentation, and filtration) (Badawy et al. 2012; Mazhar et al. 2020).

To minimise the development of DBPs, models can guide decision-making for operational control during the treatment process. Examples of such controls include adjusting the pH and disinfectant dose or the hydraulic residence duration (contact time) in reservoirs (Chowdhury et al. 2009). Additionally, DBP models can be a helpful to determine where to increase the residual levels of chlorination to guarantee the eradication of bacteria and minimise the development of DBPs. To choose the sampling locations for the distribution system's water quality control, additionally, residual disinfectant models can be combined with the DBP models (Haddad et al. 2014). Considering this, appropriate management, oversight, and control procedures are needed to prevent THM development in drinking water supplies and shield people from their harmful consequences. To lower the content of THMs in water supplies, DWTP utilities must apply appropriate treatment techniques and control

Table 5 Pearson correlation between water quality parameters and THM

	pH	Temperature (°C)	TOC (mg/l)	Res. Chlorine	TTHM (µg/l)
pH	1				
Temperature (°C)	0.401	1			
TOC (mg/l)	0.259	0.343	1		
Res. chlorine (mg/l)	0.195	0.195	0.767	1	
TTHM (µg/l)	0.188	0.226	0.934	0.801	1

procedures. To lessen the production of THMs, disinfection must be done under carefully monitored circumstances. Future efforts should focus on reducing the total organic carbon content of the water supply sources. Reduced generation of DBPs might result from improved control over operational parameters (pH, disinfection contact duration, etc.). Finally, to analyse trends and supervise the water treatment and distribution system, regular THM monitoring, and regulation is necessary.

Correlation between water quality parameters and THMs

Correlation matrix was established by Pearson's correlation to evaluate the most prominent parameter accountable in THMs formation and the correlation is shown in Table 5.

From the Pearson correlation matrix, a weak relationship was obtained between total THMs formation with pH (r = 0.188) and temperature (r = 0.226). A strong relationship was obtained between THM formation with TOC (r = 0.934) and residual chlorine (r = 0.801). This indicates that pH and temperature are not much influential parameter in the formation of THM. For an increase in the amount of TOC and residual chlorine in treated water will lead to rise in THM formation. This indicates that TOC and residual chlorine will be the influential parameter in the formation of THM.

Conclusion

In this research 216 water samples from the command areas of nine WTP's of Delhi were analysed during January 2019 to December 2019 for THMs monitoring. It was observed that the treated water samples at various locations showed higher total THMs concentrations ranging from 11.41 to 175.54 µg/l. The concentration of THMs species were found to be varying between: CF (7.53–66.80 µg/l), BDCM (3.88–56.46 µg/l), DBCM (ND-45.68 µg/l) and BF (ND-6.60 µg/l). At some locations the total THM levels were higher than the permissible limits given by WHO (100 µg/l) and USEPA (80 µg/l). Water sample showed higher CF levels followed by BDCM, DBCM and BF levels. Seasonal variation of THMs showed higher concentrations in summer and monsoon and lower concentrations in winter and post-monsoon. Formation of THMs is higher because of high rate of chemical reactions in summer and higher dose of chlorination is usually practiced in monsoon. High, moderate and weak correlations were observed between water quality parameters and THMs. Seasonal variation of THMs also depends on TOC (r = 0.934) and residual chlorine (r = 0.801) which are the prominent parameter responsible for THM formation. From the results obtained it is observed that the population living in the command areas of six WTPs are consuming water with high level of THMs. The CWTP command area showed the highest level of THMs followed by WWTP command area > HWTP command area > BWTP command area > SVWTP command area > NWTP command area. Population living in command area of BWTP command area, OWTP command area and DWTP command area are consuming water with low level of THMs. The existence of THMs in water suggests the presence of other DBPs. These uncharacterized DBPs can also be reduced by controlling the formation THMs in drinking water. Due to high levels of THMs measured in this study, necessary actions are required by the concerned authorities to monitor the THM formation.

Acknowledgements The authors acknowledge and express there thanks to the administration of Jamia Millia Islamia and are highly thankful for providing the research facilities used in this work.

Author contributions Mohd. Aamir Mazhar: Investigation, Data Curation, Writing Original Draft, Sirajuddin Ahmed: Conceptualization, Methodology, Resources, Writing Review and Editing, Geeta Singh: Project Administration, Writing Review and Editing, Azhar Husain: Supervision, Writing Review and Editing, Rahisuddin: Formal Analysis, Visualization, Writing Review and Editing. All authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received for conducting this study.

Data availability The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Declarations

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

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate Not applicable.

Consent to publish Not applicable.

References

Ahmed F, Khan TA, Fakhruddin ANM, Rahman MM, Mazumdar RM, Ahmed S, Imam MT, Kabir M, Abdullah ATM (2019) Estimation and exposure concentration of trihalomethanes (THMs) and its human carcinogenic risk in supplied pipeline water of Dhaka City. Bangladesh Environ Sci Pollut Res 26:16316–16330. https://doi.org/10.1007/s11356-019-05049-6

Albanakis C, Tsanana E, Fragkaki AG (2021) Modeling and prediction of trihalomethanes in the drinking water treatment plant of Thessaloniki. Greece J Water Process Eng 43:102252. https://doi.org/10.1016/j.jwpe.2021.102252

APHA (2017) Standard Methods for the Examination of Water and Wastewater. Am Public Health Assoc 51:1–1274. https://doi.org/10.2105/AJPH.51.6.940-a

Badawy MI, Gad-Allah TA, Ali MEM, Yoon Y (2012) Minimization of the formation of disinfection by-products. Chemosphere 89:235–240. https://doi.org/10.1016/j.chemosphere.2012.04.025

Barul C, Fayossé A, Carton M, Pilorget C, Woronoff AS, Stücker I, Luce D (2017) Occupational exposure to chlorinated solvents and risk of head and neck cancer in men: a population-based case-control study in France. Environ Health A Glob Access Sci Source 16:1–12. https://doi.org/10.1186/s12940-017-0286-5

Beane Freeman LE, Cantor KP, Baris D, Nuckols JR, Johnson A, Colt JS, Schwenn M, Ward MH, Lubin JH, Waddell R, Hosain GM, Paulu C, McCoy R, Moore LE, Huang AT, Rothman N, Karagas MR, Silverman DT (2017) Bladder cancer and water disinfection by-product exposures through multiple routes: a population-based case–control study (New England, USA). Environ Health Perspect 125:1–9. https://doi.org/10.1289/EHP89

Bove GE, Rogerson PA, Vena JE (2007) Case control study of the geographic variability of exposure to disinfectant byproducts and risk for rectal cancer. Int J Health Geogr 6:1–12. https://doi.org/10.1186/1476-072X-6-18

Brown D, Bridgeman J, West JR (2011) Predicting chlorine decay and THM formation in water supply systems. Rev Environ Sci Biotechnol 10:79–99. https://doi.org/10.1007/s11157-011-9229-8

CDWQ (2006) National Standard of the People's Republic of China Standards for Drinking Water Quality. Drink Water Qual GB 5749:1–13

Census of India (2020) Population projections for India and States 2011–2036, National Commission on Population Ministry of Health and Family Welfare India 1–270

Chaib E, Moschandreas D (2006) Modeling spatial variation of brominates trihalomethanes in a water distribution system Ontario. Canada Environ Sci Health Part A 40:2447–2464

Chang EE, Guo HC, Li IS, Chiang PC, Huang CP (2010a) Modeling the formation and assessing the risk of disinfection by-products in water distribution systems. Environ Sci Health Part A 45:1185–1194

Chang HH, Tung HH, Chao CC, Wang GS (2010b) Occurrence of haloacetic acids (HAAs) and trihalomethanes (THMs) in drinking water of Taiwan. Environ Monit Assess 162:237–250. https://doi.org/10.1007/s10661-009-0792-1

Charisiadis P, Andra SS, Makris KC, Christophi CA, Skarlatos D, Vamvakousis V, Kargaki S, Stephanou EG (2015) Spatial and seasonal variability of tap water disinfection by-products within distribution pipe networks. Sci Total Environ 506–507:26–35. https://doi.org/10.1016/j.scitotenv.2014.10.071

Chowdhury S (2013) Exposure assessment for trihalomethanes in municipal drinking water and risk reduction strategy. Sci Total Environ 463–464:922–930. https://doi.org/10.1016/j.scitotenv.2013.06.104

Chowdhury S, Champagne P, McLellan PJ (2009) Models for predicting disinfection byproduct (DBP) formation in drinking waters: a chronological review. Sci Total Environ 407:4189–4206. https://doi.org/10.1016/j.scitotenv.2009.04.006

Chowdhury S, Rodriguez MJ, Sadiq R (2011) Disinfection byproducts in Canadian provinces: associated cancer risks and medical expenses. J Hazard Mater 187:574–584. https://doi.org/10.1016/j.jhazmat.2011.01.085

Chu WH, Gao NY, Deng Y, Krasner SW (2010) Precursors of dichloroacetamide, an emerging nitrogenous DBP formed during chlorination or chloramination. Environ Sci Technol 44:3908–3912. https://doi.org/10.1021/es100397x

Dat ND, Chau VNM, Tran ATK (2023) Temporal and spatial distribution of disinfection byproducts in drinking water supplied to the mega city of vietnam and assessment of the associated risks. Expo Health. https://doi.org/10.1007/s12403-023-00542-3

Delpla I, Florea M, Pelletier G, Rodriguez MJ (2018) Optimizing disinfection by-product monitoring points in a distribution system using cluster analysis. Chemosphere 208:512–521. https://doi.org/10.1016/j.chemosphere.2018.06.009

DeMarini DM (2021) A review on the 40th anniversary of the first regulation of drinking water disinfection by-products. Environ Mol Mutagen 61(6):1–26. https://doi.org/10.1002/em.22378.A

Domínguez-Tello A, Arias-Borrego A, García-Barrera T, Gómez-Ariza JL (2015) Seasonal and spatial evolution of trihalomethanes in a drinking water distribution system according to the treatment process. Environ Monit Assess 189:1–19. https://doi.org/10.1007/s10661-015-4885-8

Dong F, Pang Z, Yu J, Deng J, Li X, Ma X, Dietrich AM, Deng Y (2022) Spatio-temporal variability of halogenated disinfection by-products in a large-scale two-source water distribution system with enhanced chlorination. J Hazard Mater 423:127113. https://doi.org/10.1016/j.jhazmat.2021.127113

Du Y, Zhao L, Ban J, Zhu J, Wang S, Zhu X, Zhang Y, Huang Z, Li T (2021) Cumulative health risk assessment of disinfection by-products in drinking water by different disinfection methods in typical regions of China. Sci Total Environ 770:144662. https://doi.org/10.1016/j.scitotenv.2020.144662

Elshorbagy WE, Abu Qdais H, Elsheamy MK (2000) Simulation of THM species in water distribution systems. Water Res 34:3431–3439. https://doi.org/10.1016/S0043-1354(00)00231-1

EUDWD (2019) Drinking water directive. Counc EU 2019:1–124

Feungpean M, Panyapinyopol B, Elefsiniotis P, Fongsatitkul P (2015) Development of statistical models for trihalomethane (THM) occurrence in a water distribution network in Central Thailand. Urban Water J 12:275–282. https://doi.org/10.1080/1573062X.2013.871042

Font-Ribera L, Gràcia-Lavedan E, Aragonés N, Pérez-Gómez B, Pollán M, Amiano P, Jiménez-Zabala A, Castaño-Vinyals G, Roca-Barceló A, Ardanaz E, Burgui R, Molina AJ, Fernández-Villa T, Gómez-Acebo I, Dierssen-Sotos T, Moreno V, Fernandez-Tardon G, Peiró R, Kogevinas M, Villanueva CM (2018) Long-term exposure to trihalomethanes in drinking water and breast cancer in the Spanish multicase-control study on cancer (MCC-SPAIN). Environ Int 112:227–234. https://doi.org/10.1016/j.envint.2017.12.031

Fooladvand M, Ramavandi B, Zandi K, Ardestani M (2011) Investigation of trihalomethanes formation potential in Karoon River water. Iran Environ Monit Assess 178:63–71. https://doi.org/10.1007/s10661-010-1672-4

Gallard H, Von Gunten U (2002) Chlorination of natural organic matter: Kinetics of chlorination and of THM formation. Water Res 36:65–74. https://doi.org/10.1016/S0043-1354(01)00187-7

Grenoble Z, Zhang CC, Ahmed S, Jeffcoat SB, Karanfil T, Selbes M, Kaplan SS, Begum S, Ahmad R (2007) Physico-chemical processes. Water Environ Res 79:1228–1296. https://doi.org/10.2175/106143007X218395

Guilherme S, Rodriguez MJ (2014) Occurrence of regulated and non-regulated disinfection by-products in small drinking water systems. Chemosphere 117:425–432. https://doi.org/10.1016/j.chemosphere.2014.08.002

Guilherme S, Rodriguez MJ (2015) Short-term spatial and temporal variability of disinfection by-product occurrence in small drinking water systems. Sci Total Environ 518–519:280–289. https://doi.org/10.1016/j.scitotenv.2015.02.069

Haddad M, Mcneil L, Omar N (2014) Model for predicting disinfection by-product (DBP) formation and occurrence in intermittent water supply systems: palestine as a case study. Arab J Sci Eng 39:5883–5893. https://doi.org/10.1007/s13369-014-1200-x

IS:10500 (2012) DRINKING WATER—SPECIFICATION (Second Revision) IS 10500 (2012): drinking water. Bur India Stand 25:1–18

Jones RR, DellaValle CT, Weyer PJ, Robien K, Cantor KP, Krasner S, Beane Freeman LE, Ward MH (2019) Ingested nitrate, disinfection by-products, and risk of colon and rectal cancers in the iowa women's health study cohort. Environ Int 126:242–251. https://doi.org/10.1016/j.envint.2019.02.010

Kalankesh LR, Zazouli MA, Susanto H, Babanezhad E (2021) Variability of TOC and DBPs (THMs and HAA5) in drinking water sources and distribution system in drought season: the North Iran case study. Environ Technol (united Kingdom) 42:100–113. https://doi.org/10.1080/09593330.2019.1621952

Kimura SY, Cuthbertson AA, Byer JD, Richardson SD (2019) The DBP exposome: development of a new method to simultaneously quantify priority disinfection by-products and comprehensively identify unknowns. Water Res 148:324–333. https://doi.org/10.1016/j.watres.2018.10.057

Krasner SW (2009) The formation and control of emerging disinfection by-products of health concern. Philos Trans R Soc A Math Phys Eng Sci 367:4077–4095. https://doi.org/10.1098/rsta.2009.0108

Kumar S, Forand S, Babcock G, Hwang SA (2014) Total trihalomethanes in public drinking water supply and birth outcomes: a cross-sectional study. Matern Child Health J 18:996–1006. https://doi.org/10.1007/s10995-013-1328-4

Kumari M, Gupta SK (2015) Modeling of trihalomethanes (THMs) in drinking water supplies: a case study of eastern part of India. Environ Sci Pollut Res 22:12615–12623. https://doi.org/10.1007/s11356-015-4553-0

Leavey-Roback SL, Sugar CA, Krasner SW, Suffet IHM (2016) NDMA formation during drinking water treatment: a multivariate analysis of factors influencing formation. Water Res 95:300–309. https://doi.org/10.1016/j.watres.2016.02.060

Lee J, Kim ES, Roh BS, Eom SW, Zoh KD (2013) Occurrence of disinfection by-products in tap water distribution systems and their associated health risk. Environ Monit Assess 185:7675–7691. https://doi.org/10.1007/s10661-013-3127-1

Legay C, Rodriguez MJ, Miranda-Moreno L, Sérodes JB, Levallois P (2011) Multi-level modelling of chlorination by-product presence in drinking water distribution systems for human exposure assessment purposes. Environ Monit Assess 178:507–524. https://doi.org/10.1007/s10661-010-1709-8

Li C, Wang D, Xu X, Wang Z (2017) Formation of known and unknown disinfection by-products from natural organic matter fractions during chlorination, chloramination, and ozonation. Sci Total Environ 587–588:177–184. https://doi.org/10.1016/j.scitotenv.2017.02.108

Li J, Zhang Z, Aziz MT, Granger CO, Qiang Z, Richardson SD, Dong H (2023) Simple and sensitive method for synchronous quantification of regulated and unregulated priority disinfection byproducts in drinking water. Anal Chem 95:10975–10983. https://doi.org/10.1021/acs.analchem.3c01013

Linge KL, Kristiana I, Liew D, Nottle CE, Heitz A, Joll CA (2017) Formation of n-nitrosamines in drinking water sources: case studies from Western Australia. J Am Water Works Assoc. https://doi.org/10.5942/jawwa.2017.109.0036

Liu J, Lujan H, Dhungana B, Hockaday WC, Sayes CM, Cobb GP, Sharma VK (2020) Ferrate(VI) pretreatment before disinfection: An effective approach to controlling unsaturated and aromatic halo-disinfection byproducts in chlorinated and chloraminated drinking waters. Environ Int 138:105641. https://doi.org/10.1016/j.envint.2020.105641

Liu H, Zhang X, Fang Y, Fu C, Chen Z (2021) Trade-off control of organic matter and disinfection by-products in the drinking water treatment chain: role of pre-ozonation. Sci Total Environ 770:144767. https://doi.org/10.1016/j.scitotenv.2020.144767

Liu T, Zhang M, Wen D, Fu Y, Yao J, Shao G, Peng Z (2023) Temporal and spatial variations of disinfection by-products in South Taihu's drinking water, Zhejiang Province. China J Water Health 21:1504–1517. https://doi.org/10.2166/wh.2023.149

Luo Y, Feng L, Liu Y, Zhang L (2020) Disinfection by-products formation and acute toxicity variation of hospital wastewater under different disinfection processes. Sep Purif Technol 238:116405. https://doi.org/10.1016/j.seppur.2019.116405

Matilainen A, Vepsäläinen M, Sillanpää M (2010) Natural organic matter removal by coagulation during drinking water treatment: a review. Adv Colloid Interface Sci 159:189–197. https://doi.org/10.1016/j.cis.2010.06.007

Mazhar MA, Khan NA, Ahmed S, Khan AH, Hussain A, Rahisuddin, Changani F, Yousefi M, Ahmadi S, Vambol V (2020) Chlorination disinfection by-products in municipal drinking water: a review. J Clean Prod. https://doi.org/10.1016/j.jclepro.2020.123159

Mazhar MA, Ahmed S, Husain A, Rahisuddi, (2022) Monitoring of trihalomethanes and its cancer risk assessment in drinking water of Delhi City India. Pollution 8:830–843. https://doi.org/10.22059/POLL.2022.334618.1267

Mercier Shanks C, Sérodes JB, Rodriguez MJ (2013) Spatio-temporal variability of non-regulated disinfection by-products within a drinking water distribution network. Water Res 47:3231–3243. https://doi.org/10.1016/j.watres.2013.03.033

Mishaqa ESI, Radwan EK, Ibrahim MBM, Hegazy TA, Ibrahim MS (2022) Multi-exposure human health risks assessment of trihalomethanes in drinking water of Egypt. Environ Res 207:2009–2012. https://doi.org/10.1016/j.envres.2021.112643

Nshemereirwe A, Zewge F, Malambala E (2022) Evaluation of formation and health risks of disinfection by-products in drinking water supply of Ggaba waterworks, Kampala Uganda. J Water Health 20:560–574. https://doi.org/10.2166/wh.2022.272

O'Driscoll C, Sheahan J, Renou-Wilson F, Croot P, Pilla F, Misstear B, Xiao L (2018) National scale assessment of total trihalomethanes in Irish drinking water. J Environ Manag 212:131–141. https://doi.org/10.1016/j.jenvman.2018.01.070

Pan S, An W, Li H, Su M, Zhang J, Yang M (2014) Cancer risk assessment on trihalomethanes and haloacetic acids in drinking water of China using disability-adjusted life years. J Hazard Mater 280:288–294. https://doi.org/10.1016/j.jhazmat.2014.07.080

Pang Z, Zhang P, Chen X, Dong F, Deng J, Li C, Liu J, Ma X, Dietrich AM (2022) Occurrence and modeling of disinfection byproducts in distributed water of a megacity in China: implications for human health. Sci Total Environ. https://doi.org/10.1016/j.scitotenv.2022.157674

Parveen N, Chowdhury S, Goel S (2023) Probabilistic approach for health hazard assessment of trihalomethanes through successive showering events. Environ Sci Pollut Res 30:24793–24803. https://doi.org/10.1007/s11356-021-17087-0

Pieri P, Andra SS, Charisiadis P, Demetriou G, Zambakides N, Makris KC (2014) Variability of tap water residual chlorine and microbial counts at spatially resolved points of use. Environ Eng Sci 31:193–201. https://doi.org/10.1089/ees.2013.0480

Quist AJL, Inoue-Choi M, Weyer PJ, Anderson KE, Cantor KP, Krasner S, Freeman LEB, Ward MH, Jones RR (2018) Ingested nitrate and nitrite, disinfection by-products, and pancreatic cancer risk in postmenopausal women. Int J Cancer 142:251–261. https://doi.org/10.1002/ijc.31055

Rahman MB, Cowie C, Driscoll T, Summerhayes RJ, Armstrong BK, Clements MS (2014) Colon and rectal cancer incidence and water trihalomethane concentrations in New South Wales, Australia. BMC Cancer 14:1–9. https://doi.org/10.1186/1471-2407-14-445

Ramírez Á, de la Morena A, Sánchez N, Peñuela L, Sánchez-Carretero A, Muñoz M, Llanos J (2023) Formation of disinfection by-products within the drinking water production system and distribution network of a real case study. Appl Water Sci 13:1–12. https://doi.org/10.1007/s13201-023-01974-7

Redondo-Hasselerharm PE, Cserbik D, Flores C, Farré MJ, Sanchís J, Alcolea JA, Planas C, Caixach J, Villanueva CM (2022) Insights to estimate exposure to regulated and non-regulated disinfection by-products in drinking water. J Expo Sci Environ Epidemiol. https://doi.org/10.1038/s41370-022-00453-6

Reyad AM, Shahat MMA, Syed MI (2021) The bacteria and the tirhalomethanes in drinking water distribution system in jazan region. Polish J Environ Stud. 30:5711–5722. https://doi.org/10.15244/pjoes/135755

Richardson SD, Postigo C (2015) Formation of DBPs: state of the science. ACS Symp Ser 1190:189–214. https://doi.org/10.1021/bk-2015-1190.ch011

Richardson SD, Postigo C (2016) Discovery of new emerging DBPs by high-resolution mass spectrometry. Comprehensive analytical chemistry. Elsevier, Amsterdam. https://doi.org/10.1016/bs.coac.2016.01.008

Richardson SD, Thruston AD, Rav-Acha C, Groisman L, Popilevsky I, Juraev O, Glezer V, McKAgue AB, Plewa MJ, Wagner ED (2003) Tribromopyrrole, brominated acids, and other disinfection byproducts produced by disinfection of drinking water rich in bromide. Environ Sci Technol 37:3782–3793. https://doi.org/10.1021/es030339w

Rodriguez MJ, Sérodes JB (2001) Spatial and temporal evolution of trihalomethanes in three water distribution systems. Water Res 35:1572–1586. https://doi.org/10.1016/S0043-1354(00)00403-6

Rook JJ (1974) Formation of haloforms during chlorination of natural waters. Water Treat Exam 23:234–243

Sadeghi H, Nasseri S, Yunesian M, Mahvi AH, Nabizadeh R, Alimohammadi M (2019) Trihalomethanes in urban drinking water: Measuring exposures and assessing carcinogenic risk. J Environ Health Sci Eng 17:619–632. https://doi.org/10.1007/s40201-019-00374-x

Scheili A, Rodriguez MJ, Sadiq R (2015) Seasonal and spatial variations of source and drinking water quality in small municipal systems of two Canadian regions. Sci Total Environ 508:514–524. https://doi.org/10.1016/j.scitotenv.2014.11.069

Selvam R, Muniraj S, Duraisamy T, Muthunarayanan V (2018) Identification of disinfection by-products (DBPs) halo phenols in drinking water. Appl Water Sci 8:1–8. https://doi.org/10.1007/s13201-018-0771-1

Serrano M, Montesinos I, Cardador MJ, Silva M, Gallego M (2015) Seasonal evaluation of the presence of 46 disinfection by-products throughout a drinking water treatment plant. Sci Total Environ 517:246–258. https://doi.org/10.1016/j.scitotenv.2015.02.070

Shah AD, Mitch WA (2012) Halonitroalkanes, halonitriles, haloamides, and N-nitrosamines: a critical review of nitrogenous disinfection byproduct formation pathways. Environ Sci Technol 46:119–131. https://doi.org/10.1021/es203312s

States S, Cyprych G, Stoner M, Wydra F, Kuchta J, Monnell J, Casson L (2013) Marcellus shale drilling and brominated THMs in pittsburgh, pa., drinking water. J Am Water Works Assoc 105:432–448. https://doi.org/10.5942/jawwa.2013.105.0093

Tafesse N, Porcelli M, Hirpessa BB, Gasana J, Padhi RK, Garie SR, Ambelu A (2023) Exposure and carcinogenic risk assessment of trihalomethanes (THMs) for water supply consumers in Addis Ababa, Ethiopia. Toxicol Rep 10:261–268. https://doi.org/10.1016/j.toxrep.2023.02.004

USEPA (2018) Edition of the drinking water standards and health advisories tables. Water Supply 25:1205–1241

USEPA (1995) Method 551.1: Determination of chlorination disinfection byproducts, chlorinated solvents , and halogenated pesticides/herbicides in drinking water by liquid-liquid extraction and gas chromatography with electron-capture detection. Environ. Prot. 1–61

Uyak V, Soylu S, Topal T, Karapinar N, Ozdemir K, Ozaydin S, Avsar E (2014) Spatial and seasonal variations of disinfection byproducts (DBPs) in drinking water distribution systems of Istanbul City,Turkey. Environ Forensics 15:190–205. https://doi.org/10.1080/15275922.2014.890145

Valdivia-Garcia M, Weir P, Frogbrook Z, Graham DW, Werner D (2016) Climatic, geographic and operational determinants of trihalomethanes (THMs) in drinking water systems. Sci Rep 6:1–12. https://doi.org/10.1038/srep35027

Wang Y, Small MJ, VanBriesen JM (2017) Assessing the risk associated with increasing bromide in drinking water sources in the monongahela river,Pennsylvania. J Environ Eng 143:04016089. https://doi.org/10.1061/(asce)ee.1943-7870.0001175

WHO (2018) Typhoid vaccine: WHO position paper - March 2018. Wkly Epidemiol Rec 13:153–172

WHO, 2017. Guidelines for Drinking-water Quality, Fourth Edition Incorporating the First Addendum, Fourth Inc. ed.

Won G, Kline TR, LeJeune JT (2013) Spatial-temporal variations of microbial water quality in surface reservoirs and canals used for irrigation. Agric Water Manag 116:73–78. https://doi.org/10.1016/j.agwat.2012.10.007

Yang X, Guo W, Zhang X, Chen F, Ye T, Liu W (2013) Formation of disinfection by-products after pre-oxidation with chlorine dioxide or ferrate. Water Res 47:5856–5864. https://doi.org/10.1016/j.watres.2013.07.010

Zamyadi A, Ho L, Newcombe G, Bustamante H, Prévost M (2012) Fate of toxic cyanobacterial cells and disinfection by-products formation after chlorination. Water Res 46:1524–1535. https://doi.org/10.1016/j.watres.2011.06.029

Zhai H, Zhang X, Zhu X, Liu J, Ji M (2014) Formation of brominated disinfection byproducts during chloramination of drinking water: New polar species and overall kinetics. Environ Sci Technol 48:2579–2588. https://doi.org/10.1021/es4034765

Zhang SH, Miao DY, Liu AL, Zhang L, Wei W, Xie H, Lu WQ (2010) Assessment of the cytotoxicity and genotoxicity of haloacetic acids using microplate-based cytotoxicity test and CHO/HGPRT gene mutation assay. Mutat Res - Genet Toxicol Environ Mutagen 703:174–179. https://doi.org/10.1016/j.mrgentox.2010.08.014

Zhou X, Zheng L, Chen S, Du H, Gakoko Raphael BM, Song Q, Wu F, Chen J, Lin H, Hong H (2019) Factors influencing DBPs occurrence in tap water of Jinhua Region in Zhejiang Province. China Ecotoxicol Environ Saf 171:813–822. https://doi.org/10.1016/j.ecoenv.2018.12.106

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

Received: 29 January 2024 | Revised: 11 March 2024 | Accepted: 17 April 2024

DOI: 10.1002/aoc.7507

FULL PAPER

Applied Organometallic Chemistry WILEY

Ultrafast catalytic reduction of organic pollutants using ternary zinc–copper–nickel layered double hydroxide

Jigyasa Pathak | **Poonam Singh**

Department of Applied Chemistry, Delhi Technological University, Delhi, India

Correspondence
Poonam Singh, Department of Applied Chemistry, Delhi Technological University, Delhi 110042 India.
Email: poonam@dtu.ac.in

Funding information
The authors are grateful to UGC BSR Startup Grant No. F.30-416/2018 (BSR) for the funds to carry out this research. Author JP would like to thank Delhi Technological University for providing a fellowship.

Water is the utmost essential commodity for the support of life process in animals. In recent decades, rapid industrialization and uncontrolled agricultural practices have led to increased level of toxic pollutants in groundwater, including organic pollutants (dyes, nitroarene compounds, antibiotics, etc.), inorganic pollutants (heavy metals, nanoparticles, etc.), and pesticides, posing a serious threat to human health and the environment. Although various technologies such as adsorption, photocatalysis, reverse osmosis, filtration, sedimentation, flocculation, and precipitation are available for the elimination of toxic pollutants from wastewater, issues such as their cost and efficacy limit their usage. Recently, catalysis has been found to be the most effective approach since it exhibits significant capability to eliminate almost all the pollutants and offers the benefit of simple design and low initial cost. In the present study, ternary zinc–copper–nickel layered double hydroxide (ZnCuNi-LDH) synthesized using facile acid hydrolysis method was employed as an efficacious heterogeneous catalyst for the reduction of NACs (para-nitrophenol and para-nitroaniline) and degradation of organic azo dyes (methyl orange, amaranth, and brilliant black). ZnCuNi-LDH exhibited superior catalytic activity for the reduction of all five model pollutants with reduction efficiency of more than 95% within a short time span of 5 min. Therefore, keeping in view the layered structure, high specific surface area, and remarkable catalytic performance, ZnCuNi-LDH holds immense potential for the large-scale catalytic degradation of refractory pollutants.

KEYWORDS
azo dyes, catalytic reduction, layered double hydroxide, nitroarene compounds, wastewater purification

1 | INTRODUCTION

The expulsion of refractory pollutants from water sources holds prime significance for the sustenance of ecosystems and bio-diversities, since the availability of potable water has become a burgeoning predicament because of increasing industrialization and urbanization. The contamination of water by a variety of toxic pollutants including dyes, phenolic compounds, volatile organic compounds (VOCs), drugs, and surfactants is a matter of grave concern as these pollutants exhibit toxic, carcinogenic, and mutagenic effects, thereby threatening human health and the ecosystem.[1] Nitroarene compounds (NACs) and azo dyes are among the most vicious organic pollutants that are responsible for deteriorating water quality. Their ubiquitous presence in water resources is a

 © 2024 John Wiley & Sons Ltd.

direct consequence of the uninhibited discharge of untreated effluents from paper, cosmetics pulp, printing, and textile industries. These pollutants often possess chemical and thermal stability, bioaccumulation capability, and high levels of toxicity, thereby causing severe detrimental effects on aquatic and non-aquatic ecosystems. Thus, the removal of these toxins from water sources has become an issue of serious concern so as to maintain ecological balance (Figure 1).[2]

Till today, a variety of methods such as adsorption, coagulation-flocculation, ultra- and nanofiltration, biological processes, and ion exchange have been utilized for pollutant sequestration from water bodies.[3–5] However, these techniques possess certain limitations, that is, high operating cost and low efficiency, thereby hindering their commercial usage at large scale. During the past few years, catalysis has emerged as a promising strategy for water remediation because of its easy and versatile nature, high efficacy, and economic viability.[6] Specifically, catalytic degradation of pollutants using $NaBH_4$ serves as an effective green method to break down highly stable azo dye and nitroarene molecules into smaller, safer organic moieties, thereby reducing their toxicity. Furthermore, this approach may also be used for conversion of 4-nitrophenol to 4-aminophenol that can be employed as a corrosion inhibitor, an intermediary building block for the manufacture of dyes and pigments, hair colors, drugs (analgesic-phenacetin, antipyretic-paracetamol, etc.), and in the textile and pharmaceutical industry. In addition, catalysis produces safer degradation products that are easily detectable using techniques such as absorption spectroscopy.

Despite this, the applicability of this method is still quite limited because of the time-intensive nature of the reaction, that is, thermodynamic favorability but kinetic unfeasibility. However this drawback can be overcome using several approaches, out of which the most favored one is the usage of the catalyst along with the reducing agent $NaBH_4$ that speeds up the reaction by overcoming the kinetic barrier. Previously, materials such as metal and metal hydroxide and oxide nanoparticles, carbon and carbon-based materials, polymer-based materials, and metal–organic frameworks (MOFs) have been used as co-catalysts for nitroarene reduction and dye degradation.[7–9] For instance, Ahmad et al. synthesized a composite consisting of iron oxide functionalized with Cu nanoparticles and applied it for $NaBH_4$-assisted reduction of several organic pollutants.[10]

These materials possess certain disadvantages such as agglomeration tendency, complex and expensive synthesis methods, and low catalytic efficiency, because of which the use of these materials is not favored. Also, for a material to act as a good catalyst, it should possess certain features such as high specific surface and large number of active sites.

Based on the above criteria, anionic layered materials, especially layered double hydroxides (LDHs), have emerged as breakthrough materials for catalytic applications by virtue of their high surface area, abundance of active sites, biocompatible nature, and low toxicity. LDHs

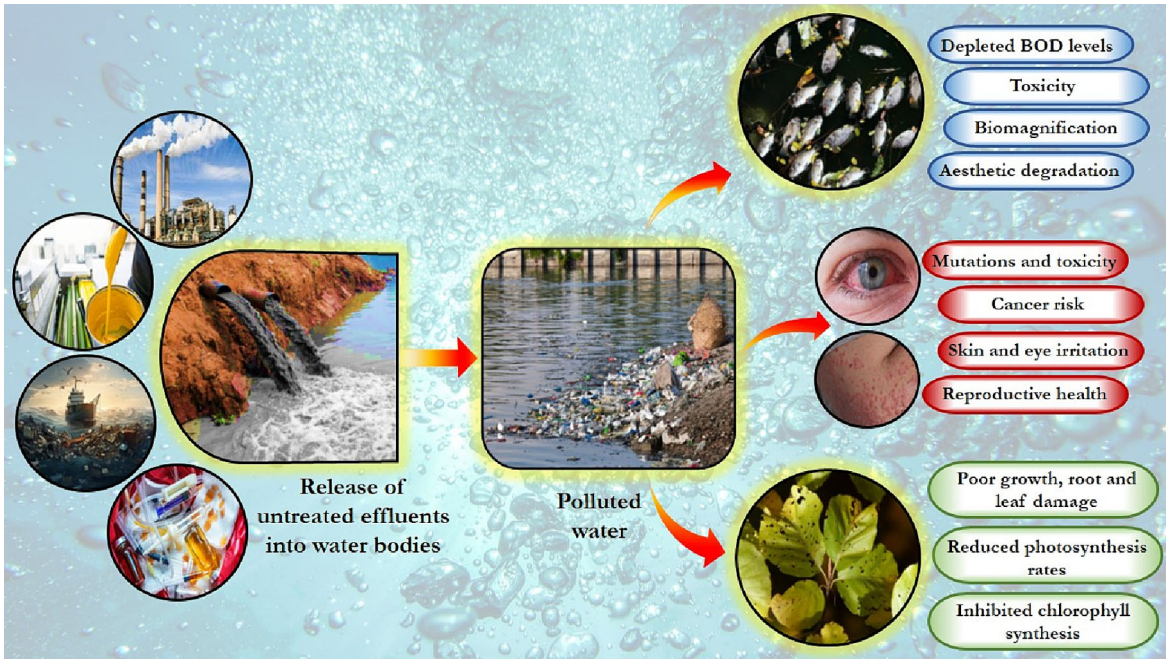


FIGURE 1 Detrimental impacts of water pollution on aquatic, human, and plant health.

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

possess a unique two-dimensional layered structure with positively charged layers composed of divalent (M^{2+} = Zn, Ni, Mg, etc.) and trivalent cations (M^{3+} = Al, Fe, Sb, etc.) existing in octahedral coordination with six hydroxyl moieties, and exchangeable anions (Cl^-, NO_3^-, SO_4^{2-} etc.) and water molecules present in the interlayer region. As a result, these materials possess several exemplary physicochemical features such as structural stability and anion exchange capability along with the unique memory effect property and exhibit a myriad of electronic, magnetic, and optical properties.[11] The use of transition metals (e.g., Cu, Ni, and Co) as the intralayer cations in LDHs further augments these properties because of their special electronic structures, high surface area, and large number of active sites, which endows them with the ability to adsorb various molecules on their surface and activate them. Moreover, since transition metals can attain variable oxidation states, they can be easily paired with a variety of anionic, cationic, and neutral ligands to form homogenous (e.g., organometallic complexes) and heterogeneous catalysts (e.g., transition metal-based LDHs). Several transition metal-containing binary LDHs have been previously employed for the degradation of organic compounds, but they exhibit low catalytic activity.[12] The introduction of another transition metal in the lattice is quite desirable since it warrants improved catalytic activities because of unique geometrical and synergistic interactions between the transition metals. Therefore, transition metal-based-ternary LDHs may be considered as excellent potential candidates for catalytic applications since incorporation of additional transition metal ion in the existing lattice can help fulfill all the characteristics of good catalyst material.[13] Still, research focusing on synthesis and application of transition metal-based-ternary LDHs for $NaBH_4$-assisted reduction of NACs and azo dyes is still quite limited, owing to the difficulty in governing the oxidation states of transition metal ions. Hence, the development of facile synthesis method for transition metal-based ternary LDHs and their usage in catalytic applications is extremely desirable.

Therefore, the objective of this study was to evaluate the efficacy of ternary ZnCuNi-LDH as a catalyst for the reduction of organic pollutants. For this purpose, first, the ZnCuNi-LDH lattice was successfully synthesized using acid hydrolysis approach.[14] Later, a systematic investigation of the catalytic activity of the synthesized ZnCuNi-LDH toward the hydrogenation of 2 NACs (*p*-Nitrophenol and *p*-Nitroaniline) and reductive degradation of different azo dyes (amaranth, brilliant black, and methyl orange) was conducted. The impact of various influencing parameters (catalyst dosage, pollutant concentration) on the catalytic activity of synthesized ZnCuNi-LDH for the reduction of *p*-NP and MO was also studied. Additionally, the universality and the recyclability of ZnCuNi-LDH in the catalysis process were also assessed. The results suggested that the NACs and organic azo dyes underwent rapid and efficacious reduction when $NaBH_4$ was employed as reductant and ZnCuNi-LDH was used as the catalyst.

2 | EXPERIMENTAL SECTION

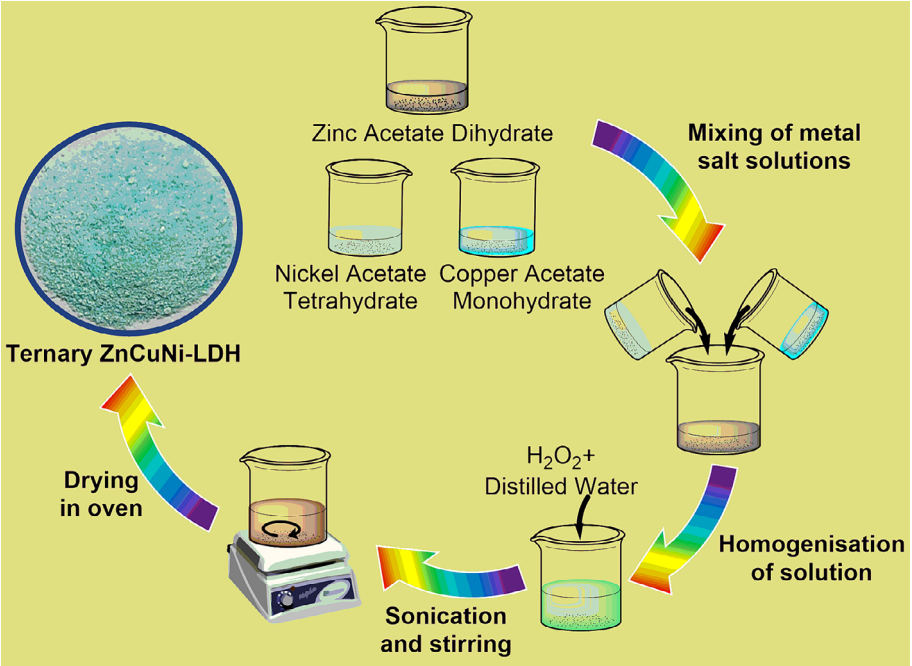
2.1 | Materials and methods

ZnCuNi-LDH was first prepared using acid hydrolysis method as reported in our previous study.[14] For the synthesis, first, 5-mL aqueous solutions of zinc acetate dihydrate (0.4664 g), copper acetate monohydrate (0.0374 g), and nickel acetate tetrahydrate (0.0466 g) were prepared separately. Afterward, copper acetate monohydrate and nickel acetate tetrahydrate solutions were added to zinc acetate dihydrate solution and then stirred for 15 min so as to obtain complete homogeneity. To initiate the acid hydrolysis process, 30% H_2O_2 solution (2.5 mL) was added to reaction solution, followed by distilled water (15 mL), and a shift in the solution color from greenish-blue to brown was observed. Reaction mixture was then ultrasonically treated for 30 min to achieve uniformity in the size of LDH particles, followed by stirring for 1 day at ambient temperature. After completion of stirring, the solution was completely dried at 60°C in oven. The schematic representation of ZnCuNi-LDH synthesis is shown in Scheme 1.

2.2 | Characterization

Crystallographic properties of the sample were studied by the use of powder X-ray diffraction (PXRD) diffractograms collected using a high-resolution Bruker diffractometer (D8 Discover) using Cu Kα radiation. Data were recorded at 298 K over the range of 2θ = 5–70° at a scan rate of 1.0 s/step and step size 0.02. The various functional groups present in synthesized material were analyzed using Fourier-transform infrared (FTIR) spectrum collected with the help of a Perkin Elmer 2000 FTIR spectrometer by the KBr disk technique. Thermal degradation behavior was evaluated by thermo gravimetric analysis (TGA) data that was recorded using the Pyris 1® thermo gravimetric analyzer in the range 50–900°C at a heating rate of 5°C min^{-1} under flowing nitrogen. SEM imaging and EDX analysis of the synthesized samples was done using a Zeiss Gemini SEM microscope. The specific surface area (S_{BET}) and average pore size distribution were

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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



SCHEME 1 Synthesis of ternary ZnCuNi-LDH via acid hydrolysis method.

obtained by BET-BJH (Barrett–Joyner–Halenda) N_2 adsorption–desorption technique on a Quanta chrome Nova-1000 instrument at liquid nitrogen temperature 77 k.

2.3 | Evaluation of catalytic activity

2.3.1 | Catalytic hydrogenation of NACs

10 mM stock solutions of NACs and 50 mM stock solution of $NaBH_4$ were prepared with deionized water, and dilutions were made as required.

2.3.2 | *p*-Nitrophenol

The catalyst-assisted hydrogenation of *p*-nitrophenol (*p*-NP) was accomplished using the reducing agent $NaBH_4$ and ZnCuNi-LDH as the catalyst. Initiation of the reaction was done by adding freshly prepared $NaBH_4$ solution (50 mM, 10 mL) and distilled water (10 mL) to aqueous solution of *p*-NP (1 mM, 10 mL), and it was stirred well to ensure homogeneity. This was followed by the addition of ZnCuNi-LDH catalyst (5 mg) to reaction mixture at room temperature and recording the absorbance at pre-fixed time intervals using Shimadzu UV–Vis spectrophotometer. The kinetic progress of the hydrogenation was assessed by determining the variation in absorbance of reaction mixture at a wavelength of 400 nm at 1-min intervals till the peak completely disappeared and the solution became colorless.

2.3.3 | *p*-Nitroaniline

Similarly, the reduction experiment for *p*-nitroaniline was carried out by first stirring *p*-nitroaniline (*p*-NA) aqueous solution (1 mM, 10 mL) with freshly prepared $NaBH_4$ solution (50 mM, 10 mL) and distilled water (10 mL) and then followed by addition of ZnCuNi-LDH (5 mg) at room temperature. The progression of catalytic reduction was evaluated by measuring the absorbance of reduction mixture at a wavelength of 380 nm at 1-min intervals till the solution was completely decolorized.

2.3.4 | Catalytic degradation of methyl orange (MO), amaranth (Ama), and brilliant black (BB) dyes

To evaluate the decolorization ability of LDHs toward azo dyes, catalytic reduction was carried out using Ama, BB, and MO as model molecules. First, 100 mg L^{-1} stock solutions of Ama, BB, and MO dyes and 50 mM stock solution of $NaBH_4$ were prepared. For the catalytic degradation of MO dye, MO solution (100 mg L^{-1}, 10 mL) was stirred with freshly prepared $NaBH_4$ solution (50 mM, 10 mL) and distilled water (10 mL) to ensure complete mixing. Following this, ZnCuNi-LDH catalyst (5 mg) was added to the reaction mixture. The progress of reaction was assessed by measurement of the absorbance at $\lambda_{max} = 463$ nm for MO at 1-min intervals till the peak disappeared and the solution was completely decolorized. The catalytic reduction of amaranth and BB dyes

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

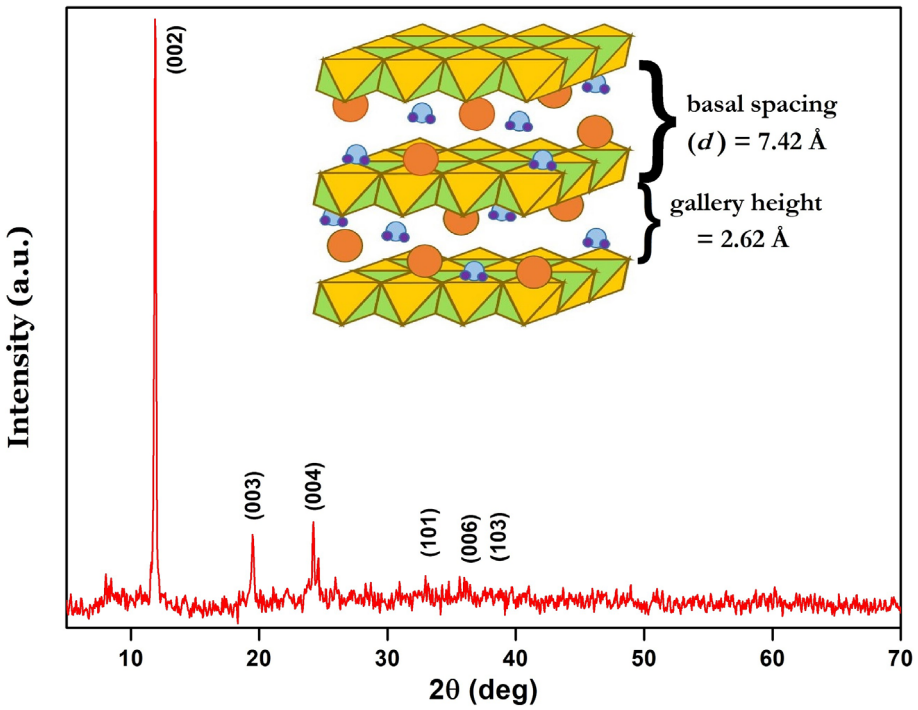


FIGURE 2 PXRD diffractogram of ternary ZnCuNi-LDH.

was studied in a similar manner. Change in absorbance was measured at $\lambda_{max} = 520$ and 490 nm for amaranth and BB, respectively.

The percentage degradation of chosen toxic moieties was calculated using the formula:

$$Percentage\ degradation = \frac{(C_O - C_f)}{C_O} \times 100 \qquad (1)$$

where C_O and C_f denote the initial and terminal concentrations of NAC and azo dyes.

The influence of various parameters such as catalyst dosage and pollutant concentration on the catalytic prowess of ZnCuNi-LDH were also assessed. Further, the universality and recyclability of ZnCuNi-LDH were also studied.

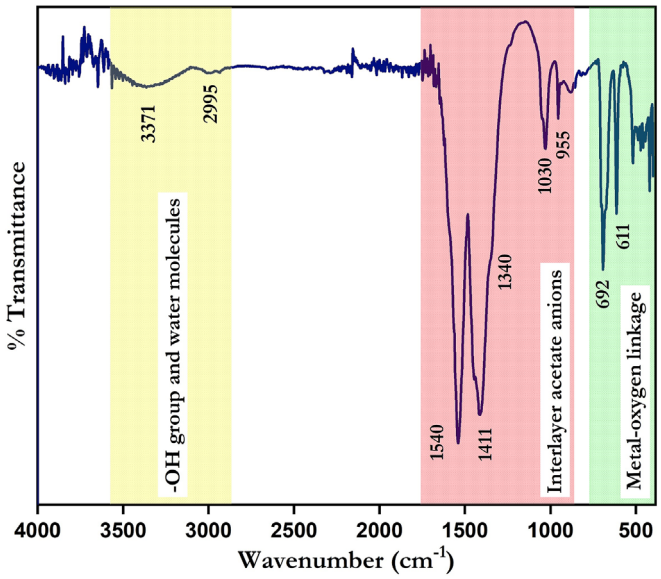


FIGURE 3 FTIR spectrum of ternary ZnCuNi-LDH.

3 | RESULTS AND DISCUSSION

3.1 | Characterization results

The PXRD diffractogram of ZnCuNi-LDH, presented in Figure 2, shows the characteristic diffraction peaks of crystalline layered materials, consisting of high-intensity reflections at low 2θ and low-intensity peaks at higher 2θ. The diffraction peaks can be indexed under hexagonal symmetry having lattice parameters $a = 3.19$ Å and $c = 14.9$ Å.[14] The sharp and symmetric nature of peaks suggests the formation of highly crystalline LDH with coherently stacked 2D layers. The d-spacing for the (002) basal plane, found to be 7.42 Å, and the gallery height, calculated to be 2.62 Å, are indicative of the monolayer alignment of acetate anions in the interlamellar region, keeping in view the thickness of brucite sheet (4.8 Å) and edge to edge length of acetate anion (3.6 Å).[15] Furthermore, d-spacing value of synthesized ZnCuNi-LDH was found to be completely different from those of zinc hydroxy acetate, copper hydroxy acetate, or nickel hydroxy acetate, thereby confirming the formation of

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

lattice in pure phase with isomorphically substituted Cu and Ni cations in the LDH layers.[16–18]

The FTIR spectrum of LDH (Figure 3) exhibits a broad band at 3371 cm^{-1} that can be assigned to νO–H and interlayer water molecules, and the shoulder appearing at 3013 cm^{-1} indicates hydrogen bonded νO–H --- O. The bands at 2995, 1540, and 1411 cm^{-1} can be attributed to C–H stretching vibrations and antisymmetric and symmetric vibration modes of COO– group, respectively. Fingerprint vibration modes of acetate anions were observed at 1340, 1030, 955, 692, and 611 cm^{-1} that can be ascribed to bending/rocking modes of $–CH_3$ group, C–C stretching, and bending and rocking modes of $–CO_2$ moiety, respectively. The bands observed below 700 cm^{-1} are representative of νM–O linkage and indicate the links of metal ions of the –OH groups (O–M–O, M–O, and M–O–M) present in the LDH layers.[16]

The thermal stability and weight loss of prepared ZnCuNi-LDH were determined from the TGA thermogram presented in Figure 4. The thermal decomposition of ZnCuNi-LDH was found to be a three-tier process, including the first stage of weight loss from room temperature to 220°C, attributed to the elimination of physisorbed and interlayer water molecules, with a mass loss of 6.82%. The second stage from 220 to 385°C involves decarboxylation, dehydroxylation, and removal of interlayer acetate ions from the LDH lattice and showed the maximum mass loss of 41.73%. The final stage occurs above 385°C wherein further dehydroxylation and decarboxylation result in collapse of layered arrangement and formation of pure phase mixed metal oxide.[19] The aggregate mass loss during the thermal degradation of ZnCuNi-LDH was found to be 49.9%.

SEM imaging was conducted to identify the morphological and surface textural features of ZnCuNi-LDH. SEM micrograph (Figure 5a) clearly shows the presence of coherent lamellar orientation of LDH, corroborating the formation of layered structure of LDH. In addition, the EDX analysis was carried out to ascertain the elemental composition. The presence of Zn, Cu, Ni, C, and O elements in the EDX profile (Figure 5b) confirms the homogenous distribution of compositional elements in the synthesized heterogeneous ZnCuNi-LDH.

The specific surface area and porosity of ZnCuNi-LDH were determined using N_2 adsorption–desorption study by BET analysis. The specific surface area (S_{BET})

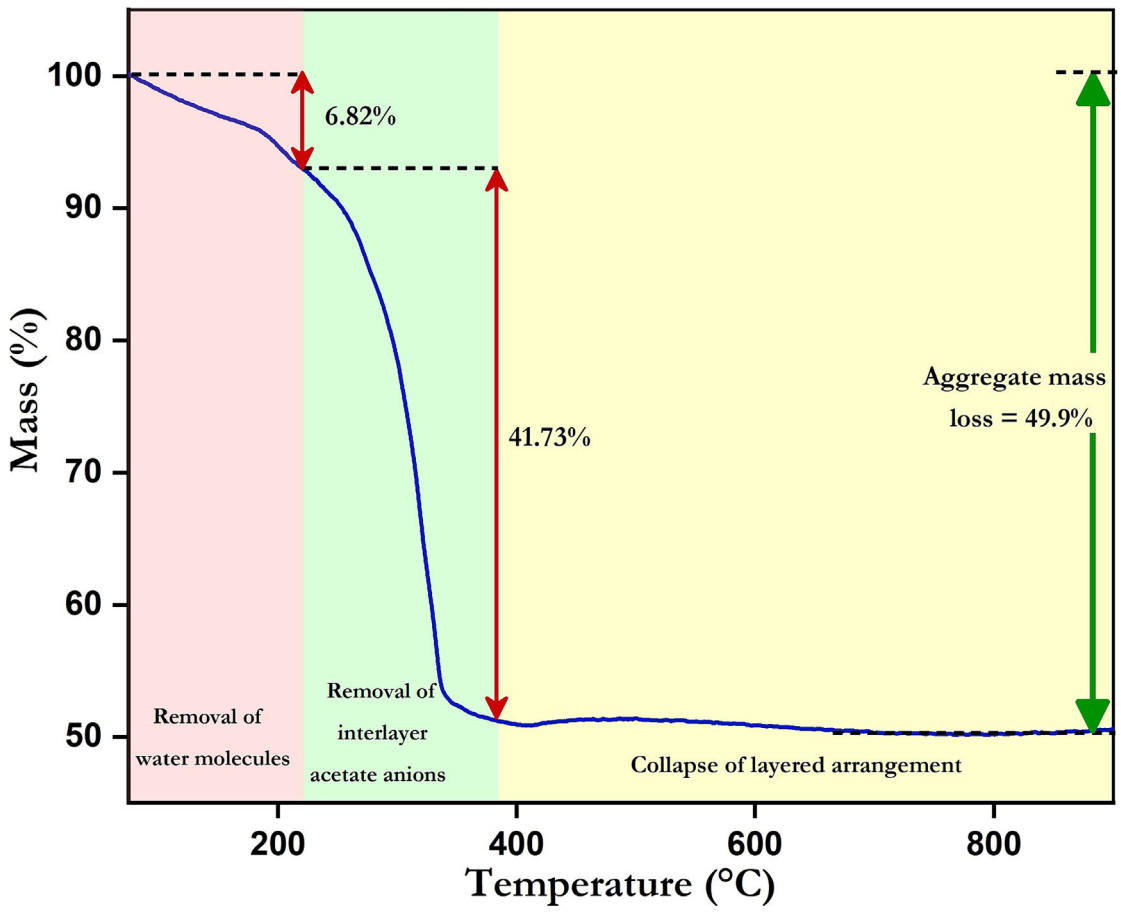


FIGURE 4 TGA thermogram of ternary ZnCuNi-LDH.

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

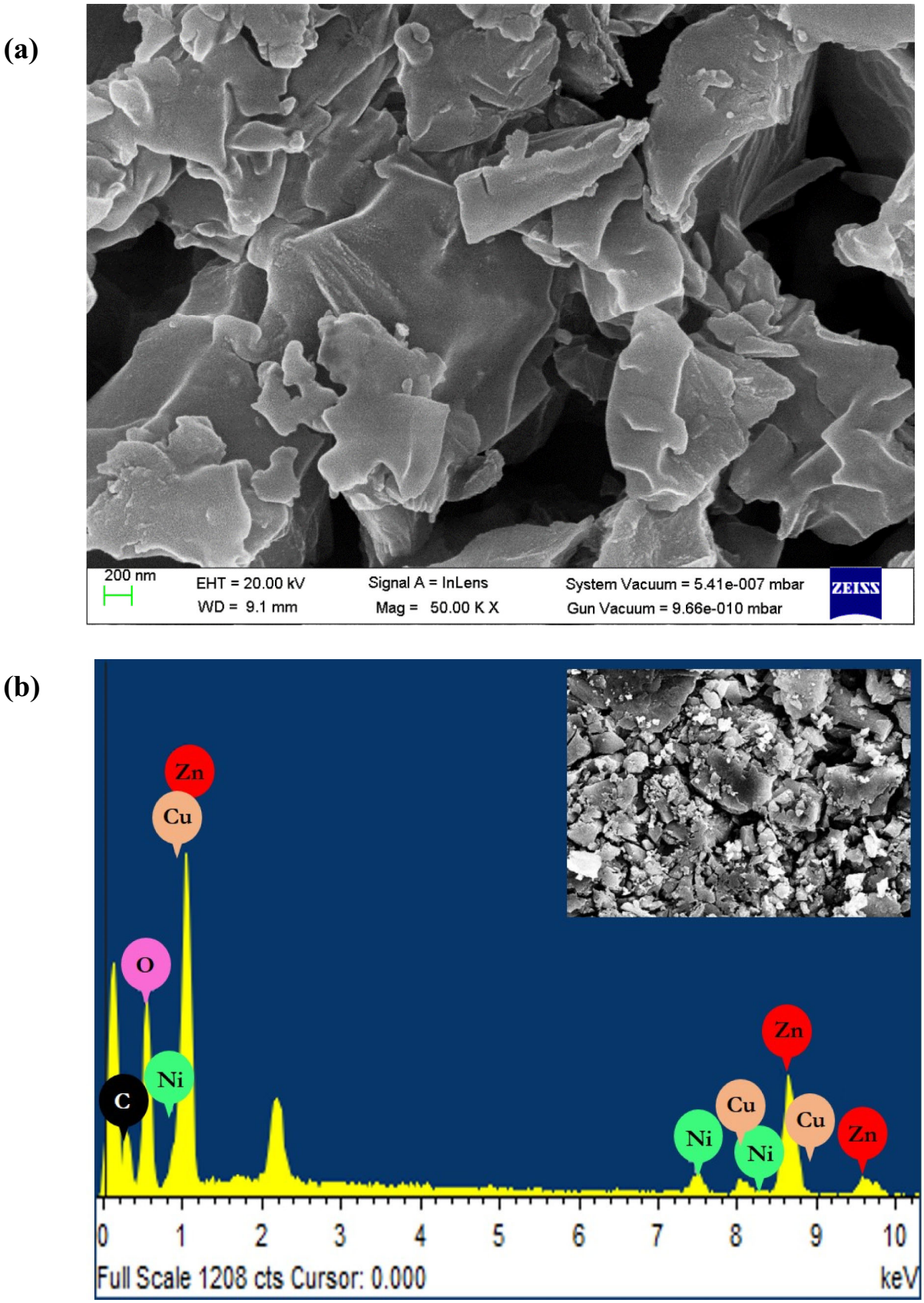


FIGURE 5 (a) SEM micrograph and (b) EDX spectrum of ternary ZnCuNi-LDH with inset SEM image.

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

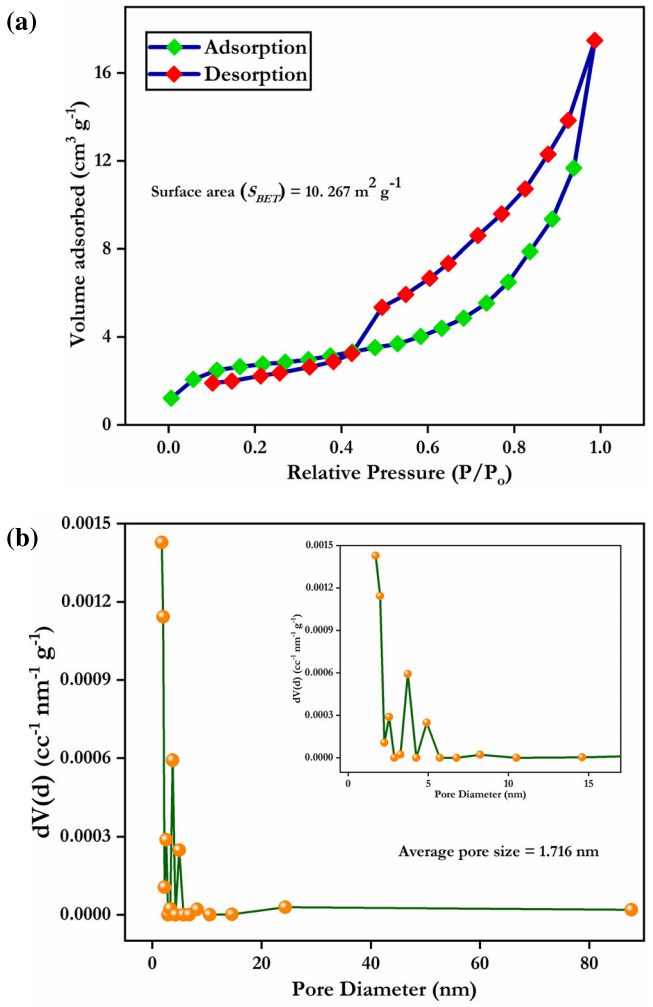


FIGURE 6 (a) N_2 adsorption–desorption isotherm and (b) corresponding pore size distribution plots with inset of enlarged view of pore size distribution between 0 and 20 nm of ternary ZnCuNi-LDH.

was found to be 10.267 m^2g^{-1}, and the pore volume was 0.004 cc g^{-1}. The average pore diameter, calculated from the N_2 desorption data using the BJH method, was found to be 1.716 nm. The shape of the isotherm in the range $0.2 < P/P_o < 0.8$ (Figure 6a) resembles the Type-IV isotherm according to IUPAC classification and indicates the presence of non-cylindrical mesopores in the synthesized LDH. Furthermore, a small amount of N_2 adsorption is observed below $P/P_o = 0.2$ and above $P/P_o = 0.8$, indicating the appearance of micropores and macropores, respectively.[20] The few macropores can be attributed to the existence of inter-particle spaces in the layered structure, usually observed in topotactic materials. The H4-type hysteresis loop also points to the existence of micropores and narrow slit-shaped mesopores in the synthesized ZnCuNi-LDH that appear because of the lamellar arrangement.[21] Additionally, the pore diameter data presented in Figure 6b demonstrate the extensive scattering of micropores and mesopores, as well as macropores throughout the synthesized product, indicating prominent non-uniformity in ZnCuNi-LDH. The presence of large number of mesopores may be a contributing factor in the high catalytic activity of ZnCuNi-LDH.

3.2 | Catalytic activity

The efficacy of synthesized ZnCuNi-LDH as a co-catalyst with $NaBH_4$ reductant for the hydrogenation/reduction of nitroarenes and azo dyes was examined using UV–Visible spectrophotometer. The progress of catalytic reduction was assessed by observing the change in absorbance values of the solution of the selected model molecule. Further, since the concentration of $NaBH_4$ (50 mM) is much greater than that of NACs (1 mM) and azo dyes (100 mg L^{-1}), the reaction rate is considered to be free of the influence of BH_4^- concentration. Hence, the hydrogenation of NACs and reduction of azo dyes was considered to follow pseudo-first-order kinetics, as presented in Equation (2).[22]

$$\ln\left(\frac{C_f}{C_O}\right) = K \tag{2}$$

where K denotes the rate constant and C_O and C_f are the initial and final concentrations of NAC and azo dyes.

Since the evaluated molecules possess distinct colors and yield peaks in the visible region, the reaction rate is evaluated in terms of comparative absorptive intensity. Therefore, using Beer–Lambert law (Equation 3)

$$A = \varepsilon Cl \tag{3}$$

Kinetics of reaction are expressed in terms of absorption by the equation[23]:

$$\ln\left(\frac{C_f}{C_O}\right) = \ln\left(\frac{A_f}{A_O}\right) = kt \tag{4}$$

where A_O and A_f are the initial and final absorbance of NAC and azo dyes.

3.2.1 | Catalytic hydrogenation of *p*-NP and *p*-NA

NACs such as *p*-NP and *p*-NA pose tremendous risk to human health and aquatic life even at trace levels because of their high toxicity and high stability. Therefore, their elimination from water is quite essential.

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

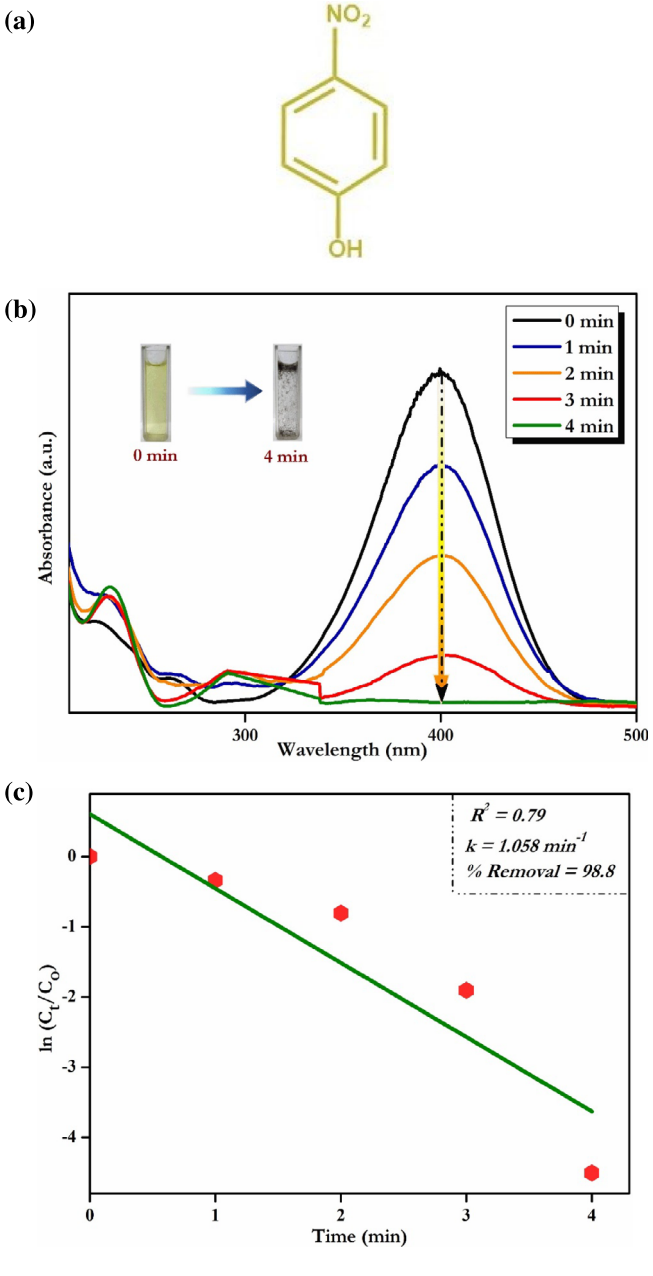


FIGURE 7 (a) Molecular structure of *p*-NP, (b) UV–Vis absorption spectrum, and (c) linear fitting of pseudo-first kinetics model for the $NaBH_4$-assisted catalytic hydrogenation of *p*-NP using ZnCuNi-LDH.

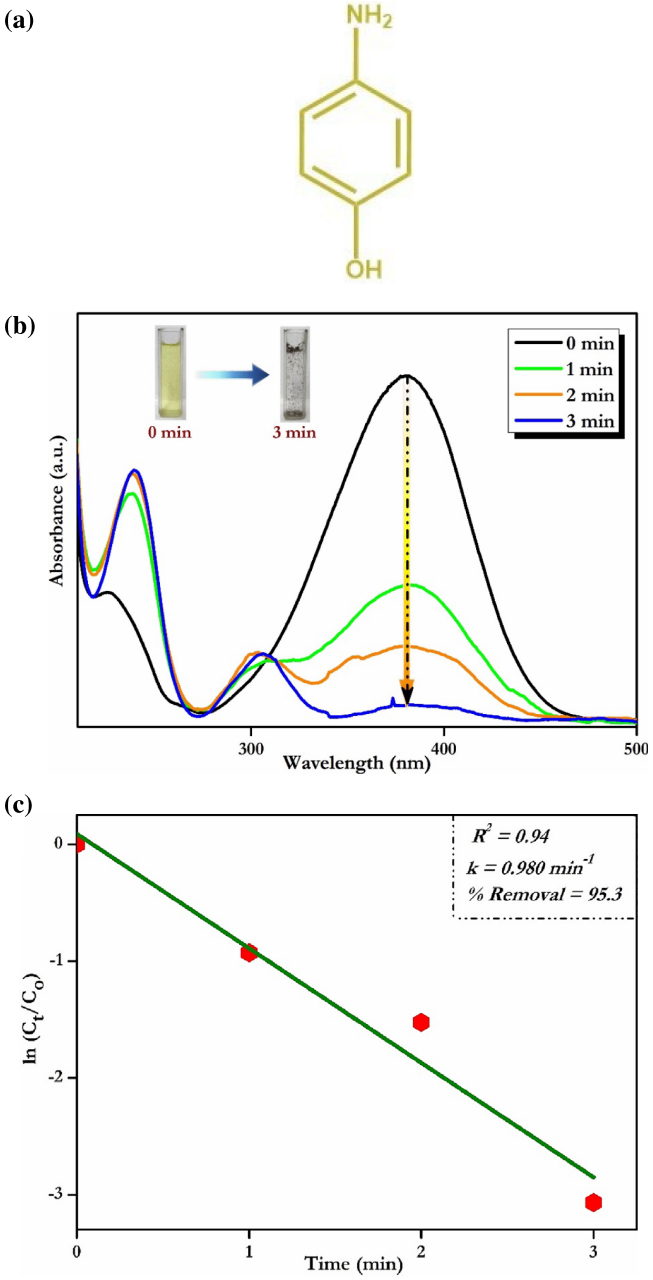


FIGURE 8 (a) Molecular structure of *p*-NA, (b) UV–Vis absorption spectrum, and (c) linear fitting of pseudo-first kinetics model for the $NaBH_4$-assisted catalytic hydrogenation of *p*-NA using ZnCuNi-LDH.

Catalytic hydrogenation using $NaBH_4$ with the assistance of a co-catalyst is an effective way to reduce these NACs to obtain safer reduction products having industrial merits. In the current study, the catalytic hydrogenation of *p*-NP and *p*-NA was studied as a model reaction.

In the case of *p*-NP (Figure 7a), the mixing of the solution of $NaBH_4$ and *p*-NP produces an absorbance peak at 400 nm as a consequence of the generation of nitrophenolate ion, as seen in Figure 7b. Then, with the addition of synthesized LDH, the intensity of peak at 400 nm begins decreasing along with the appearance of a new peak at 302 nm. The appearance of the new peak confirms the production of *p*-AP on *p*-NP reduction and is confirmed by the reduction of the peak at 400 nm. The peak disappears completely after 4 min, indicating the ultrafast hydrogenation of *p*-NP. The percentage reduction of *p*-NP was evaluated using Equation (1) and was found to be 98.8%. The rate of reaction (k_{app}) was evaluated to be 1.058 min^{-1} ($R^2 = 0.79$) in accordance with the pseudo-first-order model (Figure 7c).

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

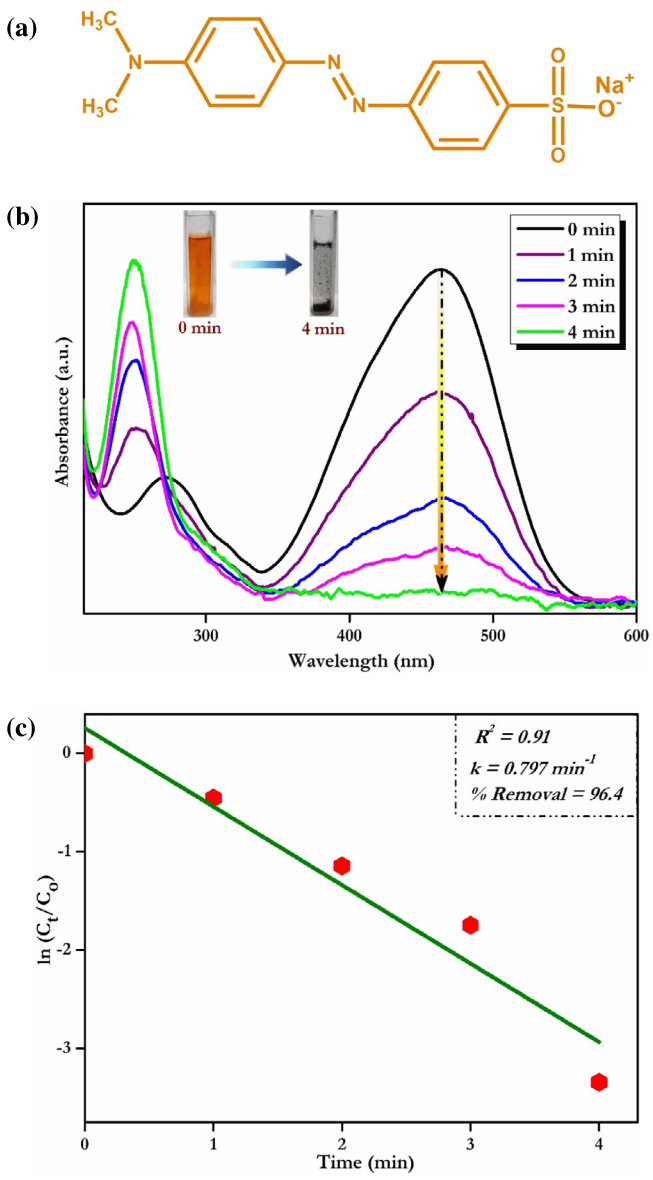


FIGURE 9 (a) Molecular structure of methyl orange, (b) UV–Vis absorption spectrum, and (c) linear fitting of pseudo-first kinetics model for the $NaBH_4$-assisted catalytic reduction of MO using ZnCuNi-LDH.

The catalyst-assisted hydrogenation of *p*-NA (Figure 8a) was evaluated in a similar manner. The plot of UV–Vis spectrum of *p*-NA (Figure 8b) shows prominent peak at 380 nm on mixing of $NaBH_4$ solution and *p*-NA solution. On the introduction of ZnCuNi-LDH to the reaction mixture, the peak intensity at 380 nm decreases rapidly before ultimately disappearing after 3 min. Meanwhile, a new peak appears at 304 nm on the addition of LDH and may be attributed to the formation of *p*-phenylenediamine. The percentage reduction of *p*-NA, calculated using Equation (1), was found to be 95.3%. Using pseudo-first-order kinetics, the rate constant of reduction (k) was calculated to be 0.98 min^{-1} ($R^2 = 0.94$) (Figure 8c).

3.2.2 | Catalytic reduction of azo dyes

Methyl orange degradation

Methyl orange, having the molecular formula $C_{14}H_{14}N_3NaO_3S$ (Figure 9a), is an anionic azo dye commonly used as a pH indicator in laboratories as well as widely used in various industries, including textile, pulp and paper, and printing, among others. In azo dyes such as MO, the appearance of color is due to the existence of the main azo (–N=N–) chromophore group. However, MO is classified as an acutely toxic compound and has been reported to be mutagenic. Excessive exposure to MO may even cause skin and eye irritation, dermal corrosion, respiratory and gastrointestinal irritation, as well as nausea, vomiting, and diarrhea. Hence, the removal of MO from water sources has received major interest.

Degradation of MO using $NaBH_4$ is an effective and thermodynamically favorable approach to achieve its degradation, but it is rendered kinetically unfavorable due to the electrostatic repulsion between BH_4^- moiety and anions of the dye molecules. Hence, $NaBH_4$ is able to only partially reduce the azo group but is unable to break the N–N bonding. The addition of ZnCuNi-LDH aids the degradation by adsorbing the active species on its surface and facilitating electron transfer between the adsorbed species. Thus, LDH can effectively act as an excellent catalytic material for the $NaBH_4$-assisted degradation of azo dyes.

The UV–Vis absorbance versus wavelength plot for the catalytic reduction of MO is presented in Figure 9b. Absorbance change was noted at $\lambda_{max} = 463$ nm for MO. No change in the peak intensity and position was observed on the addition of $NaBH_4$ alone. Then, on the addition of ZnCuNi-LDH catalyst, the intensity of the peak present at 463 nm began to decrease rapidly, whereas a new peak emerged at 250 nm, along with a visible decrease in the intensity of the color of the reaction mixture. As the reaction progressed, the peak at 463 nm kept decreasing rapidly until, eventually, it disappeared completely, and the intensity of the peak at 250 nm kept increasing. The solution was completely decolorized within the short time span of 4 min. The percentage degradation of dye was calculated using Equation (1) and was calculated to be 96.4%.

The rate constant k_{app}, obtained by linearly regressing experimental data to the pseudo-first-order model, was evaluated to be 0.797 min^{-1} ($R^2 = 0.91$) (Figure 9c). The action of $NaBH_4$ and LDH catalysts causes the cleavage of the chromophoric azo (–N=N–) group of MO, leading

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

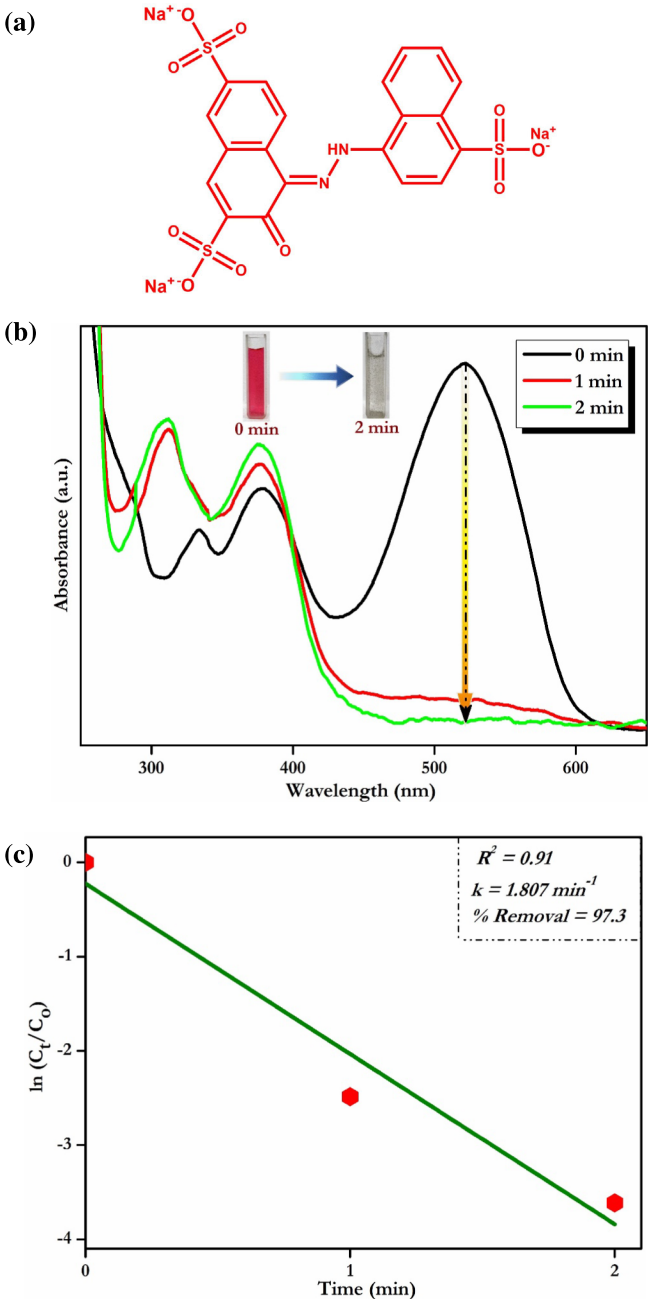


FIGURE 10 (a) Molecular structure of amaranth, (b) UV–Vis absorption spectrum, and (c) linear fitting of pseudo-first kinetics model for the $NaBH_4$-assisted catalytic reduction of amaranth using ZnCuNi-LDH.

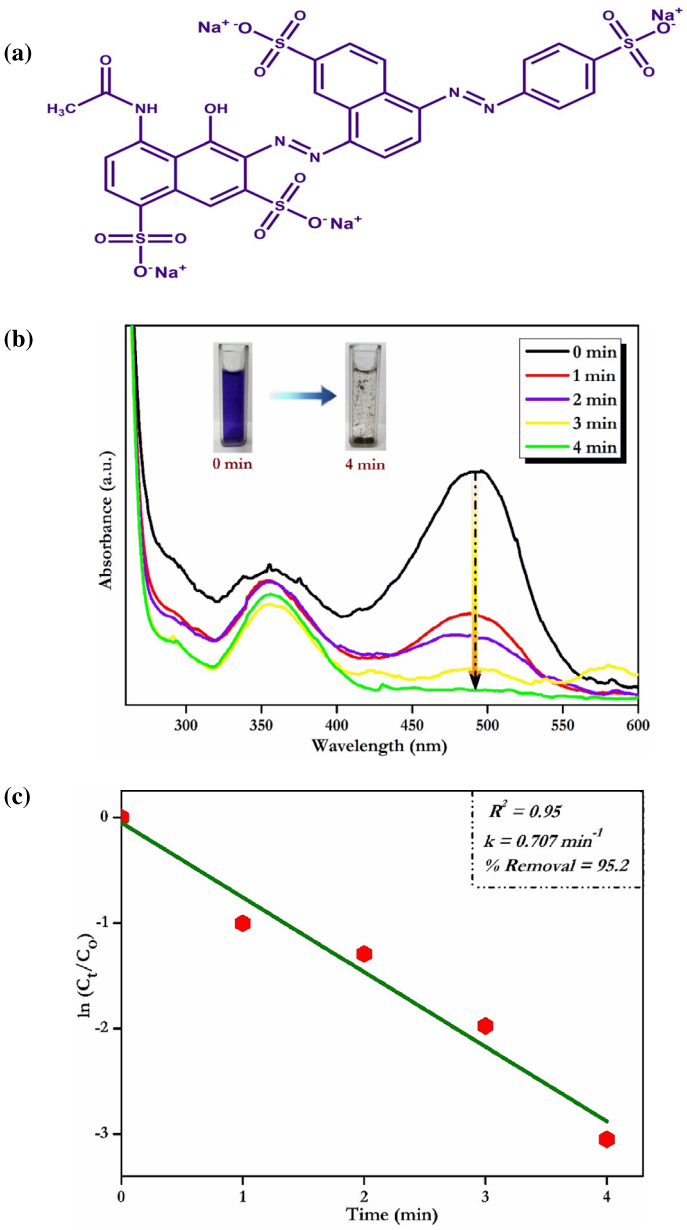


FIGURE 11 (a) Molecular structure of brilliant black BN, (b) UV–Vis absorption spectrum, and (c) linear fitting of pseudo-first kinetics model for the $NaBH_4$-assisted catalytic reduction of brilliant black BN using ZnCuNi-LDH.

to the formation of new degradation products possibly including sulfanilic acid and N,N-dimethyl-p-phenylenediamine, that exhibit absorbance peaks around 248, and 242 nm, respectively.[24,25] The emergence of a new peak at 250 nm in the recorded absorbance spectra may be attributed to these species and along with the disappearance of peak at 463 nm gives evidence for the rapid degradation of MO by LDH and $NaBH_4$, rather than adsorption of LDH.

Amaranth

Amaranth, or Acid Red 27, is an anionic mono azo dye having molecular formula $C_{20}H_{11}N_2Na_3O_{10}S_3$ (Figure 10a). This reddish-brown dye is widely used as a colorant in the food and cosmetics industry. However, its use has now been prohibited in several countries because of suspected carcinogenicity. In vivo combination of amaranth with benzoate and salicylate can cause several severe health issues, such as impaired liver functioning. Hence, the elimination of amaranth from water sources needs to be seriously addressed.

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

To study the catalytic reduction of Ama, absorbance change was noted at $\lambda_{max} = 520$ nm. The absorbance plot of amaranth dye black solution consisted of a singular peak at 520 nm (Figure 10b). This peak does not undergo any change on the addition of $NaBH_4$, rather the emergence of a new peak is observed at 377 nm. This new peak may be attributed to the reduction of azo (–N=N–) group to –NH–NH– by $NaBH_4$. However, the appearance of peaks at both 520 and 377 nm indicates the partial reduction and no cleavage of the azo group. Further, on the addition of ZnCuNi-LDH, rapid discoloration of the solution is observed along with a rapid reduction in peak intensity at 520 nm and the appearance of a new peak at 328 nm that could possibly be attributed to the formation of naphthionic acid as a degradation product.[26] The reaction mixture was completely decolorized within 2 min, suggesting an ultrafast reduction of amaranth dye by ZnCuNi-LDH catalyst. The absorbance of the solution was recorded at 520 nm at 1-min intervals, and the percentage degradation was calculated to be 97.3%. The rate constant (k_{app}), calculated by performing linear regression analysis on the experimental data, was found to be 1.807 min^{-1} ($R^2 = 0.91$) (Figure 10c).

Brilliant black

Brilliant black BN, having molecular formula $C_{28}H_{17}N_6Na_4O_{14}S_4$ (Figure 11a), is a black-colored diazo dye that finds prevalent use in the food industry. This dye is suspected to be genotoxic and mutagenic and has been linked to worsening of asthma and DNA damage. Although $NaBH_4$-assisted catalytic reduction of several toxic azo dyes has been widely reported, virtually, no reports are available about the catalytic reduction of BB.

The λ_{max} for BB dye typically appears at 571 nm (Figure 11b) but the addition of $NaBH_4$ to BB solution causes the peak to shift from 571 to 490 nm, and this shift may be attributed to the partial cleavage of –N=N– azo bond and its reduction to –NH–NH– bond.[27] Further, on the addition of ZnCuNi-LDH, the peak at 490 nm becomes less intense, along with the appearance of a new peak at 251 nm. Complete decolorization of BB was attained within 4 min of reaction initiation. The percentage degradation efficiency of BB was calculated to be 95.2%, and the linear regression of experimental data to the pseudo-first-order model yielded a k_{app} value of 0.707 min^{-1} ($R^2 = 0.95$) (Figure 11c).

3.2.3 | Effect of influencing parameters

The performance of a catalyst is influenced by various parameters. Therefore, the effect of catalyst dosage and pollutant concentration on the catalytic activity of ZnCuNi-LDH was studied by choosing *p*-NP and MO as model pollutants.

Effect of catalyst dosage

The impact of catalyst dosage on the reduction process was investigated using three different amounts, that is, 2.5, 5, and 10 mg. The plots presented in Figure 12 exhibit the influence of ZnCuNi-LDH dosage on the reduction of *p*-NP and MO. Figure 12a–c represents the catalytic reduction of *p*-NP with 2.5-, 5-, and 10-mg LDH, whereas the reduction of MO dye by 2.5, 5, and 10 mg of LDH catalyst is presented in Figure 12d–f.

The absorption spectrum of *p*-NP presents a prominent peak at 400 nm that shows a pronounced decrease in intensity on the addition of catalyst to 1 mM *p*-NP solution. The addition of 2.5 mg catalyst led to the complete hydrogenation of *p*-NP within 6 min. An increase in dosage of ZnCuNi-LDH to 5 and 10 mg leads to a decrease in time required for the complete *p*-NP hydrogenation to 4 and 2 min, respectively. In the case of MO dye, the absorption spectrum presents an intense peak at 463 nm that decreased with the addition of catalyst to 100 mg L^{-1} MO solution. The dye solution was completely decolorized within just 7 min on the addition of 2.5 mg catalyst. An increase in ZnCuNi-LDH dosage to 5 and 10 mg promotes faster reduction, and the dye was completely decolorized within 4 and 2 min, respectively.

The observed trend may be attributed to the fact that at lower catalyst dosage, fewer active sites are available to catalyze the reduction reaction, leading to higher reaction time. An increase in dosage of ZnCuNi-LDH leads to the availability of more number of active sites for reaction, thereby leading to acceleration of the reduction process and faster culmination of reaction.[28]

Effect of pollutant concentration

The impact of pollutant concentration was studied using three concentrations for *p*-NP and MO. The plots presented in Figure 13 provide insights into the catalytic reduction process.

In the case of *p*-NP, the absorption spectrum presented in Figure 13a–c exhibits a peak at 400 nm that shows a pronounced decrease on the addition of 5 mg catalyst. Complete decolorization of 0.25 mM *p*-NP was observed within just 1 min of catalyst addition. On increasing *p*-NP concentration to 0.5 and 1.00 mM, reduction time increased to 2 and 4 min, respectively. A similar trend was observed for the reduction of MO dye, as presented in Figure 13d–f. An addition of 5 mg catalyst to 25 mg L^{-1} MO solution leads to complete reduction within 2 min. An increase in dye concentration to 50 and 100 mg L^{-1} leads to increase in reduction time to 3 and 4 min, respectively. It may be attributed to the fact that at

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

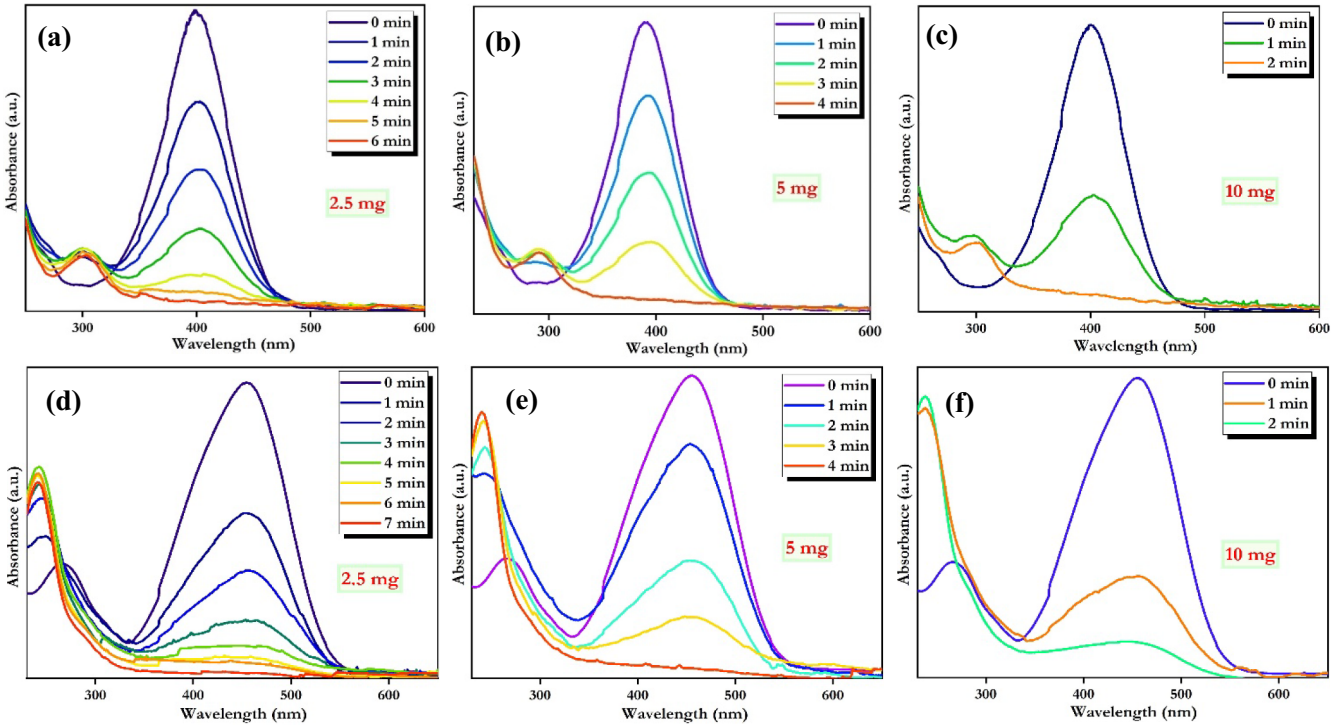


FIGURE 12 UV–Vis absorption spectra for the catalytic reduction of (a–c) *p*-NP and (d–f) MO by varying dosage of ZnCuNi-LDH.

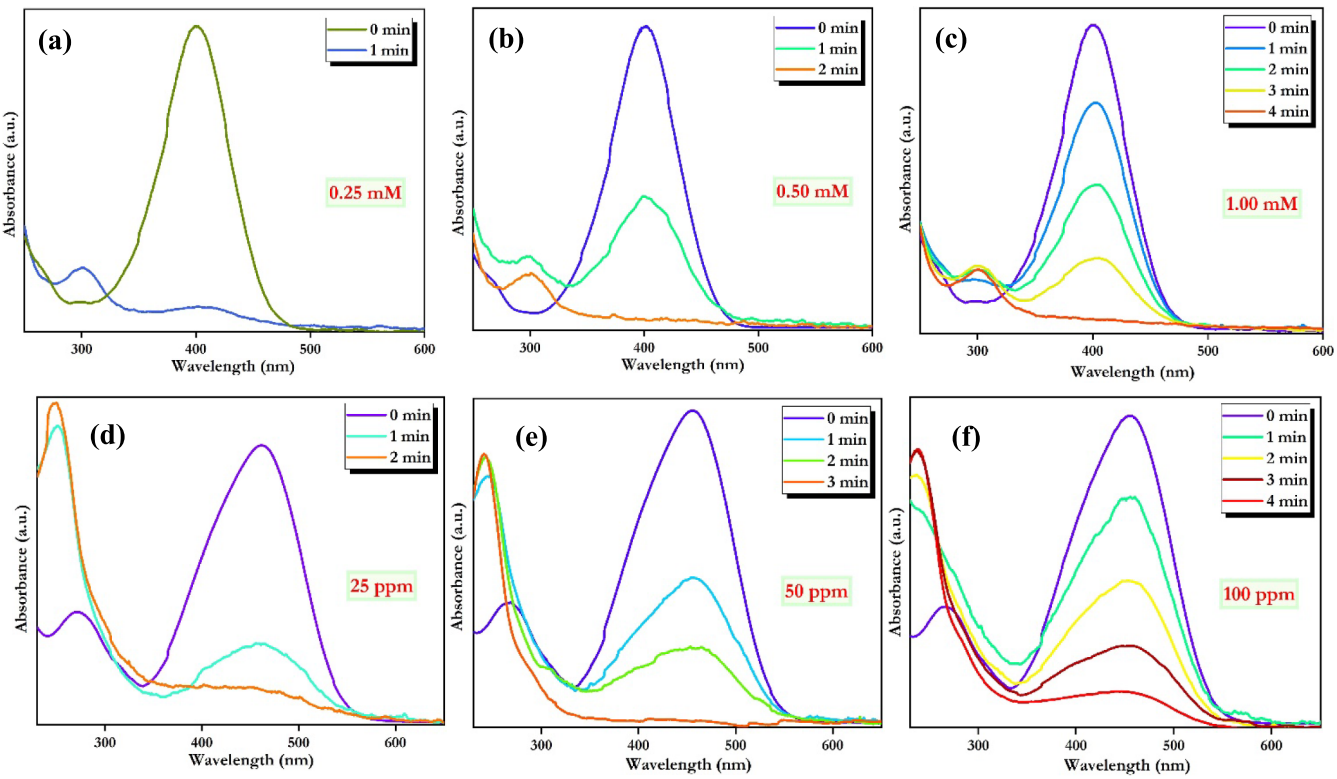


FIGURE 13 UV–Vis absorption spectra for the catalytic reduction of various concentrations of (a–c) *p*-NP and (d–f) MO by ZnCuNi-LDH.

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

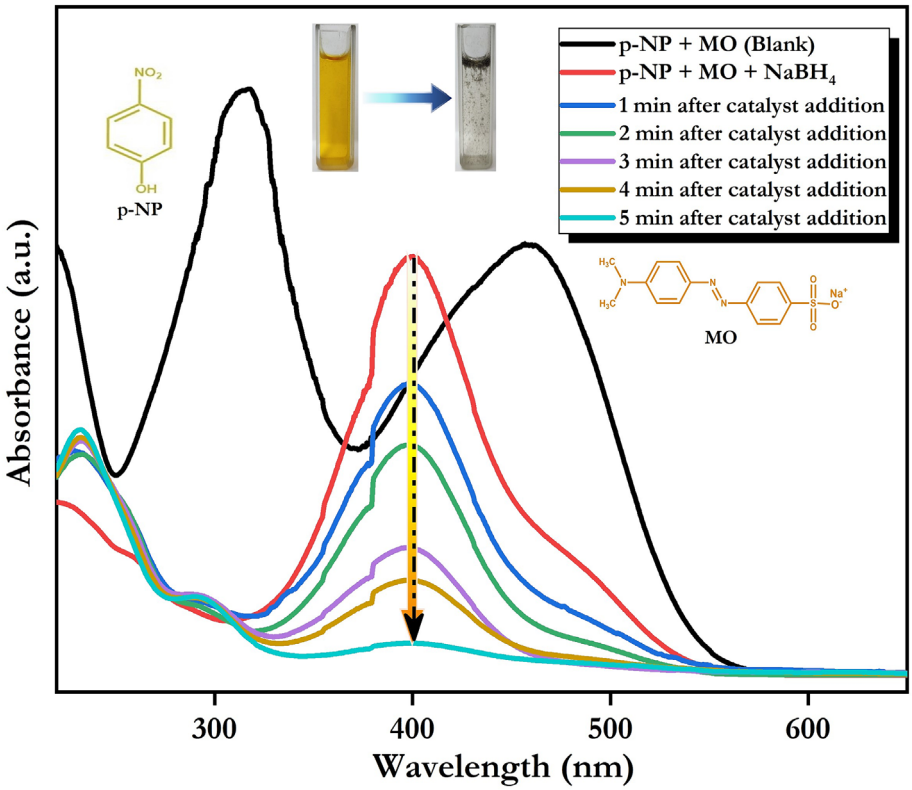


FIGURE 14 UV–Vis absorption spectrum for $NaBH_4$-assisted simultaneous reduction and degradation of *p*-NP and MO by ZnCuNi-LDH.

lower concentration, fewer number of pollutant molecules need to be catalyzed, thereby leading to faster reduction of all pollutant molecules in less time. Conversely, at higher concentrations, more number of molecules are required to be reduced, and thus, more time is expended for the complete reduction of the pollutants.[29]

3.2.4 | Universality of ZnCuNi-LDH as catalyst

The universality of a catalyst is a fundamental parameter in determining the practical and profitable usability of the material in large-scale industrial applications. An efficient catalyst must possess the ability to simultaneously reduce/degrade various types of molecules into simple and benign by-products. Therefore, in order to ascertain the catalytic prowess of synthesized ZnCuNi-LDH toward the simultaneous removal of *p*-NP and MO, 5 mL of *p*-NP solution (1 mM) was combined with 5 mL MO solution (100 mg L^{-1}), following which 10 mL freshly prepared $NaBH_4$ solution and 10 mL distilled water were added to the mixture. After the addition of ZnCuNi-LDH, the reaction mixture was stirred and the absorbance was recorded at pre-fixed time intervals.

It was observed that the absorbance spectrum of the blank mixture of *p*-NP and MO exhibits a peak at 318 nm that can be attributed to *p*-NP molecules, whereas the peak at 464 nm appears because of MO molecules. On the addition of $NaBH_4$, the peak at 318 nm disappeared and a new peak appeared at 400 nm due to the conversion of *p*-NP molecules to phenolate anions (Figure 14). The peak at 464 nm does not shift in position but only decreases in intensity and appears as a neck in the hybrid spectrum because of the closeness of the λ_{max} values of *p*-NP and MO.

Further, on the addition of ZnCuNi-LDH, the intensity of the hybrid spectrum decreased rapidly and eventually disappeared within a short span of 5 min, along with the decolorization of the reaction mixture. The rapid decolorization of the reaction mixture elucidates the fact that ZnCuNi-LDH can be successfully used as a universal catalytic material for the simultaneous reduction of NACs and anionic azo dyes.

3.3 | Plausible mechanism for catalytic activity of ZnCuNi-LDH

The surface catalytic reduction of nitro compounds and organic dyes by ZnCuNi-LDH potentially proceeds through the Langmuir-Hinshelwood mechanism, involving three steps: (1) dissociative adsorption of reactants onto LDH surface, (2) interaction between adsorbed

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

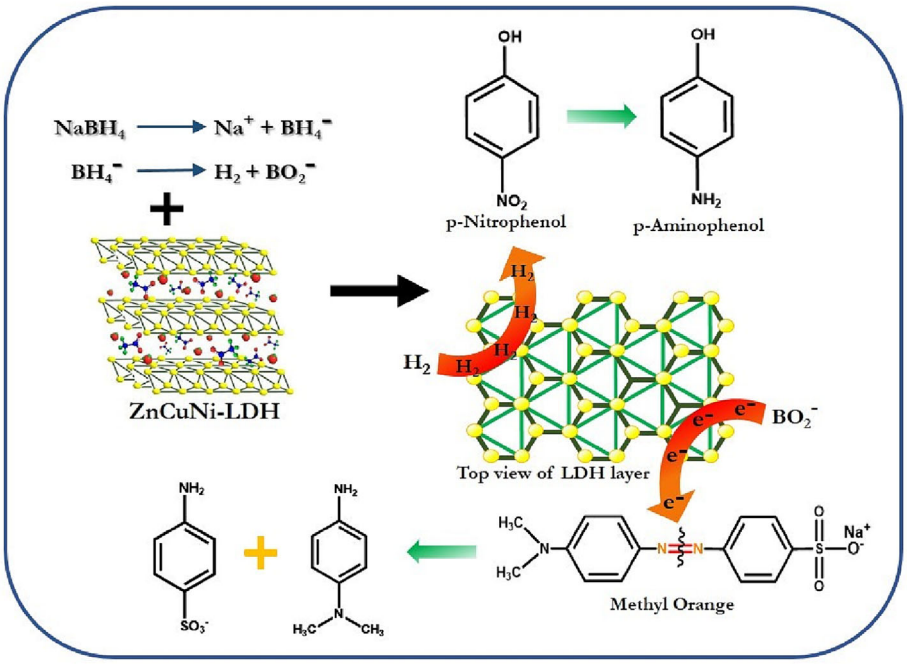


FIGURE 15 Potential mechanism for the $NaBH_4$-assisted catalytic degradation of *p*-NP and MO by ZnCuNi-LDH.

moieties, and (3) desorption of reduction products from LDH surface. In aqueous medium, $NaBH_4$ breaks down into Na^+ and BH_4^- ions. BH_4^- anions further hydrolyze into BO_2^- and H_2, and these moieties are adsorbed onto LDH since they possess a high surface area (Figure 15).

For the degradation of *p*-NP and *p*-NA, H_2 is adsorbed on ZnCuNi-LDH surface, resulting in metal-hydride complex formation. The action of this complex as a hydrogen-mediator facilitates H_2 transfer to the $-NO_2$ group along with electrons. This aids in the hydrogenation of $-NO_2$ group, converting it to $-NH_2$ and resulting in the formation of p-aminophenol and p-phenylenediamine as reduction products following the hydrogenation of *p*-NP and *p*-NA respectively.

In the case of organic dyes, ZnCuNi-LDH acts as a redox catalyst and facilitates the transfer of electrons to the dye molecules. In the reaction mixture, initially, the anionic dye molecules get adsorbed onto the positively charged LDH surface via electrostatic forces of attraction. Then, because of the electron relay behavior of LDH, electrons are transferred from the electron-donating BO_2^- moiety to the electron-accepting dye molecules, leading to their reduction. Lastly, the reduction products present on LDH surface get desorbed and diffused into the solution. The synergistic interactions between the copper and nickel atoms present in the lattice may also contribute and augment the electron relay ability of the ZnCuNi-LDH, and result in the ultrafast reduction of NACs and organic dyes.

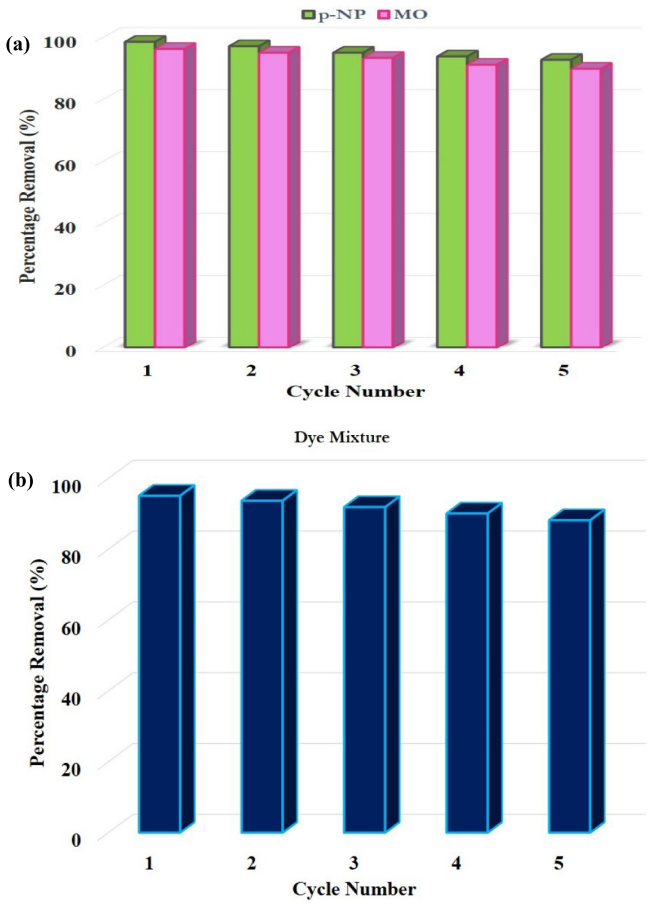


FIGURE 16 Recyclability of ZnCuNi-LDH for catalytic reduction of (a) *p*-NP and MO and (b) mixture of dyes.

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

Pollutant	Catalyst	Rate constant (k_{app})	Reference
p-NP	Au NPs-LMW chitosan	0.2537 min^{-1}	[30]
	Magnetite-Ag/LDH/Starch	0.30 s^{-1}	[31]
	Fe3O4@PPy-MAA/Ag	0.08 min^{-1}	[32]
	ZnCuNi-LDH	1.058 min^{-1}	This work
p-NA	CoAl-LDH	0.40 min^{-1}	[12]
	Ag-PNiM microgel	0.0852 min^{-1}	[33]
	Au Nanorods	0.0564 s^{-1}	[34]
	ZnCuNi-LDH	0.98 min^{-1}	This work
MO	Ag-Cu Rhizome Powder	0.0029 s^{-1}	[8]
	Asp-Au NPs	0.314 min^{-1}	[9]
	Au NPs-LMW chitosan	0.2701 min^{-1}	[30]
	ZnCuNi-LDH	0.797 min^{-1}	This work
Ama	Ag-Cu Rhizome Powder	0.0009 s^{-1}	[8]
	Asp-Au NPs	0.228 min^{-1}	[9]
	ZnCuNi-LDH	1.807 min^{-1}	This work

TABLE 1 Comparison of rate constants of ZnCuNi-LDH with previously reported catalysts.

3.4 | Reusability and recyclability

The recyclability of heterogeneous catalysts is a significant factor that impacts their profitable commercial use. Therefore, the reusability of ZnCuNi-LDH for the reduction of *p*-NP, MO, and mixture of dyes was investigated for 5 cycles. After each reduction reaction, catalyst was recovered and washed with distilled water, dried at 60°C, and then used for subsequent catalytic cycles.

The results presented in Figure 16a indicate that ZnCuNi-LDH sustains up to 90% removal efficacy in just 4 min for the reduction of *p*-NP and MO even after 5 cycles. Moreover, ZnCuNi-LDH also exhibited excellent catalytic reduction ability of up to 88% within 5 min toward a mixture of MO, Ama, and BB dyes even after 5 cycles, as shown in Figure 16b. These results suggest that ZnCuNi-LDH is a robust catalyst that can be easily recovered and utilized for consecutive catalytic cycles without any significant decrease in catalytic performance. The slight decrease in removal efficacy may be attributed to the loss and agglomeration of catalytic material during the recovery and reuse stages.

3.5 | Comparison with reported catalysts

The comparison of the catalytic activity of ZnCuNi-LDH with previously reported catalysts toward the reduction of *p*-NP, *p*-NA, MO, and Ama is presented in Table 1. Virtually no reports were available for the catalytic reduction of BB dye. The rate constants for reactions catalyzed by ZnCuNi-LDH were found to be comparable or even better than the rate constants of most reported catalysts. Therefore, it may be concluded that ZnCuNi-LDH can act as an exemplary catalyst for the catalytic reduction of both nitroarene compounds and azo dyes. The superior catalytic activity may be attributed to the small size, high surface area, and homogeneous distribution of Zn, Cu, and Ni cations in the ZnCuNi-LDH matrix.

4 | CONCLUSION

In summary, ternary ZnCuNi-LDH was successfully synthesized through a facile acid hydrolysis route and was later used as an efficient catalyst for the rapid hydrogenation of *p*-NP and *p*-NA, as well as for the reductive degradation of MO, Ama, and BB dyes. Percentage degradation efficiency for NACs was found to be 98.8% (*p*-NP) and 95.3% (*p*-NA) respectively, whereas organic azo dyes follow the trend Ama (97.0%) > MO (96.4%) > BB dyes (95.2%). The kinetics of the reduction reactions best aligned with the pseudo-first-order model, and rate constant values (k_{app}) for the catalytic hydrogenation of *p*-NP and *p*-NA were found to be 1.058 and 0.98 min^{-1}, respectively, whereas rate constant values (k_{app}) for reductive degradation of MO, Ama, and BB were calculated to be 0.797, 1.807, and 0.707 min^{-1}, respectively. In addition, catalyst dosage and pollutant concentration were found to be important parameters that influence the catalytic performance of the synthesized lattice. It was observed that reduction time increased with increasing catalyst dosage and decreasing concentrations of the model pollutants. Furthermore, ZnCuNi-LDH rapidly decolorized a

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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

binary solution consisting of *p*-NP and MO, thereby exhibiting its universality as a catalyst. The synthesized catalytic material also exhibited more than 90% reduction efficiency even after five regeneration cycles. Thus, the facile synthesis, excellent catalytic activity, and recyclability reiterate the use of ternary ZnCuNi-LDH as effective catalytic materials for the degradation of NACs and azo dyes.

AUTHOR CONTRIBUTIONS

Jigyasa Pathak: Investigation; writing—original draft; methodology; validation; software; formal analysis. **Poonam Singh:** Conceptualization; investigation; methodology; validation; visualization; writing—review and editing; project administration; supervision; resources.

ACKNOWLEDGMENTS

The authors would like to express their immense gratitude to Delhi Technological University for providing the resources and infrastructure to conduct this research work. Author JP would like to thank Delhi Technological University for providing a fellowship.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

Data are available upon request.

ETHICS STATEMENT

Not applicable.

REFERENCES

[1] United Nations Environment Programme, **2016**, 1–252.
[2] X. Qu, J. Brame, Q. Li, P. J. J. Alvarez, *Acc. Chem. Res.* **2013**, *46*, 834.
[3] J. Pathak, P. Singh, *J. Polym. Environ.* **2022**, *31*, 327.
[4] B. Pandey, P. Singh, V. Kumar, *Environ. Nanotechnol. Monit. Manag.* **2021**, *16*, 100596.
[5] B. Pandey, P. Singh, *Appl. Organomet. Chem.* **2023**, *37*, 7072.
[6] Y. Deng, R. Zhao, *Curr. Pollut. Reports* **2015**, *1*, 167.
[7] A. Ucar, M. Findik, I. H. Gubbuk, N. Kocak, H. Bingol, *Mater. Chem. Phys.* **2017**, *196*, 21.
[8] E. M. Bakhsh, M. Ismail, U. Sharafat, K. Akhtar, T. M. Fagieh, E. Y. Danish, S. B. Khan, M. I. Khan, M. A. Khan, A. M. Asiri, *J. Mater. Res. Technol.* **2022**, *18*, 769.
[9] N. Garg, S. Bera, L. Rastogi, A. Ballal, M. V. Balaramakrishna, *Spectrochim. Acta - Part a Mol. Biomol. Spectrosc.* **2020**, *232*, 118126.
[10] I. Ahmad, G. Aalam, M. Amir, A. Chakravarty, S. W. Ali, S. Ikram, *Sci. Total Environ.* **2022**, *846*, 157154.
[11] J. He, M. Wei, B. Li, Y. Kang, D. G. Evans, X. Duan, *Interface Sci. Technol.* **2006**, *119*, 89.
[12] S. R. Leandro, I. J. Marques, R. S. Torres, T. A. Fernandes, P. D. Vaz, C. D. Nunes, *Inorganica Chim. Acta* **2021**, *521*, 120336.
[13] S. Mallakpour, H. Tabebordbar, *Layered Double Hydroxide Polymer Nanocomposites for Catalysis*, Elsevier Ltd **2020**.
[14] J. Pathak, P. Singh, *Polym. Compos.* **2022**, *43*, 7836.
[15] H. C. Greenwell, W. Jones, D. N. Stamires, P. O'Connor, M. F. Brady, *Green Chem.* **2006**, *8*, 1067.
[16] L. Poul, N. Jouini, F. Fievet, *Chem. Mater.* **2000**, *12*, 3123.
[17] S. P. Newman, W. Jones, *J. Solid State Chem.* **1999**, *148*, 26.
[18] A. Kasai, S. Fujihara, *Inorg. Chem.* **2006**, *45*, 415.
[19] T. S. Kazeem, M. Zubair, M. Daud, N. D. Mu'azu, M. A. Al-Harthi, *Korean J. Chem. Eng.* **2019**, *36*, 1057.
[20] R. Valéro, B. Durand, J. L. Guth, T. Chopin, *Microporous Mesoporous Mater.* **1999**, *29*, 311.
[21] G. Carja, R. Nakamura, T. Aida, H. Niiyama, *Microporous Mesoporous Mater.* **2001**, *47*, 275.
[22] M. Thomas, G. A. Naikoo, M. U. D. Sheikh, M. Bano, F. Khan, *New J. Chem.* **2015**, *39*, 9761.
[23] S. Wunder, F. Polzer, Y. Lu, Y. Mei, M. Ballauff, *J. Phys. Chem. C* **2010**, *114*, 8814.
[24] J. Gu, S. Kan, Q. Shen, J. Kan, *Int. J. Electrochem. Sci.* **2014**, *9*, 6858.
[25] Y. Sha, I. Mathew, Q. Cui, M. Clay, F. Gao, X. J. Zhang, Z. Gu, *Chemosphere* **2016**, *144*, 1530.
[26] M. H. Mohamed, L. D. Wilson, J. V. Headley, K. M. Peru, *J. Environ. Sci. Heal. - Part a Toxic/Hazardous Subst. Environ. Eng.* **2008**, *43*, 1700.
[27] Y. Zhang, F. Gao, B. Wanjala, Z. Li, G. Cernigliaro, Z. Gu, *Appl. Catal. B Environ.* **2016**, *199*, 504.
[28] A. Shahzaib, S. Khan, I. Ahmad, S. M. Alshehri, T. Ahamad, N. Nishat, *Chem. Inorg. Mater.* **2024**, *2*, 100037.
[29] H. Dan, Y. Song, Y. Xu, Y. Gao, W. Kong, Y. Huang, Q. Yue, B. Gao, *J. Cleaner Prod.* **2021**, *283*, 124578.
[30] H. I. Salaheldin, *J. Adv. Phys.* **2018**, *14*, 5388.
[31] M. Dinari, F. Dadkhah, *Carbohydr. Polym.* **2020**, *228*, 115392.
[32] R. Das, V. S. Sypu, H. K. Paumo, M. Bhaumik, V. Maharaj, A. Maity, *Appl. Catal. B Environ.* **2019**, *244*, 546.
[33] Z. H. Farooqi, R. Khalid, R. Begum, U. Farooq, Q. Wu, W. Wu, M. Ajmal, A. Irfan, K. Naseem, *Environ. Technol. (United Kingdom)* **2019**, *40*, 2027.
[34] S. G. Jiji, K. G. Gopchandran, *Colloids Interface Sci. Commun.* **2019**, *29*, 9.

How to cite this article: J. Pathak, P. Singh, *Appl Organomet Chem* **2024**, e7507. https://doi.org/10.1002/aoc.7507

10990739, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aoc.7507 by Delhi Technological University, Wiley Online Library on [21/05/2024]. 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



Journal of Economic Policy Reform

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/gpre20

Understanding the internationalization of Indian state-owned enterprises

Aakanksha Kaushik

To cite this article: Aakanksha Kaushik (16 May 2024): Understanding the internationalization of Indian state-owned enterprises, Journal of Economic Policy Reform, DOI: 10.1080/17487870.2024.2353301

To link to this article: https://doi.org/10.1080/17487870.2024.2353301

Published online: 16 May 2024.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=gpre20

JOURNAL OF ECONOMIC POLICY REFORM
https://doi.org/10.1080/17487870.2024.2353301

Routledge
Taylor & Francis Group

Check for updates

Understanding the internationalization of Indian state-owned enterprises

Aakanksha Kaushik

University School of Management and Entrepreneurship, Delhi Technological University, Delhi, India

ABSTRACT

The article examines the global activities of high-value Indian state-owned enterprises (SOEs) classified by the government as "Maharatnas" and "Navratnas." Using the FESPET framework, the paper assesses their performance and international presence, revealing their commitment to a public mission beyond financial gains. These SOEs prioritise social spending and often surpass regulatory environmental targets, serving as valuable assets in India and globally. Profitable SOEs demonstrate positive impacts abroad, while financially unsustainable ones fall short in contributing to the non-financial "ESPET" dimension.

ARTICLE HISTORY
Received 3 November 2022
Accepted 6 May 2024

KEYWORDS
State-owned enterprises; internationalization; emerging markets; India; public enterprises; State-owned multinational enterprises

JEL CLASSFICATION
L21; L25; L32; L33; L38

Introduction

State-Owned Multinational Enterprises (SOMNEs) from developing nations have expanded globally, playing pivotal roles in their economies. International organisations now recognise their significance for both developing countries and OECD countries (IMF 2020; OECD 2020; UNCTAD 2008, 2010, 2022; World Bank 2021). UNCTAD's World Investment Report, tracking the top 100 global non-financial enterprises, consistently shows a rising share of SOMNEs – from 10% in 2013 (Kowalski et al. 2013) to 15% in 2017 (Kalotay 2018), doubling in the last decade (IMF 2020).

The (UNCTAD 2022) report shows that 43 of the top 100 non-financial multinational enterprises from developing and transition economies are state-owned. Of these 43 State-owned Enterprises (SOEs), 32 are based in China, six are in Singapore, and one enterprise each from Malaysia, Saudi Arabia, Qatar, Kuwait, and India made the list.

The limited presence of Indian SOMNEs on the list, compared to China, stems from divergent policies. India restricts SOE from overseas investments, with only energy and natural resource firms allowed to invest abroad through the automatic route.[1] In contrast, China strategically employs its SOEs, enhancing international political and economic influence (Carlos Zalaf Caseiro and Masiero 2014).

India's mixed economy model has endured economic downturns better than many other emerging economies. However, Western pushes for more liberalization and transparency risk hindering growth and social progress in emerging economies. The state-capitalist approach challenges global neoliberalism, reshaping institutional arrangements

CONTACT Aakanksha Kaushik aakanksha.kaushik@dtu.ac.in
© 2024 Informa UK Limited, trading as Taylor & Francis Group

(Nölke et al. 2022; Petry, Koddenbrock, and Nölke 2023). Indian SOEs have significantly contributed to India's development, especially in resource-based industries and manufacturing. However, they encounter challenges stemming from political interference and demonstrate comparatively lower performance in services compared to their private counterparts (Auty 1988; Bala 2006; Chatterjee 2017; Gupta and Kumar 2020; Khanna 2012; Kumar 2018)

Indian SOEs are classified as "Ratna" (meaning gem or jewel) by the Government of India based on financial, operational performance, and global presence. In descending order, the categories include Maharatna, Navratna, Mini Ratna I, and Mini Ratna II. Only Maharatna and Navratna possess the authority to invest up to 15% of their net worth abroad; others require government approval. The degree of corporatization varies between Maharatnas and Navratnas. Maharatna Companies possess increased autonomy in decision-making, particularly in areas such as investments and financial matters. The government employs Memoranda of Understanding (MoUs) as a tool to maintain control over SOEs (Clifton and Díaz-Fuentes 2019; Kumar 2018). India's approach to categorizing state-owned enterprises (SOEs) shares similarities with Norway's system. Both nations classify their SOEs based on their strategic significance and ensure that their operations align with governmental goals. However, the criteria and categories used for classification may vary between the two countries (Christensen 2023; Rygh and Benito 2022)

The Energy sector, with a focus on resource security, boasts the highest number of Maharatnas. Approximately 100 Indian SOEs have undergone internationalization, primarily through exports. While most foreign investments originate from Maharatnas and Navratnas, some Mini Ratnas also establish an international footprint. In its first section, this paper will delve into the internationalization of Indian SOEs.

The evolution of SOEs has been ongoing in India. Notably, government-owned oil and gas companies have historically prioritized fulfilling the nation's energy needs over profit generation. These entities expanded their activities from a limited scope in upstream and downstream sectors to a significant playing field. They discovered, explored, and acquired oil and gas resources in India and internationally. With heightened awareness of Global Warming and a comprehensive understanding of carbon emissions' impact, enterprises are now actively innovating and transitioning towards environmentally friendly products. Significant investments in renewable energy sources are discussed in detail in Section II. The contribution of these organizations to economic development extends beyond conventional assessments based solely on operational and financial efficiency parameters.

Moreover, these SOEs are complex entities with a rich heritage. They are constantly trying to strike a balance between the market-oriented vision and the public mission under which they were incorporated. Many SOEs in India are hybrid entities that are role models of social-value creation. It is challenging to compare private and state-owned enterprises because of the complex nature of the public mission that SOEs operate under, as noted by Sorrentino (2020).

SOEs have been assigned various roles, including Macroeconomic Stability, National Interest, Redistribution of Wealth, Model Employer, and Reduction of Externalities, as identified by Heath and Norman (2004). India Ratings and Research, a rating agency, reported that SOEs in India contributed 19.2% of gross value added (GVA) over a ten-

year period from FY2011–12 to FY20–21, with wages accounting for 39.2% (Sinha 2022). Additionally, a CRISIL study revealed that in 2020, SOEs accounted for 8% of companies engaged in Corporate Social Responsibility (CSR) expenditure, while they spent 28% of total CSR expenditure (CRISIL Foundation 2020).

For this study, the definition of SOMNE is taken from the generally accepted definition by Cuervo-Cazurra et al. (2014) in the limited literature on the subject. The study defined SOMNE as:

> . . . a legally independent firm with direct ownership by the state that has value-adding activities outside its home country. These value-added activities can be downstream activities such as production facilities or sales subsidiaries, as tend to be assumed when one thinks about a multinational company or upstream activities such as purchasing subsidiaries or design or R&D centres.

To thoroughly compare SOEs and private enterprises, Clifton and Díaz-Fuentes (2023) developed a multi-dimensional FESPET consisting of five criteria: financial, environmental, social, political, and technological. The second section evaluates the performance of SOMNEs using this framework and examines how Indian SOMNEs behave abroad.

The trend in India's Outward Foreign Direct Investment has experienced a notable evolution over the years, reflecting strategic shifts in response to changing economic landscapes and global market dynamics. A significant aspect of this evolving trend is the changing sectoral composition of Indian OFDI flows. The studies highlight a broadening diversification across sectors, with a substantial increase in the primary sector, particularly in crude petroleum and natural gas extraction. Meanwhile, the services sector has emerged as the leading contributor, surpassing manufacturing, which dominated OFDI flows in the 1990s and 2000s. There has been a shift in manufacturing from low-technology products to high-technology industries, indicating a focus on technology-intensive and knowledge-driven sectors (Joseph 2019; Pradhan 2017; Pradhan and Sauvant 2010).

The growth in OFDI, particularly in knowledge-based industries, reflects Indian firms' improved competitive capabilities. Acquisitions are becoming a preferred route for technological advancement in the liberal environment, with a focus on industries such as software, pharmaceuticals, and automotive. Indian Multinational Enterprises contribute positively to local R&D and employment in host countries, potentially shaping more competitive market structures (Pradhan and Sauvant 2010; Tiwari and Herstatt 2010)

This study is divided into two main sections. The first section delves into the internationalization of Indian SOEs. The second section analyzes the performance of the SOMNEs using the FESPET framework. The conclusion is detailed in section III.

Section I: understanding internationalisation of Indian state-owned enterprises

India's internationalization process must be understood in the context of a broad policy framework. Since 2014 in India have initiated large-scale privatization programmes through Strategic Sales[2] and Asset monetization,[3] unlike just the selling of minority shareholdings in the previous years. The policy "New Public Sector Enterprise Policy in 2021" aims for the bare minimum presence of state-owned

enterprises in the strategic sector. Moreover, the new policy has also revised and fast-tracked the closure of state-owned enterprises in the non-strategic sector. The government closed Board for Reconstruction of Public Sector Enterprises, the organization set up for the revival of loss-incurring SOEs, in November 2015. The environment is not conducive to the growth of state-owned enterprises in the country.

The profit-making enterprises are being sold off, and loss-incurring ones are being shut down in a "time-bound" manner. This impacts the performance of management and millions of employees employed in hundreds of SOEs in the country. Despite this, some enterprises remained profitable and have seen an increase in internationalization.

The number of SOEs that had earnings in foreign exchange in India is around 120 from 2011 to 2022. But for this analysis, enterprises from Hotel and Tourist Services, Aviation, Railways, Finance, and Entities for Trade promotion were excluded from the sample. These firms earned foreign earnings by operating within the country. The entities approved for closure by the government, namely Handicrafts & Handloom Exports Corporation of India Limited, Hindustan Fluorocarbons Limited, Instrumentation Limited, National Textile Corporation Limited, and Scooters India Limited, were also excluded. This has reduced the sample firms to 82 firms.

The internationalization of SOEs is measured for select state entities (Table 1) by dividing foreign revenues earned by SOEs, such as Exports of goods and services, royalties, interest, dividends and others, by the firm's total revenues. The data from 2011–2012 to 2020–21 is collected from the Public Enterprise (PE) Survey Report published annually by the Government of India and the companies' Annual Reports.

Only four years, 2011–12, 2016–17, 2019–2020, and 2020–21, are presented in Table 1 so that the broad trend in internationalization can be observed. The average internationalization of select Indian SOEs has only marginally increased by 1.5% over ten years from 2011–12 to 2021–21, with one tail test significant at 10%. The impact of the pandemic and global slowdown can be observed from 2019–20 and 2020–21 internationalization estimates, which have declined for thirty-four firms within a year. The average internationalization fell by −1.30% and was statistically significant under one tail test at a 10% significance level.

Table 2 presents the sectoral composition of the sample firms. The SOEs are internationalised across various sectors. The most significant number of SOEs are present in Heavy & medium engineering (seventeen), followed by Contract, Construction & Technology Consultancy services (fourteen), and Trading and Marketing (eleven). Most enterprises are profit-making. The highest loss-making entities belonged to the Trading and Marketing sector (three): Metals and Minerals Trading Corporation of India Limited, Project and Equipment Corporation of India Limited, and State Trading Corporation of India Limited. The government has decided to denotify the organizations from Exim transactions are considering shutting them down; therefore, they are also excluded from discussion in section II.

Fifteen of seventeen companies under the Heavy and Medium Engineering group were profit-making. All these firms have played an essential role in the country's economic development. The two loss-making SOEs are Hindustan Shipyard Limited and Rajasthan Electronics and Instruments Limited. Four companies, Bharat Earth Movers Limited, Bharat Electronics Limited, Bharat Heavy Electricals Limited, and

Table 1. Trend of Internationalisation of State-owned enterprises of India.

S.no.	Company Name	Internationalisation*			
		2011–12	2016–17	2019–20	2020–21
1	Andrew Yule & Company Limited	0.37	0.43	1.19	1.31
2	Antrix Corporation Limited	2.98	9.45	0.48	1.54
3	Artificial Limbs Manufacturing Corporation of India	0.04	0.00	0.00	0.00
4	Bharat Earth Movers Limited	3.98	0.59	13.68	6.75
5	Balmer Lawrie & Company Limited	1.19	4.65	4.41	3.19
6	Bel-Thales Systems Limited	#	34.01	74.98	45.65
7	Bharat Dynamics Limited	0.01	0.01	5.53	7.78
8	Bharat Electronics Limited	0.00	4.85	2.76	2.72
9	Bharat Heavy Electricals Limited	28.69	8.56	16.44	9.07
10	Bharat Petro Resources Limited	32.18	130.34	3.59	8.02
11	Bharat Petroleum Corporation Limited	8.68	4.04	4.63	2.19
12	Bharat Sanchar Nigam Limited	0.41	0.00	0.26	0.37
13	Brahmaputra Crackers & Polymer Limited	#	0.00	0.27	0.47
14	Broadcast Engineering Consultants India Limited	3.60	0.09	0.20	0.01
15	Central Cottage Industries Corporation of India Limited	32.38	36.61	42.90	5.55
16	Central Electronics Limited	–	0.76	0.24	0.31
17	Certification Engineers International Limited	4.45	0.89	0.47	0.06
18	Coal India Limited	0.01	0.00	0.05	0.00
19	COCHIN SHIPYARD LIMITED	47.48	5.05	0.10	2.53
20	Educational Consultants India Limited	6.37	0.32	14.84	2.69
21	Electronics Corporation of India Limited	0.48	0.00	0.00	0.00
22	Engineering projects India Limited	0.00	33.89	53.54	34.20
23	Engineers India Limited	2.98	20.76	11.20	11.01
24	Gas Authority of India Limited	0.00	1.61	0.18	25.39
25	Garden Reach Shipbuilders & Engineers Limited	0.00	0.78	0.00	7.67
26	Goa Shipyard Limited	0.04	38.62	3.67	0.04
27	HIL(India) Limited	0.00	0.91	3.01	6.15
28	Hindustan Aeronautics Limited	2.54	2.55	1.00	1.06
29	Hindustan Antibiotics Limited	2.22	0.08	0.00	0.00
30	Hindustan Copper Limited	4.05	18.86	56.98	44.93
31	Hindustan Petroleum Corporation Limited	4.11	0.56	2.90	1.27
32	Hindustan Shipyard Limited	0.60	0.00	0.07	0.09
33	HLL Infra Tech Services Limited	#	4.35	0.00	0.00
34	HLL Lifecare Limited	8.01	12.15	6.20	1.76
35	HMT (International) Limited	14.35	26.69	22.62	7.82
36	HMT Limited	0.14	2.24	0.88	2.04
37	HMT Machine Tools Limited	0.00	0.14	0.01	0.83
38	Indian Telephone Industries Limited (ITI Limited)	0.00	0.00	0.00	0.00
39	Indian Oil Corporation Limited	5.30	3.26	3.84	2.99
40	IRCON Infrastructure & Services Limited	0.00	11.91	0.70	1.17
41	IRCON International Limited	52.61	8.83	10.39	11.32
42	Indian Rare Earths Limited	29.47	35.28	47.26	47.74
43	Karnataka Antibiotics & Pharmaceuticals Limited	9.46	5.74	5.49	6.59
44	Kudremukh Iron Ore Company Limited	32.48	0.08	81.40	78.00
45	Metals and Minerals Trading Corporation of India Limited	3.12	13.25	4.59	6.85
46	MSTC Limited	0.00	0.00	0.00	0.00
47	Mahanadi Coalfields Limited	0.00	0.00	0.00	0.00
48	Mahanagar Telephone Nigam Limited	0.17	0.12	0.08	0.03
49	Mangalore Refinery & Petrochemicals Limited	40.93	16.88	27.25	11.01
50	Mazagon Dock Shipbuilders Limited	0.00	0.00	0.00	0.00
51	Metallurgical & Engineering Consultants (India) Limited	0.73	0.24	0.73	0.51
52	Mishra Dhatu Nigam Limited	0.00	0.00	1.46	2.38
53	National Aluminium Company Limited	36.92	41.31	40.64	56.51
54	National Fertilizers Limited	2.09	0.00	0.03	0.19
55	National Film Development Corporation Limited	0.86	0.63	0.50	0.57
56	National Seeds Corporation Limited	0.28	0.30	0.00	0.00
57	National Small Industries Corporation Limited	0.08	0.00	0.00	0.00
58	National Buildings Construction Corporation Limited (NBCC)	0.00	0.00	0.00	0.27
59	National Hydroelectric Power Corporation Private Limited	0.00	0.00	0.00	0.00
60	National Thermal Power Corporation	0.00	0.01	0.00	0.00

(*Continued*)

Table 1. (Continued).

S.no.	Company Name	Internationalisation*			
		2011–12	2016–17	2019–20	2020–21
61	NTPC Vidyut Vyapar Nigam Limited	0.00	0.00	46.27	36.60
62	Numaligarh Refinery Limited	0.00	0.00	1.19	0.81
63	Oil & Natural Gas Corporation Limited	0.82	3.94	1.69	0.35
64	Oil India Limited	0.02	0.50	0.00	3.19
65	ONGC Mangalore Petrochemicals Limited	0.00	71.17	85.49	90.57
66	ONGC Videsh Limited	35.94	66.19	77.98	85.23
67	Project and Equipment Corporation of India Limited	9.37	1.37	0.00	–
68	Pawan Hans Limited	38.20	0.00	0.00	0.00
69	Power Grid Corporation of India Limited	0.09	0.09	0.10	0.04
70	Project & Development India Limited	8.04	5.68	1.07	0.17
71	Rajasthan Electronics and Instruments Limited	0.21	0.10	0.06	0.00
72	Rashtriya Chemicals and Fertilizers Limited	0.03	0.00	0.00	0.00
73	Rashtriya Ispat Nigam Limited	3.76	8.25	9.33	22.70
74	Rail India Technical and Economic Service Limited	35.94	29.27	26.29	9.55
75	Shipping Corporation of India Limited	0.00	103.14	101.70	105.07
76	State Trading Corporation of India Limited	1.00	9.49	0.37	0.00
77	Steel Authority of India Limited (SAIL)	2.42	3.52	5.90	8.88
78	Telecommunications Consultants India Limited	29.48	5.59	2.20	5.00
79	The Cotton Corporation of India Limited	1.87	0.00	0.00	0.18
80	The Jute Corporation of India Limited	0.00	0.00	0.00	0.00
81	Uranium Corporation of India Limited	0.00	0.00	0.01	0.00
82	Water and Power Consultancy Services (India) Limited	32.59	19.88	11.09	11.58

The Public Enterprise (PE) Survey Report published annually by the Government of India and the Annual Reports of the respective companies.

Table 2. The sectoral composition of the sample firms that internationalized.

Main Activity	No. of Profitable* Companies	No. of Loss Incurring Companies	Total No. of Companies
Agro Based	1	0	1
Chemicals & pharmaceuticals	4	1	5
Coal	2	0	2
Contract & construction and technology consultancy services	13	1	14
Crude oil	3	1	4
Fertilisers	2	0	2
Heavy & medium engineering	15	2	17
Industrial and consumer goods	3	0	3
Other minerals & metals	5	0	5
Petroleum (refinery & marketing)	6	0	6
Power Generation	2	0	2
Power transmission	1	0	1
Steel	2	1	3
Telecommunication & information technology	1	2	3
Trading & marketing	8	3	11
Transport and logistic services	2	0	2
Transport and vehicles	1	0	1
Grand Total	71	11	82

The Public Enterprise (PE) Survey Report published annually by the Government of India and the Annual Reports of the respective companies.

*Profitability is measured by profit after tax (PAT), and if data was unavailable for the financial year 2021–22, the latest available year value was taken.

Hindustan Aeronautics Limited, which have investments abroad, are discussed in detail in the next section. All the firms under the following main activities – Agro based, Chemicals & Pharmaceuticals, Coal, Fertilizers, Industrial and Consumer goods, other minerals & metals, Petroleum (refinery & marketing), Power Generation, Power Transmission, Transport & Logistic services, and Transport and Vehicles – are profitable.

The top three sectors in terms of internationalization are Transport and Logistics services, Other Minerals and Metals, and Crude Oil. There are only two firms under Transport and Logistics – Shipping Corporation of India Limited and Pawan Hans Limited.[4] The latter has a minuscule percentage of revenues from outside the country. The high rate of foreign exchange earnings of Shipping Corporation of India Limited is due to the nature of its business. It is the largest shipping company in India, conducting business on national and international lines.

The heavy and medium engineering sector has a higher number of international firms. Still, the share of foreign revenue compared to total revenues has remained low, consistently below 10% over the years.

The Other Minerals and Metals group has a substantial percentage of revenue in foreign exchange, which has grown over the last ten years. In 2011–12, the average percentage of internationalisation constituted 20.6%, which increased to 45.4% in 2020–21. The four companies of the group are Hindustan Copper Limited (44.93%), Indian Rare Earths Limited (47.74%), Kudremukh Iron Ore Company Limited (78.00%), National Aluminium Company Limited (56.51%), and Uranium Corporation of India Limited (0.00%). All these companies have internationalised mainly through exports. According to the 2021–22 annual report of Kudremukh Iron Ore Company Limited, the export revenue now constitutes 97.7% of its total revenue. The company deals in Iron ore, does not operate from any international location, and has operations restricted to India.

The companies under the groups Crude oil and Petroleum & Refinery Markets have acquired more foreign assets than any other listed company. The foreign investments are global, including developed and developing countries (Table 3). But the internationalisation is 24.2% because the entities cater primarily to the domestic market. The energy sector SOEs of India operate with a public mission to secure the country's energy needs. This is why the share of foreign revenue from the power generation and transmission groups is negligible in firms' total revenues. However, the power transmission group had projects in collaboration with friendly neighboring countries.

All three sample firms in the steel sector, Mishra Dhatu Nigam Limited, Rashtriya Ispat Nigam Limited, and Steel Authority of India Limited, have consistently increased internationalization. Their primary source of foreign earnings is exports. Steel Authority of India Limited's exports exceed those of Rashtriya Ispat Nigam Limited, but the latter has more revenue from abroad.

The other two sectors showing a consistent rise in internationalization are Chemicals & Pharmaceuticals and Contract, Construction & Technology Consultancy. Internationalization in Chemicals & Pharmaceuticals has increased massively in ten years. It was only 2.9% in 2011–12 and has risen to 20.8% in 2020–21. All the firms are internationalized through exports. Except for Hindustan Antibiotics Limited, all have shown a rise in foreign exchange revenues in proportion to total revenues. But there is significant variation within the group in the extent of

Table 3. Details of international markets served by the SOMNEs.

S.no.	Company	Geographical Area
1	Bharat Petroleum Corporation Limited	Oman, Russia, Australia, UAE, Brazil, Singapore, Netherlands, Indonesia, Mozambique
2	Balmer Lawrie & Company Limited	China, Kenya, Nepal, Sri Lanka, New Zealand, Qatar, Kuwait, Spain, Italy, Germany, France, Ireland, Malaysia, Australia, and Others
3	Bharat Earth Movers Limited	The major markets are Middle East countries, African countries and Southeast Asian countries. (Around 68 countries)
4	Bharat Electronics Limited	USA, France, China, Germany, Singapore, Oman Switzerland, ASEAN, Mauritius, Republic of Armenia, Sri Lanka, Sweden, Israel Kenya, Chile, Suriname, Malaysia, Nepal and Bangladesh
5	Bharat Heavy Electricals Limited	In all the continents (Around 88 Countries)
6	Coal India Limited	Mozambique
7	ENGINEERING PROJECTS (INDIA) Limited	Kuwait, Iraq, Saudi Arabia, Maldives, Thailand, Yugoslavia, Bhutan, UAE, Oman, Sri Lanka
8	Engineers India Limited	Major Market Middle East, Africa, South Asia & Central Asia.
9	Gas Authority of India Limited	Singapore, US, Russia and Mayanmar
10	Hindustan Aeronautics ltd.	USA, Mauritius, Ecuador, Israel, Mauritius, Namibia, Nepal, Russia, UK, Oman, Malaysia, Thailand, Germany, Russia, Suriname, France Vietnam
11	Hindustan Petroleum Corporation Limited	Major Markets are in Asia, the Middle East, Australia, and a few African countries.
12	Indian Oil Corporation Limited	Around 32 countries across the globe
13	IRCON International Limited	128 projects in 25 Countries, majorly from Africa, Asia and the Middle east
14	Mahanagar Telephone Nigam Limited	Mauritius and Nepal
15	Metals and Minerals Trading Corporation of India Limited	Asia, Europe, Africa, Australia, North and South America
16	National Buildings Construction Corporation Limited	Asia, Africa and the Middle east
17	National Thermal Power Corporation Limited	Asia, Africa, South America, Oceania
18	Oil & Natural Gas Corporation Limited	Asia, Africa, Europe, and South America.
19	Oil India Limited	Asia, Africa, Europe, South America, and North America
20	Power Grid Corporation of India Limited	Asia, Asia-Pacific, Europe, and Africa
21	Rail India Technical and Economic Service (RITES) Limited	Asia, Africa, Latin America, South America and the Middle East (over 55 countries)
22	Telecommunications Consultants (India) Limited	Asia, Africa, the Middle East and Europe and North America (Around 65 countries)
23	Water and Power Consultancy Services (WAPCOS) Limited	Asia, Africa, CIS, Eurasia, Europe, North America, Oceania, Pacific Islands and South America. (Around 51 countries)

internationalization. Brahmaputra Crackers & Polymer Limited's internationalization estimate is 0.43%, and on the other hand, the estimate of the step-down subsidiary of Oil & Natural Gas Corporation Limited (ONGC), ONGC Mangalore Petrochemicals Limited is 90.57%. It is a loss-making entity located in the Mangalore Special Economic Zone.

The companies under the Contract, Construction & Technology Consultancy group have offices, subsidiaries, and investments in international locations besides internationalisation through exports. But the extent of internationalisation of the group has been consistently declining over the years. From 11.3% in 2011–12, it has reduced to only 5.9%. Out of fourteen sample firms, internationalisation was facilitated for nine firms in the sector: Broadcast Engineering Consultants India Limited, Certification Engineers International Limited, Educational Consultants India Limited, IRCON International Limited, Metallurgical & Engineering Consultants (India) Limited, National Small

Industries Corporation Limited, Project & Development India Limited, Rail India Technical and Economic Service Limited, and Water and Power Consultancy Services (India) Limited.

For the sectors Fertilisers, Trading & Marketing, and Industrial and Consumer Goods, the magnitude of foreign exchange earnings has remained low and inconsistent. The decline in the Industrial and Consumer goods group is due to HLL Lifecare Limited. The proportion of foreign earnings for the group has reduced to 1.76% from a peak of 12.15% in 2016–17.

Only a few organizations invested abroad. They are not discussed in the literature; less is known about them, and lesser about their operations abroad. The next section of the study will examine the performance of such SOEs using the multidimensional framework.

Section II how Indian SOMNEs behave abroad

The SOMNEs with value-adding activities abroad were selected. The firms that internationalized through exports or royalties were excluded. As stated earlier, only Maharatna and Navaratna firms have the autonomy to invest a certain amount. But the list also includes SOEs with Mini Ratna Status. These companies are not as large (in terms of domestic and foreign assets) as oil and gas companies but are more internationalized and dependent on foreign earnings.

Eleven companies in India have been accorded Maharatna status, and nine of them are part of the sample. Only two companies, Power Finance Corporation and Steel Authority of India Limited, are not on the list. The Power Finance Corporation is a non-banking finance company that serves only Indian markets. Steel Authority of India Limited has internationalised through exports only and has no investments abroad except International Coal Ventures Private Limited, which is already being discussed in the study. The share of foreign revenue to total revenues in Steel Authority of India Limited has increased from 2.42% to 8.88% from 2011–12 to 2020–21. Of the thirteen companies under the list of Navratna companies, only six have investments or have taken projects abroad. Seven companies are Category-I Mini-Ratna companies, and one is from Category-II Mini-Ratna.

The majority of companies are from the Contract & Construction and Technology consultancy services group (six), followed by Heavy and medium Engineering (four), Petroleum (Refinery and marketing), Crude Oil (two), and telecommunication and information technology (two). The other remaining groups had one company each.

The oil and gas SOMNEs play a crucial role in managing the global exploration and production (E&P) blocks, serving as key components in the domestic hydrocarbon value chain for their respective governments. Despite having substantial foreign investments, these companies, particularly exemplified by India's ONGC, exhibit low internationalisation. UNCTAD data from 2020 reveals that ONGC's foreign assets constitute 24% of its total assets, and this ratio aligns with the proportion of foreign employees. However, the share of foreign sales in total sales is merely 3%, marking ONGC as having the lowest percentage among the top 100 non-financial multinational enterprises from developing and transition economies. In contrast, Chinese SOMNEs, notably China National Petroleum Corporation, have a different focus, with foreign investments geared towards

securing energy needs. China National Petroleum Corporation ranks 24th among the top 100 non-financial multinational enterprises, displaying a higher ratio of foreign sales (37%) and foreign assets (21%) compared to ONGC.

Table 3 provides insights into the global footprint of SOMNEs, showcasing their widespread presence in both developed and developing nations. These companies, spanning oil and gas as well as heavy and medium engineering sectors, consistently generate foreign exchange earnings across Asia, Africa, the Middle East, Russia, the USA, and Australia. Europe and America receive relatively limited attention in terms of investment flows.

The study employs a comprehensive multi-dimensional framework by Clifton and Díaz-Fuentes (2023), focusing on Financial, Environmental, Social, Political, and Technological (FESPET) criteria to evaluate the performance of Indian SOMNEs. This framework is modified for this study, and qualitative & quantitative indicators are added to the framework for the performance evaluation of the Indian SOMNEs. Under the Financial criterion, the study assesses the profitability of business activities and investigates whether these enterprises undertake riskier investments to fulfill public missions. The Environmental criterion examines the extent to which government mandates related to environmental goals are achieved both domestically and internationally. Social criteria explore the public mission of SOMNEs and their commitment to social spending, utilizing Corporate Social Responsibility (CSR) expenses in India as a proxy. The final dimension, Technology, evaluates the Research and Development (R&D) intensity of these enterprises, considering R&D expenses relative to total sales as a key variable. The study aims to provide a nuanced understanding of the performance and motives of Indian SOMNEs in the global landscape.

Financial dimension

The overall financial health of SOMNEs is assessed through profitability measure: Profit After Tax of the consolidated results of the enterprises are presented in Table 4. The data is compiled from the Public Enterprise Survey Report of the Department of Public Enterprises, Ministry of Finance, Government of India, and Companies' Annual Reports. The pandemic harmed the financial health of public and private enterprises. But most companies in the sample, like Bharat Heavy Electricals Limited, that had gone red during the pandemic, have returned to profitability from the current financial year. Therefore, firms are marked profitable if they are earning profit in the current financial year 2021–22, and an average of past years is not taken as it will be a biased estimate.

It is also important to note that the prices of petroleum and diesel have been deregulated and are market determined in India. But the rising oil prices have considerable political implications, and the government intervenes whenever necessary to insulate the economy against unprecedented rises in oil prices. The recent increase in oil prices has adversely affected the financials of state-owned oil marketing companies Bharat Petroleum Corporation Limited, Hindustan Petroleum Corporation Limited, and Indian Oil Corporation Limited. The government and the SOEs share the under-recovery. A simple measure of profitability, Profit after Tax, of the consolidated financial

Table 4. Financial performance of Indian SOMNEs in 2021–22.

Company Name	Main Activity	If, Profitable?
Bharat Earth Movers Limited	Transport and vehicles	YES
Balmer Lawrie & Company Limited	Heavy & medium engineering	YES
Bharat Electronics Limited	Heavy & medium engineering	YES*
Bharat Heavy Electricals Limited	Heavy & medium engineering	YES
Bharat Petroleum Corporation Limited	Petroleum (refinery & marketing)	YES
Coal India Limited	Coal	YES
Engineering Projects (India) Limited	Contract & construction and tech. consultancy services	No*
Engineers India Limited	Contract & construction and tech. consultancy services	YES
Gas Authority of India Limited	Petroleum (refinery & marketing)	YES
Hindustan Aeronautics Limited	Heavy & medium engineering	YES
Hindustan Petroleum Corporation Limited	petroleum (refinery & marketing)	YES
Indian oil Corporation limited	Petroleum (refinery & marketing)	YES
IRCON International Limited	Contract & construction and tech. consultancy services	YES
Mahanagar Telephone Nigam Limited	Telecommunication & information technology	NO
National Buildings Construction Corporation Limited	Contract & construction and tech. consultancy services	YES
National Hydroelectric Power Corporation Limited	Power generation	YES
National Thermal Power Corporation Limited	Power generation	YES
NTPC Vidyut Vyapar Nigam limited	Trading & marketing	YES*
Oil & Natural Gas Corporation Limited	Crude oil	YES
Oil India Limited	Crude oil	YES
Telecommunications Consultants India Limited	Telecommunication & information technology	No
Water and Power Consultancy Services (WAPCOS) Limited	Contract & construction and tech. consultancy services	YES*

Own compilation from Public Enterprise Survey Report of Department of Public Enterprise (DPE), Ministry of Finance (MoF), Government of India (GoI), Companies'Annual Reports.
*If the financial year 2021–22 data was unavailable, 2020–21 data was collected.

results of the enterprise is adopted to assess the financial performance of SOMNEs across sectors.

All the companies are profit-making except Mahanagar Telephone Nigam Limited (MTNL), Engineering projects India Limited and Telecommunications Consultants India Limited. The MTNL has exited the loss incurring foreign joint venture UTL in Nepal. The other investment of MTNL, the subsidiary MTML is a profit-making entity. Engineering projects India Limited has three international offices, with more than 50% of revenues from abroad before the pandemic. The Indian credit rating Agency ICRA improved it's rating because of its strong book position in 2020. After that, the company's financials worsened during the pandemic, and losses mounted in the financial year 2020–21 and became the highest in the last ten years.

Large enterprises like ONGC and Bharat Petroleum Corporation Limited have wholly owned subsidiaries that manage overseas operations. ONGC Videsh Limited is a subsidiary of ONGC and is accorded Category I, Mini-Ratna company status. Through it, ONGC owns participating interests in E&P blocks in thirty-five countries. The SOMNE also has several other investments and is a profitable entity. Bharat Petroresources Limited is a wholly owned subsidiary of Bharat Petroleum Corporation Limited, which holds participating interests in E&P blocks in five countries and has

consistently shown losses. The company will remain red until production commences in Brazil and Mozambique.

Risky investments

Investing overseas for resource-seeking is risky because it depends on host countries' energy policies. The host government can change it at any time or ban exports of resources from the country. Indian SOMNEs have invested in risky investments motivated by public missions or in geographical areas where there is unrest continuously. These investments will not be taken by many private enterprises that are governed by principles of profit maximization and market seeking. In the quest to secure energy resources, the Indian SOMNEs have invested in areas like Syria, South Sudan, Sudan, and Mozambique. The unrest is due to several factors, such as unresolved political and territorial conflicts, insurgency attacks, protests, etc.

India has diplomatic ties with the Mozambique government, and Indian SOMNEs have invested billions of US dollars in Mozambique. The country is resource-rich, but the insurgency attacks have jeopardized investments. Indian SOMNEs had to evacuate employees from Mozambique and suspend operations. Moreover, most investments in Mozambique have been financially draining for the Indian SOMNEs, either coal mining or oil and gas E&P activities.

Another example is Bharat Petro Resources JPDA (Joint Petroleum and Development Area) Limited, a wholly-owned subsidiary of Bharat Petroresources Limited, which had 20% PI in Block JPDA 06–103 in Timor Leste. The exploration activity was suspended due to arbitration proceedings between the government of Timor Leste and Australia.

Over the years, SOMNEs have gained experience and do due diligence before investing. Due to operational infeasibility, the oil and gas SOMNEs have relinquished blocks in Australia and Israel. Moreover, the oil and gas SOMNEs have strategically diversified their portfolios in different geographical areas of the World and in blocks in different production phases to mitigate risks.

The ICVL was incorporated to secure coal reserves outside India. The Joint Venture of Steel Authority of India Limited, Rashtriya Ispat Nigam Limited, National Mineral Development Corporation Limited, Coal India Limited, and National Thermal Power Corporation. Two enterprises, Coal India Limited and National Thermal Power Corporation, exited because they considered it as a financial burden. The steel companies in India continued because they needed coking coal for the furnaces, which is critical for smooth production. National Thermal Power Corporation exit was also because the company required thermal coal for operations, not coking coal. The firm could not achieve success and sophistication like oil and gas companies. The functions are restricted to Mozambique, and it has failed to acquire profitable coal reserves overseas.

Environmental dimension

The SOEs in India are not mandated to achieve targets above and beyond regulatory requirements. SOEs and private enterprises must comply with the same environmental regulations. This is the dimension where many Indian SOMNEs have outperformed. The government uses the instrument of MoU to motivate SOEs to achieve their financial and non-financial goals. The multifarious contributions of the SOEs are guided through

MoUs by the government. Sustainable development was a mandatory category to incentivize efforts by SOEs, but the weightage was only 5%. Moreover, the category was withdrawn in 2019 from MoU rating guidelines. Under the sustainable category, the SOEs were assessed on energy management, carbon management, and increased use of cleaner alternative energy sources.

The study's objective is to evaluate to what extent the SOMNEs have achieved the environmental mandate and beyond. Furthermore, the second objective is to find evidence if the practices were transferred abroad. SOMNEs' performance is evaluated on five criteria to test the research objective. The first is to check if the SOMNE has any sustainable policy or energy management system at the company level and is certified by any international organization. The Indian government encourages certification from International Organization for Standardization (ISO). Waste management is the second criterion. The criterion is more relevant for manufacturing and energy sector enterprises requiring steps for Solid Waste Management. Hazardous Waste Management, e-Waste Management, and Waste Oil Management System. Waste management can reduce illness and diseases and assist enterprises in reducing production costs. The third criterion is efforts towards environment conservation such as water conservation, energy conservation and use of renewable sources, efforts towards reduction of carbon emissions and green initiatives taken by the SOMNE. Green initiatives include plantation drives or afforestation, green innovation, or any other.

The summary of the first criterion under the environmental dimension is if the SOMNEs comply with ISO 14,001 or ISO 50,001 energy management system. ISO certification 14,001 is broad in its scope and covers all environmental aspects. ISO 50,001 only focuses on energy performance but sets specific targets, making it more difficult to achieve ISO 50,001 certification.

National Buildings Construction Corporation Limited and Water and Power Consultancy Services (India) Limited, because of their nature of business, are excluded from his criterion. All SOMNEs except MTNL are ISO14001 certified. The best practices have been transferred abroad also. The Indian SOMNE, with the largest foreign assets ONGC and its subsidiary ONGC Videsh Limited, is also ISO 14,001 certified. Very few SOMNEs comply with ISO 50,001 energy management system standards. Only Bharat Heavy Electricals Limited, Bharat Petroleum Corporation Limited, ONGC, Coal India Limited, and Power Grid Corporation of India Limited have certification.

The following criterion is regarding waste management. The waste management system is expected to be in place for all SOMNEs that have obtained ISO 14,001 certification. The percentage of recycled waste will likely differ from company to company and will depend upon the nature of the business.

As Clifton and Díaz-Fuentes (2023) noted, SOEs in the energy sector were active in this dimension. Coal India Limited reused 80.6% of discharged mined water; the remaining was retained for future use. The company has adopted several measures to ensure reclaiming of mined areas. Indian Oil has set benchmarks for the sector. The SOMNE recycles 99.5% of waste generated, 88% of wastewater, and 100% slop oil. The performance is, above all, that of public and private enterprises in India. Vedanta recycles 30.71% of wastewater, and Reliance recycles 46%. ONGC recycles 39.6% of wastewater. Bharat Petroleum Corporation Limited's all retail outlets, liquefied petroleum gas bottling plants, and Mumbai Refinery are Zero Waste to Landfills.

The ONGC's subsidiary that overlooks overseas operations, ONGC Videsh Limited, recycles wastewater more than the parent company, 67%. ONGC Videsh Limited complies with the host country's environmental regulations and has undertaken projects beyond what is required. It has also established water treatment facilities in South Sudan, Venezuela, and Russia. Bharat Petroresources Limited operates participating interests in only two blocks, domestic and overseas. It complies with the regulations of Abu Dhabi's environmental regulatory agency, the only overseas block jointly owned in Abu Dhabi.

The SOMNEs under the main activity group Heavy and Medium engineering from defence, i.e. Bharat Earth Movers Limited and Hindustan Aeronautics Limited, accomplished less than the energy companies in this segment. The products are not recycled because they are of strategic or national importance. The waste recycled was only 2% for Hindustan Aeronautics Limited. Bharat Earth Movers Limited recycled 5 to 10% of waste. The other two companies from Heavy and medium engineering companies, Bharat Electronics Limited and Bharat Heavy Electricals Limited, have excelled. Bharat Electronics Limited recycled 100% of the waste. Food waste was also used to generate biogas. Bharat Heavy Electricals Limited recycled 25% of the waste generated.

The SOMNEs that do not generally produce process wastes have also taken initiatives to process other solid wastes. Rail India Technical and Economic Service Limited and Water and Power Consultancy Services (India) Limited recycle 100% of the waste generated. The consultancy enterprises Engineers India Limited, Engineering projects India Limited, National Buildings Construction Corporation Limited, Telecommunications Consultants India Limited and Water and Power Consultancy Services (India) Limited, by business, are involved in projects related to waste management. National Buildings Construction Corporation Limited is implementing the zero-waste concept in its construction projects to manage 100% of waste at the source, which is not planned and practiced widely in India during construction.

The following criterion is efforts made by the SOMNEs for environment conservation in India and abroad. The government of India has stated the target to achieve 500 GW of green energy by 2030 and has committed to cut its carbon emissions net zero by 2070. The Indian SOEs have aligned efforts with the goal, and many SOMNEs have committed to net zero. Most energy sector SOEs have announced and committed to Net Zero. Coal India Limited is to achieve the target by as early as 2024. Bharat Petroleum Corporation Limited, Hindustan Petroleum Corporation Limited, and ONGC committed to Net Zero by 2040 and Indian Oil Corporation Limited by 2046. Engineers India Limited has committed to achieving the target by 2035. Gas Authority of India Limited and National Thermal Power Corporation are preparing their roadmap to reach net zero.

The energy sector SOEs are now foraying into green energy. ONGC plans to develop 1 GW of green energy by 2025. Similarly, other SOEs, such as Gas Authority of India Limited, Bharat Petroleum Corporation Limited, Coal India Limited, Bharat Heavy Electricals Limited, National Thermal Power Corporation Limited, and others, have invested large sums in developing the production capacities of green energy in the country. Oil and Gas companies are increasing their production of Biofuels, Natural Gas, Green Hydrogen and Compressed Biogas from Biomass waste.

Other enterprises are yet to declare net zero targets, and all play a crucial role in the transition. For example, Bharat Heavy Electricals Limited is the market leader in the nuclear and emission control business. It is also a pioneer in Engineering, Procurement

and Construction solutions for solar photovoltaic plants. The current scenario is not so green; the country meets most of its power demand from coal, which causes significant carbon emissions. To solve the immediate problem, the enterprise with National Thermal Power Corporation have devised a technology[5] that could reduce carbon emissions by 20%. Bharat Heavy Electricals Limited has commissioned the largest floating solar power plant in India at National Thermal Power Corporation. The enterprise plans to increase investments in floating solar at reservoirs of its projects because it increases the production of renewable energy and has the additional advantage of preventing water evaporation of reservoirs.

National Thermal Power Corporation has also used innovative ideas for environment conservation. The SOMNE is converting shut downed station into an eco-park. It will be constructed on the ash disposal area of Badarpur Thermal Power station. It has also fired 77,000 MT of biomass pellets, preventing nearly 45,000 acres of farm fire. Power Grid Corporation of India has minimized power transmission loss to 3%, which is at par with global standards.

All SOMNEs have contributed towards environment conservation in different ways except loss incurring MTNL. Except for MTNL, every SOMNE had rainwater harvesting at various locations and took steps to ensure groundwater recharging and recycled water. Most have Sewerage Treatment Plants, and all oil and gas companies installed state-of-the-art effluent treatment plants. Energy companies took initiatives at all levels to conserve water and reduce water consumption. Indian Oil Corporation Limited excelled and achieved a 7.17% year-on-year reduction in total water consumption. National Thermal Power Corporation was successful in reducing it by 4%.

Except for MTNL, all SOMNEs met energy requirements from alternate sources and took steps to conserve energy and reduce their carbon footprint. Bharat Earth Movers Limited and Bharat Electronics Limited outperformed in the criteria. Bharat Earth Movers Limited met 85%, and Bharat Electronics Limited met 49.55% of its energy requirement internally through green energy projects. The consultancies, by nature of business, did not have considerable investments. Still, all took steps to conserve energy, such as energy audits, rooftop solar plants, plantation drives, energy-efficient lighting, rainwater harvesting systems, and others.

The Indian SOMNEs that have excelled in this dimension also transferred their green practices abroad. Several sustainability measures were taken at Indian Oil Corporation Limited's foreign subsidiary in Sri Lanka. The many outlets in Sri Lanka are solarised, the solar facility at Trincomalee terminal reduced electricity by 5%, rainwater harvesting systems, water recycling, increased production of eco-friendly products, energy efficient lightening, etc. The SOMNE took green initiatives in Mauritius's subsidiary to increase green cover and energy-efficient lighting.

ONGC Videsh Limited took several initiatives in the host countries. In the Lower Zakum Concession region, "Project Lightening" was undertaken in Abu Dhabi. The project supplies electricity from clean energy sources to offshore platforms. The company has implemented a biodiversity project in Colombia to strengthen the ecological structure and connectivity of the forest remnants (fragmented forests) in the Moriche and Jazmin fields. Prize Petroleum International Pte Limited, a subsidiary of Hindustan Petroleum Corporation Limited, incorporated in Singapore, also promotes sustainable development and takes green initiatives.

The SOMNE consultancies, such as Engineers India Limited, by nature of business, are contributing to sustainability in the host countries. Engineers India Limited provides solutions for water management in overseas markets. The SOMNE has ongoing projects in UAE, Nigeria, and Bahrain in the segment. The enterprise also offers technological solutions for energy security, carbon-neutral technologies, and similar measures to assist enterprises across the globe in meeting their Environmental, Social, and Governance goals.

But the transfer of best green practices abroad is dependent on several factors. Indian SOMNEs from the oil and gas sector are not the operators in many investments abroad. Wherever Indian SOMNEs have managerial controls, green initiatives were taken. All SOEs are clubbed together but are a heterogeneous group, and the performance depends on the leadership. The SOMNEs that outperformed in environmental parameters in the home country have also transferred their best practices to foreign subsidiaries. The best examples are Indian Oil Corporation Limited and ONGC.

Social dimension

The market reforms such as corporatization, commercialization, and liberalization have been implemented in India. The prices across sectors are market-determined, even for petroleum products. The government intervenes only during the unprecedented rise in oil prices and freezes the selling price temporarily. The government directs CSR expenditures of SOEs through publishing guidelines. During the pandemic, the government made Health the central theme and directed CSR funds of the SOEs toward relief efforts. But the CSR expenditure mandate is the same for state and private enterprises. The data on social expenses and details of projects undertaken abroad is limited. So as a proxy, the CSR expenditures of the SOEs in India are taken as the proxy variable to test the commitment of SOMNEs towards social goals. If the SOMNEs act like private enterprises and spend more.

The government of India mandates that 2% of the average net profit of the past three years should be spent on CSR activities. The definition, by default, exempts loss-incurring enterprises from spending on social expenditures. All private and state-owned enterprises are to spend on CSR if they have a net worth of Rupees 5000 million or more, a turnover of Rupees 10,000 million or more, or a net profit of Rupees 50 million. As discussed earlier, according to the CRISIL Foundation report, in 2020, SOEs accounted for 8% of companies engaged in CSR activities, but SOEs spent 28% of total CSR expenditure. The oil companies accounted for 17% of total CSR contributions.

Table 5 presents the details of the CSR expenditures of the select SOMNEs. The CSR policy was implemented in 2014; therefore, data is available from the financial year 2014–15. If the SOMNE spends more than required in any financial year, the corresponding cell is marked "Y", implying that it spends more than required by law. If the SOMNE spends less than required in any financial year, the corresponding cell is marked "N", implying that it does not spend more than required. "NA" means Not Applicable because the firms were incurring losses and were not required by law to spend on CSR.

It can be observed from the above table that all the profit-making SOMNEs mostly spend more than required. Some of them have been spending more consistently for the

Table 5. presents the details of the CSR expenditures of the select SOMNEs.

Company Name	2014–2015	2015–2016	2016–2017	2017–2018	2018–2019	2019–2020	2020–2021	2021–2022
Bharat Earth Movers Limited	NA	NA	Y	Y	Y	Y	Y	Y
Balmer Lawrie & Company Limited	Y	Y	Y	N	Y	Y	Y	Y
Bharat Electronics limited	Y	Y	N	N	N	N	Y	–
Bharat Heavy Electricals limited	N	Y	N	N	Y	Y	Y	N
Bharat Petroleum Corporation Limited	N	Y	Y	N	N	Y	Y	N
Coal India limited	Y	Y	Y	Y	Y	Y	Y	N
Engineering Projects (India) limited.	N	Y	Y	NA	NA	NA	Y	NA
Engineers India limited	Y	Y	Y	Y	Y	Y	Y	Y
Gas Authority of India Limited	N	Y	Y	Y	Y	Y	Y	Y
Hindustan Aeronautics limited	N	Y	Y	Y	Y	Y	Y	N
Hindustan Petroleum Corporation Limited	Y	Y	Y	Y	Y	Y	Y	N
Indian Oil Corporation Limited	–	Y	N	Y	Y	Y	Y	Y
IRCON International Limited	Y	Y	N	Y	Y	Y	Y	Y
Mahanagar Telephone Nigam Limited	NA	NA	NA	NA	NA	NA	NA	NA
National Buildings Construction Corporation Limited	Y	Y	Y	Y	Y	N	Y	N
National Thermal Power Corporation Limited	N	Y	Y	Y	Y	Y	Y	Y
Oil & Natural Gas Corporation Limited	N	Y	N	Y	Y	Y	Y	Y
Oil India Limited	Y	Y	Y	Y	Y	Y	Y	Y
Power Grid Corporation of India Limited	N	Y	Y	Y	N	Y	N	Y
Rail India Technical and Economic Service (RITES) Limited	Y	Y	N	N	N	Y	Y	N
Telecommunications Consultants India Limited	NA	NA	NA	NA	NA	NA	Y	NA
Water and Power Consultancy Services (WAPCOS) Limited.	Y	Y	Y	Y	Y	Y	Y	–

Annual report of the SOMNEs from the period 2015 to 2022.
Y: Spent more than required; N: Spent Less than required; NA: Not Applicable; "–": Data Unavailable.

past five years: Bharat Earth Movers Limited, Engineers India Limited, Gas Authority of India Limited, Indian Oil Corporation Limited, IRCON, National Thermal Power Corporation, ONGC, Oil India Limited, and Water and Power Consultancy Services (India) Limited. Coal India Limited was also spending consistently above the mandate, except last year. Telecommunications Consultants India Limited and Engineering projects India Limited aren't required to spend because they are incurring losses, but they did spend in the financial year 2020–21. The CSR expenditure was on health theme during the peak of the COVID-19 pandemic.

Political dimension

The study's political dimension assesses how these enterprises align their activities with government priorities and programs. The institutional and political context significantly shapes the strategies and operations of Indian SOEs. Research indicates that the extent of state control, via ownership and political ties, has a considerable impact on SOEs' decisions regarding globalization and their overall level of global engagement. This is particularly relevant in emerging economies, where changing institutional landscapes affect the dynamics between the state and SOE executives (Liang, Ren, and Sun 2015). Indian SOMNEs, while not apolitical, are closely connected to their related ministries and consistently align their actions with current government policies and priorities.

Throughout history, SOMNEs have played a crucial role in promoting national development.

SOEs operate in both strategic and non-strategic sectors, alongside private enterprises, which have grown in number over the years. The Indian government uses Memorandums of Understanding (MoUs) to direct SOMNEs towards national policies. Recent guidelines incentivize SOMNEs to procure goods from micro and small enterprises, companies owned by women and minorities, and increase Corporate Social Responsibility (CSR) expenditures while reducing imports and boosting exports. SOMNEs have also aligned their activities with the government's flagship programs, "AatmaNirbhar Bharat Abhiyan" (meaning "Self-Reliant India") and "Make in India" policies. The SOMNEs, along with the government, run campaigns to encourage indigenous manufacturers. Companies at the enterprise level have also taken similar steps.

The policy framework encourages investment within India. However, it is vital to understand why Indian SOMNEs invest abroad. Energy sector enterprises' overseas investments aim to secure the country's energy needs. Other investments can be considered efficiency-seeking, but some have political and social motives attached to them. Some investments were also made to solve market imperfections in neighboring countries. For instance, infrastructure projects undertaken by Power Grid Corporation of India in Nepal facilitated cross-border transmission of power, which is more efficient than sharing any other resource with neighboring countries.

Some SOMNEs provide consultancy services for solarization projects in 47 Least Developed Countries (LDCs) and Small Island Developing States (SIDS) under the International Solar Alliance (ISA). These SOMNEs have extensive experience implementing projects in a vast country like India. Some of the international offices established by SOMNEs are market-seeking and resemble most private entities.

Technological dimension

The final dimension on which the firms are evaluated is Technology. According to the last R&D statistics report of the government, Central Government incurred the highest share of R&D expenditures, 45.4%, 93% of which was incurred by 12 major scientific agencies. The case of India is in line with existing literature (Jamasb and Pollitt 2011; Mazzucato 2011; Sorrentino 2020); agencies do high-risk basic and applied research and are still under the state's control because privatization wasn't allowed in the strategic sector. Choudhury and Khanna (2014), in the study of 42 laboratories of the Council of Scientific and Industrial Research, argued that it internationalized and filed more foreign patents over the years for resource independence.

The R&D expenditure of business enterprises has remained low despite all direct and indirect incentives. The generous tax incentives for R&D expenses and innovation initiatives include Super Tax Deductions, Customs Tax Exemptions, GST Exemptions, Tax Holidays, and the recently introduced Patent Box Regime, but the business sector R&D expenditures remain low compared to all large economies like India (Kaushik 2022). The overly generous incentives given to private firms led to firms relabeling non-R&D costs as R&D expenses for profiteering (Kaushik 2023). At the aggregate level, SOEs' R&D expenses remained low but were high for critical sectors such as Drugs and

Pharmaceuticals, Transportation, Information Technology, Chemicals (other than fertilizer), and Defense Industries.

The variable for testing is the R&D Intensity, calculated by dividing the R&D expenses of the enterprises by the total sales of the SOE. The revenue includes both domestic and foreign earnings. The government mandates in MoU that 2% of Profit before tax should be spent on R&D expenses or innovation and initiatives. This study calculated the R&D intensity of the sample Indian SOMNEs and compared it with the R&D intensity of its respective sectors. The motive is to examine if SOMNEs spend more on innovation than other firms in the industry. As discussed before, the trend of R&D intensity is discussed in the segment. R&D Intensity is calculated by dividing the R&D expenses of the enterprises by the total sales of the SOE. The revenue includes both domestic and foreign earnings. Not all SOMNEs were in business that required R&D expenditures; only relevant ones are presented in Table 6.

The government has excluded R&D intensity in the recent MoU, leading to a decline in this area. However, some SOMNEs such as Bharat Earth Movers Limited, Bharat Heavy Electricals Limited, Bharat Petroleum Corporation Limited, Hindustan Aeronautics Limited, ONGC, National Thermal Power Corporation, Indian Oil Corporation Limited, Hindustan Petroleum Corporation Limited, and Engineers India Limited have shown an increase in R&D intensity. The top five SOMNEs with high R&D intensity are Bharat Earth Movers Limited, Bharat Heavy Electricals Limited, Coal India Limited, Engineers India Limited, and Hindustan Aeronautics Limited, with SOMNEs from the defense sector spending more on R&D. While SOMNEs can be used by the government to drive innovation, there is little evidence that they have undertaken such activities abroad.

Section III conclusion

In conclusion, the study highlights the distinctive operational approach of India's SOMNEs, showcasing their ability to sustain profitability through diversified investments, even in the face of certain risky ventures. Notably, most SOMNEs have been actively engaged in environmental conservation initiatives, barring the loss-making MTNL. The profitable entities have consistently exceeded corporate social responsibility obligations, with some surpassing mandated thresholds. Surprisingly, even the loss-making Engineering projects India Limited and MTNL have supported COVID-19 relief funds, demonstrating a commitment to societal welfare.

Despite the challenges, none of the SOMNEs resorted to employee layoffs during the pandemic, emphasizing their unique commitment to economic and social stability. While some SOMNEs exhibit high research and development intensity, there is limited evidence of these activities being outsourced internationally. This distinct operating paradigm sets SOMNEs apart from their private counterparts in India, underscoring their dual focus on creating social value and ensuring profitability.

The internationalization journey of Indian SOMNEs emerges as a global precedent, showcasing an integration of profit-oriented objectives with a public mission. The study reveals key findings, including the consistent international growth trajectory of SOMNEs, blending market-driven tactics with a broader social mission. These

Table 6. The R&D expenditures of the SOMNE.

	FY 2018–19		FY 2019–20		FY 2020–21	
Company Name	R&D Expenditure (million in rupees)	R&D as % of Sales	R&D Expenditure (million in rupees)	R&D as % of Sales	R&D Expenditure (million in rupees)	R&D as % of Sales
Bharat Earth Movers Limited	707.2	2.03	1035	3.42	1327	3.73
Balmer Lawrie & Company Limited	110.3	0.62	77.8	0.51	83	0.55
Bharat Electronics Limited.	8866.6	7.34	0	0	0	0
Bharat Heavy Electricals Limited	8196.9	2.69	7665	3.57	7265	4.2
Bharat Petroleum Corporation Limited	1007.2	0.03	921.8	0.03	866.4	0.04
Coal India Limited	1036.9	11.1	86.5	1.02	270.6	4.23
Engineers India Limited	185.3	0.76	243.1	0.76	263.6	0.85
Gas Authority of India Limited	371.7	0.05	664.1	0.09	92	0.02
Hindustan Aeronautics limited	14644	7.32	12319.6	5.74	16870	7.42
Hindustan Petroleum Corporation Limited	2538.5	0.09	2579	0.1	2957	0.13
Indian Oil Corporation Limited	3001.9	0.06	2670	0.05	2828	0.07
National Thermal Power Corporation	1385.9	0.15	1833	0.18	1858	0.19
Oil & Natural Gas Corporation Limited	5832.8	0.53	5558	0.58	5541	0.81
Oil India Limited	861.9	0.63	0	0	0	0
Power Grid Corporation of India Limited	126.7	0.04	95.4	0.03	455	0.01

Public Enterprise Survey Report published by Department of Public Enterprise, Government of India.

enterprises have successfully displayed resilience and adaptability, navigating global economic fluctuations and the recent pandemic.

The history of privatization in India has seen the significant sale of profitable enterprises at substantial concessions during fiscal stress. The current government's new policy also aims to minimize its presence in SOEs. The increasing inclination of the Indian government and other national authorities towards disinvestment in all SOEs, both strategic and non-strategic, calls for a strategic reevaluation in light of these findings.

Disinvestment, often pursued to infuse capital efficiency and foster private sector participation, may inadvertently undermine the unique strengths and contributions of SOMNEs and SOEs. As this study indicates, these enterprises are significant contributors to the national economy and champions of social and environmental initiatives.

In this context, governments must consider a balanced approach that nurtures the growth and international presence of SOMNEs. Instead of disinvestment, the emphasis should be on creating an enabling environment that supports these entities in their dual mission of economic growth and social responsibility. This involves reinforcing regulatory frameworks, ensuring access to international

markets, and providing strategic support in areas like technology acquisition and sustainable practices.

For SOMNEs, the study underscores the importance of aligning internationalization strategies with global best practices in sustainability and corporate governance. This alignment maintains India's competitive edge and enhances its global reputation. The narrative of Indian SOMNEs' internationalization becomes a compelling model for other emerging economies, offering insights into leveraging state-owned enterprises for economic, social, and environmental progress. In summary, the internationalization of Indian SOMNEs is a multifaceted narrative encompassing strategic growth, societal contribution, and environmental commitment.

Notes

1. For more details, refer to the Reserve Bank of India Notification: Master Direction – Direct Investment by Residents in Joint Venture (JV)/Wholly Owned Subsidiary (WOS) Abroad https://www.rbi.org.in/Scripts/BS_ViewMasDirections.aspx?id=10637#C5 (RBI 2021)
2. Strategic Sale means the sale of large government shareholding (not necessarily major) to a strategic investor. In strategic sales, the management control is transferred to the investor, even if the strategic partner is not the major shareholder.
3. Asset monetization is leasing out of state-owned assets.
4. The government has cancelled the strategic sale of Pawan Hans after the disqualification of the winning bidder.
5. The technology is called Advanced Ultra Super Critical (AUSC) technology. It will be used to transform inefficient old power-generating units into more efficient and less polluting units.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Aakanksha Kaushik http://orcid.org/0000-0003-2552-7442

References

Auty, R. M. 1988. "State Enterprise and Resource Based Industry in Oil Exporting Countries." *Resources Policy* 14 (4): 275–287. https://doi.org/10.1016/0301-4207(88)90036-0.

Bala, M. 2006. "Economic Policy and State Owned Enterprises: Evolution Towards Privatisation in India." Available at SSRN 944423. https://doi.org/10.2139/ssrn.944423.

Carlos Zalaf Caseiro, L., and G. Masiero. 2014. "OFDI Promotion Policies in Emerging Economies: The Brazilian and Chinese Strategies." *Critical Perspectives on International Business* 10 (4): 237–255. https://doi.org/10.1108/cpoib-03-2014-0023.

Chatterjee, E. 2017. "Reinventing State Capitalism in India: A View from the Energy Sector." *Contemporary South Asia* 25 (1): 85–100. https://doi.org/10.1080/09584935.2017.1290583.

Choudhury, P., and T. Khanna. 2014. "Charting Dynamic Trajectories: Multinational Enterprises in India." *Business History Review* 88 (1): 133–169. https://doi.org/10.1017/S000768051300144X.

Christensen, S. A. 2023. ""Explaining State Ownership in Listed Companies in Norway." *Enterprise & Society* 1–26. https://doi.org/10.1017/eso.2023.19.

Clifton, J., and D. Díaz-Fuentes. 2019. "The State and Public Corporations." In *Handbook of the International Political Economy of the Corporation*, edited by A. Nölke and C. May, 106–119. Edward Elgar. https://doi.org/10.4337/9781785362538.00013.

Clifton, J., and D. Díaz-Fuentes. 2023. "How Do State Owned Enterprise Multinationals Behave Abroad? A Multi-Dimensional Framework for Analysis." *Journal of Economic Policy Reform* 26 (1): 67–81. https://doi.org/10.1080/17487870.2022.2079509.

CRISIL Foundation. 2020. *Doing Good in Bad Times: CRISIL CSR Yearbook 2020*. CRISIL Ratings Limited. https://www.crisil.com/content/dam/crisil/our-analysis/reports/corporate/documents/2020/06/doing-good-in-bad-times.pdf.

Cuervo-Cazurra, A., A. Inkpen, A. Musacchio, and K. Ramaswamy. 2014. "Governments As Owners: State-Owned Multinational Companies." *Journal of International Business Studies* 45 (8): 919–942. https://doi.org/10.1057/jibs.2014.43.

Gupta, A., and S. Kumar. 2020. "Comparing the Performance of Public and Private Enterprises: Case for a Reappraisal – Evidence from India." *International Journal of Public Sector Management* 34 (1): 87–100. https://doi.org/10.1108/ijpsm-04-2020-0117.

Heath, J., and W. Norman. 2004. "Stakeholder Theory, Corporate Governance and Public Management: What Can the History of State-Run Enterprises Teach Us in the Post-Enron Era?" *Journal of Business Ethics* 53 (3): 247–265. https://doi.org/10.1023/B:BUSI.0000039418.75103.ed.

IMF. 2020. "Fiscal Monitor: Policies to Support People During the COVID-19 Pandemic." https://www.imf.org/en/Publications/FM/Issues/2020/04/17/Fiscal-Monitor-April-2020-Policies-to-Support-People-During-the-COVID-19-Pandemic-49278.

Jamasb, T., and M. G. Pollitt. 2011. "Electricity Sector Liberalisation and Innovation: An Analysis of the UK's Patenting Activities." *Research Policy* 40 (2): 309–324. https://doi.org/10.1016/j.respol.2010.10.010.

Joseph, R. K. 2019. "Outward FDI from India." *Institute for Studies in Industrial Development Working Paper 214*. Delhi, India: Institute for Studies in Industrial Development. https://isid.org.in/wp-content/uploads/2022/07/WP214.pdf.

Kalotay, K. 2018. "State-Owned Multinationals: An Emerging Market Phenomenon?" *The Journal of Comparative Economic Studies* 13:13–38. https://doi.org/10.2139/ssrn.3091541.

Kaushik, A. 2022. "The Misdirected Policy Changes in R&D Tax Incentives." *Economic and Political Weekly* 57 (13): 14–16. https://www.epw.in/journal/2022/39/commentary/misdirected-policy-changes-rd-tax-incentives.html .

Kaushik, A. 2023. "The Effectiveness of Research and Development Tax Incentives in India: A Quasi-Experimental Approach." *International Journal of System Assurance Engineering & Management* 14 (6): 2329–2336. https://doi.org/10.1007/s13198-023-02077-x.

Khanna, S. 2012. "State-Owned Enterprises in India: Restructuring and Growth." *The Copenhagen Journal of Asian Studies* 30 (2): 5–28. https://doi.org/10.22439/CJAS.V30I2.4237.

Kowalski, P., M. Büge, M. Sztajerowska, and M. Egeland. 2013. *State-Owned Enterprises: Trade Effects and Policy Implications*. Paris: OECD Publishing. https://doi.org/10.1787/5k4869ckqk7l-en.

Kumar, S. 2018. "MoU System in State-Owned Enterprises in India: A Tool in Strategic Management." *Indian Journal of Public Administration* 64 (1): 36–48. https://doi.org/10.1177/0019556117735460.

Liang, H., B. Ren, and S. Sun. 2015. "An Anatomy of State Control in the Globalization of State-Owned Enterprises." *Journal of International Business Studies* 46 (2): 223–240. https://doi.org/10.1057/JIBS.2014.35.

Mazzucato, M. 2011. "The Entrepreneurial State." *Soundings* 49 (49): 131–142. https://doi.org/10.3898/136266211798411183.

Nölke, A., C. May, D. Mertens, and M. Schedelik. 2022. "Elephant Limps, but Jaguar Stumbles: Unpacking the Divergence of State Capitalism in Brazil and India Through Theories of Capitalist Diversity." *Competition & Change* 26 (3–4): 311–333. https://doi.org/10.1177/10245294211015597.

OECD. 2020. *The COVID-19 Crisis and State Ownership in the Economy: Issues and Policy Considerations*. OECD Publishing. https://www.oecd.org/coronavirus/policy-responses/the-covid-19-crisis-and-state-ownership-in-the-economy-issues-and-policy-considerations-ce417c46/#boxsection-d1e31 .

Petry, J., K. Koddenbrock, and A. Nölke. 2023. "State Capitalism and Capital Markets: Comparing Securities Exchanges in Emerging Markets." *Environment and Planning A: Economy and Space* 55 (1): 143–164. https://doi.org/10.1177/0308518X211047599.

Pradhan, J. P. 2017. "Indian Outward FDI: A Review of Recent Developments." *Transnational Corporations* 24 (2): 43–70. https://unctad.org/system/files/official-document/diaeia2017d3a2_en.pdf.

Pradhan, J. P., and K. P. Sauvant. 2010. "Introduction: The Rise of Indian Multinational Enterprises: Revisiting Key Issues." In *The Rise of Indian Multinationals: Perspectives on Indian Outward Foreign Direct Investment*, edited by K. Sauvant, J. Pradhan, A. Chatterjee, and B. Harley, 1–23. New York: Palgrave Macmillan US. https://doi.org/10.1057/9780230114753_1.

RBI (Reserve Bank of India). 2021 June 24. "Master Direction – Direct Investment by Residents in Joint Venture (JV)/Wholly Owned Subsidiary (WOS) Abroad." Reserve Bank of India. Accessed November 2, 2022. https://www.rbi.org.in/Scripts/BS_ViewMasDirections.aspx?id=10637.

Rygh, A., and G. R. Benito. 2022. "Governmental Goals and the International Strategies of State-Owned Multinational Enterprises: A Conceptual Discussion." *Journal of Management & Governance* 26 (4): 1155–1181. https://doi.org/10.1007/s10997-021-09595-5.

Sinha, S. K. 2022. *Public Sector Contributes Below 20% to National Income, but Commands Nearly 40% of Total Wages*. India Ratings & Research. https://www.indiaratings.co.in/rating-details/2/59348 .

Sorrentino, M. 2020. "State-Owned Enterprises and the Public Mission." In *The Routledge Handbook of State-Owned Enterprises*, edited by L. Bernier, P. Bance, and M. Florio, 73–90. New York: Routledge. https://doi.org/10.4324/9781351042543.

Tiwari, R., and C. Herstatt. 2010. "The Emergence of Indian Multinationals: An Empirical Study of Motives, Current Status, and Trends of Indian Investment in Germany." In *The Rise of Indian Multinationals: Perspectives on Indian Outward Foreign Direct Investment*, edited by K. Sauvant, J. Pradhan, A. Chatterjee, and B. Harley, 233–253. New York: Palgrave Macmillan US. https://doi.org/10.1057/9780230114753_8.

UNCTAD. 2008. "World Investment Report 2008." In *Transnational Corporations, and the Infrastructure Challenge*, Geneva: UNCTAD. https://d#x00AF;i.#x00AF;rg/unctad.publication/world-investment-report-2008 .

UNCTAD. 2010. *World Investment Report 2010: Investing in a Low-Carbon Economy*. Geneva: UNCTAD. https://investmentpolicy.unctad.org/publications/58/world-investment-report-2010—investing-in-a-low-carbon-economy .

UNCTAD. 2022. *World Investment Report 2022: International Tax Reforms and Sustainable Investment*. Geneva and New York: UNCTAD. https://unctad.org/publication/world-investment-report-2022.

World Bank. 2021. *Building SOE Crisis Management and Resilience: Emerging Practices and Lessons Learned During the COVID-19 Crisis*. Washington, DC: World Bank. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099335003032221171/p1754710c299e40f30b6cd09605ca186e55.

Educational Administration: Theory and Practice
2024, 30(5), 4798 - 4808
ISSN: 2148-2403
https://kuey.net/

Research Article

User Experience Design And Its Impact On Human Psychology

Shahid Ahmad[1*], Ranganath M Singari[2], S.L. Bhandarkar[3]

[1*,2]Department of Design, Delhi Technological University, Delhi, India, *Email: shahidahd10@gmail.com, https://orcid.org/0000-0001-9137-4566
[3]Guru Gobind Singh Indraprastha University, Delhi, India

Citation: Shahid Ahmad, et.al (2024), User Experience Design And Its Impact On Human Psychology, *Educational Administration: Theory and Practice, 30(5), 4798 - 4808*
Doi: 10.53555/kuey.v30i5.3702

ARTICLE INFO

ABSTRACT

In the digital age, the discipline of User Experience Design (UXD) has become increasingly important, influencing people's interactions with digital interfaces and goods. This extensive study paper delves into the many dimensions of user experience design (UXD), covering everything from its historical development and foundational ideas like usability, information architecture, visual design, and user-centered design to its significant impact on human psychology. Examining emotions, cognitive load, behaviour, trust, and satisfaction as crucial characteristics touched by UXD, it delves into the psychology of user experiences. Moreover, the research offers perceptive case studies that demonstrate how UXD concepts propel user delight and corporate success in a variety of industries. These case studies include UXD in E-Commerce, Mobile Applications, and Healthcare Technology. Design rules, emotional design, cognitive load reduction, trust enhancement, and loyalty fostering are outlined as practical implications for designers. The study also discusses the difficulties and moral issues that come with user experience design, highlighting the significance of inclusiveness, user privacy, and responsible design methods. Lastly, it explores cutting-edge themes including AI-powered personalization and the moral application of UXD to protect user privacy. This offers a forward-looking viewpoint on how UXD is changing in the digital era and emphasizes its continuing importance. Through an analysis of the convergence of design, psychology, and technology, research demonstrates how UXD influences user cognition, behaviour, and emotions, which in turn affects people and society as a whole.

Introduction

In an increasingly digital world, user experience has emerged as a crucial element in the design and development of products and services (Pucillo & Cascini, 2014). The multidisciplinary field of user experience design (UXD) blends elements of human-computer interaction, design, psychology, and technology (Luther et al., 2020) to give users intuitive and pleasurable experiences. This research paper aims to provide a comprehensive analysis of user experience design and how it affects human psychology.

In a world that is networking and changing quickly, the way individuals interact with technology (Cheng et al., 2012) is affecting the passage of time. From the smartphones we carry in our pockets to the websites and services we use on a daily basis, technology has become an essential part of our lives. In this digital world, the concept of User Experience Design (UXD) has become essential, impacting not just our interactions with technology but also our thoughts, feelings, and behaviours.

The engine of this technological revolution is user experience design. In order to provide intuitive and pleasurable user experiences, this multidisciplinary field combines design, technology, psychology, and human-computer interaction (Olson & Olson, 2003). The aim of user experience design (UXD) is to comprehend people's needs, expectations, and behaviours when they interact with digital goods and services. The objective is to make technology simple and easy to use, whether you're using a smartphone app, an online store, or a complex piece of software. Designing encounters that are both enjoyable and functional is the aim of user experience design (UXD). It involves ensuring that you feel good about yourself after using a digital product, whether that feeling is joy, satisfaction, or a sense of success.

Copyright © 2024 by Author/s and Licensed by Kuey. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

UXD's influence, however, extends beyond usability and enjoyment on the surface. It delves into many facets of human mind. A crucial element of UXD's design is human psychology, which is the study of human cognition, emotion, and behaviour. Our in-depth knowledge of how the human mind processes information, how emotions influence our decisions, and how our cognitive capacities affect how we engage with technology has an impact on everything from website layout to mobile app colour schemes, information flow, and software interface design.

This study article aims to explore the intricate relationship that exists between user experience design and human psychology. It seeks to provide a comprehensive understanding of UXD, covering its definition, historical evolution, tenets, and important psychological ramifications. As we examine the foundations of psychology, we will discover how emotions, behaviour, and cognition are intertwined with the production of digital experiences. Along with difficulties and ethical considerations, real-world case studies, future trends in the quickly evolving field of UXD, and practical implications for designers, the paper will also cover these topics. Readers will gain a greater understanding of the significance of user experience design (UXD), which goes beyond creating intuitive interfaces to understanding and influencing human thought, emotion, and behaviour in the digital age.

1.1 Background

When developing digital or physical products and services, user demands and preferences are prioritized over those of functionality and aesthetics. This approach is known as user experience (UX) design. In today's oversaturated digital world, the quality of the user experience (Tesařová, 2022) often determines whether a digital product succeeds or fails. There is no shortage of possibilities for users. Because User Experience Design (UXD) aims to create meaningful and enjoyable experiences that deeply resonate with people, it is an indispensable component in the competitive digital world.

1.2 Research Objectives

This research paper sets out to explore the multifaceted world of User Experience Design and its profound impact on human psychology. Our main goals are as follows:

(a) Defining User Experience Design: We will provide a thorough definition of what User Experience Design comprises, considering its historical development, guiding principles, and essential ideas.

(b) Analysing the Effect on Human Psychology: We'll explore the complex ways that UXD affects psychology in humans. We will specifically look at the emotional reactions it arouses, how it affects cognitive functions, how it affects user behaviour, credibility and trust, and how it affects user happiness and loyalty.

1.3 Methodology

Our research will use a multimodal approach in order to meet its goals.

(a). Literature Review: A comprehensive analysis of the corpus of existing research on user experience design and human psychology will be carried out, utilizing academic studies, business publications, and expert judgment.

(b) Case Studies: A variety of case studies from various sectors will be looked at in order to show how UXD affects human psychology and to provide a real-world perspective.

(c) Ethical Considerations: As the ethical facets of UXD are investigated, user privacy, persuasive design, and accessibility concerns will be looked at.

2. What is User Experience Design?

User Experience Design (UXD) is a multidisciplinary area that aims to develop systems, services, or products that give users remarkable and meaningful experiences (Borriraklert & Kiattisin, 2021). It is a thorough approach to design that considers not just a product's utility and efficiency but also its emotional and psychological impact on its customers. By making interactions with digital or physical objects as simple, entertaining, and productive as possible, user experience design (UXD) aims to enhance user pleasure and engagement.

2.1 Definition of User Experience Design:

User Experience Design is he process of creating products and services with the needs, goals, and perspective of the user in mind (Vahdat, 2023). It encompasses not just a product's technical specifications and aesthetic appeal, but also the entire user experience. UXD aims to be not only user-friendly but also emotionally charged, memorable, and ultimately helpful to the user.

2.2 Historical Evolution:

The field of User Experience Design (UXD) has its origins in the Human-Computer Interaction (HCI) movement of the 1960s and 1970s, which aimed to make computer systems more user-friendly in the workplace. The phrase "user experience" was made popular by cognitive scientist and designer Don Norman during his tenure at Apple in the 1990s. Since then, processes and technologies have evolved in tandem with the rapid expansion of UXD. Furthermore, Norman's seminal book "The Design of Everyday Things," which was released in 1988, is where the idea of UX design first emerged. Because of Norman's emphasis on the intricacy of developing often-overlooked things, UX design has become the de facto standard approach to user-centered design (Norman, 2013). Figure 1. An illustration by Orange-crush demonstrates how a product's UX design makes it fun and simple to use.

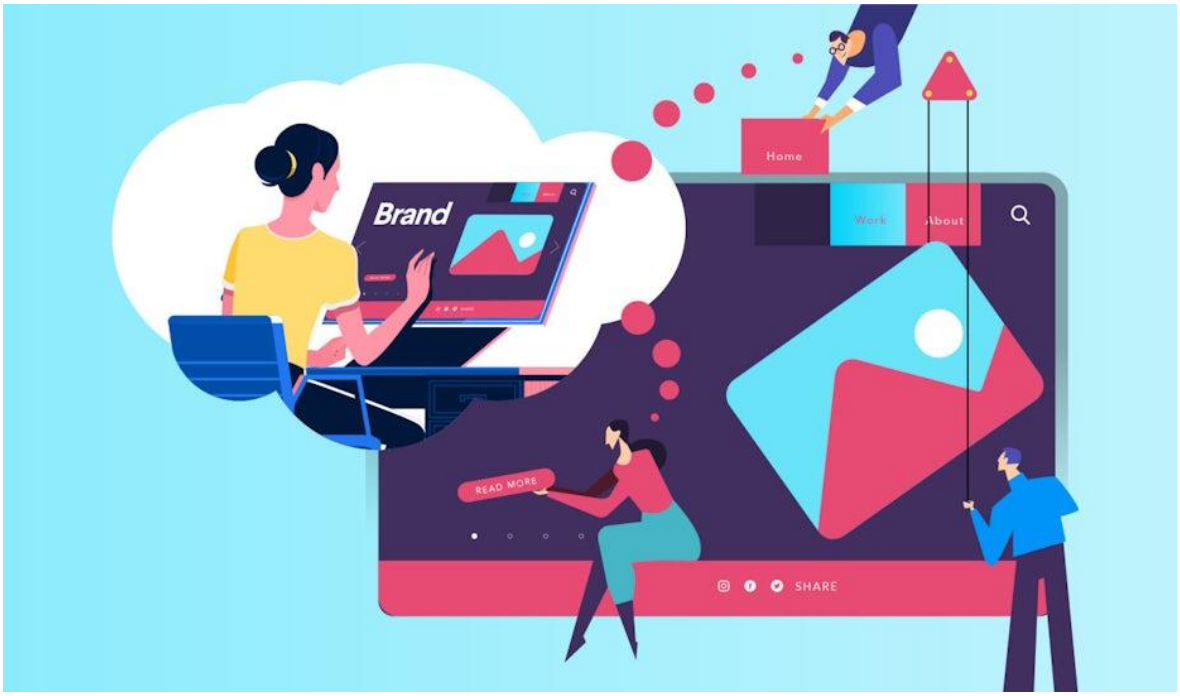


Fig 1.

2.3 Key Principles of UXD:

Any product or service must prioritize user experience. It includes all of a user's interactions with a good or service, from the beginning stage of research until the point of sale or use. It's crucial to adhere to certain standards, such simplicity, clarity, and consistency, in order to provide the best possible user experience. The following basic concepts serve as the foundation for the main tenets of user experience design (Stevens & Stevens, 2022):

Empathy: UX designers must place themselves in the shoes of their customers and consider their needs, frustrations, and motivators in order to fully comprehend and sympathize with them.

Consistency: Consumers are better able to form mental models of how a product works when design elements like layout, navigation, and vocabulary are constant. This helps to eliminate uncertainty.

Simplicity: In UXD, designing with simplicity is essential to creating intuitive, user-friendly products that give customers clear, simple instructions. A minimalist design reduces cognitive load, allowing users to execute tasks more quickly. It means getting rid of unnecessary complexity and clutter.

Feedback: By integrating feedback methods into UX design, it is possible to continuously update the design in response to user interactions and feedback, which enhances the user experience. Giving users immediate, succinct feedback on their actions helps them understand the consequences of their actions and speeds up error recovery.

Information Hierarchy: By creating a distinct information hierarchy, a page's usability and comprehension are improved by directing users' attention to the most crucial content. Information should be presented hierarchically, with the most important or often requested content being readily accessible, to ensure that visitors can find what they need fast.

Usability: A key element of UXD, usability is the degree to which tasks within a system or product can be

completed by users quickly and easily. Errors (the frequency with which users make mistakes), memorability (the ease with which users can recall how to use the product), speed (the speed at which users can pick up new skills), and satisfaction (the degree to which users are content with a product) (Rusu et al., 2015) are all included in the category of usability.

2.4 User-Centered Design:

User-Centered Design (UCD) is a fundamental aspect of UXD. Research is necessary to comprehend user behaviours, needs, and preferences. This study informs the design process, ensuring that end users are considered while developing new products and services. Common steps in UCD include brainstorming, user research, prototyping, testing, and iteration.

User-Centered Design (UCD) is an iterative design process that prioritizes understanding users and their context throughout all stages of design (Jurca et al., 2017). By including users at every stage of the design process, this method guarantees that goods are not only aesthetically beautiful but also practical and easy to use. UCD concentrates on developing products that satisfy users' requirements, goals, and input while developing special relationships with the communities they support. Through user-centered design (UCD), products are validated by end users at every level, resulting in efficient solutions that meet unique needs. Phases of the UCD process typically include gathering information about user behaviours and wants, designing solutions, testing them, and iterating designs to improve usability and satisfaction. Requirements are also specified based on user insights.

3. What is Human Psychology?

Human psychology is the scientific study of the human mind and conduct, with an emphasis on the causes and mechanisms behind people's thoughts, emotions, and behaviours (Pérez-Álvarez, 2017). It investigates many facets of human emotion, cognition, and behaviour, offering insights into the reasons behind people's decisions, responses to stimuli, and interactions with their environment. Designers need to understand human psychology in order to make products that offer outstanding user experiences.

To learn more about human behaviour, psychologists do empirical research and apply systematic approaches. To collect information and make conclusions, they could carry out surveys, observations, experiments, and clinical investigations. Studying social dynamics, societal impacts on behaviour, and individual cognitive processes are just a few of the many areas covered by psychology.

Origin and History of Psychology:

The origins of psychology can be found in the thinking of human behaviour and the secrets of the mind discovered by ancient societies such as the Greeks, Chinese, and Egyptians (lumenlearning, 2024). However, the first psychology laboratory was established in Leipzig, Germany, by German psychologist Wilhelm Wundt in 1879, which signalled the start of modern psychology as a scientific discipline in the late 19th century. This cleared the path for additional research and established psychology as an experimental and scientific field. Wundt's investigation of consciousness and his establishment of structuralism as a psychological method had a major influence on the discipline.

Influential Early Figures in Psychology

The journey of psychology's development is deeply influenced by the work of early figures who made significant contributions to the field. Among these significant forerunners are:

Wilhelm Wundt: Widely recognized as the father of psychology, Wundt's contributions paved the way for subsequent scientific investigations in the area. His self-reflective techniques aimed to reveal the composition of conscious experience.

William James: James, an American psychologist and philosopher, focused on comprehending the goals and functions of the mind through his functionalism, which shifted attention to psychology's larger picture. He is well known for writing "The Principles of Psychology," which is now regarded as a classic in the subject.

Sigmund Freud: Freud's psychoanalytic theory shed light on the complexity of human thought and behaviour by emphasizing the unconscious mind and early childhood experiences. His writings established ideas like the id, ego, and superego as well as the significance of dreams in providing insight into the unconscious.

John B. Watson: A behaviourist, Watson promoted the idea that psychology should be empirical and objective and excluded the study of mental processes in favour of an emphasis on observable behaviour. His well-known "Little Albert" experiment illustrated how classical conditioning works.

B.F. Skinner: Another significant behaviourist, Skinner developed the theory of operant conditioning, which

has real-world uses in behaviour modification and understanding. In his research, Skinner highlighted the influence of reinforcement and the environment on behaviour.
Psychology has evolved into many different schools of thought over time, such as behaviourism, psychoanalysis, functionalism, and humanistic psychology. These many viewpoints examined various facets of human psychology, ranging from analysing unconscious processes to highlighting personal development and potential to comprehending behaviour in terms of its function.

Key Elements of Psychology:
Several key elements of psychology are relevant to design:

Cognition: This component deals with mental functions like perception, memory, focus, and problem-solving. In order to build interfaces that are simple to browse and understand, designers take cognitive characteristics into account.

Emotions: It's important to comprehend how emotions affect user behavior. While unpleasant emotions can turn people away, positive emotions can improve user experiences.

Perception: When designing, designers consider how users' senses interpret and experience information. This covers haptic feedback, aural perception, and visual perception.

Motivation: In order to engage users, motivation is necessary. In order to persuade customers to perform desired actions, such finishing a purchase or signing up for a service, designers integrate motivational ideas. recognizing the needs, interests, and objectives that influence human behaviour.

Development: Studying how people evolve and change from early childhood to old age.

Social Interaction: Studying interpersonal relationships, encompassing subjects like communication, group dynamics, and social influence.

Learning and Memory: Information retention and learning are factoring that designers consider. They try to design experiences that are simple to understand and retain, reducing the need for ongoing relearning.

Decision-Making: When utilizing items, users make decisions. Reducing choice fatigue and optimizing product interfaces are two benefits of understanding decision-making processes for designers.

User Input: Obtaining user input and carrying out usability testing are crucial components of design psychology. This aids in problem identification and data-driven design enhancement.

The Importance of Psychology for Design:
A key component for designers looking to build products with exceptional user experiences is human psychology. Design heavily relies on psychology, especially when it comes to comprehending users and developing user-centered solutions (Alves, 2023). Designers may make products more intuitive and human-centered by enhancing the user experience through the integration of psychological principles into the design process. Designers that are knowledgeable about psychology are better able to relate to users, understand their wants, behaviours, and motivations, and create solutions that effectively address those needs. This information aids in the development of user interfaces that direct users toward intended actions and results, which eventually result in a satisfying user experience. Designers may use visual components to create coherent designs that consumers see as cohesive wholes by utilizing psychological ideas like Gestalt Principles (Abhishek, 2023).
Key factors that highlight the significance of psychology in design include (Lawerteh, 2022):

Human-Centered Design (HCD): HCD is a design methodology that gives users' wants and experiences first priority. Psychology aids in its implementation.

Perception and Reality: Since people interpret designs depending on contextual cues and cognitive pre-processing, it is critical for designers to comprehend the subjective nature of human perception.

Predictive Perception: In order to solve design problems and reduce human error, human cognition entails making predictions and revising mental models.

Gestalt Principles: Creating designs that are successful, logical, and in line with human experiences and expectations requires the application of Gestalt principles in design.

User-Friendly Products: By bridging the knowledge gap between design and user experience, psychology helps make goods that are easy to use, pleasurable for users, and intuitive.

Colour psychology: Colour psychology is a useful technique in design that helps elicit desired responses from users since colours have the power to inspire particular emotions and impact user behaviour.

Hick's Law: This idea stresses making options easier for consumers to understand so they can make better decisions and have less cognitive strain overall.

Mental Models: By utilizing users' pre-existing knowledge and familiarity with design patterns, designing based on their mental models improves the user experience (Nielsen & Chan, 2024).

Priming: By matching messaging to consumers' expectations, psychological concepts like priming can affect user trust and engagement with a brand or product.

Relevance of Psychology to User Experience Design

Understanding human psychology is essential for designing products and interfaces that are both efficient and user-friendly in the rapidly changing fields of technology and product design. User experience (UX) designers use psychological concepts to create intuitive, captivating experiences that ultimately satisfy the needs of the user (Kohler, 2024). The following are some significant ways that psychology and user experience design interact:

Cognitive Psychology: This branch supports interface designers in building user interfaces that match consumers' expectations and mental models. Through an understanding of human information processing, designers may create user-friendly products. Design professionals are guided in creating more intuitive user interfaces by the concepts of cognitive psychology, which include information processing, memory, and perception.

Emotional Design: Emotions have a critical influence in the user experience. Emotional design concepts are used by designers to arouse pleasant emotions in the viewer, which can improve user engagement and loyalty. Interface designers are able to produce user experiences that really connect with users when they have a thorough understanding of how people feel about visual design, information, and interactions.

Behavioural Psychology: By utilizing the concepts of operant conditioning and reinforcement, behavioural psychology principles can be utilized to promote desired user activities, such as signing up for a service or completing a purchase. In order to influence user behaviour and raise satisfaction and engagement, designers can utilize incentives and feedback (Luther et al., 2020).

User-Centered Design: User-centered design requires extensive user research and feedback gathering. The goal of design is to produce goods that satisfy the requirements, tastes, and expectations of people. Designers may better adjust product designs to customer needs by gathering insights into user behaviour and preferences through user research, questionnaires, and usability testing.

Visual Perception: Gestalt psychology, colour psychology, and typography are a few examples of concepts that designers may use to create visually appealing and functional interfaces by having a better understanding of how people see and interpret visual information. These guidelines help designers create aesthetically pleasing and intuitive user interfaces.

Decision-Making: By utilizing the concepts of behavioural economics and decision-making psychology, designers may lessen decision fatigue and simplify their designs, which will enhance user experiences. Designers can make interfaces that help users make decisions and feel less frustrated by knowing the cognitive biases that affect decision-making.

Ethical Considerations: Designers may tackle current ethical difficulties by prioritizing the well-being and autonomy of users through an understanding of user trust and privacy concerns, as well as ethical principles. Ethical concerns such data privacy, openness, and how their designs affect consumers' well-being must be considered by designers.

Understanding human cognition and behaviour is made possible by human psychology, which is the basis for many other disciplines, including user experience design. By utilizing psychological notions, designers can produce visually arresting, functionally efficient, and emotionally captivating objects. Also the history, key figures, subfields, and critical significance of psychology to user experience design is essential for navigating the intricacies of human cognition and its practical implications. These insights can be used by designers to

create products that have a strong emotional bond with users, enhancing their experiences and making a big difference in a range of industries.

Case Studies

4.1 User Experience Design in E-Commerce:

One industry where User Experience Design (UXD) is crucial is e-commerce. In this case study, we look at how good UXD may have a big impact on an online store's ability to succeed.

Background:

E-commerce platforms are online stores where customers may peruse, choose, and buy goods and services. There is a lot of rivalry in the e-commerce space, and customers have a lot of options. For this reason, offering a first-rate user experience is crucial to drawing in new business and keeping existing ones.

Key Aspects of UXD in E-Commerce:

User Interface Design: It's critical to create a visually appealing and user-friendly interface. This entails creating intuitive and aesthetically pleasing product lists, search capabilities, and navigation menus.

Product Discovery: Users may identify products that suit their tastes with the aid of efficient processes for product discovery, such as filters, sorting choices, and personalized suggestions.

Checkout Process: A crucial component of UXD is streamlining the checkout procedure to lower friction and increase conversions. This entails cutting down on the amount of steps, providing guest checkout choices, and guaranteeing safe payment processing.

Mobile Optimization: It is imperative to optimize the e-commerce platform for mobile customers, given the growing trend of using mobile devices for online purchasing. It is imperative to have responsive design, mobile-friendly navigation, and quick loading times.

Customer Reviews and Ratings: Including user-generated content can help people make wise judgments about what to buy by fostering trust. Examples of this type of content include product reviews and ratings.

Impact of Effective UXD:

Effective UXD in e-commerce can lead to several outcomes:

Increased Conversion Rates: A website that is easy to use and intuitive to navigate encourages customers to finish their transactions, which results in greater conversion rates.
Reduced Cart Abandonment: Clear pricing and an efficient checkout procedure can lower cart abandonment rates.
Enhanced User Loyalty: Happy customers are more likely to come back and make more transactions, which increases user loyalty.
Positive Word of Mouth: Happy customers can spread the word about their wonderful experiences to their friends and family, which helps with word-of-mouth advertising.
Data-Driven Improvements: By addressing user pain points and adjusting to shifting consumer preferences, ongoing UX research and data analysis enable continuous development.

4.2 User Experience Design in Mobile Applications:

Mobile apps are become a necessary component of everyday life. This case study investigates the relationship between UXD and mobile app success.

Background:

There are many different types of mobile applications, ranging from productivity and healthcare to social media and entertainment. A mobile app's usability, engagement, and user pleasure are frequently what determine its success.
Key Aspects of UXD in Mobile Applications:

User Onboarding: A smooth onboarding procedure makes it easier for users to rapidly grasp the benefits and capabilities of the app.
Navigation: Clear call-to-action buttons and intuitive menus improve user engagement and task completion.
Performance: A good user experience is influenced by quick loading times, responsive design, and effective data processing.

User Engagement Features: Notifications, gamification components, and personalization are examples of features that improve user engagement and retention.
Accessibility: It is both legally required and a user-centered design practice to make sure the software is usable by people with disabilities.

Impact of Effective UXD: Effective UXD in mobile applications can lead to:

Higher Adoption Rates: A more expansive audience is more likely to acquire an app that is easy to use and has a seamless onboarding procedure.
Increased User Retention: Enhanced user experiences, intuitive navigation, and captivating features all help to increase retention rates.
Positive Reviews and Ratings: Happy customers are more likely to give apps a high visibility and credibility in app stores by leaving positive reviews and ratings.
Monetization Opportunities: Applications with high levels of user pleasure and engagement are better suited to profit from monetization techniques like in-app purchases and subscriptions.
Brand Loyalty: A well-made app can help users stay engaged with a brand or service and strengthen their brand loyalty.

4.3 User Experience Design in Healthcare Technology:

Digital tools for healthcare include everything from telemedicine platforms to electronic health records (EHRs). We examine the impact of UXD on the healthcare technology industry in this case study.

Background:
There has been a major digital revolution in the healthcare sector, with technology being essential to patient care, diagnosis, and data management. For patients and healthcare professionals to use healthcare technology safely and effectively, effective user experience design is essential.

Key Aspects of UXD in Healthcare Technology:

Ease of Use: The technology should be simple enough for patients and healthcare professionals to use without requiring a lot of training or annoyance.
Data Security and Privacy: It is critical to guarantee patient data security and privacy. Robust security measures should be incorporated into UXD without sacrificing usability.
Accessibility: To guarantee equitable access to care, healthcare technology should be usable by people with disabilities.
Clinical Workflow Integration: Technology should not interfere with clinical workflows; rather, it should be a seamless part of them for healthcare practitioners.
Patient Engagement: Effective communication between patients and healthcare providers, adherence to treatment programs, and engagement should all be promoted via patient-facing healthcare technology.

Impact of Effective UXD: Effective UXD in healthcare technology can lead to:

Improved Patient Outcomes: Patient empowerment through effective health management can result in better health outcomes thanks to user-friendly tools.
Reduced Healthcare Errors: Clinical workflows that are integrated and intuitive can aid healthcare providers in making well-informed decisions and lowering errors.
Increased Adoption: Technology that improves patient care and fits well with their workflow is more likely to be adopted by healthcare providers.
Compliance with Regulations: Adhering to accessibility and usability guidelines guarantees adherence to healthcare laws and minimizes legal liabilities.
Enhanced Telemedicine: Effective remote healthcare delivery in the context of telemedicine depends on a good user experience.

These case studies highlight the various ways that UXD is applied in various industries and show how these applications can significantly affect user pleasure, engagement, adoption, and overall success. Because effective UXD methods and ideas can be applied to a wide range of industries, they highlight the significance of user-centered design in the digital age.

Results

Emphasizing the concrete results that User Experience Design (UXD) case studies may produce across a range of sectors is crucial. These outcomes show how UXD methods may be effectively invested in and have a major influence on user pleasure, engagement, and business success. The results highlight the revolutionary potential of user-centric design principles, from higher conversion rates and better user retention in e-commerce to

improved patient experiences and better clinical outcomes in healthcare technology. Organizations can attain tangible advantages like optimized processes, decreased attrition rates, and favourable customer feedback by placing a high priority on usability, user-friendly interfaces, and unambiguous communication. The results of UXD case studies are concisely presented in the following table, which shows Table 1 the practical applications of successful design approaches in many domains.

Case Study	**Results**
E-Commerce:	
- Increased Conversion Rates	- Streamlined checkout processes
	- Reduced cart abandonment rates
	- Higher conversion rates
- Improved User Retention	- Higher user satisfaction and loyalty
	- Increased repeat purchases
- Enhanced Mobile Experience	- Improved engagement and sales on mobile
	- Optimized mobile shopping experience
Mobile Applications:	
- Higher User Engagement	- Improved usability and navigation
	- Longer app sessions
- Reduced Churn Rate	- Enhanced onboarding experience
	- Addressed user pain points
- Positive User Reviews	- Increased app ratings and favourable reviews
	- Attracted new users
Healthcare Technology:	
- Enhanced Patient Experience	- Improved satisfaction with medical services
	- Intuitive interfaces for appointments
	- Easy access to medical records
- Better Clinical Outcomes	- Empowered patients for self-management
	- Improved health outcomes
- Increased Efficiency for	- Streamlined workflows for healthcare
- Healthcare Providers	Professionals
	- Reduced cognitive load

Conclusion

The result of our research highlights how crucial User Experience Design (UXD) is to influencing digital interactions and propelling corporate success in a variety of sectors. We have examined the fundamental ideas of UXD—such as the significance of comprehending user demands, intuitive navigation, emotional involvement, and ongoing iteration—through a thorough literature analysis. Using a range of case studies in mobile apps, e-commerce, and healthcare technologies, we have direct experience with the real-world effects of good UXD techniques.

The literature review revealed that effective UXD begins with a deep understanding of user needs, behaviours, and preferences. Through the implementation of user-centric design principles and the integration of emotional engagement aspects, organizations may craft meaningful user experiences. Furthermore, UXD's iterative approach enables ongoing improvement in response to user input and changing market trends, guaranteeing that goods and services stay competitive and relevant in a constantly shifting environment (Ahsan, 2024).

Furthermore, the case studies offered specific illustrations of how UXD concepts are translated into real-world results. Effective UXD practices have a clear influence on everything from higher conversion rates and greater user retention to better clinical results and improved patient experiences. By placing a high value on user happiness, engagement, and loyalty, businesses can stand out from the competition and cultivate enduring bonds with their clientele.

It's evident that the fundamentals of UXD will continue to be vital in determining the nature of digital experiences in the future. Organizations need to stay flexible and agile as technology and user expectations change. They should embrace new trends and adjust their design strategies accordingly. Organizations may develop creative and engaging experiences that satisfy the changing needs of their users and propel economic success in an increasingly digital environment by adhering to the fundamental principles of UXD and utilizing insights from both literature and real-life case studies.

The findings emphasize how crucial it is to comprehend customer requirements, habits, and preferences when creating goods and services that appeal to consumers. Organizations may design experiences that satisfy user expectations and build loyalty and trust by emphasizing components like emotional engagement, intuitive

navigation, and usability. Furthermore, the study emphasizes how crucial it is to continuously iterate and improve UXD because user preferences and technical environments change over time. Organizations can modify their design strategies to stay competitive and satisfy the changing needs of their target audience by keeping an eye on user input and developing trends.

In conclusion, the study highlights that making an investment in UXD involves more than simply improving user interfaces; it also entails developing memorable and significant experiences that increase user pleasure, engagement, and loyalty. The fundamentals of UXD will continue to be crucial in building the digital experiences of the future as technology develops, guaranteeing that goods and services deeply connect with consumers.

Reference:

1. Pucillo, F., & Cascini, G. (2014, March). A framework for user experience, needs and affordances. *Design Studies*, *35*(2), 160–179. https://doi.org/10.1016/j.destud.2013.10.001
2. Luther, L., Tiberius, V., & Brem, A. (2020, May 16). User Experience (UX) in Business, Management, and Psychology: A Bibliometric Mapping of the Current State of Research. *Multimodal Technologies and Interaction*, *4*(2), 85–98. https://doi.org/10.3390/mti4020018
3. Cheng, R., Hong, J., Kyrola, A., Miao, Y., Weng, X., Wu, M., Yang, F., Zhou, L., Zhao, F., & Chen, E. (2012, April 10). Kineograph. *Proceedings of the 7th ACM European Conference on Computer Systems*, 85–98. https://doi.org/10.1145/2168836.2168846
4. Olson, G. M., & Olson, J. S. (2003, February). Human-Computer Interaction: Psychological Aspects of the Human Use of Computing. *Annual Review of Psychology*, *54*(1), 491–516. https://doi.org/10.1146/annurev.psych.54.101601.145044
5. Tesařová, K. (2022, September 1). In the pixel zone: Perception of digital design. *Architecture Papers of the Faculty of Architecture and Design STU*, *27*(3), 36–41. https://doi.org/10.2478/alfa-2022-0018
6. Borriraklert, A., & Kiattisin, S. (2021, August 31). User Experience Design (UXD) Competency Model: Identifying Well-Rounded Proficiency for User Experience Designers in the Digital Age. *Archives of Design Research*, *34*(3), 61–79. https://doi.org/10.15187/adr.2021.08.34.3.61
7. Vahdat. (2023, June 5). [UX Design Strategies for Building Trust with Users]. https://medium.com/. Retrieved May 7, 2024, from https://medium.com/@barnoteam/ux-design-strategies-for-building-trust-with-users-7d884e22f80f
8. Norman, D. (2013, November 5). *The Design of Everyday Things*. Hachette UK. http://books.google.ie/books?id=I1o4DgAAQBAJ&printsec=frontcover&dq=The+Design+of+Everyday+Things&hl=&cd=2&source=gbs_api
9. Stevens, E., & Stevens. (2022, June 22). *7 fundamental UX design principles all designers should know - UX Design Institute*. UX Design Institute. Retrieved May 7, 2024, from https://www.uxdesigninstitute.com/blog/ux-design-principles/
10. Rusu, C., Rusu, V., Roncagliolo, S., & González, C. (2015, July 1). Usability and User Experience. *International Journal of Information Technologies and Systems Approach*, *8*(2), 1–12. https://doi.org/10.4018/ijitsa.2015070101
11. Jurca, G., Hellmann, T. D., & Maurer, F. (2017, December 28). Agile User-Centered Design. *The Wiley Handbook of Human Computer Interaction*, 109–123. https://doi.org/10.1002/9781118976005.ch6
12. Pérez-Álvarez, M. (2017, October 24). Psychology as a Science of Subject and Comportment, beyond the Mind and Behavior. *Integrative Psychological and Behavioral Science*, *52*(1), 25–51. https://doi.org/10.1007/s12124-017-9408-4
13. lumenlearning. (2024, May 8). *The History of Psychology | Introduction to Psychology – Lindh*. https://courses.lumenlearning.com/. Retrieved May 8, 2024, from https://courses.lumenlearning.com/suny-fmcc-intropsych/chapter/outcome-the-evolution-of psychology/#:~:text=Philosophical%20interest%20in%20behavior%20and,in%20German%20and%20American%20labs.
14. Alves, D. (2023, July 3). *The Importance of Psychology in UX Design*. https://www.linkedin.com/pulse/importance-psychology-ux-design-daniele-alves
15. Abhishek. (2023, December 6). *Role of Psychology in Design - Strate School of Design*. Strate School of Design. Retrieved May 8, 2024, from https://strate.in/role-of-psychology-in-design/.
16. Lawerteh, A. (2022, October 9). *The role of psychology in design - Adinorkie Lawerteh - Medium*. Medium. https://medium.com/@adinorkielawerteh/the-role-of-psychology-in-design-7e1505a8c41d.
17. Nielsen, J., & Chan, M. (2024, April 19). *Mental Models*. Nielsen Norman Group. https://www.nngroup.com/articles/mental-models/
18. Kohler, T. (2024, January 25). *Psychology for UX: Study Guide*. Nielsen Norman Group. https://www.nngroup.com/articles/psychology-study-guide/
19. Luther, L., Tiberius, V., & Brem, A. (2020, May 16). User Experience (UX) in Business, Management, and Psychology: A Bibliometric Mapping of the Current State of Research. *Multimodal Technologies and Interaction*, *4*(2), 18. https://doi.org/10.3390/mti4020018

20. Ahsan, W. B. (2024). The EDIT UX Framework: A UserCentered Approach to Effective Product Redesign. Userhub. https://doi.org/10.58947/PBSY-TGJR

2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI) | 979-8-3503-6052-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/IATMSI60426.2024.10502825

Visual Servoing using FPGA-based Hardware accelerated Deep Learning Solution

Prabhav Saxena
Department of Mechanical Engineering
Delhi Technological University
Delhi, India
prabhavsaxena_ch21b3_43@dtu.ac.in

Arjun Gupta
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
arjun222gupta@gmail.com

Shivansh Chadha
Department of Mechanical Engineering
Delhi Technological University
Delhi, India
shivanshuav350@gmail.com

Suhani Grover
Department of Engineering Physics
Delhi Technological University
Delhi, India
suhani1077@gmail.com

Sujal Gupta
Department of Electrical Engineering
Delhi Technological University
Delhi, India
guptasujal2003@gmail.com

Dr. Anup Kumar Mandpura
Department of Electrical Engineering
Delhi Technological University
Delhi, India
kanup@dtu.ac.in

***Abstract*—A novel deep-learning time-series architecture is proposed for visual servoing to control an unmanned aerial vehicle (UAV). This UAV is capable of quasi-stationary flight and is equipped with a mounted camera to track a target composed of a finite set of stationary points located in a plane. The proposed solution utilizes a time-series deep learning architecture to learn temporal data from sequential values. It then generates control cues to be sent to the flight controller, enabling the UAV to perform its tracking task. We also present an optimized implementation of the novel proposed architecture on FPGA. When it comes to implementing Neural Networks, FPGAs have demonstrated their efficiency. Specifically, our implementation is quicker than top-tier central processing units (CPUs) when it comes to carrying out the inference of the suggested model. Moreover, the proposed system also achieved improvements in terms of energy efficiency, which is important for UAV's longer endurance.**

Keywords— Visual Servoing, Deep Learning, Long Short-Term Memory, Hardware Acceleration, FPGA.

I. INTRODUCTION

Visual servoing refers to a control method for robotic maneuvers that utilizes optical data from present sensors. Its objective is to guide a robot from one point to another, through space whether specified or unspecified, by minimizing the error iteratively between its present position and the target position. There are several uses for visual servoing, such as robotic surgery and autonomous navigation [1].

Two traditional approaches to the visual servoing problem are Position-Based Visual Servoing (PBVS) and Image-Based Visual Servoing (IBVS). With an inertial frame of reference, PBVS specifies the error in Euclidean space and produces simple control laws. It does, however, entail the error-prone Perspective-n-Point problem, which necessitates accuracy-enhancing result filtering [2]. With Image-Based Visual Servoing (IBVS), visual feature extraction is used to define the inaccuracy inside the image plane. Control strategies are developed based on the image Jacobian/Interaction matrix, with its pseudo-inverse approximated. Despite their efficiency, both IBVS and PBVS are reliant on camera-specific factors, such as the intrinsic parameters of the camera and the need for precise calibration [3].

Our research introduces a novel approach to visual servoing, treating it as a time-varying neural network problem analyzed sequentially. We employ Long Short-Term Memory (LSTM) networks and generate a set of time-series data using a custom environment within the Robotic Operating System (ROS). This approach addresses the challenges of visual servoing. Similar to IBVS, we define the error in the image plane and calculate the Image Jacobian. Control is achieved using distinct computer vision algorithms.

LSTM is capable of accurately propagating initial information from previous time steps to the current process. Furthermore, research, such as [17][18], has demonstrated that Field Programmable Gate Arrays (FPGAs) exhibit great potential in efficiently implementing Recurrent Neural Network (RNN) and LSTM architectures. Therefore, to implement the developed model, programmed in Python and TensorFlow API, on the Xilinx Alveo U200, the program is configured into a hardware-friendly C++ program using the Vitis HLS tool.

The format of the paper is as follows: In Section II, deep learning techniques for visual servoing are thoroughly reviewed, along with the FPGA's potential for executing RNNs and LSTMs. Section III discusses the conventional IBVS algorithm for dataset collection, followed by the introduction of a new DLVS architecture that optimizes the neural network to accelerate RNN model inference on an FPGA. Section IV details the training setup, test results, and algorithm implementation on FPGA. Section V provides an overview of future research and concludes the study.

979-8-3503-6052-3/24/$31.00 ©2024 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:00:04 UTC from IEEE Xplore. Restrictions apply.

II. Related Works

The study referenced in [4] employed techniques that required the manual design of visual features for image representation. Control laws were determined using backpropagation of gradient descent on spatial information errors. Using this method, the camera and interaction matrix were estimated analytically or numerically. The algorithm was subsequently modified for reinforcement learning, a common technique in deep reinforcement learning, and applied to a deep learning-based architecture. The results suggested that the learned parameters of the model may not apply to tasks or surroundings that are unfamiliar to it since they are context-specific.

Another study utilized FlowNet Convolutional Neural Networks for visual servoing [5]. In order to predict optical flow, this method uses pairs of pictures as inputs and the accompanying x-y flow fields as real data. After which, the data is processed using several ConvNets and ReLU activations. Practical implementation of the architecture requires training in diverse environments and camera settings [9].

Furthermore, some recent academic articles have shown that FPGAs have enormous capability in both efficiency and energy consumption while carrying out RNNs and LSTMs. In citation [13], a speech recognition LSTM model was accelerated by an FPGA-based accelerator, resulting in a 20.18x faster speed compared to a software implementation using a single thread, and 5.41x faster than an implementation using 16 threads. The reference [15] introduces GRNN, an RNN-LSTM inference library based on GPUs that surpasses the most optimized and efficient inference library designed for multi-core server CPUs, with a latency reduction of up to 17.5×. A platform for optimizing deep RNNs and LSTM/LSTMP cells using multi-FPGAs is presented in [16] by the authors. The use of a single-layer LSTM-RNN on the FiC platform enabled the authors to speed up their system significantly. This resulted in a 31-fold increase in speed compared to an Intel CPU and a 2-fold increase compared to a GPU. Moreover, the system attained 61 times and 5 times faster processing speed than a CPU and a GPU, respectively, when a four-layer LSTM-RNN was used. When processing 8-bit numbers, it achieved a power efficiency of 12 GFLOPS/W.

III. Methodology

C. Image-Based Visual Servoing

In image-based visual servoing, similar to other vision-based control methods, the objective is to minimize the error e(t), which is defined as:

$$e(t) = s - s* \tag{1}$$

The features in the image are represented by the vector s, and their desired placements on the image plane are represented by the vector s*. The error term is derived to design the velocity controller. The spatial velocity of the UAV/Camera is denoted by $V_c = (v_c, w_c)$, where v_c is the instantaneous linear velocity and w_c is the instantaneous angular velocity. Only the linear velocities are calculated and controlled as the UAV is an underactuated system,

The relationship is established as

$$de/dt = L_s v_c \tag{2}$$

Where $L_s \in R^{k*6}$ is the interaction matrix/feature Jacobian matrix. We apply the following control law to reduce the error term exponentially.

$$v_c = -\lambda L^+_e e \tag{3}$$

In a real visual servo system, L^+_e is the approximated Moore-Penrose pseudoinverse of L_e, as their exact values cannot be precisely determined.

The following equations describe the relationship between the feature points in the image plane with coordinates $\mathbf{x} = (x, y)_{in\ pixels}$ and their projection from a 3D point of view with respect to the camera frame with coordinates $\mathbf{X} = (X, Y, Z)$ is defined by the following equations:

$$x = X/Z = (u - c_u)/f\alpha \tag{4}$$

$$y = Y/Z = (v - c_v)/f \tag{5}$$

In this context, c_u, c_v, f and α represent a set of intrinsic camera parameters. C_u and c_v are the principal coordinate points, f is the camera's focal length, and α is the pixel dimension ratio (usually equal to 1). For gathering and testing datasets, these parameters were defined in the ROS simulation environment. We determined the interaction matrix to be:

$$Lx = \begin{bmatrix} -1/Z & 0 & x/Z & xy & -(1+x^2) & y \\ 0 & -1/Z & y/Z & 1+y^2 & -xy & -x \end{bmatrix}$$

where Z represents the depth of the point from the camera.

B. Dataset Collection

Sequential deep learning models necessitate a substantial dataset for training and model inference. Due to the absence of a publicly available dataset for visual servoing applications, data for this problem statement is manually collected. This is achieved by executing IBVS control in a custom simulation environment built on ROS and Gazebo 3D simulator [6,7]. The data collection involves varying the target's (ArUco marker) location, velocity, and the UAV's altitude. The input vector comprises the error components (ex, ey) in both the x and y dimensions of the image plane, along with the UAV's altitude. Conversely, the output vector includes the UAVs' control cues, specifically the desired velocity computed using the IBVS control algorithm.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:00:04 UTC from IEEE Xplore. Restrictions apply.

C. Proposed Architecture



Fig. 1: Proposed network architecture

The suggested architecture follows a single-phase regression approach. Here, the model's primary objective is to extract temporal information from the sequential input data (error e(t)) and use it to predict the UAV's control cues, specifically its velocity.

As illustrated in Figure 1, the input layer of the model receives the error inputs and altitude as a stacked array with a shape of (32x3). Subsequently, the LSTM layers process the input with units of 16, 64, 256, and 512 to extract key features from the input sequence. The output from the LSTM layers, with a shape of (32x512), is then passed to the encoding portion of the neural network. Here, the output is further processed with units of 256, 64, and 16 to consolidate useful information for predicting the velocity. Finally, the output layer completes the model by predicting the required output in the shape of (32x2), representing the control cues for the UAV.

The input, forget, and output gates of the LSTM layers are recurrently activated using the Sigmoid activation function during the decision-making process. Meanwhile, the memory state activation function is a customized Modified_tanh function, which helps to alleviate the vanishing gradient problems that are frequently seen in recurrent neural networks. It has been observed empirically that using asymmetric activation functions in a large neural network can lead to faster convergence during back-propagation learning. Therefore, the proposed network architecture introduces non-linearity with a modified version of the hyperbolic tangent activation function.

$$y = 5 \times ((e^{x-5} - e^{-x/5})/(e^{x/5} + e^{-x/5})) \quad (6)$$

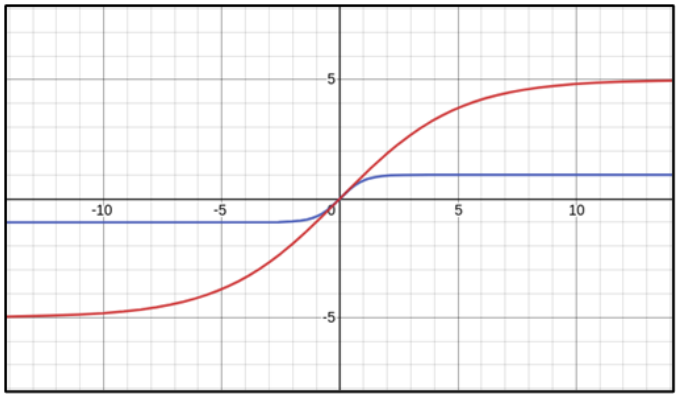


Fig 2:Plots of the activation functions: "modified_tanh" (red) and regular "tanh" (blue).

The UAV's capacity to reach higher velocities is constrained by the normal tanh activation function, which generates output values in the interval [-1, 1]. To overcome this constraint, a new modified tanh activation function is suggested, which permits the output value to lie between [-5, 5], as explained in (6). The behavior of this altered function is consistent with that of the standard tanh function, as depicted in Figure 2.

D. Neural Network Optimization

This study's ultimate objective is to expedite the inference of the RNN model on an FPGA (Xilinx Alveo U200). To reconfigure the model efficiently for the FPGA, we used the C++ functions of TensorFlow modules developed by the researchers of [12] and tailored it to use in our application. After the development of the C++ codes, they were tested against the same input data as the Python/TensorFlow model. Both of the configurations got similar outputs. Further optimization of the program was done for efficient implementation on the FPGA by pruning the model, Quantizing the weights and biases, and hardware-optimized mapping of the equations and weights.

1) Pruning

In order to reduce the computational cost of our neural network, we applied weight pruning to the LSTM layers. Specifically, we used a structured pruning approach that preserved the connectivity and structure of the LSTM layer while still reducing the number of parameters. We set a threshold for each weight value and zeroed out any weights that fell below the threshold, but we also ensured that the gates and memory cells remained intact. Through experimentation, we found that pruning up to 20% of the weights in the LSTM layers did not significantly degrade performance while reducing the number of parameters by nearly one-fifth. These results improved the computational efficiency of LSTM-based neural networks without sacrificing much accuracy.

2) Quantization

Quantization converts the weights and biases of a neural network from its original floating-point representation to a lower precision representation. This reduces the memory requirements and computational cost of the model, making it more efficient. FPGAs leverage their ability to handle numbers with arbitrary bit lengths efficiently in order to create an optimal programmable module. To achieve a high throughput rate without compromising much on the accuracy we quantized our model using Vitis-AI.

3) LSTM Hardware optimization

After obtaining the quantized values of weights and intermediate values from Vitis-AI, we allocated a register with the same bit length to store the values instead of storing them on BRAMs to reduce the latency of the model.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:00:04 UTC from IEEE Xplore. Restrictions apply.

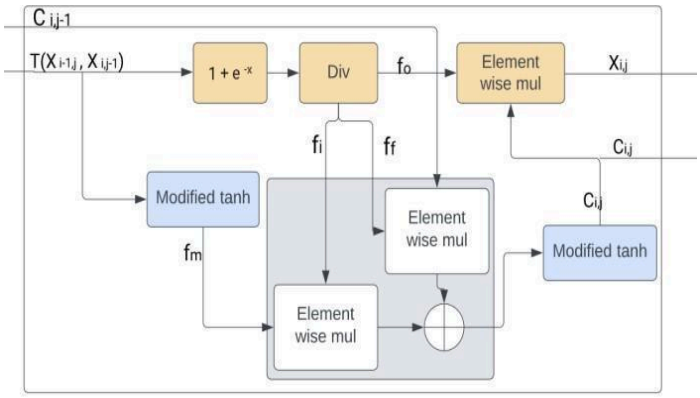


Fig. 3: Block diagram of modules implementing the basic LSTM operations

Each symbol corresponds to one of the four gates of the LSTM block, which are the input gate (f_i), forget gate (f_f), output gate (f_o), and input modulation gate (f_m). The elementwise multiplication is denoted by the usual '$\odot$' symbol. $C_{i,j}$ denotes the state or recollection of the ith layer's j^{th} time step, while b, and W and W' correspond to relevant biases and weights respectively. The dimensions of the four gates and $C_{i,j}$ at every time step are [1, Ki]. Demonstrating the employment of every term indicated in Figure 3 presents a streamlined depiction of the unit that executes the fundamental LSTM layer, featuring equations as illustrated underneath.

$$f_i = sigmoid(T(X_{i-1,j}, X_{i,j-1})) \tag{7}$$

$$f_f = sigmoid(T(X_{i-1,j}, X_{i,j-1})) \tag{8}$$

$$f_o = sigmoid(T(X_{i-1,j}, X_{i,j-1})) \tag{9}$$

$$f_m = modified\ tanh\ (T(X_{i-1,j}, X_{i,j-1})) \tag{10}$$

$$c_{i,j} = f_f \odot c_{i,j-1} + f_i \odot f_m \tag{11}$$

$$X_{i,j} = f_o \odot modified\ tanh\ (c_{i,j}) \tag{12}$$

$$T(X_{i-1,j}, X_{i,j-1}) = X_{i-1,j} \odot W + X_{i,j-1} \odot W' + b \tag{13}$$

The nomenclature utilized in this context designates the weights connecting a specific i^{th} layer and the subsequent $(i + 1)^{th}$ layer as $W_{i,i+1}$. The biases for a particular layer (layer i) are referred to as bi. The input vector for that layer is represented by Xi, and its size is [nbs, 1]. The weight matrix connecting layers i and i+1 is represented by $W_{i,i+1}$, and its size is $[K_i, K_{i+1}]$. The bias vector for layer i is represented by [nbs, 1]. Layer i has a dimension of K_i, while nbs denotes the number of input signals that will be categorized.

IV. Experimental Results

A. Training Details

The proposed architecture is developed, implemented, and trained using the TensorFlow library. In noisy issues, the adaptive momentum optimizer handles sparse gradients [11]. An internal dataset with dynamic target data and a range of elevations is used to train, infer, and validate the model. The input batch size is set to 32, representing 32 consecutive errors in x and y, along with UAV altitude values fed into the network. Training is conducted over 150 epochs on a Quadro P5000 GPU. Approximately 2.9 million parameters make up the network, crucial for collecting trainable features and predicting UAV velocities.

The architecture utilizes LSTM-based encoding and decoding to learn temporal data from 32 consecutive error values, leading to the desired outcomes. To determine the optimal length of the stacked values, experiments were conducted with 8, 16, 32, and 64 stack values. Results showed that temporal details were insufficient for training the model with 8 and 16 stacked frames. Nonetheless, the highest level of identification accuracy was attained with 32 frames. Although 64 frames performed similarly in terms of relative accuracy, handling larger stacks was necessary (Fig 4). For the model, a stack of 32 frames was selected as a result.

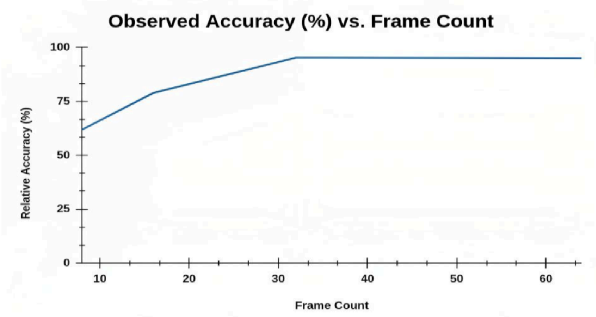


Fig. 4: The impact of input size on the performance of the proposed algorithm

B. Real-Time Testing

The method is assessed in real-world and simulation contexts. Using ROS, an open-source robotics middleware that offers a variety of software frameworks for emulating robot hardware abstractions, the algorithm is tested in simulation in a customised environment. In this arrangement, Gazebo acts as the 3D simulator. In the simulation, there is an ArUco marker placed atop a ground vehicle that serves as the tracking target, and an Iris UAV with a gimbal-stabilized camera positioned underneath. Figure 5 shows the environment used for the simulation.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:00:04 UTC from IEEE Xplore. Restrictions apply.

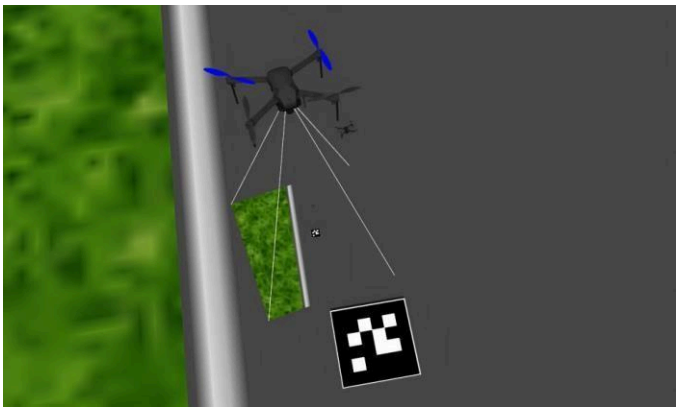


Fig. 5: ArUco marking in a frame captured by a drone in the simulation environment.

For real-world testing, depicted in Figure 6, we utilized a Tarot 650 UAV, a Pixhawk Cube flight controller, and a Jetson Nano 2.0 as the onboard computer. A Git2P camera for tracking and detection was also a part of the system; it was stabilized with a Tarot TL3T05 gimbal. The camera's frames are sent to the onboard computer (OBC) for ArUco marker detection when the program is activated. The deep learning model receives the errors and UAV height upon detection and uses those data to estimate velocities to operate the UAV. The UAV is then able to maneuver and follow the ArUco target thanks to the transmission of these output velocities as MAVLink messages to the Pixhawk flight controller.



Fig. 6: Real-life tracking of ArUco Marker using Tarot-650 UAV

C. FPGA implementation

The power consumed by the original Python/TensorFlow model executed on an Intel i9 processor is calculated using the Intel pcm-power utility to assess the efficiency of the presented system in comparison to the reference Python model. The Python program took 41 ms to be executed along with a power efficiency of 0.024 GFLOPS/W. Regarding the suggested system that employs hardware acceleration, the Xilinx Power Estimator (XPE) has indicated that the power usage on the chip when running at 200 MHZ is 12.5 watts. Based on the power consumption the power efficiency of the hardware-optimized implementation of the architecture was approximately 180 times higher energy efficient than the reference Python model with a power efficiency of 4.3 GFLOPS/W.

The resources used to implement the optimized model on the Xilinx Alveo U200 FPGA board, as shown by the Vitis HLS Tool, are presented in the table below.

Table 1: Estimated Resource utilization on Xilinx Alveo U200 FPGA

Logic Utilization (Qty)	200MHz
DSP48E	80% (5,536)
Flip Flops	34% (832,166)
Slice LUTs	42% (509,078)

V. CONCLUSION

By merging computer vision for object detection with an LSTM-based deep learning neural network for autonomous tracking, the research presents a novel method for visual servoing. The suggested method creates UAV control signals by encrypting error data that is obtained from calculating the separation between the frame center and the target ArUco marker's pixels. To train the time series model, a custom dataset was made using the Gazebo Simulator and ROS. The algorithm was validated in real-world and simulation settings. After attaining favorable outcomes, the system was refined for hardware acceleration using a top-tier Xilinx Alveo U200 FPGA.

To implement the deep learning model on the FPGA, optimized C++ functions were developed for TensorFlow modules for LSTM cells, leveraging open-source code provided by researchers [12]. Additional optimizations were incorporated into the methodology. The optimized Python program was found to be 180 times more energy-efficient in terms of GFLOPS/W compared to the original program when executed on a high-end system.

In terms of the research's future outlook, we strive to

- Implementing the program on a lighter FPGA i.e. the Xilinx Zynq 7000 SoC to replace the Jetson Nano onboard computer with higher power efficiency i.e., essential for UAV.
- Using an RL-based model to learn and estimate the control cues for target tracking with higher accuracy, efficiency, and reduced time complexity.

VI. REFERENCES

[1] W. Fu, H. Hadj-Abdelkader and E. Colle, "Visual servoing based mobile robot navigation able to deal with complete target loss," 2013 18th International Conference on Methods & Models in Automation & Robotics (MMAR), 2013, pp. 502-507, doi: 10.1109/MMAR.2013.6669961.

[2] Marinela G. Popova and Hugh H. Liu. "Position-Based Visual Servoing for Target Tracking by a Quadrotor UAV," AIAA 20162092. AIAA Guidance, Navigation, and Control Conference. January 2016

[3] F. Chaumette and S. Hutchinson, "Visual servo control. I. Basic approach," in IEEE Robotics & Automation Magazine, vol. 13, no. 4, pp. 82-90, Dec. 2006, DOI: 10.1109/MRA.2006.250573.][F. Chaumette and S. Hutchinson, "Visual servo control. II. Advanced approach [Tutorial]," in IEEE Robotics & Automation Magazine, vol. 14, no. 1, pp. 109-118, March 2007, DOI: 10.1109/MRA.2007.339609.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:00:04 UTC from IEEE Xplore. Restrictions apply.

[4] Q. Bateux, E. Marchand, J. Leitner, F. Chaumette and P. Corke, "Training Deep Neural Networks for Visual Servoing," 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018, pp. 3307-3314, doi: 10.1109/ICRA.2018.8461068

[5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 3431-3440, doi: 10.1109/CVPR.2015.7298965.

[6] Quigley, Morgan, et al. "ROS: an open-source Robot Operating System." ICRA Workshop on open-source software. Vol. 3. No. 3.2. 2009.

[7] N. Koenig and A. Howard, "Design and use paradigms for Gazebo, an open-source multi-robot simulator," 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566), 2004, pp. 2149-2154 vol.3, DOI: 10.1109/IROS.2004.1389727.

[8] Sepp Hochreiter, Jürgen Schmidhuber; Long Short-Term Memory. *Neural Comput* 1997;9(8):1735-1780. doi: https://doi.org/10.1162/neco.1997.9.8.1735

[9] Anurag Sai Vempati, Vipul Choudhary, Laxmidhar Behera, Quadrotor: Design, Control and Vision Based Localization, IFAC Proceedings Volumes, Volume 47, Issue 1, 2014,Pages 1104-1110,https://doi.org/10.3182/20140313-3-IN-3024.00229. (https://www.sciencedirect.com/science/article/pii/S1474667016327926)

[10] Virmani, Aman, Arjun Gupta, Vayam Jain, Nilesh Aggarwal, and Anup Kumar Mandpura. "MTSO: Multi-Target Search Optimisation based on Probability Map."

[11] Kingma, D.P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. CoRR, abs/1412.6980.

[12] Tsantikidou, K.; Tampouratzis, N.; Papaefstathiou, I. A Novel FPGA-Based Intent Recognition System Utilizing Deep Recurrent Neural Networks. Electronics 2021, 10, 2495. https://doi.org/10.3390/electronics10202495

[13] Guan, Y.; Yuan, Z.; Sun, G.; Cong, J. FPGA-based accelerator for long short-term memory recurrent neural networks. In Proceedings of the 22nd Asia and South Pacific Design Automation Conference (ASP-DAC), Jeju Island, Korea, 22–25 January 2018; pp. 629–634.

[14] N. Aggarwal, V. Jain, JP Kesari “Image-based visual servoing for tracking and landing on a moving target using a multirotor” vol. 3 issue 11, November 2021

[15] Holmes, C.; Mawhirter, D.; He, Y.; Yan, F.; Wu, B. GRNN: LowLatency and Scalable RNN Inference on GPUs. In EuroSys ’19: Proceedings of the Fourteenth EuroSys Conference 2019; Association for Computing Machinery: New York, NY, USA, 2019

[16] Yuxi, S.; Hideharu, A. FiC-RNN: A multi-FPGA acceleration framework for deep recurrent neural networks. IEICE Trans. Inf. Syst. 2020, E103D, 2457–2462.

[17] Han, S.; Kang, J.; Mao, H.; Hu, Y.; Li, X.; Li, Y.; Xie, D.; Luo, H.; Yao, S.; Wang, Y.; et al. ESE: Efficient Speech Recognition Engine with Compressed LSTM on FPGA. arXiv 2016, arXiv 1612.00694. [16] Wang, S.; Lin, P.; Hu, R.; Wang, H.; He, J.; Huang, Q.; Chang, S. Acceleration of LSTM With Structured Pruning Method on FPGA. IEEE Access 2019, 7, 62930–62937.

[18] N. Aggarwal, Anunay, V. Jain, T. Singh and D. K. Vishwakarma, "DLVS: Time Series Architecture for Image-Based Visual Servoing," 2023 8th International Conference on Control and Robotics Engineering (ICCRE), Niigata, Japan, 2023, pp. 101-107, doi: 10.1109/ICCRE57112.2023.10155585.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 09,2024 at 07:00:04 UTC from IEEE Xplore. Restrictions apply.